Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1765–1776. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1827–1838.

NUCLEI
Experiment

Comparative Analysis of the ^{178m_2}Hf Yield at Reactions with Different Projectiles *

S. A. Karamian**

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Received August 4, 2004; in final form, December 9, 2004

Abstract—The long-lived high-spin ^{178m_2}Hf K isomer can be produced in nuclear reactions with different projectiles. The reaction yields and cross sections have been measured in a series of experiments and the results are now reviewed. The systematics of isomer-to-ground-state ratios are drawn and real production capabilities are estimated for the best reactions. Such a summary is relevant to the significance of isomer studies, both for nuclear-science knowledge and for possible applications. Potential isomer applications have been stressed previously in popular publications with probably overestimated expectations. The real possibilities are restricted in part by the production yield and by other shortcomings as well.

INTRODUCTION

Nuclear isomers in the mass range close to $A = 180$ are of special interest because they are characterized by unique combinations of high excitation energy, high spins, and K-quantum numbers with long lifetimes. Such features make these isomers extremely attractive for applications to γ-ray pulsed sources because they may store the nuclear excitation energy for long time and also provide a high density of energy. In the classical example of the 31-year ^{178m_2}Hf isomer, the energy density reaches 1.3 GJ/g.

An excited nuclear state manifests itself as a metastable isomer when its decay is significantly retarded due to some kind of mismatch between the wave functions of initial and final states. Such hindrances to the decay typically arise because of collective deformation of a nucleus in excited state and either of the angular-momentum step at the decay transition or of a structure inhibition. In the region of statically deformed nuclei, the structure hindrance may play an important role, in addition to the selection rules by the spin and parity (I^π) for the multipole electromagnetic transitions. For instance, a change in the orientation of the angular momentum vector generates special structure inhibition known in the literature as K hindrance. Recall that K is a quantum number of the I-vector projection onto the deformation axis. The axial symmetry of the nuclear deformation is a major assumption. Many K-hindered isomers are known, and their properties are described in the literature (see, for instance, [1–4] and references therein). ^{178m_2}Hf is one of the clearly identified K isomers because the K-hindrance factor is as high as about 10^9 in this case. Without the latter value, the isomeric state should decay fast, with a half-life of 0.8 s, instead of the real $T_{1/2} = 31$ yr. Thus, the methods available for unique isomer production are of interest in different aspects, in particular, for extension of the exotic-nuclei phenomenology.

The field of the present work is intentionally narrowed to the nuclear reaction features only. The nuclear structure properties are touched as far as they influence the isomer-production cross section. The discussion of applications is excluded because we have no intention of supporting a "sensation" that has appeared in some newspapers and magazines, for instance, in the *Washington Post* and *Der Spiegel*. One can find a more professional and responsible approach in the review articles published in physical journals [1–5]. They contain the explanation that the isomer research is at a very basic stage, far from real use of the isomer sample as an energy or weapon unit. In the summary of the present work, the restrictions are discussed for the amount of isomer material that can really be produced.

Decay retardation, being useful for energy accumulation, is accompanied at the same time with the suppression of the isomer-production cross section because of the similar factor of the wave-function mismatch. In theory, the conservation of K-quantum number is not an absolute imperative because it is conserved until the axial symmetry of nuclear shape is perturbed. After the experiments of [6, 7], it was evident that the K-hindrance factor decreases with the

*The text was submitted by the authors in English.

**e-mail: karamian@nrmail.jinr.ru

excitation energy growth. The long-lived states (isomers) are populated in nuclear reactions through the cascade of γ quanta emitted by the excited reaction residue. At typical residual excitations, the K hindrance must be significantly diminished and the isomer yield should fortunately be increased. But yet, the isomer-to-ground-state ratio in many cases remains not high, $\sigma_m/\sigma_g \ll 1$, as known from experiments. This is because of high spin of the isomeric states. In γ cascades, the most probable are the transitions of low multipolarity, $E1$, $M1$, $E2$, and they cannot directly supply the required spin deficit if the isomer spin is much higher than the residual angular momentum. Many stretched transitions in the cascade are needed and the probability is decreased. There is no possibility of violating the angular-momentum conservation because it is an integral of motion and it should be conserved absolutely in both classical and quantum mechanics.

The correlation of isomer cross section σ_m with the spin deficit was qualitatively clear even before the experiments reviewed here. However, it does not mean that the isomer yields and m/g ratios could be reliably calculated in theory and used for practical estimations. In reality, the spin distribution of the residual nucleus cannot be easily predicted for many reactions. On the other hand, the measured m/g ratio sometimes serves as a basis for estimates of the mean angular momentum of the residual nuclei, for instance, in a spallation reaction with intermediate-energy protons. Another uncertainty is due to the structure peculiarities of the level scheme and the γ-cascade branching for some individual nucleus. Simplified statistical-model calculations may not be very accurate, especially if they are applied to excited levels of below 3 MeV. So, as usual in nuclear physics, experimental measurements are needed to obtain reliable values of the reaction cross section and yield. The experiments are described below.

One introductory remark concerns also the discussion of the isomer application in a mode of the controlled source of energy and radiation, assuming that the isomer decay can be artificially stimulated (triggered) by the external radiation. Within the "up-conversion" scheme, a photon is absorbed and provides the transition from the isomer to some higher lying level. The latter one should decay fast to the ground state and the isomeric energy is released. But the efficiency of such a process should be again restricted by the wave-function mismatch between the isomer and other levels. The triggering cross section can be too small, even if such a lucky intermediate level exists. Extensive experimental studies may clarify the triggering efficiency. In the present paper, triggering experiments are out of the discussion, and a review of experimental attempts for triggering known to date is given in [1].

1. ^{178m_2}Hf PRODUCTION WITH BREMSSTRAHLUNG

Photon-induced nuclear reactions were systematically studied in [8, 9] at the irradiations of $^{\rm nat}$Ta, $^{\rm nat}$Hf, and ^{178m_2}Hf targets with bremsstrahlung at the end-point energy of 23.5 MeV. The activation technique was applied, and as many as 18 yields of the (γ,γ'), (γ,n), (γ,p), $(\gamma,2n)$, and (γ,α) reactions were successfully measured. Among them, there were reactions leading to the population of isomeric and ground states; thus, the isomer-to-ground-state ratios were deduced. Most original was the observation of reactions with the isomeric ^{180m}Ta and ^{178m_2}Hf nuclei and the demonstration that the high-spin isomers are easy populated in the reaction product when the target nucleus is also a high-spin isomer. This does not contradict the correlation of yield with the nuclear spin difference of the target and product.

At the attempts of ^{178m_2}Hf production with bremsstrahlung, its low activity could not be distinguished in the presence of other radionuclides induced in the irradiated $^{\rm nat}$Hf target [8, 9]. But later, at the 22-MeV end-point bremsstrahlung irradiation, higher sensitivity of measurements was reached [10] and the yield of ^{178m_2}Hf was observed and attributed to the ^{179}Hf(γ,n) reaction. The latter reaction was most productive because of the highest yield of (γ,n) products at such an energy and because of the highest spin value of the ^{179}Hf nuclei among the other stable Hf isotopes. The isomer-to-ground-state ratio was found to be

$$Y_m/Y_g = (3.5 \pm 1.0) \times 10^{-5}. \qquad (1)$$

Respectively, the absolute yield should not be high when one uses the 22-MeV bremsstrahlung for the accumulation of the ^{178m_2}Hf isomer. Assume that 10 g of isotopically enriched ^{179}Hf is irradiated close behind the converter and the electron beam has an intensity of 100 μA. The estimated yield of ^{178m_2}Hf of about 4×10^7 atom/s is not of interest for the production of the isomer in an amount needed for the target preparation.

It would be very probable that the yield must be higher with an electron beam of higher energy, and a new experiment has been performed using a 4.5-GeV electron beam at the Yerevan Synchrotron. A stack of Ta foils was exposed to the bremsstrahlung generated in the W converter. The long-based collimation system and relatively thin converter and target samples were used for the better definition of experimental conditions. After a long cooling time, the

activity of Ta foils was measured with a Ge gamma spectrometer having a 20% efficiency. Only long-lived products survived to the time of measurements, such as ^{178m_2}Hf, ^{172}Hf, ^{150}Eu, and ^{133}Ba, and they could be quantitatively determined. The experiment is described in [11] in some detail.

The yield of ^{178m_2}Hf has been measured and the isomer-to-ground state ratio is found to be

$$Y_m/Y_g = 0.032 \pm 0.010, \qquad (2)$$

i.e., much higher than the value given above in (1), for the reaction induced by bremsstrahlung at 22 MeV. The transmutation of ^{181}Ta into ^{178m_2}Hf requires the emission of a proton and two neutrons; however, the $^{181}\mathrm{Ta}(\gamma, p2n)^{178m_2}$Hf reaction can be written formally because at high energy not only nucleons but also mesons are generated and emitted. Respectively, a variety of reactions leading to the same product arise. Photon absorption at $E_\gamma \geq 200$ MeV involves the mechanism of meson generation and corresponding peaks are strongly manifested in the excitation function. Above 1200 MeV, the absorption cross section reaches an almost constant asymptotic value of about 0.12 mb/nucleon.

At low energies, $E_\gamma \leq 200$ MeV, the quasideuteron mechanism provides the highest contribution, and even lower, the tails of $E1$ and $E2$ giant multipole resonances are of importance. The yield of the $^{181}\mathrm{Ta}(\gamma, p2n)$ reaction at $E_\gamma \leq 50$ MeV should be deteriorated by the energy deficit. Thus, one may conclude that, in the irradiations with 4.5-GeV bremsstrahlung, the ^{178m_2}Hf isomer is produced mostly due to the absorption of photons in the range from 50 to 1200 MeV. After understanding this, it would be easy to accept ratio (2), which is comparable with the values determined in [12] for the Ta spallation by protons of intermediate energy of 600–300 MeV. In the range of a few hundred MeV, the photon absorption transfers to the nucleus sufficient energy for the emission of many particles, as in the case of proton-induced spallation.

The estimated m/g ratio also leads to the conclusion that a reasonably high angular momentum is acquired by the residual nucleus in the reactions with high-energy photons, not only with protons. The measured yield of ^{178m_2}Hf allows estimating the maximum achievable productivity of the reaction with 4.5-GeV bremsstrahlung. In a view toward optimization, thicknesses of the target and the converter should be enlarged. In this way, more quanta can be created and used for nuclear reactions, although obviously the absorption will also be increased. A reasonable compromise would be to combine both converter and target, which would then be a rather thick sample of Ta.

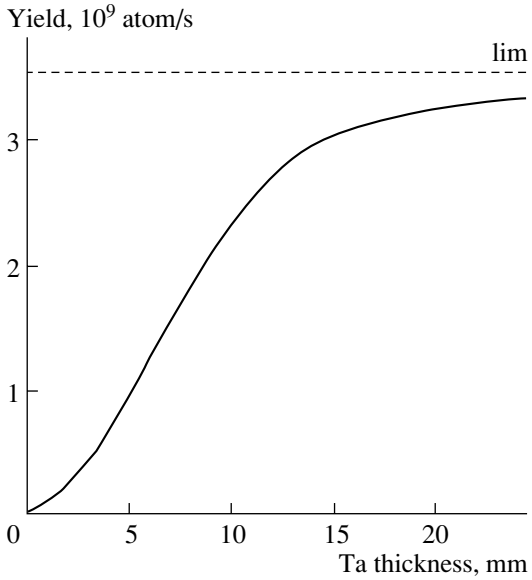

Fig. 1. Yield of the ^{178m_2}Hf as a function of thickness of the Ta sample exposed to 100-μA electron beam at 4.5 GeV.

The problem of optimization of such a unified assembly was solved analytically in [11]. As is known, the radiation energy losses of electrons in the ultrarelativistic case are proportional to the electron energy

$$\left(\frac{dE}{dx}\right)_{\mathrm{rad}} = cE_e. \qquad (3)$$

The total yield of photonuclear reaction in a sample of thickness d exposed to an electron beam can be evaluated as follows:

$$Y_{\mathrm{nr}} \sim \int_0^d N_\gamma(t)\sigma dt = \frac{\bar{\sigma}c}{\mu^2}\left(\mu d + e^{-\mu d} - 1\right), \qquad (4)$$

where $\bar{\sigma}$ is a mean cross section of the reaction, and μ is a linear attenuation coefficient for the bremsstrahlung photons. In Fig. 1, the absolute production yield of ^{178m_2}Hf is plotted as a function of the thickness of the Ta sample when it is irradiated with a 100-μA electron beam at 4.5 GeV. In addition to the major production by bremsstrahlung, the reactions induced directly by electrons and the secondary processes are taken into account in the simplified approximation. With a Ta sample 20 mm thick, the number of ^{178m_2}Hf atoms reaches the value of

$$Y = 3.2 \times 10^9 \ [\mathrm{atom/s}]/100\ \mu\mathrm{A}. \qquad (5)$$

The photon-conversion efficiency and, respectively, the nuclear reaction yield must be increased by some factor under channeling conditions, when electrons are directed along the crystal axes or planes. But with a thick sample, such a factor should not be significant.

2. NEUTRON-INDUCED REACTIONS

Production cross sections in neutron capture reactions with thermal neutrons are typically low for high-spin isomers with $I \geq 10$. The isomer ^{177m}Lu($I^\pi = 23/2^-$) is an exception that confirms the general tendency, because the high spin of the target ^{176g}Lu($I^\pi = 7^-$) nucleus provides a rather moderate spin deficit $\Delta I = 4\hbar$ in the ^{176}Lu$(n,\gamma)^{177m}$Lu reaction. In contrast with neutron capture, fast neutron reactions provide additional possibilities.

The yield of the high-spin ^{178m_2}Hf isomer in reactions with neutrons was tested in experiments [13–15]. A relatively high cross section was found [14, 15] in the ^{179}Hf$(n,2n)^{178m_2}$Hf reaction induced with 14.5-MeV neutrons. But productivity is restricted by the neutron flux available when the T$(d,n)^4$He reaction is used for neutron generation. Much higher fluxes are created in reactors, but the spectrum is soft and needed energies of $E_n \geq 10$ MeV have very low probability. Unfortunately, slow neutrons are not effective for ^{178m_2}Hf production because the neutron capture ^{177}Hf$(n,\gamma)^{178m_2}$Hf reaction is characterized by a very low isomer-to-ground-state ratio of $\sim 0.5 \times 10^{-9}$, according to [13].

At the same time, the isomer population in the ^{178}Hf$(n,n'\gamma)^{178m_2}$Hf reaction has never been experimentally tested. In addition, the low yield observed in the (n,γ) reaction [13] can be a result of burnup of the produced isomeric nuclei. The burnup process may have been significant at the high fluence applied in [13], but its cross section was not known. Even now, the data on the ^{178m_2}Hf burnup in reactor irradiations are not complete. Only the branch of the ^{178m_2}Hf$(n,\gamma)^{179m_2}$Hf transmutation was experimentally characterized in [16], but the total ^{178m_2}Hf(n,γ) cross section was not yet measured. In such a context, a new experiment has enough motivation to be performed testing both burnup and $(n,n'\gamma)$ processes.

Metal $^{\text{nat}}$Hf foils 1 mm thick were activated in an external channel of the IBR-2 reactor at FLNP (JINR, Dubna) and were then studied using a Ge gamma detector having 20% efficiency. This was accomplished by the spectrometric electronics, which allowed a count rate up to 20 kCs/s with a reasonable dead time and conservation of spectral resolution. The neutron spectrum at the location of the target was known from previous experiments. But in addition, NiCr-alloy samples were used as spectators for the calibration of the thermal and fast neutron fluence. The Hf samples were irradiated with and without Cd shields and the method of Cd difference allowed isolating the effect of thermal neutrons and deduction of the thermal cross section.

In measured spectra of activated Hf, the γ lines were observed and quantitatively determined for the following radionuclides: ^{175}Hf, ^{179m_2}Hf, ^{180m}Hf, and ^{181}Hf. The bulk of the activity was defined by ^{175}Hf and ^{181}Hf formed in (n,γ) reactions. They served as the intrinsic calibration of the thermal and resonance neutron fluxes in the presence of flux attenuation due to the self-absorption in the 1-mm Hf samples.

In this way, the thermal cross section σ_{th} and resonance integral I_γ values were figured out for the ^{180m}Hf isomer formation and the results are in accordance with the tabular values [17]. The yield of the high-spin ^{179m_2}Hf isomer was newly obtained and attributed to the ^{179}Hf$(n,n'\gamma)^{179m_2}$Hf reaction with neutrons of the fission spectrum. The isomer-to-ground-state ratio $\sigma_m/\sigma_g \approx 1.6 \times 10^{-3}$ does not contradict the systematics of [18].

The activity of ^{178m_2}Hf was too low, and it could not be distinguished or estimated even after long (1.5 yr) "cooling" of the sample after irradiation. Only an upper limit was established for the number of produced ^{178m_2}Hf nuclei. Respectively, a limit for the cross section of ^{178}Hf$(n,n'\gamma)^{178m_2}$Hf could be evaluated immediately. This reaction may be productive only at neutron energies of above 3 MeV because of the isomer excitation energy 2.45 MeV plus 0.5 MeV spent for the ejected neutron and gammas. The number of such neutrons was estimated using the known spectrum of fast neutrons at the location of the irradiated sample and the measured activity of ^{58}Co in the NiCr spectator. As a result, the cross section of ^{178m_2}Hf production in the $(n,n'\gamma)$ reaction has been restricted to an upper limit of $\sigma_m \leq 0.04$ mb, and the corresponding σ_m/σ_g value has been found to be as low as $\leq 1.5 \times 10^{-5}$.

More complicated would be an estimation of the definite value of a limit for the neutron capture cross section leading to the ^{178m_2}Hf isomer. Complications are due to the contribution of both thermal and resonance neutrons and due to the unknown burnup cross section for the produced isomeric nuclei. The approach of [13] was obviously too simplified because a rather low burnup cross section of $\leq$20 b was assumed, and the effect of resonance neutrons was neglected. In the present experiment with total fluence $\leq 10^{18}$ neutron/cm^2, the burnup can be completely neglected (but not at 3–4 orders of magnitude higher fluences as in [13]).

However, the isolation of individual yields induced by thermal and resonance neutrons is still a problem in our case too. The Cd-difference method is not applicable when the product yield is not really observed and only the upper limits restrict the numbers of produced nuclei. Thus, we propose to operate with

Table 1. Cross sections of the ^{178m_2}Hf production in reactions with neutrons

Reaction	Projectile energy	ΔI, $\hbar$	References	Cross section, mb	σ_m/σ_g
^{177}Hf(n,γ)	Thermal	12	Recalculated result of [13]	1.1×10^{-3}	1.5×10^{-9}
	Thermal	12	Present study	$\leq 2 \times 10^{-2}$	$\leq 1.1 \times 10^{-8}$
	Resonance	12	Present study	$I_\gamma \leq 0.2$	$\leq 2.8 \times 10^{-8}$
^{178}Hf$(n,n'\gamma)$	$E_n \geq 3$ MeV	31/2	Present study	$\leq 4 \times 10^{-2}$	$\leq 1.5 \times 10^{-5}$
^{179}Hf$(n,2n)$	$E_n = 14.5$ MeV	11	[14]	7.3	3.5×10^{-3}

the effective cross section σ_{eff}. As is known, after irradiation for t_{irr} in a reactor, the product yield in a linear approximation is proportional to

$$Y \sim t_{\text{irr}} \left(\sigma_{\text{th}} F_{\text{th}} + \frac{I_\gamma}{\ln(E_2/E_1)} F_{\text{res}} \right) = t_{\text{irr}} F_{\text{th}} \sigma_{\text{eff}}. \quad (6)$$

Here, F_{th} and F_{res} are the fluxes of thermal and resonance neutrons, respectively, and σ_{eff} is defined as follows:

$$\sigma_{\text{eff}} = \left(\sigma_{\text{th}} + \frac{I_\gamma}{c} \right), \quad (7)$$

$$c = \frac{F_{\text{th}}}{F_{\text{res}}} \ln(E_2/E_1), \quad (8)$$

where c is a constant for definite irradiation position in a definite reactor, E_1 and E_2 define the "resonance" range of energies, and for heavy nuclei the typical value is known to be

$$\ln \frac{E_2}{E_1} \approx 8\text{–}10. \quad (9)$$

The $F_{\text{th}}/F_{\text{res}}$ ratio should be below unity, at least, for the Dubna IBR-2 reactor, and the recommended value $c = 5$ can be realistic in our case.

Such an approach expressed in Eqs. (6)–(9) has the advantages that:

(i) the σ_{eff} value for the isomer can be deduced simply from the measured number of produced nuclei using the measured thermal neutron flux and

(ii) the isomer-to-ground-state ratio can be evaluated exploiting the tabular values of σ_{th} and I_γ for the production of the ground-state nuclei. Equation (7) allows one to calculate σ_{eff} for the ground-state products and to compare it with that determined in an experiment for the isomer.

Thus, the results in the present experiment were deduced for slow neutrons. The upper limits for (n,γ) and $(n,n'\gamma)$ production reactions are compared in Table 1 with the cross sections known from the literature.

In order to specify the resonance integral I_γ, we apply another method for processing the results. An additional sample was irradiated in the straight channel of the reactor (behind the neutron mirror) within a container surrounded with a CB_4 layer 3 mm thick.

The thermal neutrons were completely screened out, and low-energy resonances at $E_n \leq 10$ eV were suppressed too. After γ-spectra measurements, the number of ^{178m_2}Hf in the sample was restricted to the upper limit and the flux calibration was done by such products as ^{95}Zr, ^{175}Hf, ^{181}Hf, and ^{182}Ta. The resonance integral values for such products were corrected taking into account the neutron spectrum after the CB_4 filter. However, such a correction is not enough because in a Hf sample 1 mm thick the self-absorption near the resonance energies may be significant. The self-absorption can be neglected for ^{95}Zr and ^{182}Ta; it appears for ^{175}Hf and is significant for ^{181}Hf. Recall that ^{95}Zr is produced due to the small admixture (3%) of Zr in the Hf material, and ^{182}Ta is formed in the two-step capture process, the same as in the astrophysical s process: ^{180}Hf$(n,\gamma)^{181}$Hf $\beta^- \rightarrow ^{181}Ta(n,\gamma)^{182}$Ta. The ^{181}Hf yield is measured directly and the result includes the self-absorption factor. In the second reaction, the yield of ^{182}Ta is not influenced by self-absorption because of very low concentration of the ^{181}Ta nuclei in the target during the irradiation. Finally, the real flux values are estimated using ^{95}Zr and ^{182}Ta activities; both are in agreement, while the calibrations by ^{175}Hf and ^{181}Hf show a noticeable reduction of the flux. Using such quantitative measurements, we could estimate the resonance flux attenuation for the ^{177}Hf(n,γ) reaction, as well. Finally, the upper limits for I_γ and for the isomer-to-ground-state ratio were deduced. They correspond to the production of ^{178m_2}Hf with resonance neutron capture and are also given in Table 1.

It seems that [13] provided a higher sensitivity of measurements. But, as mentioned above, the burnup effect of the produced ^{178m_2}Hf could be strong at fluences above 10^{21} neutron/cm^2 and was probably underestimated in [13]. In addition, the resonance

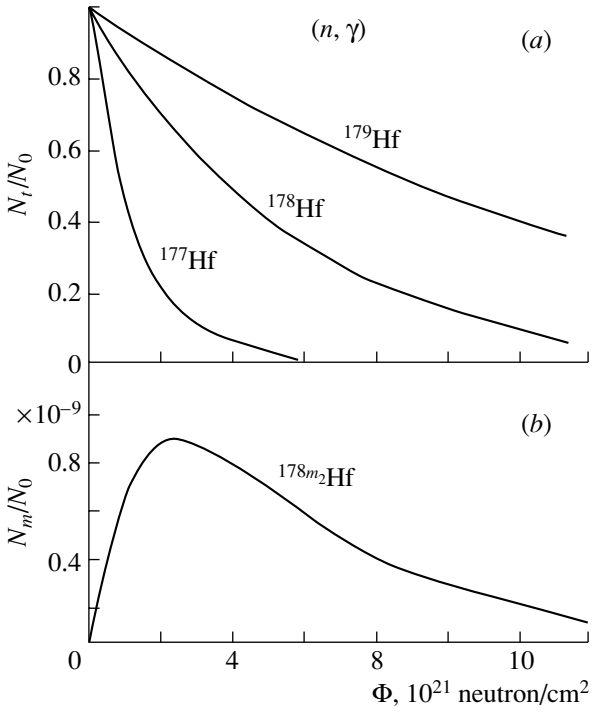

Fig. 2. Calculated transmutation functions for the Hf target stable isotopes in (n,γ) reaction (a) and the ^{178m_2}Hf accumulation (b). For the stable isotopes, burnup cross sections are taken from [17], and for the isomer, they are estimated using the results of [16]. The condition of irradiations corresponds to that in [13].

flux was neglected, and that could not be correct for irradiation inside any reactor. Even assuming the best thermalization, one has to use Eq. (7) for σ_{eff} with the numerical coefficient $c \approx 20$. On this basis, we recalculated the results [13] in more realistic approximations including the resonance neutron contribution and the burnup effect.

It would be clear that radiative neutron capture is the most destructive process, because the cross section σ_{eff} can be as high as thousands of barns. Respectively, in the fluence range of $\Phi \geq 10^{21}$ neutron/cm^2, isomer burnup can be manifested. Other nuclear processes in reactor irradiations are characterized by much lower cross sections, but due to neutron capture, even the feedstock (target) isotopes are in danger of useless depletion. The transmutation functions for stable Hf isotopes are calculated and shown in Fig. 2. σ_{th} and I_γ values are taken from [17], and they are combined to get the unified parameter σ_{eff} using Eq. (7) and numerical coefficient $c = 20$. The latter choice should be adequate to the conditions of the experiment in [13]. One can see that ^{177}Hf is the most unstable in neutron flux among other Hf nuclei.

In the presence of burnup of the target and isomer nuclei, the isomer accumulation function is expressed as follows:

$$N_m(\Phi) = \frac{N_0\sigma_{mp}}{(\sigma_{tb} - \sigma_{mb})} \times \{\exp(-\sigma_{mb}\Phi) - \exp(-\sigma_{tb}\Phi)\}, \quad (10)$$

where N_m and N_0 are the numbers of the isomer and target atoms, Φ is fluence, σ_{mp} is the isomer-production cross section, and σ_{mb} and σ_{tb} are the burnup cross sections of the isomer and target. In (10), we use again σ_{eff} quantities. The production and destruction σ_{eff} values are needed for the calculation of the accumulation function for ^{178m_2}Hf. One can estimate the destruction cross section using the results of the experiment in [16]. The values of $\sigma_{\text{th}} = 45$ b and $I_\gamma = 1070$ b were obtained for the partial branch of the $^{178m_2}\text{Hf}(n,\gamma)$ reaction with the population of the ^{179m_2}Hf isomeric state. Combined together with $c = 20$, they lead to $\sigma_{\text{eff}} \approx 100$ b. But the total destruction cross section should also include that for the (n,γ) branch leading to the ^{179g}Hf ground state. In a rough approximation, we took $\sigma_{\text{eff}} = 200$ b as the total destruction cross section for ^{178m_2}Hf due to (n,γ) capture.

With such a choice, the processing of the experimental results in [13] should be revisited, and finally, the production cross section is increased significantly. More intense destruction requires, respectively, a higher production probability to get the same number of produced atoms [see Eq. (10)]. The recalculated cross section is given in Table 1 and the corresponding σ_m/σ_g value is given as well. They characterize the (n,γ) reaction observed in [13] and evaluated in the present work under more realistic assumptions. In addition, the accumulation function is shown in Fig. 2 for the ^{178m_2}Hf isomer as is calculated in a similar approach with the same numerical parameters. One can see that, at fluences above 10^{21} neutron/cm^2, the accumulation curve deviates strongly from a linear function and then decreases. This is due to both the transmutation of the ^{177}Hf target nuclei and the burnup of accumulated ^{178m_2}Hf.

The experiment of the present work also does not promise a much higher yield of the ^{178m_2}Hf isomer. Finally, the conclusion follows that the reactor irradiation cannot serve as a high-efficiency method for ^{178m_2}Hf production.

3. SPALLATION BY INTERMEDIATE ENERGY PROTONS

It is established that the largest quantity of ^{178m_2}Hf was produced at Los Alamos with 800-MeV protons from a high-current accelerator (formerly LAMPF). The advantage of this method was the ability to

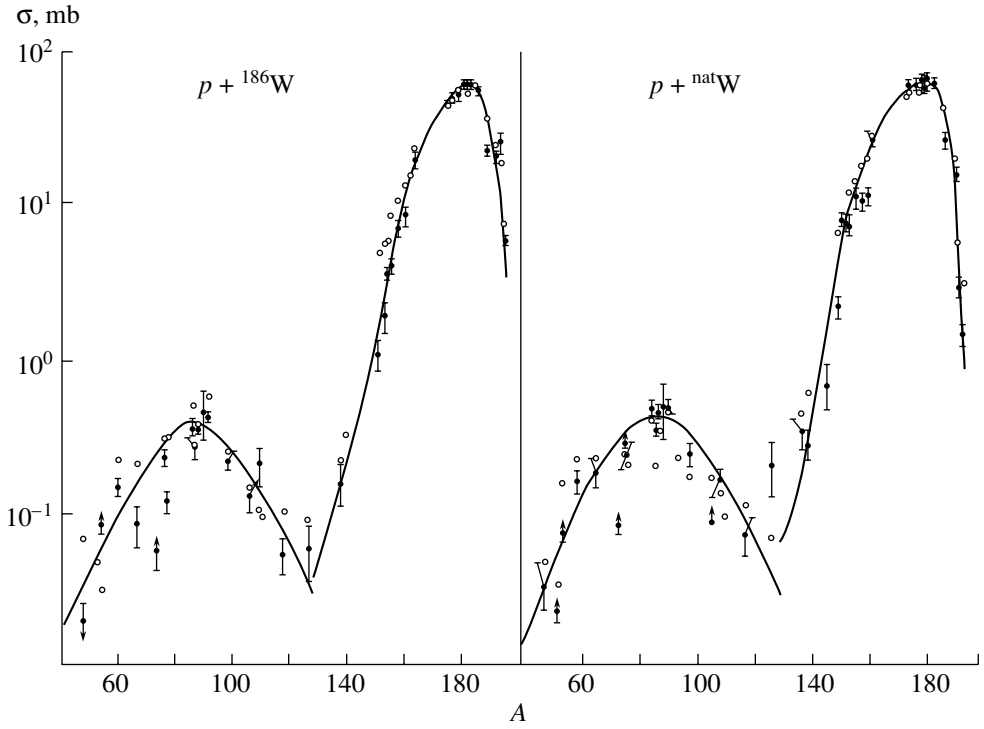

Fig. 3. Mass distribution of the nuclides produced after the irradiation of ^{186}W and ^{nat}W targets with protons of 630-MeV mean energy. Circles with error bars correspond to experimental results, while circles without bars correspond to the calculation using the LAHET code. The curves are used to guide the eyes.

accumulate the isomer as a by-product within a massive Ta beam dump during the operation of the accelerator for other experiments. The shortcoming was due to a very high activity of other radionuclides produced in Ta fragmentation. The yield of ^{178m_2}Hf was reported in [19], but the experimental details were described schematically and the productivity was only estimated. Recently, the reactions of proton-induced spallation were systematically studied for Ta, W, and Re targets of natural isotopic composition and for an enriched ^{186}W target ([12, 20]) using the 660-MeV Synchrocyclotron at Dubna. The yields of the long-lived high-spin isomers of ^{179m_2}Hf, ^{178m_2}Hf, and ^{177m}Lu are quantitatively determined and the measured values can be used for productivity optimization in some future irradiations.

Let us characterize briefly these recent experiments. Metal foil targets fixed to a cooled Al backing were inserted for irradiation into the internal beam of protons of the Dubna Synchrocyclotron (Phasotron). By choosing the position of a target inside the accelerator, it was possible to vary the beam energy from 100 to 650 MeV on the basis of known calibrations.

Gamma-spectroscopy measurements of the irradiated samples were performed after a "cooling" period of one month because of the high activity of short-lived radionuclides accumulated during the activation. After the measurements, the samples were dissolved for chemical processing and isolation of the Hf fraction. The decay activity of the long-lived ^{178m_2}Hf state is rather low as compared to the activity of other nuclides and chemical isolation was necessary to achieve good accuracy for the ^{178m_2}Hf yield. The gamma spectra were measured for the chemically isolated elemental fractions, in addition to the full activity spectra measured before.

In total, as many as about 70 radionuclides were identified, and as a result, the mass distribution of fragmentation products could be plotted. For example, it is shown in Fig. 3 for the case of the p + ^{186}W reaction at 630 MeV. Master tables of the radionuclide yields and cross sections are given in [12, 20] for different targets and energies. Many details of the γ-spectroscopic measurements, the calibration and evaluation procedures, etc., are also given there. Figure 3 characterizes here only basic properties common for the variety of studied fragmentation reactions. Two peaks in Fig. 3 correspond to the fission and spallation reaction mechanisms. For our purpose, most important are the production cross sections for nuclides and isomers near and below $A = 180$.

We focus, in the following, on the discussion of the spallation yield of the long-lived high-spin iso-

Table 2. Cross sections and isomer-to-ground-state ratios for the formation of high-spin isomers after the spallation of different targets at a proton beam energy of 650 MeV (some products with high gamma activity are also listed for comparison)

Nuclide	Target			
	^{nat}Ta	^{nat}W	^{186}W	^{nat}Re
Cross section σ, mb				
^{179m_2}Hf	0.52	0.36	0.80	0.12
^{178m_2}Hf	0.31	0.18	0.48	0.13
^{177m}Lu	0.15	0.13	0.26	0.04
^{178}W	5.9	23	21.8	36
^{175}Hf	56	55	55.6	59
^{172}Hf	47	53.5	57.4	55
^{173}Lu	61	61	60	61
σ_m/σ_g Ratio				
^{179m_2}Hf	0.040	0.14	0.25	0.24
^{178m_2}Hf	0.021	0.044	0.092	0.14
^{177m}Lu	0.103	0.21	0.29	0.40

mers of ^{177m}Lu, ^{178m_2}Hf, and ^{179m_2}Hf. The cross sections and isomer-to-ground-state ratios σ_m/σ_g are compared in Table 2 for the production of isomers with a 650-MeV proton beam using different targets. The cross sections of neighboring radionuclides with strong gamma activity are also reported in Table 2. They define the radioactive contaminations of the final isomeric material. It turns out that background isotopes are produced with very similar cross sections when different targets are used, while the productivity of isomers has large variations. As expected, the best material for isomer production is a ^{186}W target; its use leads to the increase in the cross section by a factor of about 2.5 is relation to the natural W target for the high-spin isomers of Hf and Lu. This is valid not only at 650-MeV beam energy but also at 450 MeV.

In Table 2, the σ_m/σ_g values of the order of 0.1–0.2 are the highest known in the literature for the production of the high-spin isomers. Consequently, the angular momentum of the spallation residues cannot be low; it is probably as high as $10\hbar$ or more. The dependence of the σ_m/σ_g values on the target mass number indicates a growth of the residual spin as the number of emitted nucleons increases. For instance, the total cross section for ^{178}Hf decreases from Ta to Re, while the σ_m/σ_g ratio, on the contrary, increases. With the ^{186}W target, the total cross section for the ^{179}Hf-, ^{178}Hf-, and ^{177}Lu-nuclide formation is higher than with Re, while the σ_m/σ_g ratio is much better than with the Ta target. Thus, enriched ^{186}W is the best target because both σ_m/σ_g and total cross section $(\sigma_m + \sigma_g)$ are optimal in this case. The experimental results [12, 20] also show that σ_m and σ_m/σ_g values are flatly reduced with decreasing proton energy. Such a weak dependence allows one to use the results in Table 2 for calculations of the isomeric yield in the case of thick targets when the proton-energy is noticeably degraded within the target. The highest productivity irradiation with intermediate-energy protons requires a massive target [19]. However, a few-kilogram amount of enriched ^{186}W material would be too expensive. Instead, let us estimate the maximum production yield of ^{178m_2}Hf in the regular Ta target. From Table 2, it follows that the cross section with Ta should be lower by a factor of 0.65 than with ^{186}W, but such a reduction is reimbursed due to the gain in the price of the material.

For estimation, assume now the most powerful proton beam of the Los Alamos accelerator with energy of 800 MeV and with intensity up to 1 mA. Passing through a 10-cm Ta target, a proton loses energy from 800 to 560 MeV. The ^{178m_2}Hf-formation cross section should be about 0.3 mb (Table 2). Thus, one can figure out immediately the yield value for ^{178m_2}Hf:

$$Y_{\max} \approx 10^{12}\ [\text{nucleus/s}]/1\ \text{mA}, \qquad (11)$$

i.e., about 10 mg of isomeric material per year of effective irradiations with the high-power beam. Such a brute-force approach is hardly realistic, but the estimated value is near the absolute maximum of production with the facilities for irradiations known in the literature. A factor of 2 gain can be reached assuming a target 20 cm thick and taking into account the reactions induced by the secondary particles. But this leads to a very large amount of Ta for chemical processing and to severe radiation safety conditions under irradiation and processing. At present, a cardinal increase in the production yield much above (11) is probably impossible, unless new facilities supply some fantastically high fluxes of accelerated particles, or (and) kilogram amounts of enriched exotic ^{180m}Ta are available as a target material.

4. REACTION WITH ^{4}He IONS AT LOW ENERGY

A method of ^{178m_2}Hf isomer production using the ^{176}Yb(^{4}He, $2n$) reaction was proposed and studied (see [21, 22]). A high-purity isomeric material was

accumulated in extensive irradiations with a high-current ^{4}He-ion beam at the Dubna U-200 cyclotron. The amount of isomeric substance was enough to prepare the targets for investigation of the nuclear reactions with the high-spin exotic isomer and some of them were successfully observed and studied. However, the general deficit in the amount of material has restricted the development of such studies. In total, it was possible to accumulate only about 1 μg ^{178m_2}Hf after high-intensity long irradiations with a ^{4}He beam. Absolute productivity is not high because the energy losses of ^{4}He ions in matter allow using only about 0.2 g of ^{176}Yb-target material and no more during each irradiation. Some details of this method are described below taking into account that it was productive for the performance of many experiments with ^{178m_2}Hf, reviewed in [16].

The excitation function of ^{178m_2}Hf in the ^{176}Yb(^{4}He, $2n$) reaction was measured and the cross section showed a peak near $E_\alpha = 32$ MeV. The optimum energy range of 28–36 MeV was deduced with a mean cross section of about 7 mb. Respectively, the mean isomer-to-ground-state ratio was estimated to be $\sigma_m/\sigma_g \approx 0.05$. The stopping of ^{4}He ions from 36 to 28 MeV in Yb_2O_3 corresponds to a layer 70 mg/cm^2 thick.

Targets made of superenriched $^{176}Yb_2O_3$ material were prepared and exposed to the ^{4}He-ion beam at long irradiations for ^{178m_2}Hf accumulation. The U-200 cyclotron at Dubna was modified for operation in the mode of high-intensity α-particle beam. The extracted beam current of 36-MeV ^{4}He^{++} ions reached 100 μA; i.e., the beam power was 1.8 kW. Because of the short range of the 36-MeV α particles, the volume density of power was high, and the Yb oxide layer could have been unstable under the beam.

The target construction was specially designed to prevent losses of the Yb oxide material; a scheme is shown in Fig. 4. The layer of Yb_2O_3 was pressed down onto the Al backing; thin Al foil above the layer was also applied to improve stability of the target. The Yb_2O_3 material was distributed over a larger area than the beam cross section, and the target was inclined to the beam direction. Thus, heat removal due to heat conductivity was improved. The Al backing was cooled down with the water flow from the rear side of the backing plate. Targets of such a design were able to resist a ^{4}He^{++}-ion current of 100 μA, while at higher intensities some losses of the Yb_2O_3-target material occurred.

The produced ^{178m_2}Hf material was chemically isolated after irradiation and then mass-separated (for more detail, see [21, 22]). Only one important remark concerns the purity of materials used in the target

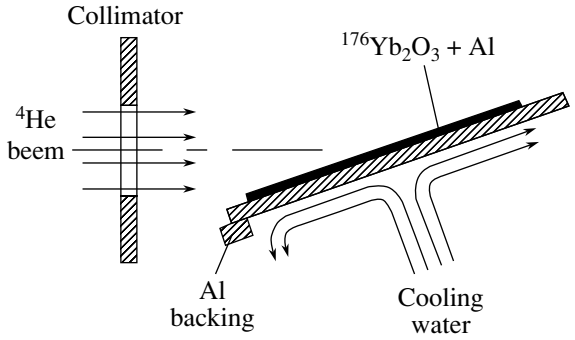

Fig. 4. Schematic drawing of the target design for irradiations of ^{176}Yb enriched material with a 36-MeV ^{4}He-ion beam of high intensity.

construction. To prevent some ballast ^{nat}Hf contamination in the isolated Hf fraction, it was necessary to take care about chemical purity of the target, backing, and chemical compounds used in the extraction of a product material. The commercially supplied 96% enriched ^{176}Yb was additionally enriched using an electromagnetic mass separator up to a purity of 99.998%. This expensive operation was needed for the suppression of the yield of the ^{175}Hf($T_{1/2} = 70$ d) and ^{172}Hf(1.87 yr) background activities that could disturb some experiments with the produced ^{178m_2}Hf samples.

Under the described conditions, the absolute yield of the ^{178m_2}Hf nuclei reached a value of

$$Y = 5 \times 10^8 \text{ [nucleus/s]}/100\ \mu\text{A}, \qquad (12)$$

which should be compared with the yield of the other reaction.

5. OTHER REACTIONS AT LOW ENERGIES

The relatively high cross section of the ^{176}Yb$(\alpha, 2n)^{178m_2}$Hf reaction leads to the idea of possible use of reactions like ^{181}Ta(p, α), ^{178}Hf(α, α'), ^{179}Hf$(\alpha, \alpha' n)$, and ^{176}Lu(^{7}Li, αn) at low energies. Production of ^{178m_2}Hf in these reactions has not been studied yet, but it is known from the nuclear-reaction phenomenology that all of them are more or less probable processes at energy well above the interaction barrier. It means that the total cross section should be hundreds of millibarns, and reasonably high angular momentum of the products provides a not very low isomer-to-ground-state ratio. Therefore, the ^{178m_2}Hf-production cross section is expected to be comparable with that known for the ^{176}Yb(^{4}He, $2n$) reaction, though not much more preferable.

Table 3. Quantitative parameters characterizing the different methods of ^{178m_2}Hf-isomer production

Projectile	E_{max}, MeV	Intensity	Target	Amount	Productivity, atom/s	σ_m, mb	σ_m/σ_g	Ref.	Rank
1	2	3	4	5	6	7	8	9	10
Photons, γ	22	100 μA*	^{179}Hf	10 g (total)	4×10^7	–	3×10^{-5}	[10]	6
	4500	100 μA*	Ta	33 g/cm^2	3×10^9	–	0.03	[11]	3
Neutrons, ^1n	Thermal	5×10^{14}/cm^2 s	^{177}Hf	1 g (total)	3.4×10^5	2×10^{-4}	0.5×10^{-9}	[13]	7
	14	10^{13}/cm^2 s	^{179}Hf	10 g (total)	2.5×10^9	7.3	3.5×10^{-3}	[15]	4
Protons, ^{1}H$^+$	650	100 μA	Ta	33 g/cm^2	2×10^{10}	0.3	0.02	[12]	1
	650	100 μA	^{186}W	5 g/cm^2	5×10^9	0.5	0.09	[20]	2
Alphas, ^{4}He^{++}	36	100 μA	^{176}Yb	0.07 g/cm^2	5×10^8	7	0.05	[21]	5

Note: σ_m is not given for the bremsstrahlung-induced reactions because of the continuous spectrum of photons. The yield ratio was measured and is given in the eighth column.

* Electron beam intensity.

Especially attractive is the ^{176}Lu(^{7}Li, αn) reaction because ^{176}Lu is a unique case of high-spin (7^-) target. Respectively, the σ_m/σ_g ratio can reach a level of 50% in this reaction, i.e., 10 times higher as compared to ^{176}Yb($\alpha, 2n$). But at the same time, the maximum current of the ^{7}Li ions is restricted due to the higher density of energy released in the target layer. In total, a factor of 3–5 can be the gain if one uses a high-current ^{7}Li beam and a 90% enriched ^{176}Lu target of the best design in the sense of heat removal. A few orders of magnitude higher productivity is not yet visible. Nevertheless, the reactions indicated above should be experimentally studied in order to operate with reliable results, instead of some realistic estimations.

6. COMPARISON OF DIFFERENT REACTIONS

In Table 3, the absolute productivities of the reactions induced by different projectiles are compared for the ^{178m_2}Hf isomer, following the measurements discussed above. The comparison is somewhat conventional, because the absolute yield depends on the beam intensity and on the appropriate amount of target material. Despite that, we want to get some ranking of reactions; therefore, they should be compared under similar conditions with respect to input parameters characterizing strength of the irradiation. For instance, the beam current is chosen to be 100 μA for all accelerators, and the same target thickness is assumed unless it is physically restricted due to flux absorption or price of the target material. The chosen parameters are absolutely real, i.e., already reached at the facilities described in the literature and remaining in operation today. No extraordinary powerful systems are involved in comparison. The quantities of enriched target isotopes are restricted to a value of 10 g because of the high price of such substances.

The ranks in Table 3 (column 10) reflect the absolute yield of the reaction (column 6) under comparable conditions. The p + Ta spallation is the most productive and its first rank could be expected. A productivity of 2×10^{10} atom/s is given in Table 3 as the best, but above, in Section 3, we discussed the much higher yield achievable in this reaction. There is no contradiction, because the absolute maximum has been estimated above assuming that the beam current can be as high as 1 mA with a target thickness of 10 cm. Even so, the production of ^{178m_2}Hf is restricted to milligram amounts, while effective applications require kilograms. The latter amount is beyond reality, at least at the modern status of experimental physics. Despite such orders of magnitude mismatch, the results reviewed in the present paper and summarized in Table 3 are of importance. They give a real basis for some speculations and estimations and also stimulate nuclear-science progress in understanding of the processes with high-spin nuclear states.

For nuclear-reaction theory, even more significant are the isomer-to-ground-state ratios, in addition to the production yields. In the σ_m/σ_g ratio, the scale factors in the reaction cross section are excluded, and the ratio value eventually has strong implications for study of the nuclear-reaction mechanism. In particular, mean angular momentum of the reaction residue has strong influence on the σ_m/σ_g ratio. Fortunately, the latter parameter is measurable, and for some reactions, the residual spin can be figured

out in theory. Thus, the correlation between σ_m/σ_g and the reaction-product spin can be verified after the measurements. Such a dependence is plotted in Fig. 5 for ^{178m_2}Hf isomer production. When the reaction product spin I_r increases, the spin-deficit parameter ΔI, respectively, decreases, and the probability of isomer population grows. Such natural behavior is experimentally confirmed and quantitatively characterized in Fig. 5.

It would not be easy to calculate in theory the $\bar{I}_r$ value for the reactions with intermediate-energy protons or with high-energy bremsstrahlung. In such cases, the systematics of Fig. 5 can be used for estimation of the $\bar{I}_r$ parameter based on the measured σ_m/σ_g ratio. In this way, unique information is deduced confirming that the reaction residue receives rather high spin, like $\bar{I}_r \sim 10\hbar$, both in proton-induced spallation and in the reaction of photon absorption at GeV energies. In addition, the systematics can be used in application to other processes for estimation of the production possibilities with reactions not yet studied.

At the end, let us discuss a somewhat fantastic idea of using the ^{180m}Ta material as a high-productivity target. Because of high spin (9^-) of this exotic nucleus, the ^{178m_2}Hf high-spin isomer can be produced in a spallation reaction with a much higher isomer-to-ground-state ratio. The productivity can be enhanced by a factor of 10 using such a target, as compared to the regular $^{\text{nat}}$Ta target.

This follows from the systematics of Fig. 5. However, a kilogram amount of 90% enriched ^{180m}Ta material is beyond reality today. Creation of a special facility for ^{180m}Ta separation and accumulation of it in large amounts should be extremely expensive, and even technical restrictions for that are not yet clear.

Ignoring the cost arguments, one can deduce the absolute maximum of productivity as:

$$Y_{\max} = 10^{13}\ \text{atom/s} \qquad (13)$$

if a 1-kg target made of 90% enriched ^{180m}Ta is exposed to 800-MeV protons at a beam current of 1 mA. Therefore, about 100 mg of ^{178m_2}Hf can be accumulated in a 1-yr effective irradiation run.

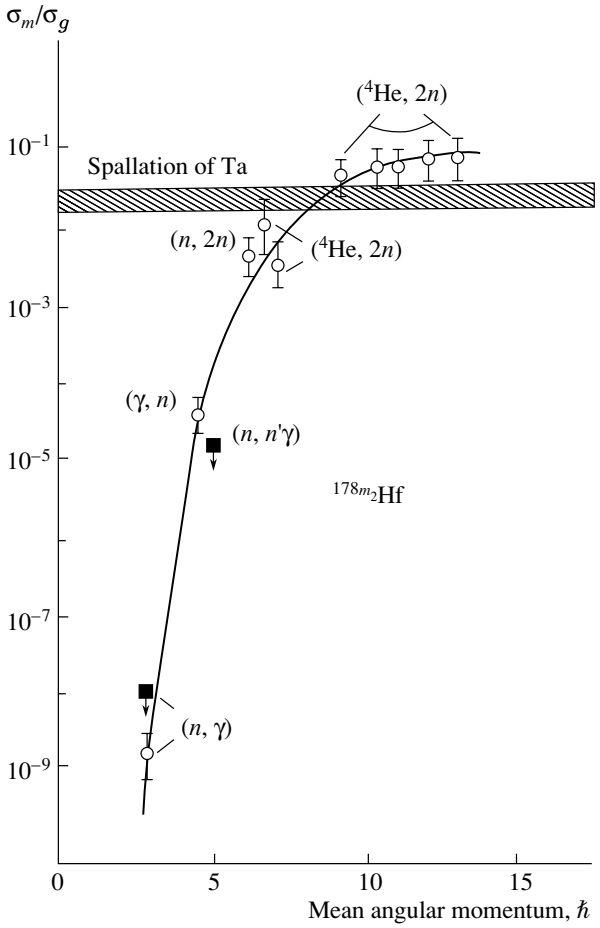

Fig. 5. Systematics of the isomer-to-ground-state ratio vs. the reaction-product spin for the ^{178m_2}Hf isomer as is measured in reactions with different projectiles.

SUMMARY

The known experimental results are reviewed for the production cross sections of the ^{178m_2}Hf exotic isomer. The productivity of different reactions is compared and they are ranked in order of decreasing yield. Respectively, the values are estimated for the amount of ^{178m_2}Hf material that can be accumulated in irradiations with different projectiles. Realistic parameters of existing experimental facilities restrict the production of large amounts, while the applications discussed in the literature require orders of magnitude higher quantities. A conceivable maximum of productivity is estimated under the assumption that the parameters of irradiations can be significantly enlarged using new facilities specially constructed for such irradiations and a new isotope separator for preparing a kilogram amount of ^{180m}Ta and ^{176}Lu isotopes.

The measured isomer-to-ground-state ratios are systematized as well, because they define the quality characteristics of the accumulated ^{178m_2}Hf material. In addition, such systematics is significant in the nuclear-reaction phenomenology and can be used for the prediction of productivity in the case of unstudied reactions.

Experiments on ^{178m_2}Hf isomer production could be carried out only within collaborations under definite financial support and the corresponding ac-

knowledgments are expressed in [10–12, 18, 20–22]. These results are used in the present paper.

REFERENCES

1. J. J. Carroll, Laser Phys. Lett. **1**, 275 (2004).
2. R. Coussement, R. Shakhmouratov, G. Neyens, and J. Ouders, Europhys. News **34**, 190 (2003).
3. P. M. Walker and G. Dracoulis, Nature **399**, 35 (1999).
4. J. J. Carroll, S. A. Karamian, L. A. Rivlin, and A. A. Zadernovsky, Hyperfine Interact. **135**, 3 (2001).
5. H. Roberts, Hyperfine Interact. **107**, 91 (1999).
6. S. A. Karamian, C. B. Collins, J. J. Carroll, and J. Adam, Phys. Rev. C **57**, 1812 (1998).
7. S. A. Karamian, C. B. Collins, J. J. Carroll, *et al.*, Phys. Rev. C **59**, 755 (1999).
8. S. A. Karamian, J. de Boer, Yu. Ts. Oganessian, *et al.*, Z. Phys. A **356**, 23 (1996).
9. Yu. Ts. Oganessian and S. A. Karamian, Hyperfine Interact. **107**, 43 (1997).
10. S. A. Karamian and J. J. Carroll, Laser Phys. **12**, 310 (2002).
11. S. A. Karamian, J. J. Carroll, J. Adam, and N. A. Demekhina, Preprint No. E1-2004-36, OIYaI (Joint Institute for Nuclear Research, Dubna, 2004); Nucl. Instrum. Methods Phys. Res. A **530**, 463 (2004).
12. S. A. Karamian, J. Adam, D. V. Filossofov, *et al.*, Nucl. Instrum. Methods Phys. Res. A **489**, 448 (2002).
13. R. G. Helmer and C. W. Reich, Nucl. Phys. A **211**, 1 (1973).
14. Yu. Weixiang *et al.*, Chin. J. Nucl. Phys. **14** (4), 326 (1992).
15. M. B. Chadwick and P. G. Young, Nucl. Sci. Eng. **108**, 117 (1991).
16. S. A. Karamian, Yu. Ts. Oganessian, J. Adam, *et al.*, in *Proceedings of the International School-Seminar on Heavy-Ion Physics* (World Sci., Singapore, 1998), p. 565.
17. S. F. Mughabghab, *Neutron Cross Sections* (Academic, New York, 1984), Vol. 1, Part B.
18. S. A. Karamian, J. J. Carroll, *et al.*, Laser Phys. **14**, 438 (2004).
19. H. A. O'Brien, Nucl. Instrum. Methods Phys. Res. B **40–41**, 1126 (1989).
20. S. A. Karamian, J. Adam, P. Chaloun, *et al.*, Preprint No. E6-2004-7, OIYaI (Joint Institute for Nuclear Research, Dubna, 2004); Nucl. Instrum. Methods Phys. Res. A **527**, 609 (2004).
21. Yu. Ts. Oganessian, S. A. Karamian, Yu. P. Gangrski, *et al.*, J. Phys. G **18**, 393 (1992).
22. Yu. Ts. Oganessian, S. A. Karamian, Yu. P. Gangrski, *et al.*, in *Proceedings of the International Symposium on Nuclear Physics of Our Times* (World Sci., Singapore, 1993), p. 521.

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1777–1786. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1839–1848.

NUCLEI
Experiment

Study of the *pd* Reaction at Ultralow Energies Using Hydrogen Liner Plasma*

V. M. Bystritsky[1)], Vit. M. Bystritskii[2)], G. N. Dudkin[3)], V. V. Gerasimov[1)], A. R. Krylov[1)], G. A. Mesyats[4)], B. A. Nechaev[3)], V. M. Padalko[3)], S. S. Parzhitsky[1)], F. M. Pen'kov[1)], N. A. Ratakhin[5)], and J. Wozniak[6)]**

Received June 21, 2004; in final form, December 28, 2004

Abstract—The *pd* reaction ($pd \to {}^3\mathrm{He} + \gamma$ (5.5 MeV)) is studied in the astrophysical energy collision range of protons with deuterons using the hydrogen liner in the inverse Z-pinch configuration at the pulsed power generator MIG (HCEI, Tomsk). Fundamental characteristics of this and other light-nucleus reactions at ultralow energies are important for problems of basic physics and astrophysics. The knowledge of the energy distribution of the nuclei participating in these reactions is important due to their exponential type of dependence on the collision energy. Two experimental techniques were designed and tested for recovering the energy distribution of liner protons incident on the CD_2 target by using optical detectors and ion collectors. It is shown that the combined use of these two techniques could provide relevant information on the energy distribution of the accelerated protons in the liner. The estimates of the upper limits for the astrophysical S factor and effective cross section of the *pd* reaction in the proton–deuteron collision energy range of 2.7–16.7 keV are obtained: $\bar{S}_{pd}(E_{pd} = 10.2\,\mathrm{keV}) \leq 2.5 \times 10^{-7}$ MeV b; $\tilde{\sigma}_{pd}(2.7 \leq E_{pd} \leq 16.7\,\mathrm{keV}) \leq 4 \times 10^{-33}$ cm^2. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Earlier experiments [1–3] carried out on the pulsed power generator SGM at the Institute of High Current Electronics in Tomsk (IHCE) with $I \approx 950$ kA, $\tau = 80$ ns [4] demonstrated the effectiveness of the inverse Z-pinch configuration [5, 6] to study reactions between light nuclei in the region of ultralow collision energies. This class of reactions is of special interest not only for testing fundamental symmetries in strong interactions but also for solving specific problems of astrophysics [7]. The mentioned study in the energy range of 1.8–3.7 keV, which is inaccessible for conventional accelerators, gave the first experimental estimates for the astrophysical S factor and effective cross sections of the *dd* reaction ($dd \to {}^3\mathrm{He} + n$) [1–3]. The use of the same technique at higher collision energies would allow one to compare it with calculated and experimental results obtained earlier at the same energy range with conventional accelerators and to estimate the method's potential for this application in general. To make it feasible, we need to increase the energy input, i.e., the current amplitude in the plasma liner, keeping its mass the same. For example, when upgrading the liner's current to ~2 MA, we could operate in the inverse Z-pinch configuration with particle collision energy in the range of 3–10 keV. It also would allow us to study quite a few nuclear reactions in the higher energy range with the values of cross sections substantially smaller than the respective values in the *dd* reaction, such as

$$pd \to {}^3\mathrm{He} + \gamma(5.5\,\mathrm{MeV}), \tag{1}$$

$$d^3\mathrm{He} \to {}^5\mathrm{Li} + \gamma(16.4\,\mathrm{MeV}), \tag{2a}$$

$$\to p(14.64\,\mathrm{MeV}) + \alpha(3.5\,\mathrm{MeV}), \tag{2b}$$

$${}^3\mathrm{He} + {}^3\mathrm{He} \to {}^5\mathrm{Li} + p(9.1\,\mathrm{MeV}), \tag{3}$$

$${}^3\mathrm{He} + {}^4\mathrm{He} \to {}^7\mathrm{Be} + \gamma(1.6\,\mathrm{MeV}). \tag{4}$$

In addition, in the inverse Z-pinch configuration with an expanding plasma flow, a reliable time discrimination between the background and "useful"

*The text was submitted by the authors in English.

[1)]Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
[2)]Department of Physics and Astronomy, University of California, Irvine, USA.
[3)]Scientific Research Institute of Nuclear Physics, Tomsk Polytechnic University, Tomsk, Russia.
[4)]Institute of Electrophysics, Russian Academy of Sciences, Yekaterinburg, Russia.
[5)]Institute of High-Current Electronics, Siberian Division, Russian Academy of Sciences, Tomsk, Russia.
[6)]Department of Physics and Nuclear Technology, AGH University of Science and Technology, Cracow, Poland.
**e-mail: bystvm@jinr.ru

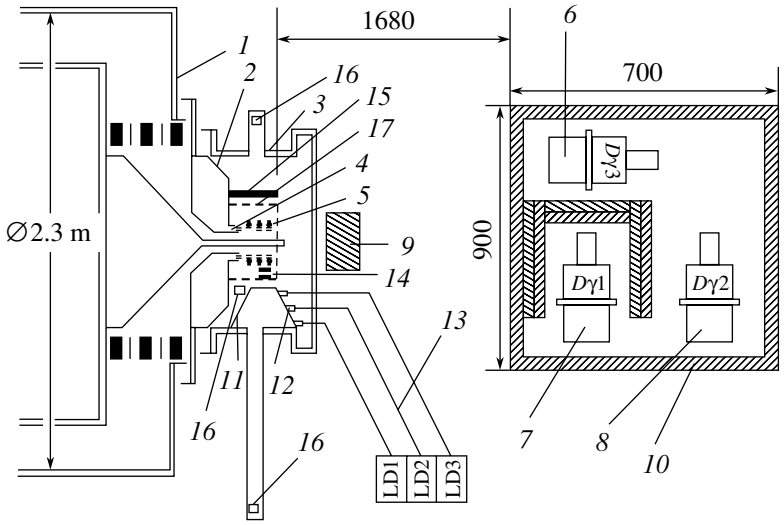

Fig. 1. Experimental layout: (*1*) generator MIG, (*2*) load module, (*3*) diagnostic chamber, (*4*) Laval nozzle, (*5*) liner, (*6*) scintillation detector $D\gamma 3$, (*7* and *8*) detectors $D\gamma 1$ and $D\gamma 2$, (*9*) suspended lead shielding, (*10*) lead shielding of detectors, (*11*) light cone, (*12*) collimators, (*13*) light guides, (*14*) magnetic dB/dt probes, (*15*) CD_2 target, (*16*) ion collectors, and (*17*) current-intercepting structure (CIS).

events (detection of the nuclear reaction products in the collision of plasma flow with the target) could be provided, in contrast with imploding plasma flow in a conventional Z pinch. The inverse Z-pinch technique could also enhance the feasibility of calculating the ion energy distribution due to a longer distance of outward plasma propagation to a target. It could be sufficient to provide ordering of the plasma flow in such a way that the energy of the ions would be uniquely coupled with their spatial position in the flow. In such a flow, the fastest ions are in the lead with the slow ions trailing behind. The calculations are based on the measured ion current densities and radiation of the plasma (in the waveband characteristic for liner ion species, for example, H_α waveband for hydrogen liner) by arrays of ion and optical fast detectors, respectively, placed at different distances from the axis of the Z pinch. It is worth noting here that information on the plasma ion energy distribution is particularly important due to exponential dependence of the considered reaction cross sections in the energy range of interest.

The reported research was focused on the estimation of

(i) the ratio of the experimental *pd* reaction yield to the background level of the γ detectors;

(ii) the upper limits for the astrophysical S factor and effective cross section of the *pd* reaction at ultralow proton–deuteron collision energy.

2. EXPERIMENTAL APPARATUS

The schematic of the experimental setup is shown in Fig. 1.

It was installed at the pulsed power generator MIG (IHCE, Tomsk) [8, 9] and consisted of the load module, deuterium target, γ-quantum detectors, and diagnostic equipment for inverse plasma liner formation and outward expansion, which included magnetic dB/dt probes, bolometers, ion collectors, and plasma radiation detectors in the optical range to measure the current sheath dynamics in the plasma liner at the stage of its acceleration, energy content of the plasma flow, ion current density, and radiation of the excited neutrals in the liner during its free expansion to the target, respectively.

Higher power and current levels of the PPG MIG [8, 9] ($\sim$2 TW, $I \geq 1.7$ MA), compared with the PPG SGM earlier used in [1–3], and a lighter gas of the liner (hydrogen instead of deuterium) made it technically more difficult to form a supersonic jet with a required mass per unit length. The Laval nozzle is located at the grounded electrode (anode) of the load module. In combination with the puff valve, the Laval nozzle generates a hollow supersonic hydrogen jet 25 mm long, with inner and outer diameters of 29 and 31 mm, respectively. The array of current-intercepting structure (CIS) has the shape of a squirrel cage cylinder 90 mm in diameter, fabricated of 1-mm-diameter by 30-mm-long wires. The CIS connects the HV wheel shaped electrode along its circumference with the grounded flange beyond the jet diameter. The load module installed at the MIG generator was earlier used in the experiments at the SGM generator [1–3] and its detailed description is given in [5, 6]. The waveform of the current pulse in the load module was measured with the Rogovsky

coil. At the generator current $I \geq 1.5$ MA, the inverse Z-pinch scheme provides transfer of the stored energy to the specific liner kinetic energy at a level of $E \approx 1.5$ kJ/cm, which corresponds to the average radial ion velocity of the expanding liner $V \sim 8.5 \times 10^7$ cm/s with the specific mass of the liner $m = 5.6$ μg/cm. The limitation of the specific liner mass from above imposes the lower limit for the average ion energy at 2.5–3.6 keV (at specified total energy stored in the liner). The measurements with magnetic probes showed that our design of the load module provided effective energy transfer from the storage unit inside the gas shell of the liner.

In shots with the hydrogen liner, the energy content in the radially diverging plasma flow was measured with a foil bolometer, whose electrode featured a bent copper foil strip 2 mm wide and 45 mm long. It was placed at a radial distance of 360 mm from the liner axis. In shots, the bolometer measured the average deposited specific energy of the order of $\sim$1 J/cm^2. With assumption of azimuthal uniformity of the plasma flow, these measurements provided data for inventory of the accelerated ions N_p bombarding the target, $N_p = N_p^0 k_1$ (here, k_1 is the geometry factor, N_p is the number of ions falling on the target, and N_p^0 is the total number of ions in the plasma liner when it passes CIS). Based on the 0D calculations and B-dot probe data on the plasma motion inside the CIS, we can conclude that nearly 40% of the liner energy is absorbed by the target. The measured energy density agrees with the results in the previous experiments at the SGM generator [1–3].

An alternative method of calculating the total number of particles falling on the target is based on direct use of the bolometric data and average energy of the ion as $N_p = E_L/\bar{E}_p$. Here, E_L is the liner energy, and the $\bar{E}_p$ is the average ion energy. They are defined from the energy distribution curve.

The deuterium target featured three sections of a hoop fabricated from a 45-mm-wide copper strip, placed at a distance of 40 cm from the axis. The inner surface of the target was covered with CD_2 disks 0.25 mm thick and 40 mm in diameter. The total area of all disks was 250 cm^2, which made up 22% of the total target area.

Three γ-quantum detectors with sizes of $\varnothing 160 \times 210$ mm ($D\gamma 1$ and $D\gamma 2$) and $\varnothing 50 \times 50$ mm ($D\gamma 3$) were used for detection of γ quanta from reaction (1) and placed at a distance of 180 cm from the liner and were surrounded with a 10-cm-thick Pb shield in the back and at the sides and a 5-cm-thick one in the front.

As follows from our earlier studies of the *dd* reaction in the inverse Z-pinch configuration [1–3] at the high-current SGM generator, each shot of the generator generated two rather large amplitude pulses (separated in time) from the neutron detectors, based on organic plastic scintillators. The first pulse, as was pointed out in [1–3], arises from the bremsstrahlung radiation, and the second pulse arises from the background neutrons generated in the *dd* reaction inside the plasma in the course of liner $I \times B$ acceleration. We were able to separate in time the "useful" events corresponding to the interaction of the deuterium plasma with the deuterium target from the neutrons of background origin by placing the deuterium target and the neutron detectors at a sufficiently large distance from the CIS, or compared with the neutron yield in the shots with a bare target without CD_2 coating.

In the case of *pd* experiments, there is only one possibility to provide a reliable time separation of two-event detection of the bremsstrahlung and of the accompanying 5.5-MeV γ quanta from the target—by organizing the distance between the deuterium target and the CIS to be big enough to meet the signal-to-noise ratio requirement. In our experiment, the time duration of the intense bremsstrahlung radiation during plasma liner crossing of the CIS was $\sim$120 ns, and the time of the expanding plasma arrival at the deuterated target placed at a distance of 40 cm was $\sim$300–600 ns; i.e., the time pause between two pulses was much longer than a typical pulse formation time of the fast plastic γ detectors ($\tau \approx 5$ ns). These detectors significantly increased the accuracy of the time resolution compared with slow NaI(Tl) scintillators used in the earlier reported *dd* experiment. Though the γ-detection efficiency of the plastic scintillators is significantly lower than that of the NaI(Tl) or CsI(Tl), the former provided a reliable time discrimination of useful and background events. Using plastic γ detectors, we were able to discriminate by the amplitude, time of arrival, and duration of the background pulse from the 5.5-MeV γ quanta pertinent to the *pd* reaction.

To suppress the background coupled with the bremsstrahlung in $D\gamma 1$ and $D\gamma 2$, the HV was applied to the photocathode–modulator gaps of their PMT with a controlled time delay after HV pulse arrival at the load. Such a technique reduced by a factor of 10^3 the background pulse amplitude at the PMT output. The required minimum time delay of the PMT was determined as plasma acceleration phase duration plus its travel time from the CIS to the target. In our experimental conditions, it equaled $\sim$360–400 ns. The linearity of $D\gamma 3$ operation in the heavy-background environment was provided by attenuation of the PMT (XP-2020) multiplication factor. This was done by using the PMT output from the tenth dynode instead of the anode.

Table 1. Main operational characteristics of the hydrogen liner

Shot	I_L, MA	m_L, μg/cm	N_p^0, 10^{18}	V_I, 10^7 cm/s	t_{CI}, ns
1	1.1	4.5	6.78	6.8	126
2	1.5	2.4	3.62	11.0	92
3	1.4	4.7	7.08	8.0	114
4	1.5	4.6	6.93	8.5	110
5	1.5	8.3	12.51	6.8	126
6	1.5	2.4	5.12	9.6	100
7	1.5	6.0	9.04	8.1	112
8	1.6	8.2	12.36	7.3	120
9*	1.6	6.0	9.04	8.1	112
10	1.5	5.6	8.44	8.3	111

Note: I_L is the liner's current, m_L is the mass per unit length of the liner, N_p^0 is the number of H^+ ions in the liner, V_I is the radial velocity of the current sheath, t_{CI} is the time interval for the current sheath to arrive at CIS. Shots 1–8 were fired without a target, shot 9* was fired with only a copper strip for a target, and shot 10 was fired with a combined target consisting of CD_2 thin disks placed on the inner surface of the copper strip.

Prior to and during the experimental runs, γ-detector's outputs were periodically calibrated with standard ^{60}Co, ^{88}Y, and ^{234}Th γ-quantum sources.

The detection efficiencies of Dγ detectors for 5.5-MeV γ quanta computed by the Monte Carlo method with the selected threshold of detection on the level of 1 MeV gave values of 5.5×10^{-5} and 8.8×10^{-7} for $D\gamma 1$, $D\gamma 2$, and $D\gamma 3$, respectively.

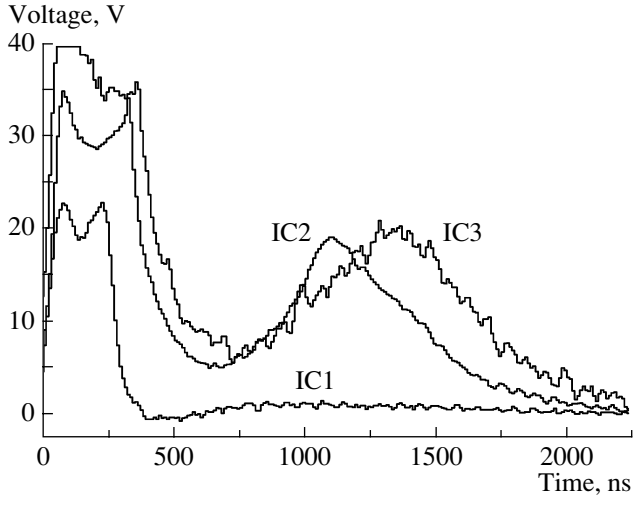

Fig. 2. Waveforms of the ion currents measured with three ion collectors IC1–IC3 in shot 10. The time origin is the high-voltage pulse at the load module of the MIG generator.

Information on the proton energy distribution in the expanding plasma flow was recovered from the ion current amplitude-time waveforms, recorded by several ion collectors (IC), placed at the radii $r_1 =$ 25 cm (IC1), $r_2 = 124$ cm (IC2), and $r_3 = 152$ cm (IC3) relative to the position of the CIS. Such a number of ICs were adequate to perform a variety of tests for optimization of the recovery process of the ion energy distribution. The criteria of applicability of this technique are listed below:

The plasma liner after the end of the acceleration stage expands in the state of a local thermal equilibrium.

All ions in the liner before they reach the positions of the ICs must form a spatially "ordered" flow in which the fastest ions lead the front and the slowest trail behind; thus, the energies of the ions are uniquely determined by their spatial positions and, respectively, by arrival time at the ion collector.

The total rate of all processes leading to recombination and deexcitation of the ions (such as collisions of protons with atoms, molecules, and molecular ions and ion–electron collisions) must be independent of or at least weakly dependent on the kinetic energy of ions in the considered energy range. With this condition realized, the ions' energy distribution in the flow stays pretty constant during expansion. Respectively, the cooling of the plasma also does not change the shape of the respective IC waveforms, resulting only in a decrease in the waveform's amplitude as a proportional addition to its attenuation due to the plasma flow spatial divergence.

The listed conditions in the present experiment were fulfilled, the shape of the IC waveforms at different radii remained similar and could be adequately normalized, accounting for IC signal attenuation with distance mainly due to the spatial expansion.

Table 1 presents the main operational characteristics of the hydrogen liner measured in ten shots at the MIG generator after testing the inverse Z-pinch formation diagnostic equipment.

Figure 2 shows the ion current waveforms from three ion detectors IC1, IC2, IC3 in shot 10. The high-voltage pulse of the MIG generator was used as a synchronizing trigger. The structure of the waveforms can be treated as follows. Its initial part in the time interval 0–35 ns is due to the photoelectron emission from the IC caused both by the bremsstrahlung arising from the operation of the

generator[7)] and by the ultraviolet radiation of the liner in the course of its radial acceleration (the peak near the boundary of the time interval mentioned). The other rather wide peak in the time interval from 600 to 2000 ns is due to detection of the ion flow. It is evident that the time separation of signals related to photoemission and ion flow becomes more evident with increasing distance between the CIS and ICs. The estimate shows that, with $l \geq 1$ m, the timing of the ion flow arrival can be measured with good accuracy.

The IC2 and IC3 signal waveforms were analyzed to recover the energy distribution of the protons incident on the target.

Ignoring the radial size of the plasma flow in the region of the CIS and taking for the origin $t = 0$ the moment of time when the current sheath passes through the CIS (which was calculated with a 0D model starting from arrival of the HV pulse at the load), we can restore the energy distribution of the protons in the liner solving the following system of equations:

$$\frac{dN_p}{dE_p} \sim I(E_p)E_p^{-3/2}, \tag{5}$$

$$\frac{dN_p}{dt} \sim I(t), \tag{6}$$

$$I(t) = I\left(E_p = \frac{m_p l^2}{2t^2}\right). \tag{7}$$

Here, dN_p/dt and dN_p/dE_p are the time and energy distributions of liner ions incident on the collector, m_p and E_p are the mass and energy of liner protons, l is the distance between the CIS and IC, and $I(t)$ is the measured IC current.

Figure 3 shows the ion energy distributions corresponding to the time distributions shown in Fig. 2. These distributions are derived from the ion current–time relations measured with the IC2 and IC3 detectors and transformed using (5)–(7). The IC1 data were not analyzed because of a rather high background (photoelectron emission) during the time interval of the proton flow arriving at it. As is evident from Fig. 3, the main group of accelerated protons has quite a wide energy distribution (3–10 keV). It should be mentioned that the current sheath velocity measured by the magnetic B-dot probes (when combined with a 0D model) is near the lower limit of the above energy range. The data obtained with the IC2 and IC3 placed at distances of 124 and 152 cm, respectively,

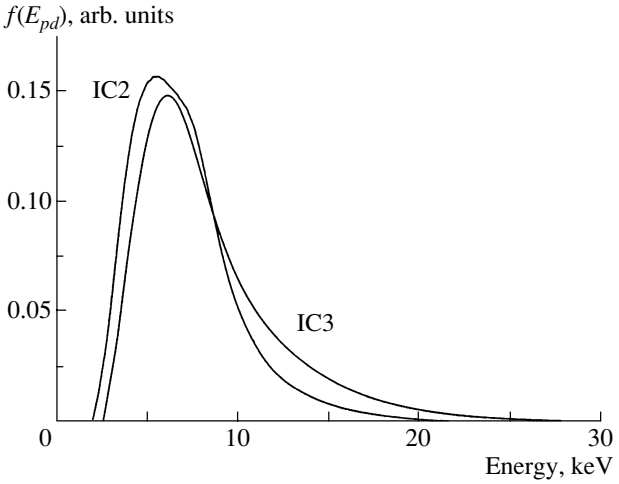

Fig. 3. Energy distributions of liner protons in shot 10 based on the IC2 and IC3 ion current waveforms shown in Fig. 2.

show the presence of a minor proton fraction with an energy as high as ~15–20 keV compared to the energy of the main group. The velocity of such protons is $(2.0–2.3) \times 10^8$ cm/s. More precise information on these high-energy proton energy spectra is necessary to carry out measurements with ICs placed at distances more than 1.5 m from the CIS. In this case, it would be feasible to separate in time the IC current pulse from this proton fraction on the background of the UV radiation pulse of the liner.

The recovery of the proton energy distribution in the diverging plasma flow was also done by a parallel diagnostics based on measuring radiation of excited neutrals H* in the region of H_α wavelength. Three optical radiation detectors LD1–LD3 were used to this end (named elsewhere in the text as "LD" technique, or method). The LDs (see Fig. 1) included a collimator 40 mm long (the aperture diameter of the collimator was 0.8 mm), a quartz fiber 1 mm in diameter and 7 m long, H_α filter, and PMT. The light detector LD1 was placed at a variable radial distance of 75–95 mm from the CIS. The distances between LD1 and LD2 and between LD2 and LD3 were equal to 50 mm each. A detailed description of the method of time waveform signal from light detectors of transformation to the proton energy distribution is presented in [3, 6].

Figure 4 shows waveforms of signals from the optical detectors LD1–LD3 and Fig. 5 shows the corresponding energy distributions of liner protons obtained by the LD method of transformation in shot 10.

3. ANALYSIS AND DISCUSSION OF THE RESULTS

In shots 1–8, both methods for recovering the proton energy distribution were tested separately and

[7)] The bremsstrahlung is due to the electron losses in the load module and in the vacuum transmission convolute during energy flow transition to the self-magnetically insulated regime.

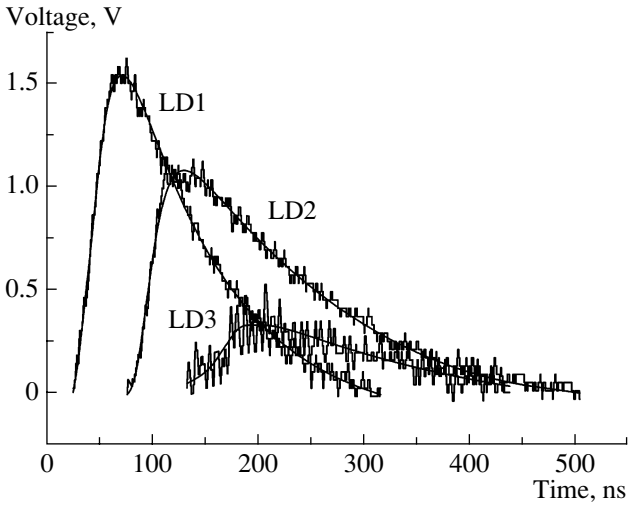

Fig. 4. Waveforms of signals from optical detectors LD1–LD3 in shot 10.

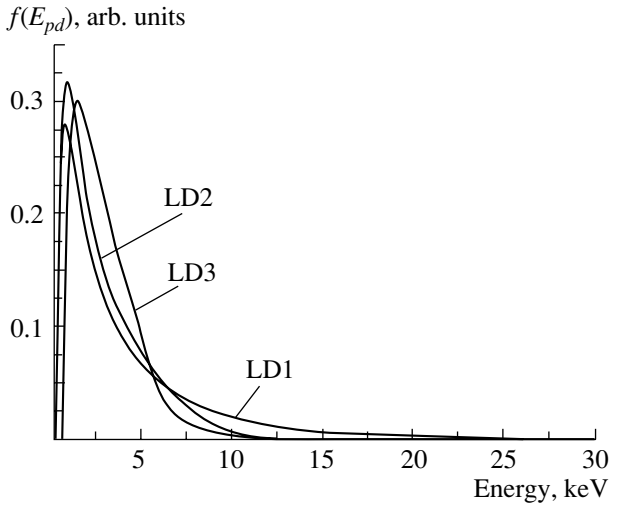

Fig. 5. Energy distributions of the liner protons based on the waveforms from optical detectors LD1–LD3 shown in Fig. 4.

in combination. Shots 9 and 10 were done to measure the γ-background level in the γ detectors and to compare it with the expected γ yield from the *pd* reaction. We also have done the estimation of the upper boundary for the astrophysical S_{pd} factor and the effective *pd* cross section. Shot 10 was taken with the deuterated target placed inside the load module, and shot 9* was taken without the target. Table 2 presents the results of analysis of the waveforms from the light detectors LD1–LD3 and the ion collectors IC1–IC3 obtained in shots 9* and 10. Figure 6 illustrates the energy distributions of the liner's protons in shot 10 based on the averaged optical data from LD1–LD3 and IC1–IC3, respectively. There is a noticeable difference between these two curves. The curves approximately match in the region of 3–4 keV with differences building up at both ends of the ion spectra.

This is likely to be caused by the violation of the conditions listed in Section 2 which define the applicability of the methods used. Specifically, as was already mentioned earlier, the violation could be connected with the cooling of the plasma flow during its expansion, with resulting nonconservation of the relative ion concentrations of different energy groups. The latter phenomenon pertains to the net decrease in the total H^+ recombination rate with the increase in their kinetic energy. Indeed, the charge-exchange cross sections σ for H^+ in the collisions with residual O and N drop from 5.9 to 4.3×10^{-16} cm^2 and from 4.9 to 3.3×10^{-16} cm^2, respectively, with the energy rise from 3 to 10 keV [10]. The H^+ recombination rate in atomic H also drops via electron capture collisions from 3×10^{-15} to 5×10^{-16} cm^2 with the energy rise from 0.1 to 10 keV [11]. At the same time, as was already discussed earlier, the cross sections for electron capture in the collisions of H^+ with N and O increase. It seems that the former phenomenon prevails, and, as a result, the maximum of the relative concentration of the excited neutrals in the flow correlates with the maximum of the lower energy ion group, whereas the maximum of the ion concentration corresponds to the higher energy ion group. Accounting for the fact that the plasma flow during its expansion is being "ordered," the highest concentration of the excited H* neutrals trails at the back front of the plasma flow, and the highest concentration of the ions travels in the front part of the flow.

Besides this, the distances of the LDs from the CIS in experiment were located at distances marginally or somewhat smaller than the respective values required by the named criteria for the ions belonging to a higher energy group.

Thus, the energy distribution recovered on the basis of the IC data would shift upward the ion energy distribution, whereas the recovered energy distribution from the LDs data would shift downward the distribution compared with the real situation.

It is also worth noting another mechanism—ambipolar ion diffusion at the front of the plasma flow, when the latter expands in vacuum. It is akin to collective ion acceleration in the electron beams [12]. This phenomenon boils down to formation of a virtual cathode at the characteristic distance of $\sim c/\omega_{\rm pl}$ (here, $\omega_{\rm pl}$ is the Lengmuir plasma frequency) ahead of the front with generation of a local axial electric field and respective collective acceleration of a small group of ions (1–2% of the total) to an energy a factor of 2–3 higher compared with the energy of the main ion bulk in the flow [12]. It agrees with

detection of such a high-energy small group of ions in the experiment.

The next stage of experimental data analysis included processing of the waveforms from the γ-quantum scintillation detectors $D\gamma1-D\gamma3$ in shots 9* and 10. Figure 7 gives the signals waveforms from the γ detectors $D\gamma1-D\gamma3$ in shot 10. Arrows indicate the boundaries of the calculated time interval corresponding to detection of "useful" events. As is evident from Fig. 7, there is a shift in the position of the time interval for $D\gamma1$, $D\gamma2$, and $D\gamma3$. This time interval shift is due to the different intrinsic delays in the spectrometric channels of the detectors $D\gamma1$, $D\gamma2$, and $D\gamma3$.

The following is the algorithm and analysis of the γ-detector data. It provided the estimates of the main pd-reaction characteristics, the astrophysical S factor, and the effective cross section. Since at the present time we do not have a final recipe for "matching" proton energy spectra recovered with ICs and LD, we analyzed γ spectra $D\gamma1-D\gamma3$ data separately for each of them.

In these shots, no γ quanta were detected in the time intervals corresponding to the pd reaction. The indicated time intervals included in each shot the liner acceleration phase (between HV pulse arrival at the load module and the moment of the liner passage through the CIS), the time interval it takes for the liner to cover the distance from the CIS to the target, with the statistical deviation in the arrival moment due to the proton energy spread. The arrows in Fig. 7 show the calculated time interval for detection of γ quanta from the pd reaction in shot 10, corresponding to the proton energy interval of 4.1 keV $\leq E_{pd} \leq$ 25.3 keV. It was calculated from the LD data. The similar energy interval based on the IC data gave 5.6 keV $\leq E_{pd} \leq$ 25.0 keV, which is quite near to the former.

The experimental information on the astrophysical S_{pd} factor and the effective pd reaction cross section was deduced by measuring the yield of 5.5-MeV γ quanta and using the parametrical dependence of the reaction cross section on the proton–deuteron collision energy

$$\sigma_{pd}(E_{pd}) = \frac{S(E_{pd})}{E_{pd}} e^{-2\pi\eta}. \qquad (8)$$

Here, E_{pd} is the proton–deuteron collision energy in the c.m.s.; $2\pi\eta = (\mu/E'_{pd}(E_{pd}, x'))^{1/2}$; η is the Sommerfeld parameter; $\mu = 2/3$ in amu is the reduced mass of the pd system; $E'_{pd}(E_{pd}, x') = (E^2 - \pi n_t e^4 L x')^{1/2}$ is the pd collision energy after the proton has passed a deuterium sheath in the target of thickness x' (provided that the initial collision energy is E_{pd}); and L is the Coulomb logarithm for the deuterium plasma: $L = 12.8$ [13].

Table 2. Characteristics of the energy distribution of the liner protons recovered via combined analysis of the IC and LD data

	Shot 9*		Shot 10	
	LD	IC	LD	IC
E^p_{pd} [keV]	0.9	7.4	1.1	5.9
V^p_{pd} [10^8 cm/s]	0.5	1.2	0.6	1.1
$\bar{E}_{pd}$ [keV]	3.6	10.6	3.4	7.8
$\bar{V}_{pd}$ [10^8 cm/s]	0.8	1.4	0.8	1.2
E [keV]	4.5	5.8	3.8	4.1

Note: E^p_{pd} and $\bar{E}_{pd}$ are the most probable energy and the average energy of the liner protons after passing through the CIS; V^p_{pd} and $\bar{V}_{pd}$ are the most probable velocity and the average velocity of the protons; "$*$" marks a shot without a deuterium target in the measuring chamber of the generator. All quantities are in the lab. system.

In this case the average S_{pd} factor and the effective pd-reaction cross section $\tilde{\sigma}^{\exp}_{pd}$ are defined as [1–3, 14, 15]

$$\bar{S}_{pd}(E_{\rm col}) = \frac{Y^{\exp}_{\gamma}}{N_p n_t \varepsilon_\gamma \int\limits_0^\infty f(E_{pd}) dE_{pd} \int\limits_0^\infty dx' e^{-2\pi\eta}/E'_{pd}(E_{pd}, x')}, \qquad (9)$$

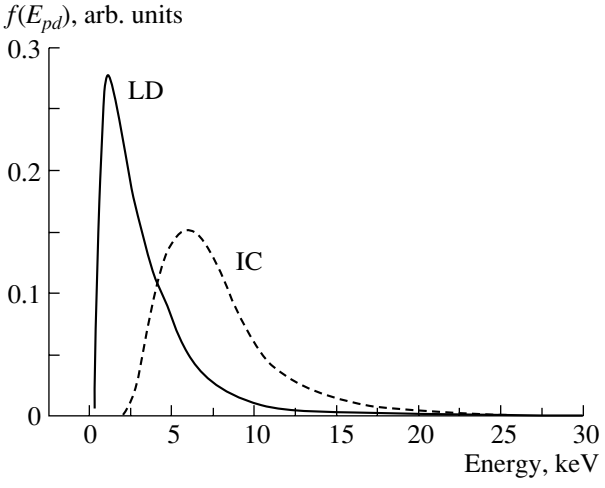

Fig. 6. Averaged energy distributions of the liner protons measured in shot 10 based on LD and IC data. The distributions are normalized to unity.

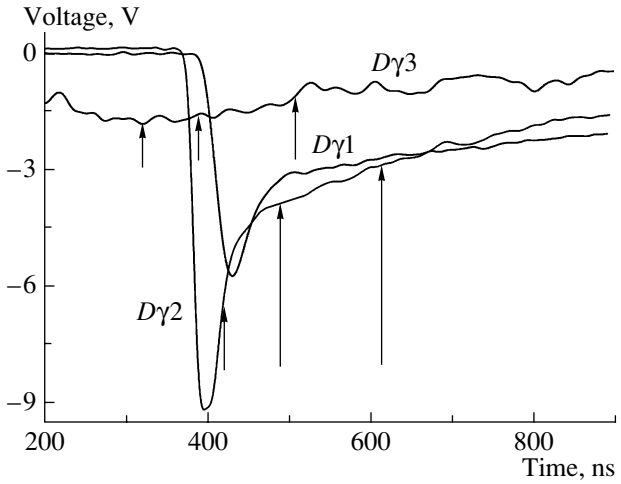

Fig. 7. Waveforms from the plastic γ detectors $D\gamma1$–$D\gamma3$ in shot 10. The solid curves are calculations. The time origin is the time when the HV pulse appears at the load module.

$$\tilde{\sigma}_{pd}^{\exp}(\Delta E_{\rm col}) = \frac{Y_\gamma^{\exp}}{N_p n_t \varepsilon_\gamma \tilde{l}},$$

$$Y_\gamma^{\rm col} = \bar{S}_{pd}(E_{\rm col}) N_p n_t \varepsilon_\gamma \tag{10}$$
$$\times \int_0^\infty f(E_{pd}) dE_{pd} \int_0^\infty \frac{e^{-2\pi\eta}}{E'_{pd}(E_{pd}, x')} dx',$$

$$\tilde{\sigma}_{pd}^{\rm cal}(\Delta E_{\rm col}) = \frac{Y_\gamma^{\rm col}}{N_p n_t \bar{S}_{pd}(E_{\rm col})},$$

$$E_{\rm col} = \int_{E_{pd}} E_{pd} P(E_{pd}) dE_{pd}, \tag{11}$$

$$P(E_{pd}) = \frac{e^{-2\pi\eta} D(E_{pd}) \int\limits_{E_{pd}}^\infty f(E_{pd}) dE_{pd}}{\int\limits_0^\infty e^{-2\pi\eta} D(E_{pd}) dE_{pd} \int_{E_{pd}}^\infty f(E_{pd}) dE_{pd}}, \tag{12}$$

$$D(E_{pd}) = -\frac{1}{E_{pd}} \frac{dx}{dE_{pd}}, \tag{13}$$

$$\frac{dE_{pd}}{dx} = -\frac{\pi n_t e^4 L}{2E_{pd}}. \tag{14}$$

Here, N_p^0 and N_p are the numbers of liner protons injected in the vacuum anode–cathode gap and hitting the CD_2 target, respectively; $Y_\gamma^{\exp}$ and $Y_\gamma^{\rm col}$ are the total experimental and calculated (at an average value of the S_{pd} factor equaling 1.2×10^{-7} MeV b [16–18]) yield of the detected γ from the pd reaction with the calculated particle energy distribution and incident on a target with infinite thickness; $E_{\rm col}$ is the average pd collision energy in the c.m.s. corresponding to the distribution function $P(E_{pd})$; $P(E_{pd})$ is the pd collision energy distribution function normalized to unity for the γ-yield probability in the pd reaction; $\Delta E_{\rm col} \equiv [E_{\rm col}^{\min}, E_{\rm col}^{\max}]$ is the pd collision energy interval; $E_{\rm col}^{\min}$ and $E_{\rm col}^{\max}$ are the minimum and maximum limits of the above pd collision energy interval defined as

$$E_{\rm col}^{\min} = E_{\rm col} - 2\sigma_{P(E_{Pd})},$$

$$E_{\rm col}^{\max} = E_{\rm col} + 2\sigma_{P(E_{Pd})},$$

where $\sigma_{P(E_{pd})}$ is the rms deviation of the pd collision energy corresponding to the $P(E_{pd})$ distribution (see Table 3); $N_p = N_p^0 k_1 k_2$ is the number of protons in the supersonic hydrogen jet (liner) injected in the vacuum anode–cathode gap, where k_1 is the fraction of the liner protons hitting the target located at a distance of 40 cm from the liner axis (according to the bolometric measurements, $k_1 = 0.4$ for shots 9* and 10), and $k_2 = 0.22$ is the fraction of the copper plate area coated with CD_2 0.25 mm thick; n_t is the deuteron density in the target; $f(E_{pd})$ is the proton energy distribution; ε_γ is the γ-quantum detection efficiency of the detectors $D\gamma1$–$D\gamma3$; dE_{pd}/dx are the specific Coulomb energy losses in the deuterated target [13]; $\tilde{l}$ is the effective thickness of the deuterated target for the real proton energy distribution, which is such that

$$Y_\gamma(\tilde{l}) = 0.9 Y_\gamma^{\rm col}, \tag{15}$$

$$Y_\gamma^{\rm col} = N_p n_t \bar{S}(E_{\rm col}) \tag{16}$$
$$\times \int_0^\infty f(E_{pd}) dE_{pd} \int_0^\infty \frac{e^{-2\pi\eta}}{E'_{pd}(E_{pd}, x')} dx',$$

where $Y_\gamma^{\rm col}$ is the calculated total pd reaction yield corresponding to the measured astrophysical factor $\bar{S}_{pd}(E_{\rm col})$ and given energy distribution of protons hitting the target of infinite thickness.

It should be mentioned that the average astrophysical factor of the pd interaction $\bar{S}_{pd}(E_{\rm col})$ is defined as

$$\bar{S}(E_{\rm col}) = \int_{E_{\rm col}^{\min}}^{E_{\rm col}^{\max}} S(E_{pd}) P(E_{pd}) dE_{pd}. \tag{17}$$

Figure 8 shows normalized distributions $P(E_{pd})$ obtained by IC and LD methods for the interval of pd collision energies 2.7 keV $\le E_{pd} \le$ 16.7 keV. Though

Table 3. Results of the experimental data analysis

	Shot 9*		Shot 10	
	LD	IC	LD	IC
$E^p_{P(E_{Pd})}$ [keV]	10.8	12.1	7.8	9.3
$E_{\rm col}$ [keV]	10.1	12.6	9.5	10.2
$\sigma_{P(E_{Pd})}$ [keV]	2.6	4.2	3.4	3.3
$\Delta E_{\rm col} \equiv [E^{\rm min}_{\rm col}, E^{\rm max}_{\rm col}]$ [keV]	4.9–15.4	4.3–20.9	2.7–16.2	3.7–16.7
$n_t\tilde{l}$ [10^{18} cm^{-2}]			1.92	13.81
$\bar{S}_{pd}(E_{\rm col})$ [MeV b]			$\leq 17.5 \times 10^{-7}$	$\leq 2.5 \times 10^{-7}$
$\tilde{\sigma}^{\rm exp}_{pd}(\Delta E_{\rm col})$ [cm^2]			$\leq 2.8 \times 10^{-33}$	$\leq 4.0 \times 10^{-33}$
$\tilde{\sigma}^{\rm cal}_{pd}(\Delta E_{\rm col})$ [cm^2]			$\leq 1.7 \times 10^{-33}$	$\leq 1.7 \times 10^{-33}$
N'_p/N^0_p			7.6×10^{-3}	4.6×10^{-2}

Note: $E^p_{P(E_{pd})}$ is the most probable proton–deuteron collision energies, corresponding to the distribution function $P(E_{pd})$; $E_{\rm col}$ is the average *pd* collision energy in the c.m.s. corresponding to the distribution function $P(E_{pd})$; $\sigma_{P(E_{Pd})}$ is the rms deviation of *pd* collision energy corresponding to the $P(E_{pd})$ distribution; $\Delta E_{\rm col} \equiv [E^{\rm min}_{\rm col}, E^{\rm max}_{\rm col}]$ is the *pd* collision energy interval; $E^{\rm min}_{\rm col}$ and $E^{\rm max}_{\rm col}$ are the minimum and maximum limits of the above *pd* collision energy interval; n_t is the target density; $\tilde{l}$ is the effective thickness of the deuterated target for the real proton energy distribution $f(E_{pd})$; $\bar{S}_{pd}(E_{\rm col})$ is the average value of the astrophysical factor of the *pd* interaction corresponding to the collision energy interval $\Delta E_{\rm col}$; $\tilde{\sigma}^{\rm exp}_{pd}(\Delta E_{\rm col})$ and $\tilde{\sigma}^{\rm cal}_{pd}(\Delta E_{\rm col})$ are the experimental and calculated boundary values of the *pd* reaction cross section, respectively; N'_p is the number of protons bombarding the deuterated target; N^0_p is the total number of protons in the plasma liner when it passes the CIS.

the normalized distributions $P(E_{pd})$ in shot 10 practically do not differ in shape for this energy interval, nevertheless, they are substantially different in the number of protons N'_p hitting the CD_2 target and contributing to the yield of γ quanta from the *pd* reaction [because of the difference in shape between the initial liner proton distributions $f(E_{pd})$ measured by the LD and IC methods, see (8) and (12)]:

$$N'_p = N_p \int\limits_{E^{\rm min}_{\rm col}}^{E^{\rm max}_{\rm col}} P(E_{pd}) dE_{pd}.$$

It follows from the aforesaid that the analyses of the spectra of the detected γ quanta with the use of the IC and LD data give slightly different upper-bound estimates of the average values for the astrophysical S factor and the effective *pd* reaction cross section $\tilde{\sigma}_{pd}$.

Based on the analysis of the data for shots 9* and 10 and using the known values for N^0_p, ε_γ, k, and $\tilde{l}$, we obtain two groups of bounded estimates of $\bar{S}_{pd}(E_{\rm col})$ and $\tilde{\sigma}^{\rm exp}_{pd}(E_{\rm col})$ corresponding to the respective energy distributions of liner protons $f(E_{pd})$.

Table 3 presents final results on calculation of $\bar{S}_{pd}(E_{\rm col})$ and $\tilde{\sigma}^{\rm exp}_{pd}(E_{\rm col})$ at a 90% confidence level based on the data in shots 9* and 10. Note that, according to the analysis, the ion density in the plasma flux incident on the target is $\sim 10^{15}$ cm^{-3}.

The obtained upper boundary values of the astrophysical S_{pd} factor for shots 9* and 10 agree with the S_{pd} values extrapolated earlier for higher *pd* collision energies ($S_{pd}(E_{\rm col} \approx 8$ keV$) = (1.28 \pm 0.08) \times 10^{-7}$ MeV b [16]; $S_{pd}(E_{\rm col} \approx 20$ keV$) = (1.09 \pm 0.10) \times 10^{-7}$ MeV b [17]; $S_{pd}(17$ keV $\leq E_{\rm col} \leq 20$ keV$) = (1.2 \pm 0.3) \times 10^{-7}$ MeV b [18]).

A comparison of the calculated boundary values of $\tilde{\sigma}^{\rm col}_{pd}(\Delta E_{\rm col})$ with the values calculated by formula (8) also shows good agreement. Here, we used the average value of the astrophysical factor $\bar{S}_{pd} = 1.2 \times 10^{-7}$ MeV b (according to [16–18]) with the real proton–deuteron collision energy distribution, corresponding to the function $P(E_{pd})$ in shots 9* and 10.

The following conclusions can be drawn from the data obtained with the scintillation γ detectors.

An increase in the *pd* reaction γ-detection efficiency by two orders of magnitude (which is feasible) will make feasible measurement of the astrophysical S_{pd} factor and effective *pd* reaction cross section in the region of ultralow proton–deuteron collision energies with defined statistical errors. Specifically,

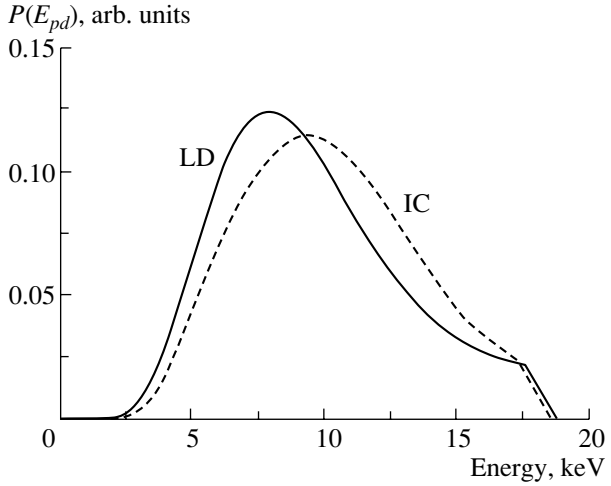

Fig. 8. The normalized distributions $P(E_{pd})$ obtained by IC and LD methods.

these S_{pd} factor and $\tilde{\sigma}_{pd}^{\exp}$ errors will be due to finite inaccuracy in recovering the proton energy distribution, via combined processing of IC and LD data. When using Monte Carlo simulation for matching two energy distributions and varying separately and jointly their shapes and respective number of events, the estimated error in the measurement of the S_{pd} factor and effective *pd* reaction cross section was $\leq 15\%$.

4. CONCLUSIONS

Finally, the tested diagnostics based on simultaneous use of two types of detectors (optical plasma radiation detectors and ion collectors) proved its applicability for measurement of the fundamental parameters in the reactions between light nuclei in the region of ultralow energies. The modified algorithm to recover the true energy distribution of the accelerated protons on the basis of the LD and IC data is needed, which will include adequate matching of both distributions in the middle part of the ion-energy distribution. It must include proportional corrections and renormalization of both distribution curves over the respective part of the full energy span.

The measurement of the astrophysical S_{pd} factor and effective *pd* reaction cross section with defined statistical errors in the 3–7 keV region could become feasible with a two-order increase in the γ-detection efficiency.

ACKNOWLEDGMENTS

We would like to express our gratitude to Prof. S.D. Korovin for his constant interest and support of the investigations, S.A. Sorokin for useful and stimulating discussions during analysis of the experimental data and participation in the fabrication of the load unit and measurements, V.A. Stolupin for participation in the experiment and fruitful discussions, A.F. Korostylev and E.N. Volkov for providing operation of the MIG generator and assistance in the measurements, and R.V. Kublikov for his help in processing experimental data.

The investigation was supported by the Russian Foundation for Basic Research (project no. 03-02-017278) and by the Fund of the Plenipotentiary of Poland at JINR.

REFERENCES

1. V. M. Bystritsky, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Laser Part. Beams **18**, 325 (2000).
2. Vit. M. Bystritskii, V. M. Bystritsky, S. A. Chaikovsky, *et al.*, Phys. At. Nucl. **64**, 855 (2001).
3. V. M. Bystritsky, V. V. Gerasimov, A. R. Krylov, *et al.*, Phys. At. Nucl. **66**, 1683 (2003).
4. A. V. Luchinsky, V. I. Makhrin, N. A. Ratakhin, *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz. **38**, 58 (1995).
5. V. M. Bystritsky, V. M. Grebenyuk, S. S. Parzhitski, *et al.*, Nucl. Instrum. Methods Phys. Res. A **455**, 706 (2000).
6. Vit. M. Bystritskii, V. M. Bystritsky, V. M. Grebenyuk, *et al.*, Zh. Tekh. Fiz. **72** (9), 29 (2002) [Tech. Phys. **47**, 1098 (2002)].
7. V. B. Belyaev, A. Bertin, Vit. M. Bystritskii, *et al.*, Nucleonika **40**, 85 (1995).
8. N. F. Kovsharov, V. F. Feduschak, and N. A. Ratakhin, Prib. Tekh. Éksp., No. 2, 114 (1990).
9. A. V. Luchinsky, Izv. Vyssh. Uchebn. Zaved. Fiz. **12**, 67 (1997).
10. *Plasma Diagnostic*, Ed. by N. G. Basov (Nauka, Moscow, 1989).
11. S. K. Allison, Rev. Mod. Phys. **30**, 1137 (1958).
12. C. F. Barnett, J. A. Ray, E. Ricci, *et al.*, ORNL-5206, A-4-2 (Feb. 1977), Vol. 1.
13. D. V. Sivukhin, *Problems of Plasma. Theory* (Atomizdat, Moscow, 1964), Vol. 4, p. 81.
14. Vit. M. Bystritskii, V. M. Bystritsky, S. A. Chaikovsky, *et al.*, Kerntechnik **66**, 42 (2001).
15. V. M. Bystritsky and F. M. Penkov, Phys. At. Nucl. **66**, 75 (2003).
16. A. Olin, A. Adamczak, G. Beer, *et al.*, Hyperfine Interact. **118**, 163 (1999).
17. G. J. Schmid *et al.*, Phys. Rev. Lett. **76**, 3088 (1996).
18. G. M. Griffiths, M. Lal, and C. D. Scarfe, Can. J. Phys. **41**, 734 (1963).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1787–1789. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1849–1851.

NUCLEI
Experiment

Determination of Nuclear Excitation Energies from the Number of Evaporated Particles

V. A. Ditlov, V. V. Dubinina, N. P. Egorenkova, V. I. Krotkova, E. A. Pozharova, and V. A. Smirnitsky

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

Received November 12, 2004; in final form, February 17, 2005

Abstract—The mean number $\langle N_b \rangle$ of particles evaporated in the interaction of ^{22}Ne, ^{32}S, and ^{56}Fe nuclei with photoemulsion nuclei was measured as a function of the number of alpha particles emitted within the fragmentation cone. It is found that $\langle N_b \rangle$ decreases with increasing number of the alpha particles and increases with increasing number of projectile nucleons involved in the interaction with a target nucleus and that $\langle N_b \rangle$ is a linear function of the excitation energy E_{ex} of the target-nucleus residue. The maximum experimental value of the mean number of evaporated particles is $\langle N_{b\,max} \rangle \cong 12-13$, which corresponds to $E_{exc} \cong 540 \pm 60$ MeV. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Relativistic nuclear physics studies processes induced by inelastic interactions of relativistic nuclei with nuclei. In a collision of high-energy nuclei, the bulk of their energy is released in the overlap region, and this is accompanied by multiparticle production processes. The nondisintegrated target-nucleus residue then receives energy, becomes hotter, and evaporates particles. In the region of low energies of nuclear reactions ($E < 50$ MeV), this evaporation process was considered in detail in [1], where respective theoretical calculations and experimental results obtained by studying the emission of particles evaporated in nucleus–nucleus interactions are described.

The results of theoretical calculations have not yet been tested experimentally at high energies. In this article, our experimental data on the production of evaporated particles in nucleus–nucleus interactions are presented with the aim of estimating the excitation energy of the target-nucleus in the region of high excitation energies (100–700 MeV) and comparing the result with the results of the calculations performed in [2].

Within photoemulsion procedures, slow target-nucleus fragments (b particles) characterized by kinetic energies for protons in the range $T_p \leq 26$ MeV (the free path in emulsion is $R \leq 3$ mm) and by $\beta \leq 0.23$ are classified as evaporated particles. We would like to emphasize, however, that nucleons do not exhaust the class of evaporated particles. According to the data reported in [2], slow charged particles emitted from a ^{109}Ag nucleus at average excitation energies (about 200 to 600 MeV) include protons, deuterons, ^{4}He nuclei, tritons, and ^{3}He nuclei, their fractions being about 65, 16, 13, 4, and 2%, respectively. Within the experimental errors, the mean number of evaporated charged particles ($\langle N_b \rangle$) is independent of (weakly dependent on) the projectile energy and atomic number [3]. However, this statement is valid only for mean values, in which case interactions are averaged over all values of the impact parameter. The energy released in the overlap region of colliding nuclei depends on the number of projectile nucleons (N_{int}) involved in the interaction with the target nucleus. One can estimate N_{int} on the basis of the relation

$$N_{int} \cong A_{pr} - \Theta_{fr}(A_{pr}/Z_{pr}), \qquad (1)$$

where A_{pr} and Z_{pr} are the projectile atomic number and charge, respectively, while Θ_{fr} is the sum of the charges of projectile fragments emitted within the fragmentation cone [4]—that is, the noninteracted charged part of the projectile. Relation (1) was obtained under the assumption that the remaining nucleons interacted with the target and that $A_{fr}/Z_{fr} \cong A_{pr}/Z_{pr}$. The last formula is valid to within 10 to 15% for projectiles not featuring a neutron excess.

2. DESCRIPTION OF THE EXPERIMENT

The emulsion chambers used were exposed to beams of ^{22}Ne and ^{32}S ions accelerated to momenta of 4.1 and 200 GeV/c per nucleon at the accelerators of the Joint Institute for Nuclear Research (JINR, Dubna) and CERN, respectively. The search for the

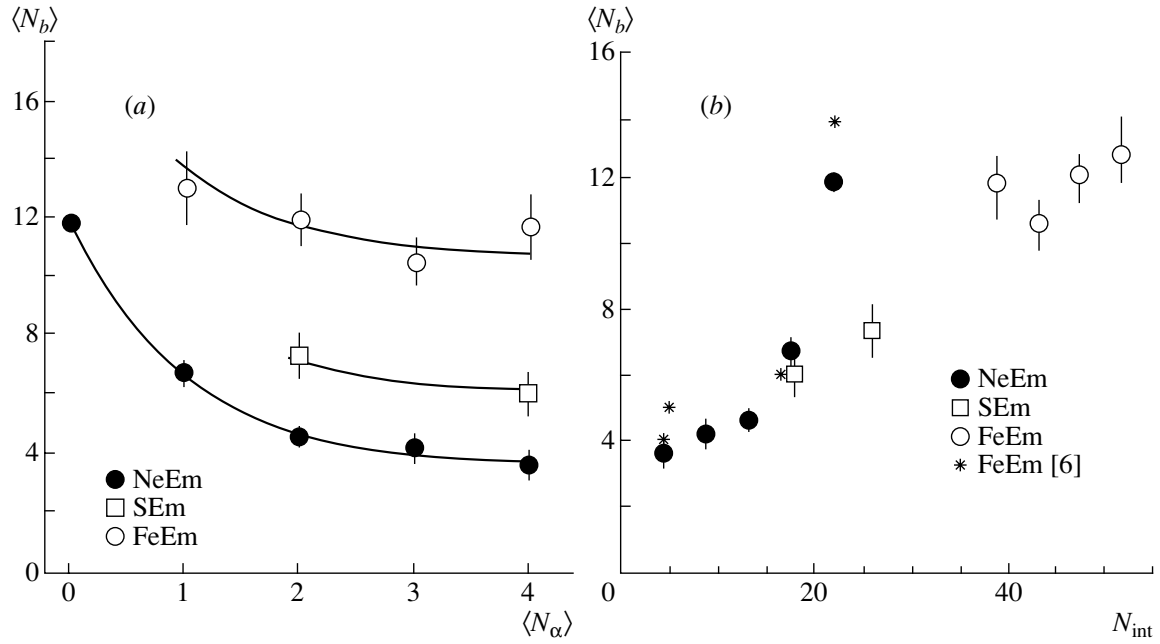

Fig. 1. (*a*) Mean number of *b* particles $\langle N_b \rangle$ as a function of the number of alpha particles (N_α) within the fragmentation cone in NeEm, SEm, and FeEm interactions (the description of the curves is given in the main body of the text); (*b*) $\langle N_b \rangle$ as a function of the number of interacted projectile nucleons ($N_{\rm int}$) in NeEm, SEm, and FeEm interactions.

interactions of the ions with photoemulsion nuclei, NeEm and SEm interactions, was performed by the method of tracing along a track. Among NeEm events found in our experiment, we selected five groups classified according to the following features of an interaction: there are no fragments within the fragmentation cone; one, two, three, or four alpha particles are emitted. Among SEm events, we selected two groups: two or four alpha particles are emitted within the fragmentation cone. In each group, we measured $\langle N_b \rangle$ and calculated $\langle N_{\rm int} \rangle$. The results of these measurements and calculations are given in Figs. 1*a* and 1*b*. Figure 1 shows our present data on NeEm and SEm interactions, data on FeEm interactions from [5], and the results of measurements for FeEm interactions from [6].[1)]

3. DISCUSSION OF EXPERIMENTAL RESULTS

From the data presented in Fig. 1, it follows that $\langle N_b \rangle$ depends on the character of projectile–target interaction: $\langle N_b \rangle$ decreases with increasing number of fragments (alpha particles) and, accordingly, increases with increasing number of projectile nucleons that interacted with the target nucleus. The value of $\langle N_b \rangle = 11.8$ was obtained for the NeEm interaction corresponding to a central collision, in which case $N_{\rm int} \sim 22$. The experimental data in Fig. 1*a* are well described by the exponential dependence

$$\langle N_b \rangle = a_1 + a_2 \exp(-a_3 N_\alpha). \qquad (2)$$

The fitted values of the coefficients a_2 and a_3 are

$$a_2(^{22}\mathrm{Ne}) = a_2(^{32}\mathrm{S}) = a_2(^{56}\mathrm{Fe}) = 8.2 \pm 0.4,$$
$$a_3(^{22}\mathrm{Ne}) = a_3(^{32}\mathrm{S}) = a_3(^{56}\mathrm{Fe}) = 1 \pm 0.2;$$

that is, these two coefficients take identical values for NeEm, SEm, and FeEm interactions. The coefficient a_1 is proportional to the projectile mass number:

$$a_1(^{22}\mathrm{Ne}) : a_1(^{32}\mathrm{S}) : a_1(^{56}\mathrm{Fe}) \sim A_{^{22}\mathrm{Ne}} : A_{^{32}\mathrm{S}} : A_{^{56}\mathrm{Fe}},$$
$$(3.6 \pm 0.2)_{^{22}\mathrm{Ne}} : (6.0 \pm 0.5)_{^{32}\mathrm{S}} : (10.6 \pm 0.5)_{^{56}\mathrm{Fe}}$$
$$\sim 22 : 32 : 56.$$

The sum $a_1 + a_2$ corresponds to a central collision, in which case there is no alpha-particle emission. This interaction is determined by the geometric cross section. Therefore, the ratio of the geometric cross sections (projectile areas) must be proportional to the ratio of the sum $(a_1 + a_2)$ for different projectiles. Indeed, we have

$$(R_{\rm Fe}/R_{\rm Ne})^2 \cong 1.84, (a_1 + a_2)_{\rm Fe}/(a_1 + a_2)_{\rm Ne}$$
$$= 1.6 \pm 0.1;$$

[1)]For FeEm, Friedlander and Heckmann [6] presented the interaction chain that they observed in a photoemulsion: $Z_{\rm pr} = 26$ ($N_b = 4$) $\rightarrow Z_{\rm pr} = 24$ ($N_b = 5$) $\rightarrow Z_{\rm pr} = 20$ ($N_b = 6$) $\rightarrow Z_{\rm pr} = 11$ ($N_b = 11$); here, $Z_{\rm pr}$ are the charges of fragments that emerge in each subsequent interaction. In the last interaction, fragments are not indicated.

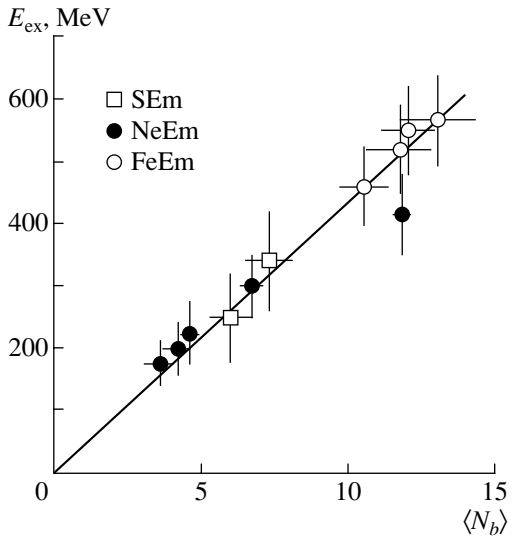

Fig. 2. Mean number $\langle N_b \rangle$ of b particles as a function of the excitation energy E_{exc} of the target-nucleus residue. The straight line is an approximation by a linear dependence.

$$(R_{Fe}/R_S)^2 \cong 1.45, (a_1 + a_2)_{Fe}/(a_1 + a_2)_S = 1.3 \pm 0.2;$$
$$(R_S/R_{Ne})^2 \cong 1.28, (a_1 + a_2)_S/(a_1 + a_2)_{Ne} = 1.2 \pm 0.2.$$

From the experimental data in Fig. 1 and from the ratio of the coefficients in (2), it follows that, in nucleus–nucleus collisions, $\langle N_b \rangle$ depends on the number of projectile nucleons that interacted with the target nucleus. We compared experimental data with the theoretical results reported by Dostrovsky *et al.* [2], who calculated the number of evaporated particles at high excitation energies of the target nucleus (100–700 MeV). Figure 2 displays our data on NeEm, SEm, and FeEm interactions. From these data, it follows that the number $\langle N_b \rangle$ of evaporated particles that originate from the interaction of high-energy nuclei with nuclei may serve as a measure of the change in the excitation energy (E_{exc}) of the target-nucleus residue. For photoemulsion nuclei ($\langle A_{Em} \rangle \cong 80$), the maximum value $\langle N_{b\,max} \rangle$ is 12 to 13, which corresponds to $E_{exc} \approx 540 \pm 60$ MeV.

ACKNOWLEDGMENTS

We are grateful to V.V. Shamanov for assistance in organizing a computer treatment of our experimental data.

REFERENCES

1. I. Dostrovsky, Z. Fraenkel, and G. Friedlander, Phys. Rev. **116**, 683 (1959).
2. I. Dostrovsky, P. Rabinowitz, and R. Bivins, Phys. Rev. **111**, 1659 (1958).
3. S. Vokal, Doctoral Dissertation in Mathematical Physics (OIYaI, Dubna, 2001); W. Wolter, M. I. Cherry, A. Dobrowska, *et al.*, Z. Phys. C **59**, 399 (1993).
4. C. F. Powell, P. H. Fowler, and D. H. Perkins, *The Study of Elementary Particles by the Photographic Method* (Pergamon, London, 1959; Inostrannaya Literatura, Moscow, 1962).
5. V. V. Dubinina, N. P. Egorenkova, V. I. Korotkova, *et al.*, Yad. Fiz. **67**, 537 (2004) [Phys. At. Nucl. **67**, 518 (2004)].
6. E. M. Friedlender and H. H. Heckmann, in *Proceedings of the Conference "Relativistic Heavy-Ion Collisions. Experiment," Lawrence Berkeley Lab., 1983*, p. 206.

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1790–1795. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1852–1857.

ELEMENTARY PARTICLES AND FIELDS
Experiment

Single-Spin Asymmetry of Inclusive Neutral-Pion Production in $pp_\uparrow$ Interactions at 70 GeV in the Region $-0.4 < x_F < -0.1$

A. N. Vasiliev[1)], V. N. Grishin[1)], A. M. Davidenko[1)], A. A. Derevshchikov[1)], Yu. A. Matulenko[1)], Yu. M. Mel'nik[1)], A. P. Meshchanin[1)], V. V. Mochalov[1)], L. V. Nogach[1)*], S. B. Nurushev[1)], A. F. Prudkoglyad[1)], P. A. Semenov[1)], L. F. Soloviev[1)], V. L. Solovianov[†1)], V. Yu. Khodyrev[1)], K. E. Shestermanov[1)], A. E. Yakutin[1)], N. S. Borisov[2)], V. N. Matafonov[†2)], A. B. Neganov[2)], Yu. A. Plis[2)], Yu. A. Usov[2)], A. N. Fedorov[2)], and A. A. Lukhanin[3)]

Received September 8, 2004; in final form, February 4, 2005

Abstract—For the kinematical region specified by the inequalities $-0.4 < x_F < -0.1$ and $0.9 < p_T < 2.5$ GeV/c, the results are presented that were obtained by experimentally determining the single-spin asymmetry of inclusive neutral-pion production in the reaction $p + p_\uparrow \to \pi^0 + X$ at 70 GeV. According to these results, the asymmetry is close to zero in the region $-0.2 < x_F < -0.1$ and grows in magnitude with decreasing x_F, amounting to $(-10.6 \pm 3.2)\%$ for $-0.4 < x_F < -0.2$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In this article, the single-spin asymmetry A_N in the reaction

$$p + p_\uparrow \to \pi^0 + X \qquad (1)$$

at 70 GeV in the Feynman variable region $-0.4 < x_F < -0.1$ is presented according to an analysis of data obtained in 1996 at the U-70 accelerator of the Institute for High Energy Physics (IHEP, Protvino). Previously, the asymmetry A_N of inclusive neutral-pion production was measured in $\pi^- p_\uparrow$ and $pp_\uparrow$ interactions. In $\pi^- p_\uparrow$ interactions, the asymmetry A_N is significant both in the central region [1] and in the region of polarized-target fragmentation [2]. In $pp_\uparrow$ collisions, the asymmetry in the central region is close to zero at energies of 70 GeV [3] and 200 GeV [4], but it is different from zero at 24 GeV [5]; a significant effect was also observed in the region of polarized-proton-beam fragmentation [6].

For the reaction in (1), measurements of A_N in the region of polarized-target fragmentation were performed for the first time. Preliminary data on a raw asymmetry (without investigating systematic errors) were reported previously in [7].

2. DESCRIPTION OF THE EXPERIMENT

Our experiment was conducted at the PROZA-M facility (beamline 14 of the U-70 accelerator complex). The layout of the experimental facility is presented in Fig. 1.

Protons of energy 70 GeV interacted with a transversely polarized frozen-type target, where propanediol ($C_3H_8O_2$) was used for a working substance [8]. In order to take into account unpolarized matter in the target, background measurements were performed with a carbon target.

Photons from neutral-pion decays were recorded by an electromagnetic calorimeter (EMC) that was shaped as a matrix of 12×12 lead-glass counters [9]. The dimensions of a counter were $3.8 \times 3.8 \times 45$ cm^3 (18 radiation-length units along the beam). The calorimeter was arranged at a distance of 2.7 m from the target center, its coverage angle in the horizontal plane being 17° to 26° in the laboratory frame. The trigger used ensured the selection of events where the energy deposition in the calorimeter exceeded 1 GeV.

A detailed description of basic units of the facility was given in [10].

The calibration of the calorimeter was performed by using a beam of 26.6-GeV/c electrons [11]. Upon the subtraction of the beam momentum spread (about 2%), the energy resolution of the calorimeter

†Deceased.
1)Institute for High Energy Physics, Protvino, Moscow oblast, 142284 Russia.
2)Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
3)Kharkov Institute for Physics and Technology, Akademicheskaya ul. 1, Kharkov, 61108 Ukraine.
*e-mail: nogach@mx.ihep.su

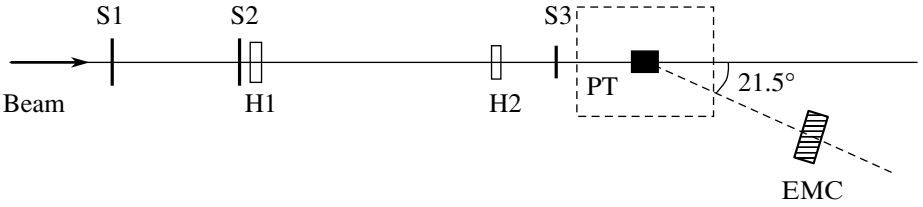

Fig. 1. Layout of the PROZA-M experimental facility: (S1–S3) scintillation counters, (H1, H2) hodoscopes, (PT) polarized target, and (EMC) electromagnetic calorimeter.

at this energy was $\sigma(E)/E \approx 3\%$. The EMC energy scale was additionally calibrated to the neutral-pion mass. The calibration accuracy of 0.1% was attained within five hours of measurements.

Around 10 million events were recorded over a 10-day run with a polarized target.

3. DATA ANALYSIS

In order to reconstruct photons in the calorimeter, we employed the algorithm that was developed in [12] and which is based on the separation of an electromagnetic shower according to a known shape. Special features of its application to analyzing data that come from the electromagnetic calorimeters of the PROZA-M facility were described in [3].

In order to obtain a physical result, we selected only those photon pairs for which (i) χ^2/NDF for a description in terms of a shower shape was less than 3.0, (ii) the asymmetry of energy satisfied the condition $\alpha < 0.7$ ($\alpha = |E_1 - E_2|/(E_1 + E_2)$, where $E_{1,2}$ are the photon energies), (iii) each photon was at a distance from the detector edges that was not smaller than half the distance from the respective counter to, and (iv) $(E_1 + E_2) > 2$ GeV. In this case, the mean multiplicity in the EMC was about 1.5 photons per event, while the photon energies were in the range between 0.5 and 10 GeV.

The distributions of photon pairs with respect to kinematical variables are displayed in Fig. 2, while the two-dimensional distribution with respect to the variables p_T and x_F is given in Fig. 3*a*. In view of a narrow acceptance of the EMC, the variables x_F and p_T are correlated.

Figure 3*b* shows a characteristic mass spectrum of photon pairs. In this spectrum, one can see a distinct peak in the region around the neutral-pion mass. In the transverse-momentum range between 0.9 and 2.5 GeV/c, the mass resolution for a neutral pion was 11 to 15 MeV/c^2.

3.1. Calculation of the Asymmetry

The single-spin asymmetry A_N is defined by the expression

$$A_N(x_\mathrm{F}, p_T) = \frac{1}{P_\mathrm{targ}} \frac{1}{\langle \cos\phi \rangle} \times \frac{d\sigma^\mathrm{H}_\uparrow(x_\mathrm{F}, p_T) - d\sigma^\mathrm{H}_\downarrow(x_\mathrm{F}, p_T)}{d\sigma^\mathrm{H}_\uparrow(x_\mathrm{F}, p_T) + d\sigma^\mathrm{H}_\downarrow(x_\mathrm{F}, p_T)}, \quad (2)$$

where P_targ is the target polarization, ϕ is the azimuthal angle between the target-polarization vector and the normal to the plane spanned by the beam axis and the momentum of the outgoing neutral pion, and $d\sigma^\mathrm{H}_\uparrow$ and $d\sigma^\mathrm{H}_\downarrow$ are the invariant differential cross sections for neutral-pion production on hydrogen for opposite directions of the target-polarization vector. In our experiment, the azimuthal angle at which we detected neutral pions was in the range $180° \pm 15°$; therefore, we set $\langle \cos\phi \rangle$ to -1. The mean degree of target polarization during data accumulation was $(80 \pm 3)\%$. Since the detection efficiency for neutral pions is identical for the two directions of the target-polarization vector, we find for the detector on the right of the beam that

$$A_N = -\frac{D}{P_\mathrm{targ}} A_N^\mathrm{raw} = -\frac{D}{P_\mathrm{targ}} \frac{n_\uparrow - n_\downarrow}{n_\uparrow + n_\downarrow}, \quad (3)$$

where A_N^raw is the raw asymmetry actually measured in the experiment, D is the target-dilution factor, and $n_\uparrow$ and $n_\downarrow$ are the normalized (to the monitor) numbers of recorded neutral pions for opposite directions of the target-polarization vector. The procedure used to calculate D was described in detail elsewhere [2]. In order to test our calculations, we employed the results obtained previously in [13] by measuring the dilution factor for the target being investigated. In assessing the asymmetry, we used the calculated values of D without allowance for errors: 8.0, 8.1, 8.2, and 9.2 for p_T from the intervals $0.9 < p_T < 1.4$ GeV/c, $1.4 < p_T < 1.8$ GeV/c, $1.8 < p_T < 2.1$ GeV/c, and $2.1 < p_T < 2.5$ GeV/c, respectively.

In measuring the asymmetry A_N, there can arise an additional asymmetry caused by trigger-electronics jitter, failures of the monitor counters, beam drift,

or some other reasons. This gives rise to a systematic bias of the true asymmetry. A method that can be used to remove this bias and which is based on the fact that the asymmetry of photon pairs off the neutral-pion mass peak is zero is described in detail in [2].

3.2. Analysis of a Spurious Asymmetry

A spurious asymmetry is determined predominantly by the drift of the calorimeter energy scale, this leading to an inaccurate reconstruction of the kinematical parameters of the photon pair. A change of 0.1% in the energy scale gives rise to a raw spurious asymmetry at a level of 0.2% (with allowance for the dilution factor and a target polarization of $D/P_{\text{targ}} \sim 10$, this corresponds to 2% in A_N).

In order to estimate the spurious asymmetry, we broke down the total data sample for the "up" direction of the target-polarization vector into two equal subsamples of events and, by convention, assigned one subsample of events a positive sign of the target-polarization vector ("+") and the other subsample a negative sign ("−"), whereupon we determined the asymmetry for them. In the same way, we calculated the asymmetry for the "down" direction of the polarization vector and for a carbon target. For each of these three sets of data, we obtained zero asymmetry in the kinematical region being studied.

Figure 4 shows the total spurious asymmetry [the errors are somewhat less than in A_N (see Fig. 5) owing to additional data accumulated with a carbon target]. By fitting the function $a + bx_{\text{F}}$ to the spurious asymmetry, we obtained the following parameter values: $a = (-1.7 \pm 5.9)\%$ and $b = (-3 \pm 32)\%$. One can see that the spurious asymmetry is independent of x_{F} and is equal to zero within the errors. From the approximation of the same data by a constant, we obtained $c = (-1.1 \pm 1.5)\%$. As was indicated above, the spurious asymmetry is due predominantly to the drift of the calorimeter energy scale, this drift being corrected by means of a permanent recalibration to the world-average value of the neutral-pion mass. From the results of our fit, we can infer that, for each value of the observed physical single-spin asymmetry A_N, the absolute systematic error does not exceed 3% upon introducing this correction.

Asymmetry A_N in the reaction $p + p_\uparrow \to \pi^0 + X$ at 70 GeV

$\langle x_{\text{F}} \rangle$	$\langle p_T \rangle$, GeV/c	A_N, %
−0.12	0.97	-1.0 ± 4.9
−0.14	1.05	-2.6 ± 4.5
−0.16	1.15	2.1 ± 4.0
−0.18	1.28	-5.8 ± 4.1
−0.22	1.49	-6.3 ± 4.3
−0.25	1.69	-11.6 ± 6.2
−0.29	1.93	-18.8 ± 8.7
−0.34	2.27	-40.0 ± 17.1

4. RESULTS AND THEIR DISCUSSION

The asymmetry A_N, which is a physical observable in our experiment, is shown in Fig. 5 and in the table. The quoted errors are purely statistical. As was indicated above, the absolute systematic error for each value of A_N does not exceed 3% and, in those intervals of x_{F} where A_N is different from zero, is much less than the statistical error. From our estimates, it follows that, for all values of A_N, the relative systematic error associated with the accuracy of determining the dilution factor and the degree of target polarization is within 10%.

The asymmetry is $(-2.5 \pm 2.0)\%$ in the region $-0.2 < x_{\text{F}} < -0.1$ ($\langle p_T \rangle \approx 1.1$ GeV/c) and $(-10.6 \pm 3.2)\%$ in the region $-0.4 < x_{\text{F}} < -0.2$ ($\langle p_T \rangle \approx 1.7$ GeV/c). The value of $x_{\text{F}} \approx -0.2$ is the threshold point for the emergence of A_N. It was indicated in [14] that, in the majority of experiments, the asymmetry of inclusive neutral-pion production is compatible with zero up to the c.m. pion energy of $E_{\text{c.m.}} = E_0 \approx 1.5$–$2.0$ GeV, whereupon it grows in magnitude. The asymmetry measured in our experiment is shown in Fig. 6 as a function of $E_{\text{c.m.}}$.

By fitting the function

$$A_N = \begin{cases} 0 & \text{for} \quad E_{\text{c.m.}} < E_0 \\ k \cdot (E_{\text{c.m.}} - E_0) & \text{for} \quad E_{\text{c.m.}} \geq E_0 \end{cases} \tag{4}$$

to our data on A_N, we obtained the value of $E_0 = 1.5 \pm 0.1$ GeV, which is in good agreement with the results reported in [14]. The fitted value of the slope parameter k is -15 ± 4.

In Fig. 7, our result (the sign of A_N is reversed for a comparison to be more convenient) is given along with data of previous experiments devoted to measuring the single-spin asymmetry of inclusive neutral-pion production in the region of polarized-proton fragmentation. One can see that, within the errors, the values of A_N in $pp_\uparrow$ interactions at 70 GeV are identical to their counterparts in $\pi^- p_\uparrow$ interactions at 40 GeV, but that, at 200 GeV, the growth of the asymmetry with increasing neutral-pion energy is slower.

As is well known, large single-spin effects cannot be explained within perturbative QCD, which is the

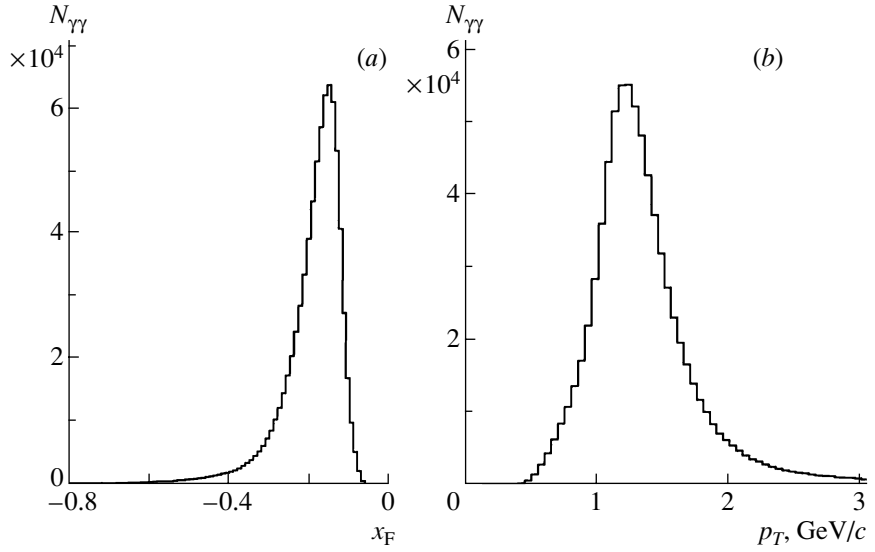

Fig. 2. Distribution of photon pairs with respect to (*a*) the Feynman variable and (*b*) the transverse momentum.

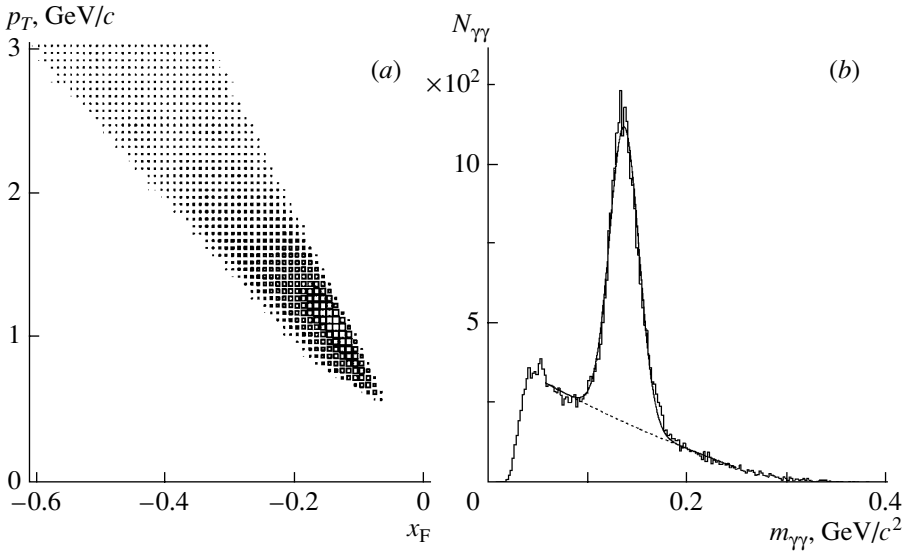

Fig. 3. (*a*) Two-dimensional distribution of photon pairs with respect to the transverse momentum (p_T) and the Feynman variable x_F and (*b*) mass spectrum of photon pairs in the region $-0.3 < x_F < -0.2$ (summation over p_T was performed). In Fig. 3*b*, the solid curve is an approximation by the sum of a Gaussian distribution and a third-degree polynomial, while the dotted curve represents the contribution of a combinatorial background.

generally accepted theory of hard interactions. For the reaction $p + p \to \pi^0 + X$, it follows from the factorization theorem that

$$d\sigma = \sum_{a,b,c} f_{a/p} \otimes f_{b/p} \otimes d\sigma(ab \to c\ldots) \otimes D_{\pi/c}, \quad (5)$$

where $f_{a/p}$ and $f_{b/p}$ are the parton distributions in colliding protons, $d\sigma(ab \to c\ldots)$ is the elementary-process cross section, and $D_{\pi/c}$ is the parton-to-pion fragmentation function. For the asymmetry $A_N \sim (d\sigma_\uparrow - d\sigma_\downarrow)$, the helicity-conservation law yields

$$A_N \sim \frac{m_q}{\sqrt{s}}\alpha_s \sim 0. \quad (6)$$

A number of models based on generalizations of the factorization theorem were proposed after the discovery of significant values of A_N in some experiments. These models assume the presence of (i) higher twist correlation functions in the distribution functions (for example, twist-3 correlation functions in the Qui–Sterman [15] and Efremov–Korotkiyan–

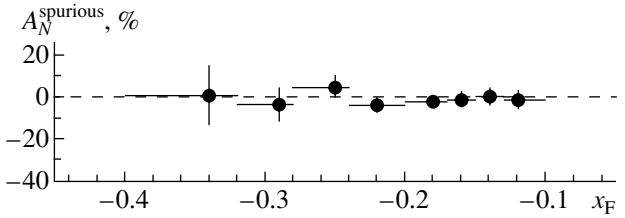

Fig. 4. Spurious asymmetry as a function of x_F. The horizontal straight-line segments show x_F intervals, while the points correspond to the average value of x_F in a given interval (summation over p_T was performed).

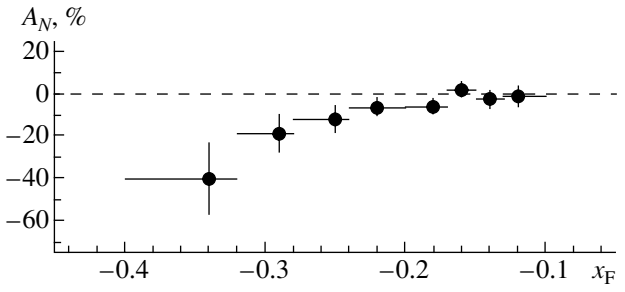

Fig. 5. Asymmetry A_N as a function of x_F.

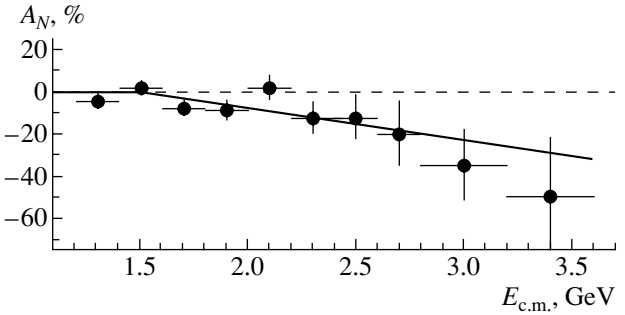

Fig. 6. Asymmetry A_N as a function of the neutral-pion energy in the c.m. frame. The solid line represents an approximation by the function in (4).

Teryaev [16] models), (ii) an intrinsic transverse momentum k_T and a spin dependence of the distribution functions (Sivers model [17]), or (iii) an intrinsic transverse momentum k_T and a spin dependence of the fragmentation function (Collins model [18]).

Also, so-called semiclassical models were developed on the basis of introducing a quark orbital angular momentum [19, 20]. Proposing various mechanisms for explaining the emergence of single-spin asymmetries, these models do not present a universal spin theory, but they describe experimental data satisfactorily.

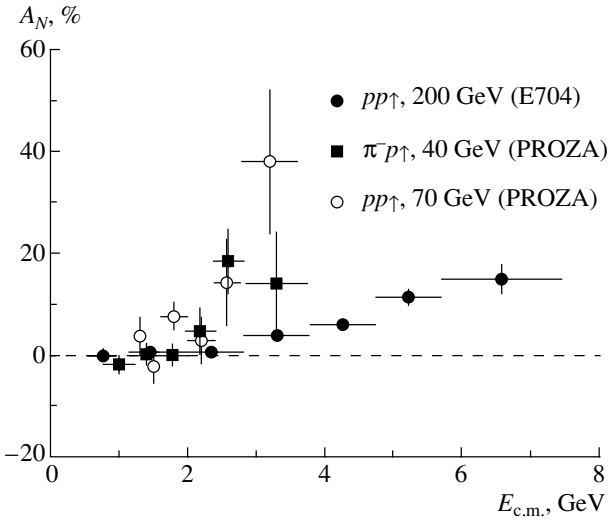

Fig. 7. Asymmetry A_N in the reaction $hp_\uparrow \to \pi^0 X$ as a function of the interaction energy for various hadron types h (π^- or p).

5. CONCLUSION

The single-spin asymmetry of inclusive neutral-pion production in the reaction $p + p_\uparrow \to \pi^0 + X$ at 70 GeV has been measured over the kinematical region specified by the inequalities $-0.4 < x_F < -0.1$ and $0.9 < p_T < 2.5$ GeV/c. The asymmetry is equal to zero for $-0.2 < x_F < -0.1$ within the errors, but it grows in magnitude as x_F decreases beyond this region, amounting to $(-10.6 \pm 3.2)\%$ in the region $-0.4 < x_F < -0.2$.

In $pp_\uparrow$ interactions, the asymmetry A_N begins deviating from zero at a c.m. neutral-pion energy of about 1.5 GeV.

In contrast to what we have in the central region, the asymmetry of inclusive neutral-pion production in the region of polarized-proton fragmentation takes the same value within the experimental errors in $\pi^- p_\uparrow$ and in $pp_\uparrow$ interactions.

Our result is compatible with the predictions of theoretical models describing spin effects in $hp_\uparrow$ interactions.

ACKNOWLEDGMENTS

We are grateful to the directorate of the Institute for High Energy Physics (IHEP, Protvino) for their support of our studies; to the personnel of the Accelerator Department and the Beam Division (IHEP), who ensured a high-quality operation of the U-70 accelerator and the beamline 14; and to N.I. Belikov, Yu.M. Goncharenko, V.A. Kormilitsyn, N.E. Mikhalin, and A.I. Mysnik for their technical assistance during the runs.

This work was supported in part by the Russian Foundation for Basic Research (project no. 03-02-16919).

REFERENCES

1. V. D. Apokin *et al.*, Phys. Lett. B **243**, 461 (1990).
2. A. N. Vasiliev *et al.*, Preprint No. 03-21, IFVÉ (Inst. High Energy Phys., Protvino, 2003); Yad. Fiz. **67**, 1520 (2004) [Phys. At. Nucl. **67**, 1495 (2004)].
3. A. N. Vasiliev *et al.*, Preprint No. 03-22, IFVÉ (Inst. High Energy Phys., Protvino, 2003); Yad. Fiz. **67**, 1512 (2004) [Phys. At. Nucl. **67**, 1487 (2004)].
4. D. L. Adams *et al.*, Phys. Rev. D **53**, 4747 (1996).
5. J. Antille *et al.*, Phys. Lett. B **94B**, 523 (1980).
6. D. L. Adams *et al.*, Z. Phys. C **56**, 181 (1992).
7. N. I. Belikov *et al.*, Preprint No. 97-51, IFVÉ (Institute High Energy Phys., Protvino, 1997).
8. N. S. Borisov *et al.*, Preprint No. 1-80-98, OIYaI (Joint Inst. Nucl. Res., Dubna, 1980).
9. G. A. Akopdjanov *et al.*, Nucl. Instrum. Methods **140**, 441 (1977); F. Binon *et al.*, Nucl. Instrum. Methods **188**, 507 (1981).
10. V. D. Apokin *et al.*, Prib. Tekh. Éksp., No. 4, 23 (1998) [Instrum. Exp. Tech., No. 4, 464 (1998)].
11. D. L. Adams *et al.*, Preprint No. 91-99, IFVÉ (Inst. High Energy Phys., Protvino, 1991).
12. A. A. Lednev, Preprint No. 93-153, IFVÉ (Inst. High Energy Phys., Protvino, 1993).
13. N. S. Amaglobeli *et al.*, Yad. Fiz. **50**, 695 (1989) [Sov. J. Nucl. Phys. **50**, 432 (1989)].
14. A. N. Vasiliev and V. V. Mochalov, Preprint No. 03-26, IFVÉ (Inst. High Energy Phys., Protvino, 2003); Yad. Fiz. **67**, 2193 (2004) [Phys. At. Nucl. **67**, 2169 (2004)]; V. V. Mochalov and A. N. Vasiliev, hep-ex/0312007.
15. J. Qui and G. Sterman, Phys. Rev. D **59**, 014 004 (1999).
16. A. V. Efremov, V. M. Korotkiyan, and O. V. Teryaev, Phys. Lett. B **348**, 577 (1995).
17. D. Sivers, Phys. Rev. D **41**, 83 (1990); **43**, 261 (1991).
18. J. C. Collins, Nucl. Phys. B **396**, 161 (1993).
19. C. Boros, L. Zuo-tang, and M. Ta-chung, Phys. Rev. Lett. **70**, 1751 (1993).
20. S. M. Troshin and N. E. Tyurin, Phys. Rev. D **52**, 3862 (1995); **54**, 838 (1996).

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1796–1811. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1858–1873.

ELEMENTARY PARTICLES AND FIELDS
Experiment

Measurements of the Total-Cross-Section Difference $\Delta\sigma_L(np)$ at 1.39, 1.69, 1.89, and 1.99 GeV*

V. I. Sharov[1)], N. G. Anischenko[1)], V. G. Antonenko[2)], S. A. Averichev[1)], L. S. Azhgirey[3)], V. D. Bartenev[1)], N. A. Bazhanov[3)], A. A. Belyaev[4)], N. A. Blinov[1)], N. S. Borisov[3)], S. B. Borzakov[5)], Yu. T. Borzunov[1)], Yu. P. Bushuev[4)], L. P. Chernenko[4)], E. V. Chernykh[1)], V. F. Chumakov[1)], S. A. Dolgii[1)], A. N. Fedorov[3)], V. V. Fimushkin[1)], M. Finger[3),6)], M. Finger, Jr.[3)], L. B. Golovanov[1)], G. M. Gurevich[7)], A. Janata[3)], A. D. Kirillov[1)], V. G. Kolomiets[3)], E. V. Komogorov[1)], A. D. Kovalenko[1)], A. I. Kovalev[8)], V. A. Krasnov[1)], P. Krstonoshich[3)], E. S. Kuzmin[3)], V. P. Ladygin[1)], A. B. Lazarev[3)], F. Lehar[9)], A. de Lesquen[9)], M. Yu. Liburg[1)], A. N. Livanov[1)], A. A. Lukhanin[4)], P. K. Maniakov[1)], V. N. Matafonov[†3)], E. A. Matyushevsky[1)], V. D. Moroz[†1)], A. A. Morozov[1)], A. B. Neganov[3)], G. P. Nikolaevsky[1)], A. A. Nomofilov[1)], Tz. Panteleev[5),10)], Yu. K. Pilipenko[1)], I. L. Pisarev[3)], Yu. A. Plis[3)], Yu. P. Polunin[2)], A. N. Prokofiev[8)], V. Yu. Prytkov[1)], P. A. Rukoyatkin[1)], V. A. Schedrov[8)], O. N. Schevelev[3)], S. N. Shilov[3)], R. A. Shindin[1)], M. Slunečka[3),6)], V. Slunečková[3)], A. Yu. Starikov[1)], G. D. Stoletov[†3)], L. N. Strunov[1)], A. L. Svetov[1)], Yu. A. Usov[1)], T. Vasiliev[1)], V. I. Volkov[1)], E. I. Vorobiev[1)], I. P. Yudin[11)], I. V. Zaitsev[1)], A. A. Zhdanov[8)], and V. N. Zhmyrov[3)]**

Received February 15, 2005

Abstract—New accurate data of the neutron–proton spin-dependent total-cross-section difference $\Delta\sigma_L(np)$ at the neutron-beam kinetic energies 1.39, 1.69, 1.89, and 1.99 GeV are presented. In general, these data complete the measurements of energy dependence of $\Delta\sigma_L(np)$ over the Dubna Synchrophasotron energy region. Measurements were carried out at the Synchrophasotron of the Veksler and Baldin Laboratory of High Energies of the Joint Institute for Nuclear Research. The quasi-monochromatic neutron beam was produced by breakup of extracted polarized deuterons. The deuteron (and hence neutron) polarization direction was flipped every accelerator burst. The initial transverse (with respect to beam momentum) neutron polarization was changed to a longitudinal one and longitudinally polarized neutrons were transmitted through the large proton longitudinally polarized target. The target polarization direction was inverted after one to two days of measurements. Four different combinations of the beam and target parallel and antiparallel polarization directions, both oriented along the neutron-beam momentum, were used at each energy. A fast decrease in $-\Delta\sigma_L(np)$ with increasing energy above 1.1 GeV and a structure in the energy dependence around 1.8 GeV, first observed from our previous data, seem to be well revealed. The new results are also compared with model predictions and with phase-shift analysis fits. The $\Delta\sigma_L$ quantities for isosinglet state $I = 0$, deduced from the measured $\Delta\sigma_L(np)$ values and known $\Delta\sigma_L(pp)$ data, are also given. The results of the measurements of unpolarized total cross sections $\sigma_{0\mathrm{tot}}(np)$ at 1.3, 1.4, and 1.5 GeV and $\sigma_{0\mathrm{tot}}(n\mathrm{C})$ at 1.4 and 1.5 GeV are presented as well. These data were obtained using the same apparatus and high-intensity unpolarized deuteron beams extracted either from the Synchrophasotron or from the Nuclotron. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The paper presents new results of the spin-dependent neutron–proton total-cross-section difference $\Delta\sigma_L(np)$ measured in 2001 with a quasi-monochromatic polarized neutron-beam and polarized proton target (PPT). The $\Delta\sigma_L(np)$ values were

*This article was submitted by the authors in English.
†Deceased.

[1)]Veksler and Baldin Laboratory of High Energies, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
[2)]Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123182 Russia.

obtained at neutron-beam kinetic energies of 1.39, 1.69, 1.89, and 1.99 GeV.

A free polarized neutron beam with sufficient intensity was produced by breakup of polarized deuterons accelerated by the Synchrophasotron of the Veksler and Baldin Laboratory of High Energies (VBLHE) at the Joint Institute for Nuclear Research (JINR) in Dubna. This accelerator provides the highest energy (3.7 GeV) polarized neutron beam that can be reached at the present moment.

The spin-dependent nucleon–nucleon (NN) observables $\Delta\sigma_L$ and $\Delta\sigma_T$ are defined as the difference in the NN total cross sections for the antiparallel and parallel beam and target polarizations oriented longitudinally (L) and transversely (T) to the beam direction. Transmission measurements of the $\Delta\sigma_L(np)$ and $\Delta\sigma_T(np)$ energy dependences over the Dubna Synchrophasotron neutron-beam energy range of 1.2–3.7 GeV have been proposed [1, 2] and started [3–6] in Dubna. The measurements were carried out within the program of the JINR project DELTA-SIGMA experiment. The aim of this experimental program is to obtain a sufficient data set on the np polarization observables over this new highest energy region of free polarized neutron beams and to do a direct reconstruction of the imaginary and real parts of the spin-dependent np forward-scattering amplitudes for the first time.

To carry out $\Delta\sigma_L(np)$ measurements, a large Argonne–Saclay PPT was reconstructed at Dubna [7–9] and a new polarized-neutron-beam line [10, 11] with suitable parameters was made and tested. A set of dedicated neutron detectors with corresponding electronics was used. The data-acquisition system was based on a CAMAC parallel branch highway controlled by an IBM PC with the branch driver [12] finished off by one of the authors. The on-line program in Pascal works under DOS. Successful data-acquisition runs were carried out in 1995 and 1997 and the $\Delta\sigma_L(np)$ values were measured at 1.19, 1.59, 1.79, 2.20, 2.49, and 3.66 GeV [3–6]. For the measurements in 1997, a new PPT polarizing solenoid [13] was developed at VBLHE.

The NN total-cross-section differences $\Delta\sigma_L$ and $\Delta\sigma_T$, together with the spin-averaged total cross section $\sigma_{0\mathrm{tot}}$ are measured in pure inclusive transmission experiments. They are linearly related with three nonvanishing imaginary parts of the NN forward-scattering amplitudes via optical theorems and allowed to reconstruct these imaginary parts directly. The data are also used to check the predictions of available dynamic models and to provide an important contribution to databases of phase-shift analyses (PSA). From the measured $\Delta\sigma_L(np)$ values, it is possible to deduce the NN isosinglet ($I = 0$) part of $\Delta\sigma_L$, using the existing pp (isotriplet $I = 1$) results.

The total-cross-section differences $\Delta\sigma_{L,T}$ for pp scattering were first measured at the ANL ZGS and then at TRIUMF, in PSI, at LAMPF, at Saturne II, and in Fermilab. Results were obtained in the energy range from 0.2 to 12 GeV and at 200 GeV. Measurements with incident charged particles need an experimental setup different from np experiments, due to the contribution of electromagnetic interactions. Existing $\Delta\sigma_{L,T}(pp)$ results are discussed in review [14] and in references therein.

The $\Delta\sigma_L(pn)$ results from 0.51 to 5.1 GeV were deduced for the first time in 1981 from the $\Delta\sigma_L(pd)$ and $\Delta\sigma_L(pp)$ measurement at the ANL ZGS [15]. These pn results were omitted in many existing PSA databases due to uncertainties in the Glauber-type rescattering corrections. They were discussed in [3, 4, 14].

Using free polarized neutrons at Saturne II, $\Delta\sigma_T(np)$ and $\Delta\sigma_L(np)$ results were obtained at 11 and 10 values of energy, respectively, in the range from 0.31 to 1.10 GeV [16–18]. The Saclay results were soon followed by PSI measurements [19] in seven energy bins from 0.180 to 0.537 GeV, using a continuous neutron energy spectrum. The PSI and Saclay sets allowed one to deduce the imaginary parts of np and $I = 0$ spin-dependent forward-scattering amplitudes [14, 18]. Only $\Delta\sigma_L(np)$ was measured at five energies between 0.484 and 0.788 GeV at LAMPF [20]. There, a quasi-monoenergetic polarized neutron beam was produced in $pd \to n + X$ scattering of longitudinally polarized protons.

At low energies, $\Delta\sigma_L(np)$ was measured at 66 MeV at the PSI injector [21] and at 16.2 MeV in Prague [22]. The $\Delta\sigma_T(np)$ values were determined in TUNL at nine energies between 3.65 and

[3)]Dzhelepov Laboratory of Nuclear Problems, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
[4)]Kharkov Institute of Physics and Technology, Kharkov, Ukraine.
[5)]Frank Laboratory of Neutron Physics, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
[6)]Charles University, Faculty of Mathematics and Physics, Praha, Czech Republic.
[7)]Institute for Nuclear Research, Russian Academy of Sciences, pr. Shestidesyatiletiya Oktyabrya 7a, Moscow, 117312 Russia.
[8)]Petersburg Nuclear Physics Institute, Russian Academy of Sciences, Gatchina, 188350 Russia.
[9)]DAPNIA, CEA/Saclay, France.
[10)]Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria.
[11)]Laboratory of Particle Physics, Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia.
[**]e-mail: sharovsunhe.jinr.ru

11.6 MeV [23] and at 16.2 MeV in Prague [24]. Recently TUNL measured $\Delta\sigma_L(np)$ at six energies between 4.98 and 19.7 MeV [25] and $\Delta\sigma_T(np)$ at three other energies between 10.7 and 17.1 MeV [26].

At high energies, the $\Delta\sigma_L(np)$ results using free polarized neutrons were obtained at the JINR Synchrophasotron only. The Dubna results smoothly connect with the existing data at lower energies. The $-\Delta\sigma_L(np)$ energy dependence shows a fast decrease to zero above 1.1 GeV and a structure around 1.8 GeV. Values of the $I = 0$ part of $\Delta\sigma_L$ are also presented. The data are compared with model predictions and the PSA fits.

Section 2 gives a brief determination of observables. Section 3 describes the method of measurements. The essential details concerning the beam, the polarimeters, the experimental setup, and the PPT are given in Section 4. Data acquisition and analyses are described in Section 5. Results and discussion are presented in Section 6.

2. DETERMINATION OF OBSERVABLES

Throughout this paper, we use the NN formalism and notation for elastic NN scattering observables from [27].

The general expression of the total cross section for a polarized nucleon beam transmitted through the PPT, with arbitrary directions of the beam and target polarizations $\mathbf{P}_B$ and $\mathbf{P}_T$, respectively, was first deduced in [28, 29]. Taking into account fundamental conservation laws, it is written in the form

$$\sigma_{\rm tot} = \sigma_{0\rm tot} + \sigma_{1\rm tot}(\mathbf{P}_B \cdot \mathbf{P}_T) + \sigma_{2\rm tot}(\mathbf{P}_B \cdot \mathbf{k})(\mathbf{P}_T \cdot \mathbf{k}), \tag{2.1}$$

where $\mathbf{k}$ is a unit vector in the direction of the beam momentum. The term $\sigma_{0\rm tot}$ is the total cross section for unpolarized particles, and $\sigma_{1\rm tot}$ and $\sigma_{2\rm tot}$ are the spin-dependent contributions. They are related to the measurable observables $\Delta\sigma_T$ and $\Delta\sigma_L$ (called total-cross-section differences) by

$$-\Delta\sigma_T = 2\sigma_{1\rm tot}, \tag{2.2}$$

$$-\Delta\sigma_L = 2(\sigma_{1\rm tot} + \sigma_{2\rm tot}). \tag{2.3}$$

The negative signs for $\Delta\sigma_T$ and $\Delta\sigma_L$ in Eqs. (2.2) and (2.3) correspond to the usual, although unjustified, convention in the literature. The total-cross-section differences are measured with either the parallel or the antiparallel beam and target polarization directions. Polarization vectors are transversely oriented with respect to $\mathbf{k}$ for $\Delta\sigma_T$ measurements and longitudinally oriented for $\Delta\sigma_L$ experiments. Only $\Delta\sigma_L$ measurements are treated below, but the formulas are similar for the both total-cross-section differences.

For $\mathbf{P}_B^{\pm}$ and $\mathbf{P}_T^{\pm}$, all oriented along $\mathbf{k}$, we obtain four total cross sections:

$$\sigma(\rightrightarrows) = \sigma(++) = \sigma_{0\rm tot} + |P_B^+ P_T^+|(\sigma_{1\rm tot} + \sigma_{2\rm tot}), \tag{2.4a}$$

$$\sigma(\leftrightarrows) = \sigma(-+) = \sigma_{0\rm tot} - |P_B^- P_T^+|(\sigma_{1\rm tot} + \sigma_{2\rm tot}), \tag{2.4b}$$

$$\sigma(\rightleftarrows) = \sigma(+-) = \sigma_{0\rm tot} - |P_B^+ P_T^-|(\sigma_{1\rm tot} + \sigma_{2\rm tot}), \tag{2.4c}$$

$$\sigma(\leftleftarrows) = \sigma(--) = \sigma_{0\rm tot} + |P_B^- P_T^-|(\sigma_{1\rm tot} + \sigma_{2\rm tot}). \tag{2.4d}$$

The signs in brackets correspond to the $\mathbf{P}_B$ and $\mathbf{P}_T$ directions with respect to $\mathbf{k}$ in this order. In principle, an arbitrary pair of one parallel and one antiparallel beam and target polarization directions determines $\Delta\sigma_L$. By using two independent pairs, we remove an instrumental asymmetry term.

Below the neutron beam and the proton target are considered. Since the $\mathbf{P}_B$ direction at the Synchrophasotron could be reversed every cycle of the accelerator, it is preferable to calculate $\Delta\sigma_L$ from the pairs $(\rightrightarrows)$, $(\leftrightarrows)$ and $(\rightleftarrows)$, $(\leftleftarrows)$, measured with the same $\mathbf{P}_T$ orientation. It helps to avoid long-time efficiency fluctuations of the neutron detectors. The spin-averaged term $\sigma_{0\rm tot}$ drops out when taking the difference, and one obtains

$$-\Delta\sigma_L(P_T^+) = 2(\sigma_{1\rm tot} + \sigma_{2\rm tot})^+ = \frac{2[\sigma(\rightrightarrows) - \sigma(\leftrightarrows)]}{(|P_B^+| + |P_B^-|)|P_T^+|}, \tag{2.5a}$$

$$-\Delta\sigma_L(P_T^-) = 2(\sigma_{1\rm tot} + \sigma_{2\rm tot})^- = \frac{2[\sigma(\leftleftarrows) - \sigma(\rightleftarrows)]}{(|P_B^+| + |P_B^-|)|P_T^-|}. \tag{2.5b}$$

Measured differences $-\Delta\sigma_L$, i.e., the asymmetry effect between the total cross sections for parallel and antiparallel orientations of the beam and target polarization, are proportional to the mean value of the beam polarizations $|P_B^+|$ and $|P_B^-|$:

$$|P_B| = \frac{1}{2}(|P_B^+| + |P_B^-|). \tag{2.6}$$

The $|P_B|$ value is well known as a function of time, because it is continuously monitored by a beam polarimeter.

Each of relations (2.5a) and (2.5b) contains a hidden contribution from the instrumental asymmetry (IA), caused mainly by a misalignment of the neutron detector counters (see below). The value of IA is given as

$$\mathrm{IA} = \frac{1}{2}[\Delta\sigma_L(P_T^+) - \Delta\sigma_L(P_T^-)]. \tag{2.7}$$

The IA disappears, giving the final results as a simple average:

$$\Delta\sigma_L = \frac{1}{2}[\Delta\sigma_L(P_T^+) + \Delta\sigma_L(P_T^-)]. \quad (2.8)$$

The scattering matrix used contains the invariant amplitudes a, b, c, d, and e, as defined in [14, 27]. The $\sigma_{0\mathrm{tot}}$, $\Delta\sigma_T$, and $\Delta\sigma_L$ are linearly related to the imaginary parts of the three independent forward scattering invariant amplitudes $a+b$, c, and d [27] via optical theorems:

$$\sigma_{0\mathrm{tot}} = (2\pi/K)\mathrm{Im}[a(0)+b(0)], \quad (2.9)$$

$$-\Delta\sigma_T = (4\pi/K)\mathrm{Im}[c(0)+d(0)], \quad (2.10)$$

$$-\Delta\sigma_L = (4\pi/K)\mathrm{Im}[c(0)-d(0)], \quad (2.11)$$

where K is the c.m. momentum of the incident nucleon. For the amplitudes in Eqs. (2.9)–(2.11) we have $a(0)-b(0) = c(0)+d(0)$. The optical theorems always provide the absolute amplitudes, as discussed in [30–32].

Using the measured $\Delta\sigma(np)$ values and the existing $\Delta\sigma(pp)$ data at the same energy, one can deduce $\Delta\sigma_{L,T}$ $(I=0)$ as

$$\Delta\sigma_{L,T}(I=0) = 2\Delta\sigma_{L,T}(np) - \Delta\sigma_{L,T}(pp). \quad (2.12)$$

3. METHOD OF MEASUREMENT

In the transmission experiment, we have measured the part of the incident beam particles which remain in the beam after its passage through the target. For the experiments with incident neutrons, this measurement is always relative. The neutron beam has a circular profile, formed by the preceding beam collimators. Out of the collimator diameter, the neutron flux is considered to be zero. The neutron-beam intensity is monitored by neutron-beam monitors placed upstream from the target. The target material consists of small beads inserted in a cylindrical container of the circular profile. The container covers the beam spot and its horizontal axis coincides with the beam axis. The transmission detectors, downstream from the target, are larger than the beam dimensions. Any unscattered beam particle is detected with the same probability.

If N_{in} is the number of neutrons entering the target and N_{out} is the number of neutrons transmitted in a counter array of solid angle Ω, then the total cross section $\sigma(\Omega)$ is related to measured quantities as

$$\frac{N_{\mathrm{out}}}{N_{\mathrm{in}}} = \exp\bigl(-\sigma(\Omega)nd\bigr), \quad (3.1)$$

where n is the number of all target atoms per cm^3, d is the target length, and $N_{\mathrm{out}}/N_{\mathrm{in}}$ is the simple transmission ratio. The number of counts of the beam monitor M and of the transmission counter T depends on the efficiency of each detector, i.e., $M = N_{\mathrm{in}}\eta(M)$ and $T = N_{\mathrm{out}}\eta(T)$. The extrapolation of $\sigma(\Omega)$ towards $\Omega = 0$ gives the unpolarized total cross section $\sigma_{0\mathrm{tot}}$.

In the $\Delta\sigma_L(\Omega)$ measurements with a completely filled target, only the number of polarizable hydrogen atoms n_{H} is important, because σ_{tot} depends on the polarizations $P_B^{\pm}$ and $P_T^{\pm}$ as shown in (2.4). If one sums over the events taken with one fixed target polarization P_T^+ or P_T^- and using Eqs. (2.5a) or (2.5b), the double transmission ratios of the measurements with the averaged P_B from Eq. (2.6) for the two $\mathbf{P}_T$ directions become

$$\frac{N_{\mathrm{out}}(++)/N_{\mathrm{in}}(++)}{N_{\mathrm{out}}(-+)/N_{\mathrm{in}}(-+)} = \exp(-\Delta\sigma_L(\Omega)|P_BP_T^+|n_{\mathrm{H}}d), \quad (3.3a)$$

$$\frac{N_{\mathrm{out}}(--)/N_{\mathrm{in}}(--)}{N_{\mathrm{out}}(+-)/N_{\mathrm{in}}(+-)} = \exp(-\Delta\sigma_L(\Omega)|P_BP_T^-|n_{\mathrm{H}}d). \quad (3.3b)$$

We use here the notation of (2.4).

Thus, the neutron detector efficiencies drop out. Further, we set $N = N_{\mathrm{out}}/N_{\mathrm{in}}$ depending on the $\mathbf{P}_B$ and $\mathbf{P}_T$ combination and Eqs. (3.3a), (3.3b) provide

$$-\Delta\sigma_L(\Omega, P_T^+) = \frac{1}{|P_BP_T^+|n_{\mathrm{H}}d}\ln\left(\frac{N(++)}{N(-+)}\right), \quad (3.4a)$$

$$-\Delta\sigma_L(\Omega, P_T^-) = \frac{1}{|P_BP_T^-|n_{\mathrm{H}}d}\ln\left(\frac{N(--)}{N(+-)}\right). \quad (3.4b)$$

A systematic uncertainty of the $-\Delta\sigma_L$ value is mainly caused by the errors in $|P_B|$, $|P_T|$, and n_{H}. A statistical error of $-\Delta\sigma_L$ is given by the formula

$$\delta_{\mathrm{stat}} = \frac{1}{|P_BP_T|n_{\mathrm{H}}d}\sqrt{\frac{1}{M^+}+\frac{1}{M^-}+\frac{1}{T^+}+\frac{1}{T^-}}, \quad (3.5)$$

where the M^+, M^- and T^+, T^- are the statistics for monitor and transmission neutron detectors with the P_B^+ and P_B^- neutron-beam polarizations, respectively.

If $\Omega \to 0$, we obtain $\Delta\sigma_L(\Omega) \to \Delta\sigma_L$.

For the np transmission measurement, we may neglect the extrapolation of $\Delta\sigma_L(\Omega)$ towards $\Omega = 0$ due to the small sizes of detectors [3–6]. The Saclay–Geneva (SG) PSA [31] at 1.1 GeV shows that, for the angles covered by our detectors, the resulting $-\Delta\sigma_L$ value decreases by 0.04 mb if we neglect the extrapolation procedure.

The ratio of $n_{\rm H}$ to other target nuclei depends on the target material. The presence of carbon in the PPT beads adds the term $\sigma_{\rm tot}({\rm C})$ in (2.4). This term is spin-independent and its contribution drops out in differences (2.5). The same occurs for ^{16}O and ^{4}He in the target and for the cryogenic envelopes.

However, there are small effects from ^{13}C and ^{3}He, which may be slightly polarized. The global contribution was estimated to be $\pm 0.3\%$ in [3–4].

4. EXPERIMENTAL SETUP

The $\Delta\sigma_L(np)$ experimental setup was described in detail in our previous publications [3–6]. Here, we briefly mention essential items which are important for the data analysis and results, as well as the items concerning modifications and improvements of the apparatus and of the experimental conditions.

Figure 1 shows both polarized-deuteron and free polarized-neutron beam lines [11], the two polarimeters [33, 34], the beryllium target (BT) for neutron production, the collimators $C1-C4$, the spin rotation magnet (SRM), the PPT [8, 9], the monitors $M1$, $M2$ of neutron-beam intensity, the transmission detectors $T1$, $T2$, $T3$, and the detectors for neutron-beam profile monitoring NP. The associated electronics was described in [3, 4]. The data acquisition system is based on a CAMAC parallel branch highway controlled by an IBM PC with the branch driver [12] finished off by one of the authors. The on-line program in Pascal works under DOS.

Accelerated deuterons were extracted at the beam momenta p_d of 4.29, 4.93, 5.36, and 5.57 GeV/c, which were known with a sufficient accuracy of $\approx 1\%$. The average intensity of the primary polarized deuteron beam was $\approx 2 \times 10^9$ d/cycle. It was continuously monitored by means of two calibrated ionization chambers placed in the two focal points upstream of the neutron-production target BT.

The beam of free quasi-monochromatic neutrons, polarized along the vertical direction, was obtained by breakup at 0° of vector polarized deuterons in the BT. Neglecting the BT thickness, neutrons have a laboratory momentum $p_n = p_d/2$ with a Gaussian momentum spread of FWHM $\simeq 5\%$ [35]. This corresponds to the neutron-beam energies $T_{\rm kin}(n)$ of 1.40, 1.70. 1.90, and 2.00 GeV, respectively. The BT contains 20 cm of Be with a cross section of 8×8 cm. Energy losses of deuterons during their passage through air, foils in vacuum tubes of the beam transport line, and the BT matter provided the decrease of 20 MeV in the deuteron-beam energy in the BT center and the mean neutron-energy decrease of 10 MeV [3, 5]. For the $\Delta\sigma_L$ results, the energies and laboratory momenta in the BT center are quoted; for the beam polarization measurements, the extracted beam energies have been used.

After the deuteron breakup, the resulting neutrons and protons have the same values and directions of the polarizations $\mathbf{P}_B(n)$ and $\mathbf{P}_B(p)$, respectively, as the vector polarization $\mathbf{P}_B(d)$ of the incident deuteron beam [36, 37].

In our previous experiments [3–6], the deuteron-beam polarization was determined by two independent asymmetry measurements, either for the dp elastic or for the pp quasielastic reactions. In the first case, the four-arm beam line polarimeter [33] was used and deuterons, scattered on the liquid hydrogen target, were analyzed by the magnetic field. This polarimeter worked in the primary deuteron-beam line. It accurately determined the dp-elastic-scattering asymmetry at $T_{\rm kin}(d) = 1.60$ GeV, where the analyzing powers of this reaction are well known [38]. In principle, the $\mathbf{P}_B(d)$ needs to be determined at one energy value only, since no deuteron depolarizing resonance at the Synchrophasotron exists [33]. On the other hand, the measurement requires changing the deuteron energy and extracting deuterons in another beam line, which is a time-consuming operation. For this reason, the dp polarimeter was not used in the 2001 run. In [5, 6], we obtained an average for positive and negative signs of the vector polarization: $|P_B(d)| = 0.524 \pm 0.010({\rm stat.}) \pm 0.010({\rm syst.})$.

Another four-arm beam polarimeter [34] with small acceptance of 7.1×10^{-4} sr continuously monitored the $\mathbf{P}_B(p)$ value during the data acquisition. The deuteron beam, considered as a beam of quasi-free protons and neutrons, was scattered on a CH_2 target at 14° lab. This polarimeter measured the pp left–right asymmetry on hydrogen and carbon at $T_{\rm kin}(p) = T_{\rm kin}(d)/2$ and the pC asymmetry was subtracted. The polarimeter was calibrated and improved [39, 40]. The average value $|P_B(p)| \approx |P_B(d)| = (|P_B^+| + |P_B^-|)/2$ of deuteron-beam polarization was continuously measured by this device during the 2001 data-acquisition run. These $|P_B(d)|$ values were taken into account for the data treatment. The weighted average $|P_B(d)|$ over the 2001 run is $|P_B(d)| = 0.528 \pm 0.004({\rm stat.}) \pm 0.008({\rm syst.})$ and agrees with the previous dp polarimeter results very well.

The dimensions and positions of the iron and brass collimators $C1-C4$ (Fig. 1) were as described in [3, 4]. Accurate measurements of the collimated neutron-beam profiles were performed in a dedicated run, using nuclear emulsions. During the data acquisition, the position and X, Y profiles were monitored by the neutron-beam profile monitor NP. It was equipped with multiwire proportional chambers and

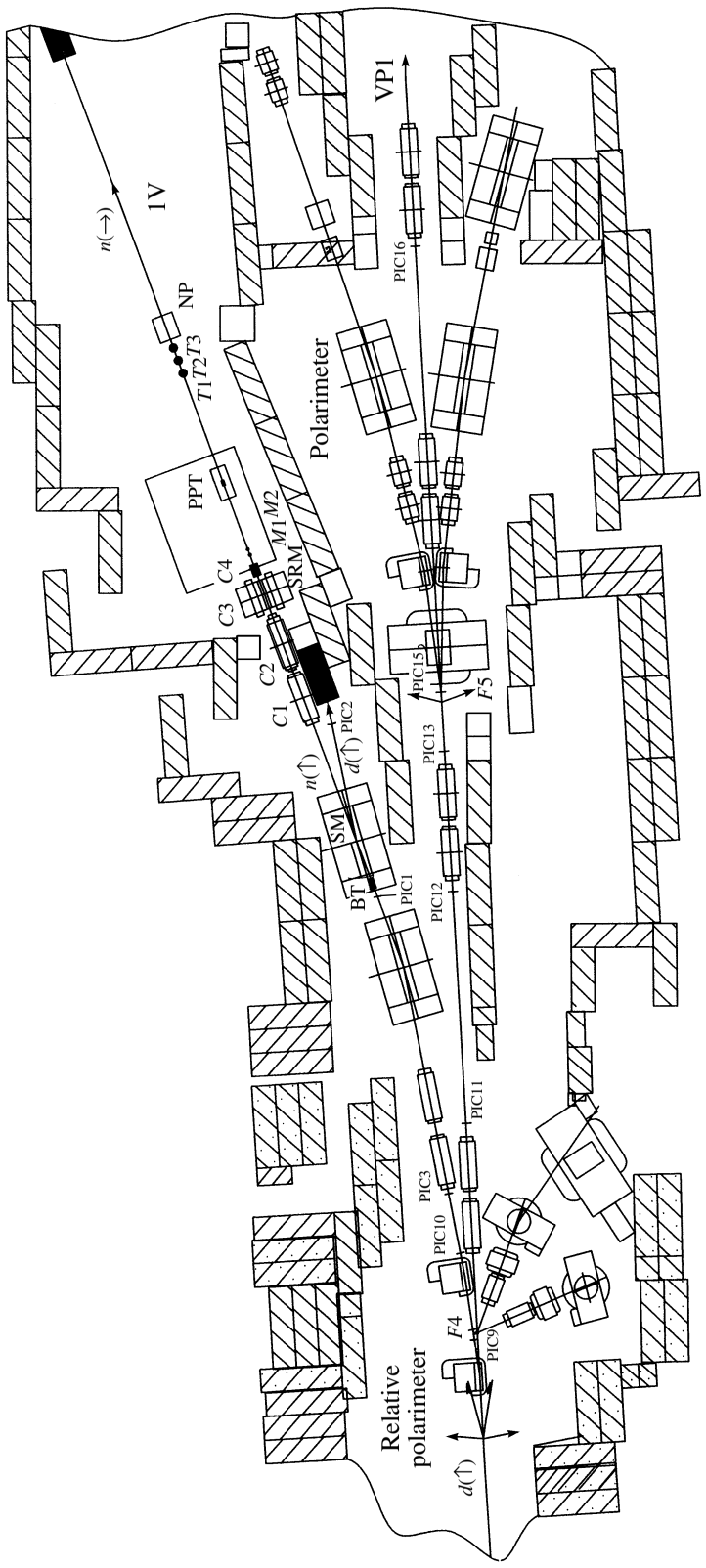

Fig. 1. Experimental setup for the $\Delta\sigma_L(np)$ measurement. Layout of the setup in the experimental hall: VP1—beam line of extracted vector polarized deuterons with $\mathbf{P}_B(d)$ oriented along the vertical direction $d(\uparrow)$; 1V—beam line to generate and form the polarized neutron beam, $n(\uparrow)$—neutrons polarized vertically, $n(\rightarrow)$—neutrons polarized longitudinally; $F4$, $F5$—foci of the extracted deuteron beam line; BT—beryllium target for neutron production; IC—ionization chamber for monitoring the deuteron-beam intensity; PIC1–PIC3, PIC9–PIC16—multiwire proportional/ionization chambers to measure the deuteron-beam profiles; SM—sweeping magnet; $C1$–$C4$—set of neutron-beam collimators; SRM—neutron spin rotating dipole; PPT—polarized proton target; and NP—neutron-beam profilometer.

Table 1

Beam polarization over the run	±0.9%
Target polarization	±5.0%
Number of polarizable hydrogen atoms	±1.1%
Polarization of other atoms	±0.3%
Magnetic field integral of the neutron spin rotator	±0.2%
Inefficiencies of veto counters	±0.1%
Total of the relative systematic errors	±5.3%
Absolute error due to the extrapolation of results towards 0°	<0.04 mb

Table 2. Unpolarized total cross sections for np and nC interactions

$T_{\rm kin}(n)$, GeV	$\sigma_{0\rm tot}(np)$, mb	$\sigma_{0\rm tot}(n{\rm C})$, mb
1.30 ± 0.013	41.35 ± 0.66	–
1.40 ± 0.014	39.18 ± 0.48	381.5 ± 2.6
1.50 ± 0.015	41.63 ± 0.83	379.6 ± 1.0

placed close downstream from the last transmission detector.

In order to change the vertical orientation of the neutron-beam polarization to the longitudinal direction, the SRM was used. The SRM magnetic field map was accurately measured, and the value of the magnetic field was continuously monitored by a Hall probe. The nonuniformity of the magnetic-field integral within the neutron-beam path area provides a small additional systematic error of ±0.2%.

The frozen-spin PPT reconstructed to the movable device [8, 9, 13] was used. The target material was 1-pentanol ($C_5H_{12}O + 5\%\ H_2O$) with a paramagnetic Cr^V impurity (EHBA) having a spin concentration of 7×10^{19} cm^{-3}. The pentanol beads were loaded in a thin-wall teflon container 200 mm long and 30 mm in diameter, placed inside the dilution refrigerator. The weight of the pentanol beads was 80.1 ± 0.05 g and the total number of polarizable hydrogen atoms on the beam neutron path was

$$n_{\rm H}d = (9.14 \pm 0.10) \times 10^{23}\ {\rm cm}^{-2}.$$

The PPT polarization $\mathbf{P}_T$ was measured using a computer-controlled nuclear-magnetic-resonance (NMR) system. These measurements were carried out during the run at the beginning and at the end of data taking for each sign of the PPT polarization. The negative and positive target proton polarizations were near the values of −0.75 and 0.60, respectively. The relaxation times were more than 5000 h for P_T^- and 7000 h for P_T^+. The relative uncertainty of the measured $\mathbf{P}_T$ values has been estimated at ±5%. This uncertainty includes the errors of polarization uniformity measurements using the NMR data from three coils placed inside the PPT container all along the target. The current values of the $\mathbf{P}_T$ were taken into account during the accumulated-data processing.

The configuration of the two neutron-intensity monitors $M1$ and $M2$ and the three transmission detectors $T1$, $T2$, and $T3$ is shown in Fig. 2. Each of the detectors was independent of the others. All the units were of similar design [16] and the electronic systems were identical, as described in [3, 4]. Each unit consisted of a CH_2 converter, 60 mm thick, placed behind a large veto scintillation counter A. The emitted forward charged particles, generated by neutron interactions in the convertor matter, were detected by two counters $S1$ and $S2$ in coincidence. The converters and $S1$, $S2$ counters for monitors $M1$ and $M2$ were 30 mm in diameter and the corresponding elements for the transmission detectors $T1$, $T2$, and $T3$ were 90, 92, and 94 mm, respectively.

The NP array, also shown in Fig. 2, was similar to the neutron detectors. The two multiwire proportional chambers behind the converter were protected by its own veto A and triggered by the dedicated $S1$, $S2$, and $S3$ counters in coincidence. The counter array used provides very good stability of the detection efficiency. The efficiencies of ≈2% for all detectors are practically constant with energy.

The result of $\Delta\sigma_L$ is independent of the neutron-beam intensity if the probability of quasi-simultaneous detection of two neutrons in one detector unit can be neglected. The small detection efficiencies decrease the probability for a converted neutron to be accompanied by another quasi-simultaneous converted neutron in the same detector. "Simultaneous" detection is to be understood within the resolution time of a scintillation counter. The probability was estimated from the results obtained with different neutron-beam intensities and radiator thicknesses [16, 17]. For the same neutron fluxes, the probability increases quadratically with increasing detector efficiency. At high efficiencies (namely, for pp transmission experiments), it represents the dominant source of systematic errors. This effect of the "simultaneous" detection of two neutrons in one detector unit was estimated to be smaller than 2×10^{-6} in the present experiment.

The misalignment of the detector components or of the entire detectors provides IA. Reaching a perfect alignment is beyond experimental possibility. In the case of misalignment, the asymmetries in each

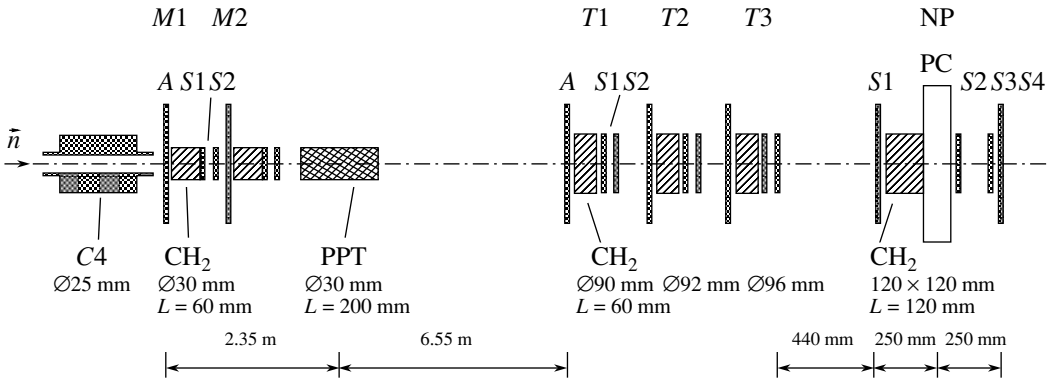

Fig. 2. Experimental setup for the $\Delta\sigma_L(np)$ measurement. Layout of the detectors for the neutron transmission measurement: $C4$—last neutron-beam collimator, 25 mm in diameter; $M1$, $M2$—monitor neutron detector modules; PPT—polarized proton target; $T1-T3$—neutron transmission detector modules; NP—neutron-beam profile monitor; CH_2—converters; A—anticoincidence scintillation counter; $S1-S4$—coincidence scintillation counters; PC—multiwire proportional chambers X, Y.

neutron detector depend on the transverse beam-polarization components only. For T detectors, the misalignment effects are practically independent of the target polarizations [16, 41]. The IA in Eq. (2.7) may be of the same order or larger than the transmission effects and provides the same contributions to each pair of measurements in Eqs. (2.5a) or (2.5b). It is obvious that this effect will be considerably stronger for the $\Delta\sigma_T$ measurement than for the $\Delta\sigma_L$ one, where only residual perpendicular $\mathbf{P}_B$ components exist. These undesirable components depend on the accuracy of the SRM current setting. As already mentioned, IA cancels out when taking the simple average of results in Eq. (2.8). The results strongly depend on the detector stabilities and their fixed positions over the data acquisition with both P_T signs.

However, there exist small instrumental random-like effects (RLE), provided, e.g., by temperature and magnetic-field fluctuations, beam-position variations, etc. They affect the stability of setup elements in an uncontrolled manner. For the final $\Delta\sigma_L$ results, uncertainties caused by these effects were taken into account using a special procedure of data treatment (see Section 5).

A possible inefficiency of the protection against charged particles by all veto counters may exist. Charged particles in the neutron beam are produced mainly in beam collimators, in CH_2 radiators of all M and T detectors, and in the target. Only a small fraction of the forward protons are polarized. They are produced in the polarized target by elastic scattering of polarized neutrons on free polarized protons close to $\theta_{\text{c.m.}} = 180°$. For the longitudinally polarized beam and target, one obtains a contribution from the spin correlation parameter $A_{ookk}(np)$ [27], which is included in the counting-rate asymmetry for the observable $\Delta\sigma_L$. This additional asymmetry was calculated in [3, 4] and may provide $\pm 0.1\%$ systematic error. Let us note that $A_{ookk}(180°, np)$ is one of the observables which determine the real parts of the forward-scattering amplitudes for the isospin $I = 0$ state [31]. Its measurement is foreseen in our future experiment.

For the measured $\Delta\sigma_L$ values, the relative normalization and systematic errors from different sources are summarized in Table 1.

Dedicated tests of the experimental setup were performed during additional runs with high-intensity unpolarized deuteron beams. The unpolarized-neutron-beam energies were 1.3, 1.4, and 1.5 GeV. We used the same transmission setup as described above, where PPT was removed and either liquid hydrogen or carbon targets were inserted in the neutron-beam line. Transmission ratios were completed by the corresponding empty target data.

These measurements allowed us to extract the total cross sections $\sigma_{0\text{tot}}(np)$ and $\sigma_{0\text{tot}}(n\text{C})$. For the nC experiment at $T_{\text{kin}}(n) = 1.5$ GeV, the transmission was measured using several carbon targets having different thickness. The obtained results for both total cross sections are summarized in Table 2.

Our results are shown in Figs. 3 and 4, where they are compared with existing data listed in compilations [42–44] for np and in [46] for nC interactions.

5. DATA ANALYSIS

For each accelerator cycle, the following main information was recorded and displayed by the data-acquisition system:

rates of the two calibrated ionization chambers used as primary deuteron-beam-intensity monitors;

rates of coincidences and accidental coincidences for the two neutron detectors $M1$ and $M2$ used as

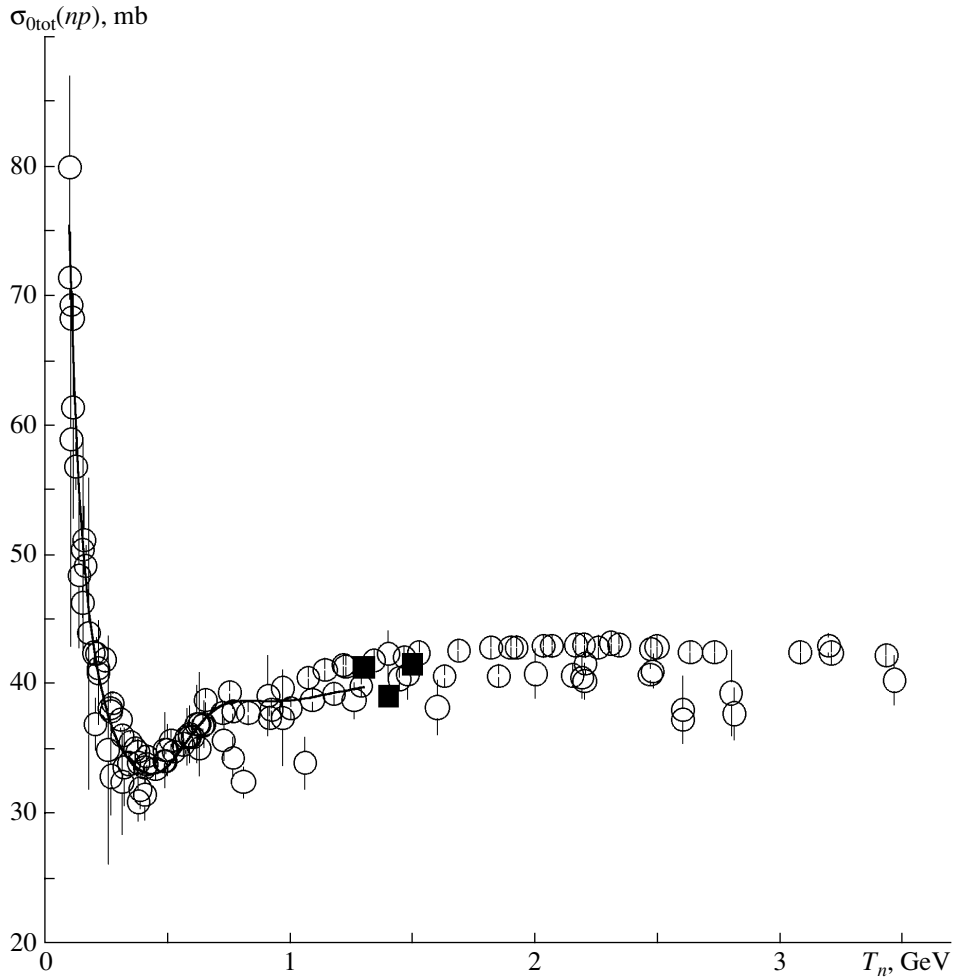

Fig. 3. Energy dependence of $\sigma_{0tot}(np)$: (○) world data from [42–44], (■) this experiment, and (solid curve) GW/VPI PSA [45] (SP03 solution).

the intensity monitors of the neutron beam incident on the PPT;

rates of coincidences and accidental coincidences for the three neutron transmission detectors $T1$, $T2$, and $T3$;

rates of the left and right arms of the pp-beam polarimeter.

At the beginning of the run, statistics at $T_{\rm kin} = 1.4$ GeV with P_T^+ and P_T^- were recorded. Then the data were taken at 1.7, 2.0, and 1.9 GeV with P_T^-. Finally, with P_T^+, the data were acquired at 1.9, 2.0, and 1.7 GeV.

The recorded data were then analyzed in three steps. In the first step, the data were cleaned of low-quality information. First, the "bad" data files were removed. Then the data files were cleaned of the "bad" cycles with an absence or incorrect sequence of labels of P_B signs. The number of such "bad" cycles for the cumulated statistics is a few tenths of percent. The data were also cleaned of the cycles with an absence ("empty" cycle) or a low-level beam intensity and/or with a possible fluctuation of the neutron detector performance. The main part of the rejected cycles ($\sim$5%) was caused by the low intensity of the neutron beam. The remaining event statistics over the run at a given energy and for each combination of the $\mathbf{P}_B$ and $\mathbf{P}_T$ directions were used to determine $\Delta\sigma_L$ and IA. The transmission ratios (3.3) averaged over the beam and target polarizations are proportional to σ_{0tot}(PPT) for all targets elements. The transmission ratios were listed in [5, 6] at three energies and are not shown here, since the conclusions are identical. With decreasing energy, σ_{0tot}(PPT) decreases slightly. A monotonic decrease in σ_{0tot}(PPT) as a function of the neutron transmission detector distance from the target is also observed, as expected.

The second and third steps of the data analysis used the previously selected events. The stabilities of the magnetic field of beam line elements and of the neutron detectors were checked, the parameters of

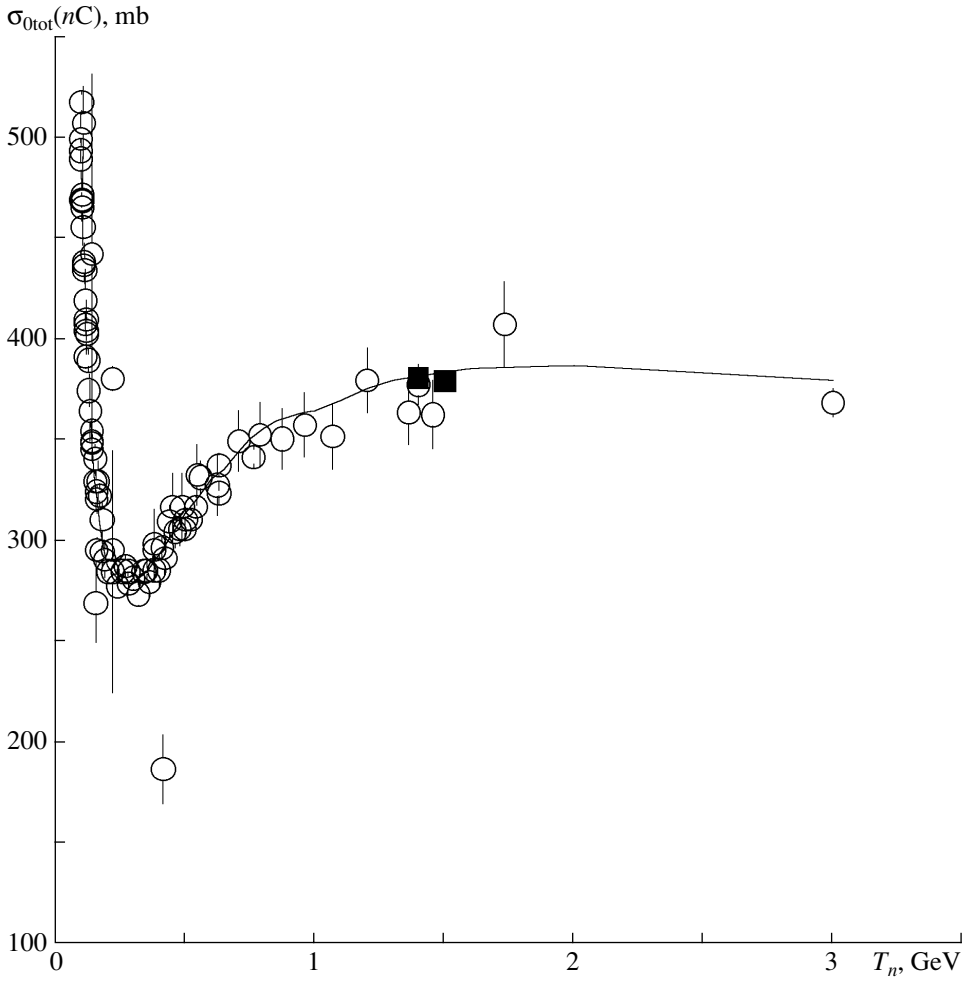

Fig. 4. Energy dependence of the $\sigma_{0\mathrm{tot}}(n\mathrm{C})$: (○) world data from [46], (■) this experiment, and (solid curve) [46].

statistical distributions of the $\Delta\sigma_L$ results for all pairs of neighboring cycles with P_B^+ and P_B^- for a sequence of data cycles accumulated over a single run were determined, and the final results were obtained. The transmission ratios as functions of time were analyzed for each combination of the individual M and T detectors, at any neutron energy and the P_T sign. No significant time dependence of checked values and no sizeable deviations from normal distributions were observed.

The $\Delta\sigma_L$ values over a given run were calculated by two different methods. In the first method, the $\Delta\sigma_L$ values were obtained using Eqs. (3.4a), (3.4b), and (3.5) for the entire statistics of each neutron detector ($M1$, $M2$, $T1$, $T2$, $T3$) accumulated over the run. For the second method, first the partial $\Delta\sigma_L$ values and their statistical errors were calculated from the individual neutron detector statistics for each "pair" of the following cycles with the opposite beam polarizations P_B^+ and P_B^-. Relations (3.4a), (3.4b), and (3.5) and the known time-dependent sequence of recorded cycles were used again. Then, obtained partial $\Delta\sigma_L$ values and their statistical errors for all "pairs" were added using an average weighting. The second method takes into account both the statistical uncertainties and the RLE.

The final results of the data treatment are presented in Table 3. The $-\Delta\sigma_L(np)$ values at four energies, measured with the individual transmission detectors $T1$, $T2$, $T3$ for both signs of P_T, their half-sums from Eq. (2.8), and half-differences, i.e., the hidden contributions of IA [16] from Eq. (2.7), are listed. The results were obtained using the combined statistics from the two monitors $M1$ and $M2$. The $T1$–$T3$ value at any energy represents the weighted average of the contributions of the three transmission detectors.

The $-\Delta\sigma_L$ values and their experimental errors presented in Table 3 have been calculated by the second ("pair") method. In order to estimate the RLE contribution to the experimental error, we calculated

Table 3. Measured $-\Delta\sigma_L(np)$ values at different neutron-beam energies T_n for the two opposite target polarizations, for the individual transmission detectors (TD), and for the cumulated statistics (the data presented were calculated using the procedure of average weighting of the results for each pair of the neighboring cycles from the sequence of all accumulated data cycles with P_B^+ and P_B^- beam polarizations; therefore, the errors quoted take into account both the statistical uncertainties and the instrumental RLE effects; since both the IA and the average $-\Delta\sigma_L(np)$ values have the same errors, they are indicated for the $-\Delta\sigma_L(np)$ values only)

T_n, GeV	TD	$-\Delta\sigma_L(P_T^+)$, mb	$-\Delta\sigma_L(P_T^-)$, mb	IA, mb	Average $-\Delta\sigma_L$, mb	R
1.39	$T1$	$+12.14 \pm 3.09$	$+3.33 \pm 2.48$	$+4.41$	$+7.73 \pm 1.98$	1.195
	$T2$	$+10.83 \pm 3.17$	$+2.92 \pm 2.56$	$+3.95$	$+6.87 \pm 2.04$	1.204
	$T3$	$+7.73 \pm 3.28$	$+1.23 \pm 2.65$	$+3.25$	$+4.48 \pm 2.11$	1.211
	$T1-T3$	$+10.32 \pm 1.83$	$+2.54 \pm 1.48$	$+3.69$	$+6.43 \pm 1.18$	1.036
1.69	$T1$	$+2.53 \pm 2.04$	$+0.68 \pm 1.58$	$+0.92$	$+1.60 \pm 1.29$	1.190
	$T2$	$+4.51 \pm 2.09$	-1.65 ± 1.62	$+3.08$	$+1.43 \pm 1.32$	1.197
	$T3$	$+6.08 \pm 2.17$	$+0.86 \pm 1.68$	$+2.61$	$+3.47 \pm 1.37$	1.203
	$T1-T3$	$+4.30 \pm 1.21$	-0.41 ± 0.94	$+2.17$	$+2.13 \pm 0.77$	1.037
1.89	$T1$	$+3.47 \pm 2.34$	$+3.08 \pm 2.46$	$+0.20$	$+3.27 \pm 1.70$	1.272
	$T2$	-0.84 ± 2.41	$+2.40 \pm 2.54$	-1.62	$+0.78 \pm 1.75$	1.284
	$T3$	$+1.17 \pm 2.49$	$+4.74 \pm 2.63$	-1.78	$+2.96 \pm 1.81$	1.295
	$T1-T3$	$+1.31 \pm 1.39$	$+3.37 \pm 1.47$	-1.03	$+2.34 \pm 1.01$	1.043
1.99	$T1$	$+0.76 \pm 2.03$	$+2.87 \pm 1.85$	-1.06	$+1.82 \pm 1.37$	1.222
	$T2$	-1.47 ± 2.09	$+3.82 \pm 1.90$	-2.64	$+1.17 \pm 1.41$	1.230
	$T3$	$+0.71 \pm 2.16$	$+2.35 \pm 1.97$	-0.82	$+1.53 \pm 1.46$	1.239
	$T1-T3$	$+0.0003 \pm 1.21$	$+3.03 \pm 1.10$	-1.51	$+1.51 \pm 0.82$	1.030

the ratio

$$R = \delta_{\text{“pair”}}/\delta_{\text{stat}} \geq 1, \qquad (5.1)$$

where the $\delta_{\text{“pair”}}$ is the error obtained by the second method and δ_{stat} is the error obtained by the first one.

The value of $R = 1$ occurs if no RLE exists and R increases with increasing RLE. In our experiment, the errors of the final results obtained by the “pair” method exceed the statistical errors obtained from Eq. (3.5) by ~4% (see last column in Table 3).

As can be seen from Table 3, the IA values at 1.39 and 1.69 GeV are positive, whereas at 1.89 and 1.99 GeV they are mostly negative. Since the elements of the neutron detectors were not moved during the run, we assume that the residual perpendicular components in $\mathbf{P}_B$ were opposite.

Table 4. Final $-\Delta\sigma_L(np)$ results (total errors are quadratic sums of the experimental and systematic errors; laboratory kinetic energies and momenta of the neutron beam in the production target center are given)

$T_{\text{kin}}(n)$, GeV	$p_{\text{Lab}}(n)$, GeV/c	$-\Delta\sigma_L(np)$, mb	Error, mb		
			exp.	syst.	total
1.39	2.13	+6.43	±1.18	±0.34	±1.23
1.69	2.46	+2.13	±0.77	±0.11	±0.78
1.89	2.67	+2.34	±1.01	±0.12	±1.02
1.99	2.77	+1.51	±0.82	±0.08	±0.82

6. RESULTS AND DISCUSSION

The final $-\Delta\sigma_L(np)$ values are presented in Table 4 and shown in Fig. 5. Statistical and systematic errors are taken into account. Total errors are the quadratic sums of experimental and systematic uncertainties.

The results from [3–6] together with the existing $-\Delta\sigma_L(np)$ data [16, 17, 19, 20], obtained with free polarized neutrons below 1.1 GeV, are also plotted in Fig. 5. In [5], we added the TUNL point at 19.7 MeV [25] and the PSI point at 66 MeV [21] in order to show the $-\Delta\sigma_L(np)$ energy dependence and existing structure at low energies. This structure (not

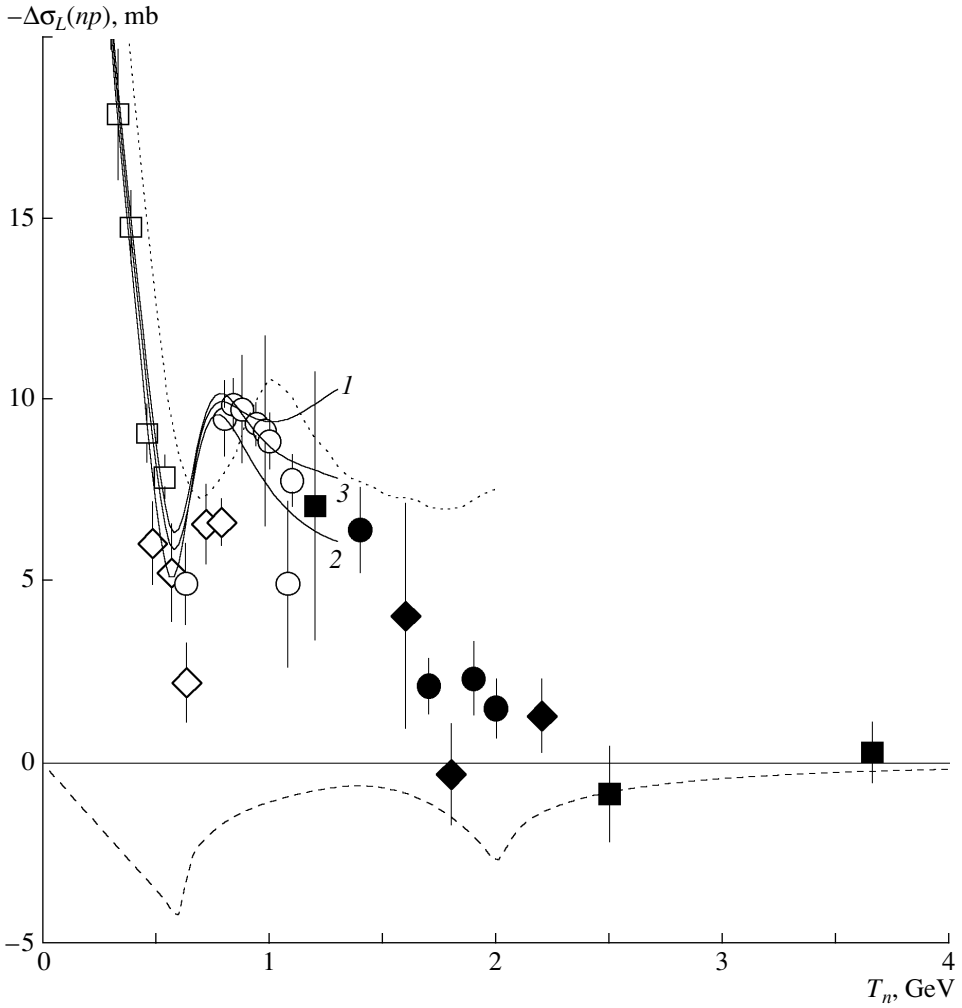

Fig. 5. Energy dependence of the $-\Delta\sigma_L(np)$: (●) this experiment; (■) JINR [3, 4]; (♦) JINR [5, 6]; (□) PSI [19]; (◊) LAMPF [20]; (○) Saturne II [16, 17]; (solid curves *1*, *2*, and *3*) FA85, SP99, and SP03 ED GW/VPI PSA solutions [45, 47], respectively; (dotted curve) meson-exchange model [48]; (dashed curve) contribution from nonperturbative QCD interaction induced by instantons [3, 4, 6].

shown here) was described by the energy-dependent (ED) GW/VPI PSA (solution SP99) [47] and is described again in the recent solution SP03 [45]. In the present paper, we discuss mainly the energy dependence over the high-energy region.

The new results agree with our previous data, confirming a fast decrease above 1.1 GeV, and suggest a minimum or a shoulder in the vicinity of 1.8 GeV. The quasielastic *pn* data from [15] (not shown here) are in good agreement with the Dubna results. The solid curves represent the evolution of the PSA fits of $-\Delta\sigma_L(np)$ (FA95, SP99, and SP03 GW/VPI PSA solutions [45, 47]) below 1.3 GeV. Above 0.6 GeV, the fits are only in qualitative agreement with the measured values.

From Eq. (2.12), one can deduce $-\Delta\sigma_L(I=0)$, using the obtained $-\Delta\sigma_L(np)$ results and the corresponding *pp* values at the same energies. In order to determine $-\Delta\sigma_L(pp)$, we have used different sources:

calculations using the ED GW/VPI PSA [45] (solution SP03);

linear interpolation of the fixed energy (FE) GW/VPI PSA [47] results, using the solutions at 1.275, 1.50, 1.70, 1.80, 1.95, and 2.025 GeV;

linear interpolation of the FE Saclay–Geneva (SG) PSA results [31], using the solutions at 1.3, 1.6, 1.8, 2.1 GeV;

interpolation of the *pp* experimental data measured at ANL ZGS [49, 50] and Saturne II [51] in the vicinity of energy values for the obtained $-\Delta\sigma_L(np)$ data.

The four calculated *pp* data sets are listed in the upper part of Table 5. Let us note that, from ED GW/VPI PSA, no errors could be calculated and,

Table 5. $-\Delta\sigma_L(pp)$ (in mb) values deduced from the four sources and four sets of $-\Delta\sigma_L(I=0)$ (in mb) results, calculated from the present np data and corresponding pp values

$T_{\rm kin}$, GeV	ED GW/VPI	FE GW/VPI	FE SG	Data interpolation
	$-\Delta\sigma_L(pp)$, mb			
1.39	$+7.83$	$+7.73 \pm 2.13$	$+7.82 \pm 0.30$	$+6.45 \pm 0.53$
1.69	$+4.17$	$+4.14 \pm 1.71$	$+4.35 \pm 0.30$	$+3.78 \pm 0.30$
1.89	$+2.60$	$+1.87 \pm 2.27$	$+3.09 \pm 0.30$	$+2.77 \pm 0.18$
1.99	$+2.07$	$+0.44 \pm 2.21$	$+2.82 \pm 0.20$	$+2.20 \pm 0.18$
	$-\Delta\sigma_L(I=0)$, mb			
1.39	$+5.03 \pm 2.38$	$+5.13 \pm 3.18$	$+5.04 \pm 2.38$	$+6.43 \pm 2.41$
1.69	$+0.09 \pm 1.58$	$+0.12 \pm 2.29$	-0.09 ± 1.58	$+0.18 \pm 1.58$
1.89	$+2.09 \pm 2.05$	$+2.82 \pm 3.04$	$+1.60 \pm 2.05$	$+1.92 \pm 2.03$
1.99	$+0.95 \pm 1.64$	$+2.59 \pm 2.75$	$+0.21 \pm 1.64$	$+0.82 \pm 1.64$

in FE GW/VPI PSA, they are obviously overestimated. SG PSA calculate errors using the error matrix, which are compatible with those obtained by the direct interpolation of neighborhood measured values.

The results of $-\Delta\sigma_L(I=0)$, using the four sets of pp values, are given in the bottom part of Table 5. We have added the SG PSA errors to the ED GW/VPI pp predictions. The results at each energy agree within the errors. Since, in general, the pp data are accurate, the $-\Delta\sigma_L(I=0)$ values have roughly two times larger errors than the np results. For this reason, an improved accuracy of np measurements is important.

New $-\Delta\sigma_L(I=0)$ results calculated with ED GW/VPI pp values are plotted in Fig. 6, together with other existing data in a large energy interval. The solid curves were calculated from np and pp ED GW/VPI PSA predictions [45, 47] (solutions FA95, SP99, SP03) below 1.3 GeV. The PSA fits described inaccurately the data starting from 0.5 GeV. In addition, for comparison, the energy dependence of isovector part, $-\Delta\sigma_L(I=1)$ calculated from ED GW/VPI (solution SP03) is shown by the dotted curve.

The new results agree again with our previous data and confirm a plateau around 1.4 GeV followed by a fast decrease and suggest a minimum in the vicinity of 1.8 GeV. The $-\Delta\sigma_L(I=0)$ values deduced from the quasi-elastic pn data [15] above 1 GeV (not shown) are in good agreement with the present results.

Some dynamic models predicted the $-\Delta\sigma_{L,T}$ energy behavior for np and pp interactions. Below 2.0 GeV, a usual meson exchange theory of NN scattering [48] gives the $-\Delta\sigma_L(np)$ energy dependence as shown by the dotted curve in Fig. 5. It can be seen that this model provides only a qualitative description at low energies and disagrees considerably with the data above 1 GeV.

The articles [3, 4, 6] discussed the model of a nonperturbative flavor-dependent interaction between quarks, induced by a strong fluctuation of vacuum gluon fields, i.e., instantons. Estimation of such nonperturbative QCD contribution to the $-\Delta\sigma_L(np)$ energy behavior is shown in Fig. 5 by the dashed curve. One can see that this prediction disagrees with the experimental data.

The investigated energy region corresponds to a possible generation of heavy dibaryons with masses $M > 2.4$ GeV (see review [52]). For example, model [53, 54] predicts the formation of a heavy-dibaryon state with a color octet–octet structure.

The possible manifestation of exotic dibaryons in the energy dependence of different pp and np observables was predicted by another model [55–59]. The authors used the Cloudy Bag Model and an R-matrix connection to the long-range meson-exchange force region with the short-range region of asymptotically free quarks. This hybrid model gives the lowest lying exotic six-quark configurations in the isosinglet and the spin-triplet state $^3S_1$ with the mass $M = 2.63$ GeV ($T_{\rm kin}(n) = 1.81$ GeV). This is close to the energy where a minimum is suggested by the energy dependence of our results.

For the $I = 0$ state, the $^3S_1$ partial wave is expected to be predominant. Since $-\Delta\sigma_T$ for an arbitrary isospin state contains no uncoupled spin-triplet state, a possible dibaryon resonance effect in $^3S_1$ may be less diluted. The measurement of $-\Delta\sigma_T$ observable for np interaction and deducing from these

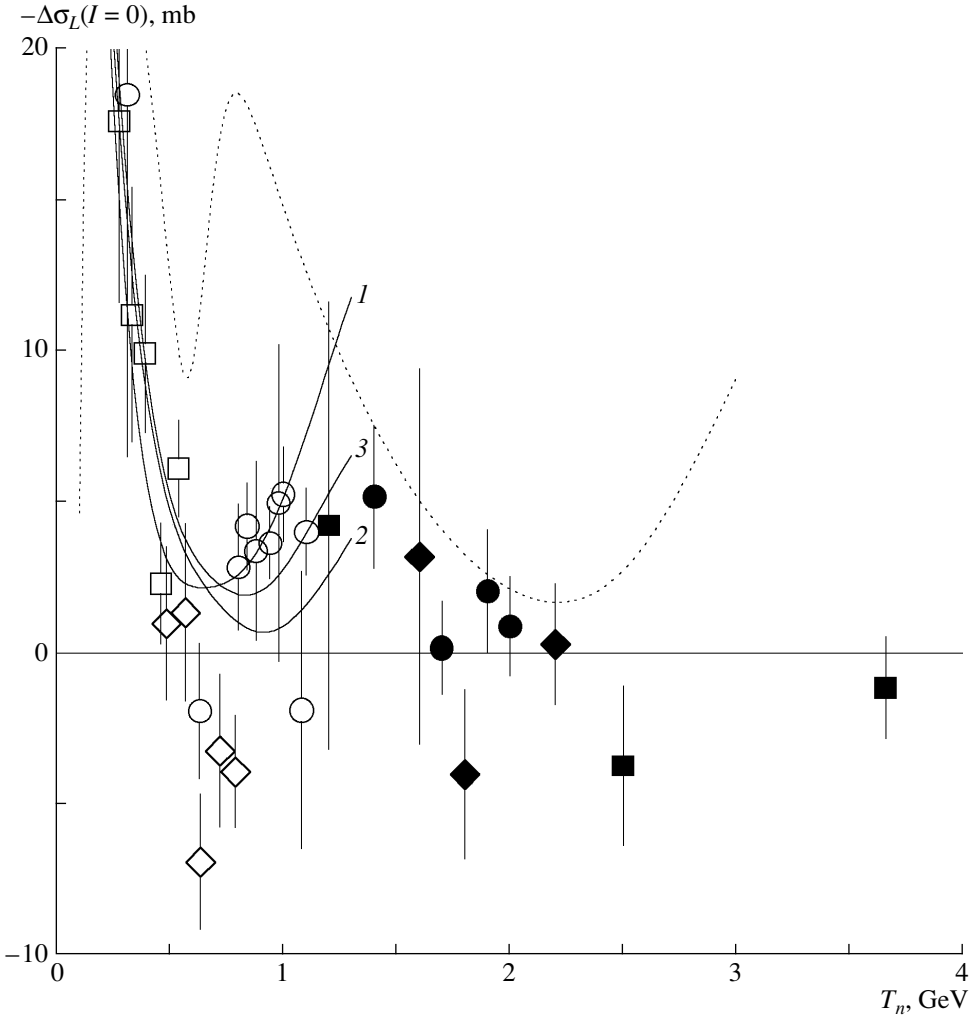

Fig. 6. Energy dependence of the $-\Delta\sigma_L(I=0)$ calculated from measured *np* data and the corresponding *pp* values from [45], solution SP03. The *np* data at the same energies as in Fig. 5 are used, and the $I=0$ results are denoted by the same symbols. All solid curves are the $I=0$ dependences calculated from the common *np* and *pp* GW/VPI PSA solutions [45, 47]: (*1*) FA95, (*2*) SP99, and (*3*) SP03. For comparison, the $-\Delta\sigma_L(I=1)$ dependence from GW/VPI PSA [45], solution SP03, is shown by the dotted curve.

data the $-\Delta\sigma_T(I=0)$ values may provide a significant more sensitive check of possible manifestation of the predicted dibaryon. Moreover, in the difference of both quantities, the spin-singlet contributions vanish. For this reason, more detailed and accurate measurements of energy dependences of $-\Delta\sigma_L(np)$ and $-\Delta\sigma_T(np)$ in the vicinity of $T_n = 1.8$ GeV are important.

The $I=0$ spin-dependent total cross sections represent a considerable advantage in seeking possible resonances. This is in contrast with the $I=1$ system, where the lowest lying exotic six-quark configuration corresponds to the spin-singlet state $^1S_0$ [59]. This state is not dominant and it is hard to separate it in the forward direction.

The three optical theorems determine the imaginary parts of the nonvanishing forward amplitudes, as shown in Eqs. (2.9)–(2.11). Extrema in $I=0$ amplitudes or in their combinations dominated by the spin-triplet states will be a necessary condition for the predicted resonance. The sufficient condition may be provided by real parts. For *np* scattering, they can be determined by measurements of observables in the experimentally accessible backward direction, as was shown in [32].

7. CONCLUSIONS

New $-\Delta\sigma_L(np)$ results, obtained in the transmission experiment, complete in the main the measurement of energy dependence at the Dubna Synchrophasotron region. Measured $-\Delta\sigma_L(np)$ values are in accordance with the existing data at lower

energies. The $-\Delta\sigma_L(np)$ energy dependence shows a fast decrease to zero above 1.1 GeV and a possible structure around 1.8 GeV. Values of the $I = 0$ part of $\Delta\sigma_L$ are also presented. The data are compared with model predictions and with the PSA fits.

The $-\Delta\sigma_L(I = 0)$ quantities, deduced from the measured $\Delta\sigma_L(np)$ values and the existing $-\Delta\sigma_L(pp)$ data, are also presented. They indicate a plateau or a weak maximum around 1.4 GeV, followed by a rapid decrease with energy and by a minimum around 1.8 GeV.

The obtained results are compared with the dynamic-model predictions and with the recent ED GW/VPI PSA fit. The necessity of more detailed and accurate $-\Delta\sigma_L(np)$ measurements around 1.8 GeV and new $-\Delta\sigma_T(np)$ data in the kinetic energy region above 1.1 GeV is emphasized.

ACKNOWLEDGMENTS

We are grateful to the Synchrophasotron staff and all the scientific and engineering groups and individuals who took part in and helped us during the $-\Delta\sigma_L$ measurement preparation, data acquisition, and data analysis. We also thank the JINR and JINR VBLHE and DLNP Directorates for support of these investigations.

For the last three years, this work was supported in part by the Russian Foundation for Basic Research (project no. 02-02-17129).

REFERENCES

1. J. Ball *et al.*, in *Proceedings of the International Workshop "Dubna Deuteron-91,"* Preprint No. E2-92-25, JINR (Joint Institute for Nuclear Research, Dubna, 1992), p. 12.
2. E. Chernykh *et al.*, in *Proceedings of the International Workshop "Dubna Deuteron-93,"* Preprint No. E2-94-95, JINR (Joint Institute for Nuclear Research, Dubna, 1994), p. 185; in *Proceedings of the V Workshop on High Energy Spin Physics, Protvino, 1993*, p. 478.
3. B. P. Adiasevich *et al.*, Z. Phys. C **71**, 65 (1996).
4. V. I. Sharov *et al.*, JINR Rapid Commun., No. 3[77]-96, 13 (1996).
5. V. I. Sharov *et al.*, Eur. Phys. J. C **13**, 255 (2000).
6. V. I. Sharov *et al.*, JINR Rapid Commun., No. 4[96]-99, 17 (1999).
7. F. Lehar *et al.*, Nucl. Instrum. Methods Phys. Res. A **356**, 58 (1995).
8. N. A. Bazhanov *et al.*, Nucl. Instrum. Methods Phys. Res. A **372**, 349 (1996).
9. N. A. Bazhanov *et al.*, Nucl. Instrum. Methods Phys. Res. A **402**, 484 (1998).
10. I. B. Issinsky *et al.*, Acta Phys. Pol. B **25**, 673 (1994).
11. A. Kirillov *et al.*, Preprint No. E13-96-210, JINR (Joint Institute for Nuclear Research, Dubna, 1996).
12. N. V. Gorbunov and A. G. Karev, in *Proceedings of the XII International Symposium on Nuclear Electronics, Varna, Bulgaria, 1988*, Preprint No. D13-88-938, JINR (Joint Institute for Nuclear Research, Dubna, 1989), p. 103.
13. N. G. Anischenko *et al.*, JINR Rapid Commun., No. 6[92]-98, 49 (1998).
14. C. Lechanoine-Leluc and F. Lehar, Rev. Mod. Phys. **65**, 47 (1993).
15. I. P. Auer *et al.*, Phys. Rev. Lett. **46**, 1177 (1981).
16. F. Lehar *et al.*, Phys. Lett. B **189**, 241 (1987).
17. J.-M. Fontaine *et al.*, Nucl. Phys. B **358**, 297 (1991).
18. J. Ball *et al.*, Z. Phys. C **61**, 53 (1994).
19. R. Binz *et al.*, Nucl. Phys. A **533**, 601 (1991).
20. M. Beddo *et al.*, Phys. Lett. B **258**, 24 (1991).
21. P. Haffter *et al.*, Nucl. Phys. A **548**, 29 (1992).
22. J. Brož *et al.*, Z. Phys. A **359**, 23 (1997).
23. W. S. Wilburn *et al.*, Phys. Rev. C **52**, 2351 (1995).
24. J. Brož *et al.*, Z. Phys. A **354**, 401 (1996).
25. J. R. Walston, PhD Thesis (North Carolina State University, 1998).
26. B. W. Raichle, PhD Thesis (North Carolina State University, 1997).
27. J. Bystrický, F. Lehar, and P. Winternitz, J. Phys. (Paris) **39**, 1 (1978).
28. S. M. Bilenky and R. M. Ryndin, Phys. Lett. **6**, 217 (1963).
29. R. J. N. Phillips, Nucl. Phys. **43**, 413 (1963).
30. J. Ball *et al.*, Nuovo Cimento A **111**, 13 (1998).
31. J. Bystrický, C. Lechanoine-Leluc, and F. Lehar, Eur. Phys. J. C **4**, 607 (1998).
32. J. Ball *et al.*, Eur. Phys. J. C **5**, 57 (1998).
33. V. G. Ableev *et al.*, Nucl. Instrum. Methods Phys. Res. A **306**, 51 (1991).
34. L. S. Azhgirey *et al.*, Prib. Tekh. Éksp., No. 1, 51 (1997) [Instrum. Exp. Tech., No. 1, 43 (1997)].
35. V. G. Ableev *et al.*, Nucl. Phys. A **393**, 941 (1983); **411**, 514(E) (1983).
36. E. Cheung *et al.*, Phys. Lett. B **284**, 210 (1992).
37. A. A. Nomofilov *et al.*, Phys. Lett. B **325**, 327 (1994).
38. V. Ghazikhanian *et al.*, Phys. Rev. C **43**, 1532 (1991).
39. L. S. Azhgirey *et al.*, Part. Nucl. Lett., No. 4 [113], 51 (2002).
40. L. S. Azhgirey *et al.*, Nucl. Instrum. Methods Phys. Res. A **497**, 340 (2003).
41. F. Perrot *et al.*, Nucl. Phys. B **278**, 881 (1986).
42. V. S. Barashenkov, *Interaction Cross Sections of Elementary Particles* (Nauka, Moscow, 1966; Israel Program for Scientific Translations, Jerusalem, 1968).
43. J. Bystrický *et al.*, *Landolt-Börnstein, New Series, Group I*, Ed. by K. H. Hellwege and H. Schopper (Springer-Verlag, Berlin, 1980), Vol. 9.
44. J. Bystrický and F. Lehar, *Nucleon–Nucleon Scattering Data*, Ed. by H. Behrens and G. Ebel (Fachinformationszentrum, Karlsruhe, 1981), Nos. 11-2 and 11-3.
45. R. A. Arndt, I. I. Strakovsky, and R. L. Workman, Phys. Rev. C **62**, 034 005 (2000).

46. V. S. Barashenkov, *Cross Sections of Interactions of Particles and Nuclei with Nuclei* (JINR, Dubna, 1993) [in Russian].
47. R. A. Arndt *et al.*, Phys. Rev. C **56**, 3005 (1997).
48. T.-S. H. Lee, Phys. Rev. C **29**, 195 (1984).
49. I. P. Auer *et al.*, Phys. Rev. Lett. **41**, 354 (1978).
50. I. P. Auer *et al.*, Phys. Rev. D **34**, 2581 (1986).
51. J. Bystrický *et al.*, Phys. Lett. B **142B**, 141 (1984).
52. I. I. Strakovsky, Fiz. Élem. Chastits At. Yadra **22**, 615 (1991) [Sov. J. Part. Nucl. **22**, 296 (1991)].
53. B. Z. Kopeliovich and F. Niedermayer, Zh. Éksp. Teor. Fiz. **87**, 1121 (1984) [Sov. Phys. JETP **60**, 640 (1984)].
54. B. Z. Kopeliovich, Fiz. Élem. Chastits At. Yadra **21**, 117 (1990) [Sov. J. Part. Nucl. **21**, 49 (1990)].
55. P. LaFrance and E. L. Lomon, Phys. Rev. D **34**, 1341 (1986).
56. P. Gonzales, P. LaFrance, and E. L. Lomon, Phys. Rev. D **35**, 2142 (1987).
57. P. LaFrance, Can. J. Phys. **68**, 1194 (1990).
58. E. L. Lomon, Colloq. Phys. (France) **51**, C6-363 (1990).
59. P. LaFrance and E. L. Lomon, in *Proceedings of the International Conference "Mesons and Nuclei at Intermediate Energies," Dubna, 1994*, Ed. by M. Kh. Khankhasaev and Zh. B. Kurmanov (World Sci., Singapore, 1995), Vol. XV, p. 97.

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1812–1829. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1874–1891.

ELEMENTARY PARTICLES AND FIELDS
Theory

Self-Consistent Description of the Inner Crust of a Neutron Star with Allowance for Superfluidity Effects

M. Baldo[1)], U. Lombardo[2)], É. E. Saperstein[3)*], and S. V. Tolokonnikov[3)]**

Received February 18, 2005

Abstract—Within the Wigner–Seitz approximation, a self-consistent fully quantum-mechanical calculation of the structure of the inner crust of a neutron star is performed over a wide range of densities with allowance for superfluidity effects. Within the approach used, the Wigner–Seitz cell consists of a nuclear-like cluster surrounded by a nearly uniform neutron gas. An effective energy functional is constructed by matching, at the cluster surface, the realistic phenomenological nuclear functional for the cluster due to S.A. Fayans and his coauthors and the energy functional calculated microscopically for neutron matter. The microscopic component of the functional is calculated within the Brueckner method by using the $v18$ Argonne interaction. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The ideas of superfluidity in neutron stars have a rather long history of about fifty years [1], but interest in this phenomenon has grown considerably in recent years in connection with the accumulation of observational data concerning irregularities in the spinning frequency of neutron stars. Various hypotheses were considered initially in order to explain this phenomenon, but the point of view prevalent at the present time is that the above irregularities are associated with the emergence and disappearance of superfluid vortices in the inner crust of a neutron star (see the review article of Pethick and Ravenhall [2] and references therein). By an inner crust, one usually means the neutron-star-envelope part that has a subnuclear density in the range $0.001\rho_0 \le \rho \le 0.5\rho_0$, where $\rho_0 \simeq 0.17$ fm^{-3} is a normal nuclear density. A crystalline "outer crust," which consists of ordinary (as a rule, neutron-rich) nuclei surrounded by electrons, has still lower densities. In the above inequality, the lower boundary is determined by the instability of highly neutron-rich nuclei against neutron emission, while the upper boundary is associated with the transition of neutron-star matter into a uniform state. According to present-day ideas, the bulk of the inner crust of a star consists of spherically symmetric nuclear-like clusters forming a crystal lattice that is submerged into a neutron sea and virtually uniform sea of ultrarelativistic electrons, which render the system electrically neutral. At some critical density of $\rho_c \simeq 0.5\rho_0$, a nonuniform configuration becomes energetically unfavorable, with the result that there arises a uniform neutron liquid featuring a small admixture of protons and electrons. In a narrow region around ρ_c, the cluster in question may lose a spherical shape, since rodlike ("spaghetti"), layer ("lasagna"), or some other exotic configurations become more favorable. Not only does the inclusion of neutron superfluidity provide an explanation for irregularities in the spinning frequency of neutron stars, this is also of importance for describing the cooling of a neutron star in the course of its evolution [2].

As a rule, one considers the superfluidity of only the neutron component, which is dominant, and de-

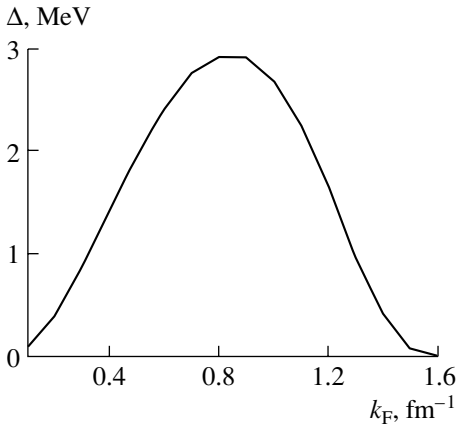

Fig. 1. Pairing gap in neutron matter.

[1)]Istituto Nazionale di Fisica Nucleare, Sezione di Catania, Catania, Italy.
[2)]Istituto Nazionale di Fisica Nucleare and Dipartimento di Fisica, Università di Catania, Corso Italia 57, I-95129 Catania, Italy.
[3)]Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, 123098 Russia.
*e-mail: saper@mbslab.kiae.ru
**e-mail: tolkn@mbslab.kiae.ru

scribes it within the local-density approximation—this means that, at each point $\mathbf{r}$, where the neutron density is $\rho_n(\mathbf{r})$, one identifies the superfluid gap $\Delta_n(\mathbf{r})$ with the gap for infinite neutron matter of the same density. The latter is usually calculated either within the Brueckner approach or within the variational method by using a realistic nucleon–nucleon potential, the results of different calculations being in good agreement. Fully quantum-mechanical calculations of the gap $\Delta_n(\mathbf{r})$ for the inner crust of a neutron star were performed in [3, 4], where the basic features of the Wigner–Seitz cell (the size R_c and the number Z of protons in the cell), as well as the form of the mean field, were taken according to the well-known study of Negele and Vautherin [5]. That study, which was performed 30 years ago, still remains the only fully quantum-mechanical self-consistent description of the inner crust of a neutron star within the Wigner–Seitz method for a wide range of densities, but we would like to emphasize that it does not take into account superfluidity. In [5], the authors determined, at a fixed value of the average density ρ, the equilibrium values of R_c and Z corresponding to the beta-stability condition

$$\delta\mu = \mu_n - (\mu_p + \mu_e) = 0, \qquad (1)$$

where μ_n, μ_p, and μ_e are the chemical potentials for, respectively, neutrons, protons, and electrons with allowance for their rest masses.

The discovery of a sizable shell effect changing the relation between ρ and Z that was predicted by semiclassical calculations that existed at that time (see, for example, [6]) was an important qualitative result of [5]. According to [5], two local minima of energy that correspond to the proton magic numbers of $Z = 40$ (giant quasinucleus of Zr) and $Z = 50$ (giant quasinucleus of Sn) and which are close to each other compete over a wide range of densities. In the calculations reported in [5], an increase in the density ρ led to the following changes in the hierarchy of the two minima in question: the former first proved to be the deeper and again became such in the vicinity of ρ_c, but the latter was the deeper at intermediate values of the density.

In recent years, there have appeared a number of quantum-mechanical calculations of the structure of the inner crust of a neutron star within the Hartree–Fock method [7–12]. Those calculations employed effective Skyrme forces—more specifically, their $SLy4$ and $Sly7$ versions, which were developed by the Lyons group precisely for describing neutron stars. In all of those studies, the authors explored phase transitions in the vicinity of ρ_c between the aforementioned phases, but we would like to emphasize that this was again done without taking into account pairing effects.

In [13–15], our group developed a method for self-consistently describing the structure of the inner crust of a neutron star, taking into account pairing correlations of neutrons and protons. The method is based on the approach proposed by Fayans and his coauthors [16, 17] in order to extend the standard Kohn–Sham method [18] of the energy density functional to superfluid systems. The generalized energy functional depends explicitly both on the usual (normal) densities $\rho_{n,p}(\mathbf{r})$ of neutrons and protons and on the anomalous densities $\nu_{n,p}(\mathbf{r})$ associated with pairing. The $\nu_{n,p}(\mathbf{r})$-dependent anomalous part of the generalized energy functional is expressed in terms of the normal-density-dependent effective pairing interaction. The parameters of the generalized energy functional were optimized in structural calculations of the properties of long isotopic chains of spherical nuclei [19]. The accuracy in describing mean nuclear features that was achieved in this way was not inferior (even being superior in some details) to the accuracy of the most successful Hartree–Fock calculations.

In [13], the generalized energy functional introduced by Fayans and his coauthors was used to study the effect of superfluidity on the structure of the inner crust of a neutron star within the Wigner–Seitz approximation. As a matter of fact, the method proposed in [5] was supplemented by taking into account the superfluidity of nucleons. The calculation was performed for only one value of the crust-matter density, that which corresponds to the Fermi momentum of $k_F = 0.7$ fm^{-1}. This choice was motivated by the following considerations: first, this density value is rather far off the critical density ρ_c, so that the spherical symmetry of the Wigner–Seitz cell is unquestionable; second, it belongs to the region where the pairing gap $\Delta(k_F)$ for neutron matter is maximal, in which case one would expect the maximum superfluidity effect. Figure 1 shows the gap calculated in the Bardeen–Cooper–Schrieffer (BCS) approximation with the $v18$ Argonne nucleon–nucleon potential. The inclusion of various many-body corrections to this approximation affects predominantly the absolute value of $\Delta(k_F)$ (reducing it by a factor of about 2), but this has virtually no effect on the shape of the curve [14].

A strong change in the equilibrium values of R_c and Z upon taking into account pairing was the main result of [13]. The analysis in [14, 15] was based on a more realistic model where the phenomenological generalized energy functional was applied only to describing the above nuclear-like cluster, while the neutron medium around it was described by the microscopic generalized energy functional calculated for neutron matter by using the Brueckner approach with the $v18$ nucleon–nucleon interaction. A smooth

matching of the two generalized energy functionals was performed at the boundary of the cluster. The calculations performed in those studies for the same value of k_F confirmed a strong effect of superfluidity on the ground state of the inner crust of a neutron star. The resulting equilibrium values of R_c and Z proved to be strongly different from those that were found on the basis of a purely phenomenological generalized energy functional. These were precisely the results that disclosed a sizable dependence of the equilibrium configuration on the model for the generalized energy functional.

In the present study, systematic calculations of the structure of the inner crust of a neutron star are performed for a wide range of densities corresponding to Fermi momenta in the range $k_F = 0.6$–1.2 fm^{-1}. The bulk of the calculations were performed with a more realistic semimicroscopic generalized energy functional [15], but, for some values of k_F, the calculations were also repeated for the phenomenological generalized energy functional used in [13].

2. GENERALIZED ENERGY FUNCTIONAL FOR THE INNER CRUST OF A NEUTRON STAR

In order to describe the inner crust of a neutron star, we will use the method that is based on a generalized energy functional in the coordinate representation with allowance for pairing and which was proposed in [16] and elaborated by Fayans and his coauthors (see, for example, [17, 19]). This approach generalizes the well-known Kohn–Sham method of an energy functional [18] to the case of superfluid systems. The correlation part of the generalized energy functional depends, on equal footing, on the normal densities of neutrons and protons, ρ_n and ρ_p, and on their anomalous analogs, ν_n and ν_p,

$$E^G_{\rm cor} = \int d\mathbf{r}\mathcal{E}^G_{\rm cor}(\rho_\tau(\mathbf{r}), \nu_\tau(\mathbf{r})), \qquad (2)$$

where $\tau = n, p$ is the isotopic index and $\mathcal{E}^G_{\rm cor}$ is the generalized-energy-functional density, which, for the sake of brevity, we will call merely the generalized energy functional. In order to render the formulas given below less cumbersome, we will also suppress, as a rule, the indices "G" and "cor."

A variation of expression (2) with respect to the densities ρ_τ yields equations for the mean potentials $U_\tau(\mathbf{r})$, while a variation with respect to the anomalous densities ν_τ leads to equations for the pairing gaps $\Delta_\tau(\mathbf{r})$ in terms of the densities ρ_τ and ν_τ. The mean fields U_τ and Δ_τ appear in the set of equations for the Bogolyubov functions $u_i(\mathbf{r})$ and $v_i(\mathbf{r})$, which in turn determine the densities ρ_τ and ν_τ according to the expressions known from [20]. Thus, there arises a set of equations that is equivalent to the set of equations in the Hartree–Fock–Bogolyubov method. It should be noted that the method of Gor'kov's equations, which yields close results, was used in [19].

In constructing the generalized energy functional that describes the Wigner–Seitz cell, we rely on the physical pattern that is obtained for the inner crust of a neutron star from the calculations reported in [5]. According to this pattern, almost all of the protons within this cell enter into the composition of a compact nuclear-like cluster occurring at the center of the cell. In many respects, the properties of this cluster are close to the properties of ordinary nuclei, but it is characterized by a high neutron excess. The medium surrounding the cluster is nearly homogeneous neutron matter. The Wigner–Seitz cell also contains ultrarelativistic electrons whose number is equal to the proton charge Z of the cell, this rendering the system electrically neutral.

In order to describe the nuclear-like cluster in question, we employ the phenomenological nuclear generalized energy functional that was introduced by Fayans *et al.* [19] and which was very successful in describing the masses and radii of spherical nuclei belonging to long isotopic chains and including nuclei far off the beta-stability valley. This functional is the sum of a normal and an anomalous (superfluid) component,

$$\mathcal{E}^{\rm ph} = \mathcal{E}^{\rm ph}_{\rm norm}(\rho_\tau) + \mathcal{E}^{\rm ph}_{\rm an}(\rho_\tau, \nu_\tau). \qquad (3)$$

The dominant, central, part of the normal component of the generalized energy functional has the form

$$\mathcal{E}^{\rm ph}_{\rm norm} = \frac{C_0}{4}\left[a_+\rho_+^2\frac{1 - h_1^+\rho_+/(2\rho_0)}{1 + h_2^+\rho_+/(2\rho_0)}\right. \qquad (4)$$
$$\left. + a_-\rho_-^2\frac{1 - h_1^-\rho_+/(2\rho_0)}{1 + h_2^-\rho_+/(2\rho_0)}\right],$$

where the normalization factor $C_0 = (dn/d\varepsilon_F)^{-1}$ is the inverse density of states at the Fermi surface; $\rho_{+,-} = \rho_n \pm \rho_p$; and $a_{+,-}$ and $h^{+;-}_{1,2}$ are dimensionless parameters. In addition to expression (4), the total normal part of the generalized energy functional [19] contains the Coulomb component, as well as the spin-orbit component and other spin-dependent components. Also, the total generalized energy functional includes a term that stems from taking into account a finite radius and which is close in physical meaning to the gradient term of the Skyrme functional.

The anomalous part of the generalized energy functional—it is responsible for superfluidity—has

the form

$$\mathcal{E}_{\rm an}^{\rm ph} = \frac{1}{2}\sum_{\tau=n,p}\mathcal{F}^{\xi,\tau}(\rho)|\nu_\tau(\mathbf{r})|^2, \qquad (5)$$

where $\mathcal{F}^{\xi,\tau}(\rho)$ is the density-dependent effective pairing interaction. In [19], it was taken in a τ-independent two-parameter form,

$$\mathcal{F}^{\xi,\tau}(\rho_+) = C_0\left(f_{\rm in}^{\xi}(\rho_+/\rho_0)^\alpha + f_{\rm ex}^{\xi}(1-(\rho_+/\rho_0)^\alpha)\right), \qquad (6)$$

where $\rho_0 = 0.165$ fm^{-3} and $\alpha = 2/3$. We note that the normalization of the effective pairing interaction in Eqs. (5) and (6) differs from the original normalization in [19] by a factor of 2, coinciding with the standard normalization adopted in the theory of finite Fermi systems [20]. The limiting case of a density-independent effective pairing interaction corresponds to the equality

$$f_{\rm in}^{\xi} = f_{\rm ex}^{\xi} = f^{\xi}. \qquad (7)$$

In [19], it was shown that fine details of the isotopic behavior of some nuclear properties (for example, the even–odd effect in root-mean-square radii) are sensitive to details in the density dependence of $\mathcal{F}^{\xi}$, favoring a pairing mechanism that involves a pronounced surface enhancement (a negative constant $f_{\rm ex}^{\xi}$ of very large magnitude versus $f_{\rm in}^{\xi}$ small in magnitude and negative in sign as well). At the same time, the majority of mean nuclear features are well described by the extremely simple one-parameter interaction (7), which corresponds to volume pairing. In this case, however, one has to introduce a weak mass-number dependence of the parameter f^{ξ}, changing it by about 20% in going over from Ca to Pb.

Strictly speaking, the effective pairing interaction is equal to the expression for $\mathcal{F}^{\xi,\tau}(\rho_+)$ multiplied by the function $\delta(\mathbf{r}_1-\mathbf{r}_2)$. As is well known, the delta-function form of the effective pairing interaction leads to a divergence in the equation for the gap, with the result that solving the problem requires introducing a model space S_0. In [19], the space S_0 was chosen in such a way that all single-particle energies satisfied the condition $|\varepsilon_\lambda^\tau - \mu^\tau| \simeq E_0^\tau$, where $E_0^n = E_0^p \simeq 35$ MeV. Since we basically follow the computational scheme adopted in [19] and retain the parameters of the generalized energy functional that were chosen there, we will use the same model space.

The physical reason why the parameter $f_{\rm ex}^{\xi}$ in (6) takes a negative value of large magnitude is that, beyond a nucleus, the effective pairing interaction is related to (asymptotically coincident with) the off-shell T matrix for free nucleon–nucleon scattering at zero value of all momenta and the negative energy $E = 2\mu \simeq -16$ MeV [21, 22],[4] this corresponding to a very strong attraction. However, the exact asymptotic relation $\mathcal{F}^{\xi} \to T(E=2\mu)$ is valid only in the case that was was considered in [22] and where the model space S_0 includes negative-energy single-particle states exclusively: $\varepsilon_\lambda < 0$; that is, $E_0 = -\mu$. Here, we are dealing with a wider model space, so that the asymptotic behavior of the effective pairing interaction beyond the nucleus must be analyzed independently. In principle, the consideration of the problem in [22] is appropriate for any model space S_0, and the result changes in an obvious way; that is,

$$\mathcal{F}^{\xi} \to T^{\rm eff}(E=2\mu), \qquad (8)$$

where the "effective" T matrix obeys a standard equation that describes free nucleon–nucleon scattering, this being so not in the whole free space but in that part of it which is the subspace S' complementary to S_0. The subspace S' includes no states of energy satisfying the condition $\varepsilon_k < (E_0+\mu)$. Thus, we have

$$T^{\rm eff}(E) = v(0,0) + \sum_{k>k_0} v(0,k)\frac{1}{E-k^2/m}T^{\rm eff}(E), \qquad (9)$$

where $E = 2\mu$ as before and where the lower limit in the sum is $k_0 = \sqrt{2m(E_0+\mu)}$.

In order to describe neutron matter surrounding the cluster, we will use the energy functional $\mathcal{E}^{\rm mi}$ that was calculated microscopically for neutron matter and whose explicit form is considered below. Our ansatz for the total generalized energy functional consists in smoothly matching, at the cluster surface, the phenomenological and the microscopic generalized energy functional,

$$\mathcal{E}(\rho_\tau(r),\nu_\tau(r)) = \mathcal{E}^{\rm ph}(\rho_\tau(r),\nu_\tau(r))F_m(r) + \mathcal{E}^{\rm mi}(\rho_\tau(r),\nu_\tau(r))(1-F_m(r)), \qquad (10)$$

where the function $F_m(r)$, which implements the matching in question is chosen in the form of a two-parameter Fermi function,

$$F_m(r) = (1+\exp((r-R_m)/d_m))^{-1}. \qquad (11)$$

The ansatz in (10) is applied with the same matching function $F_m(r)$ both to the normal and to the anomalous part of the generalized energy functional. For the neutron part of the Wigner–Seitz cell, this ansatz is actually coincident with the prescription of the local-density approximation (LDA). This imposes rather stringent constraints on the choice of $F_m(r)$: the matching function should be chosen in such a

[4] This estimate was obtained for beta-stable nuclei, in which case one can assume that $\mu_n \simeq \mu_p \simeq \mu \simeq -8$ MeV.

way as to ensure a description of the region of a sharp change in the density on the basis of predominantly the phenomenological part of the generalized energy functional. Indeed, it is well known that the local-density approximation is inapplicable in this region; at the same time, the phenomenological generalized energy functional describes successfully the surface region of ordinary nuclei. On the other hand, the microscopic part of the generalized energy functional is intended for describing the neutron part of the Wigner–Seitz cell, where the matter density changes smoothly, so that the local-density approximation is applicable. Below, we will show that the Fermi function (11) satisfies these conditions for a proper choice of parameters.

The diffuseness parameter d_m in (11) remains free. An analysis revealed that it can be chosen in a unified way for all densities and (Z, R_c) configurations being studied. At the same time, the matching radius R_m must be chosen in each specific case individually. The method for choosing R_m is the most important point in the ansatz specified by Eqs. (10) and (11), which is used here. For a given configuration, we determine R_m in terms of the proton density in such a way as to ensure fulfillment of the equality

$$\rho_p(R_m) = 0.1\rho_p(0). \tag{12}$$

On one hand, neutrons and protons coexist in the region $r < R_m$ within the nuclear-like cluster being considered; therefore, the application of the phenomenological nuclear generalized energy functional is legitimate there. On the other hand, one can disregard the exponentially decreasing tails of the proton distribution in the region $r > R_m$, treating the system there as purely neutron matter and applying to it the generalized energy functional calculated microscopically. We use the same parameters of (11) for the normal and the anomalous part of the generalized energy functional in (10). It should be noted that almost all protons are localized within the radius R_m; therefore, the matching procedure is in fact applied only to neutrons, protons being entirely described by the phenomenological nuclear generalized energy functional. Within the local-density approximation, the microscopic part of the normal component of the generalized energy functional is expressed in terms of the equation of state for neutron matter. Here, we use the equation of state that was calculated in [23] on the basis of the Brueckner approach with the $v18$ Argonne nucleon–nucleon interaction and with allowance for a small admixture of three-particle forces. In the density range being considered, the equation of state for neutron matter of high density is described by the polynomial

$$\mathcal{E}^{\rm mi}_{\rm norm}(\rho_n) = \sum_{n=1,5} b_n x^n, \tag{13}$$

where $x = \rho_n/\rho_0$ with $\rho_0 = 0.155$ fm^{-3} and the coefficients b_n are given in [23].

The matching relation (10) for the anomalous part of the generalized energy functional leads to a similar relation for the effective pairing interaction:

$$\mathcal{V}^{\rm eff}_p(r) = \mathcal{F}^{\xi,\tau}(\rho(r))F_m(r) + \mathcal{V}^{\rm mi}_p(\rho(r))(1 - F_m(r)). \tag{14}$$

3. EFFECTIVE PAIRING INTERACTION

We recall that, in fact, the matching procedure in (14) must be applied only to neutrons and that, for protons, we will use a purely phenomenological pairing interaction—specifically, its simplest form (7). The microscopic part of the effective pairing interaction of neutrons, $\mathcal{V}^{\rm mi}_p(r)$, must be found for the model subspace S_0 used here. For a fixed value of the density $\rho_n(r)$, this interaction is determined within the local-density approximation from the equation for the gap in uniform neutron matter of density $\rho = \rho_n(r)$. We will use the Bardeen–Cooper–Schrieffer approximation, which is the simplest, where the gap Δ is directly expressed in terms of the free nucleon–nucleon potential $v(k, k')$ in the $^1S_0$ channel as

$$\Delta(k) = -\sum_{k'} v(k, k')\frac{\Delta(k')}{2E(k')}, \tag{15}$$

where $E_k = \sqrt{(\varepsilon_k - \mu_n)^2 + \Delta^2(k)}$ with $\varepsilon_k = k^2/2m + U_n$, U_n being the mean potential in neutron matter. As expressed in terms of the effective pairing interaction, the equation for the gap has a similar form, but integration with respect to momenta is performed over the model subspace S_0. Specifically, we have

$$\Delta(k) = -\sum_{k'<k_0} \mathcal{V}^{\rm mi}_p(k, k')\frac{\Delta(k')}{2E(k')}, \tag{16}$$

where $k_0 = \sqrt{2m(E_0 + \mu_n - U_n)}$.

In the Bardeen–Cooper–Schrieffer approximation, the relation between the effective pairing interaction and the free nucleon–nucleon potential obviously has the form

$$\mathcal{V}^{\rm mi}_p(k, k') = v(k, k') - \sum_{k_1>k_0} \frac{v(k, k_1)\mathcal{V}^{\rm mi}_p(k_1, k')}{2E(k_1)}. \tag{17}$$

The effective pairing interaction appearing in Eq. (16) depends explicitly on momenta. In the coordinate representation, this dependence corresponds to nonlocal forces. Since we are going to use a simple local form of the phenomenological effective pairing interaction in Eq. (14), it is highly desirable to simplify

maximally, prior to applying the matching procedure, its microscopic counterpart $\mathcal{V}_p^{\rm mi}$ by reducing it to a local form as well. The simplest way consists in the following. For a fixed values of ρ, we determine a momentum-independent averaged effective pairing interaction such that, upon being substituted into Eq. (16), it leads to the same value of $\Delta(k_{\rm F})$ as the exact effective pairing interaction; that is,

$$\Delta(k_{\rm F}) = -\bar{\mathcal{V}}_p^{\rm mi}(k_{\rm F}) \sum_{k'<k_0} \frac{\Delta(k')}{2E(k')}, \qquad (18)$$

where $k_{\rm F}$ is the local Fermi momentum $k_{\rm F} = (3\pi^2\rho)^{1/3}$. On the basis of the set of $\Delta(k)$ values calculated microscopically for each $k_{\rm F}$ value being considered, we find $\bar{\mathcal{V}}_p^{\rm mi}(k_{\rm F})$ from Eq. (18) and substitute it into the matching condition (14).

For neutrons, it seems natural to employ the density-dependent two-parameter interaction (6) for the phenomenological pairing interaction in the matching condition (14). However, there is an alternative possibility, that which was proposed in [15] and which makes it possible to minimize the number of phenomenological parameters. Instead of (6), one can use the analogous ansatz

$$\mathcal{F}^{\xi,\tau}(\rho_+) = C_0 \left(f_{\rm in}^{\xi} F_m(r) + f_{\rm ex}^{\xi}(1 - F_m(r)) \right), \qquad (19)$$

where the matching function $F_m(r)$ appears as the "form factor," taking the place of the expression $(\rho_+(r)/\rho_0)^\alpha$. We retained the same notation for the phenomenological parameters as in (6), but, in fact, their values must be determined anew, since the new form factor is shifted significantly with respect to the old one toward greater r. Indeed, the radii R_n and R_p at which the neutron and proton densities take values one-half as large as their counterparts at the center of nuclei are approximately equal to each other. Therefore, the matching radius R_m is larger than the radius $R_+ \simeq R_n \simeq R_p$ by a value approximately equal to the doubled diffuseness parameter of the density distribution, $2d_n \simeq 2d_p \simeq 1$ fm.[5)] This shift is significant for the matrix elements of the effective pairing interaction that appear in the equation for the gap. The gain from going over from the original ansatz in (6) to (19) consists in that, instead of considering the quantity $f_{\rm ex}^{\xi}$ as a free parameter, one can now substitute for it the T matrix defined in (9) and calculated microscopically; that is,

$$C_0 f_{\rm ex}^{\xi} = T^{\rm eff}. \qquad (20)$$

[5)] If the proton density $\rho_p(r)$ has the form of a Fermi function, the difference of the half-density radius for protons, R_p, and the matching radius R_m defined by relation (12) is $2d_p \ln 3$.

Indeed, we recall that, at present, we consider ordinary nuclei, for which, in the region $r > R_m$, we are in fact dealing with a free space, where one can use the asymptotic relation (20). The quantity is question can readily be found from (9). By way of example, we indicate that, for $E = 2\mu = -16$ MeV and for the model subspace being considered, which is restricted by the energy $E_0 = 35$ MeV, a calculation with the Paris potential yields $T^{\rm eff} = -990.36$ MeV. Of course, this result must not depend on the specific choice of realistic nucleon–nucleon potential. Upon fixing the value of $f_{\rm ex}^{\xi}$, there remains only one free parameter, $f_{\rm in}^{\xi}$. It was found in [15] from an analysis of the properties of five tin isotopes: ^{112}Sn, ^{116}Sn, ^{120}Sn, ^{124}Sn, and ^{128}Sn. The self-consistent calculation there was performed by using the generalized energy functional from [19] and the effective pairing interaction modified according to Eqs. (19) and (20). The value of $f_{\rm in}^{\xi} = -0.42$ virtually reproduces the diagonal matrix elements of Δ that were found by using the old effective pairing interaction (6) with the parameters from [19], which were determined from a fit to experimental data. We will use this value in all calculations based on the semimicroscopic generalized energy functional.

The interpolation formula (14) will be perfectly analogous to the modified effective pairing interaction (19) if one replaces the effective T matrix by the coordinate-dependent microscopic pairing interaction $\mathcal{V}_p^{\rm mi}(r)$ calculated for each point r according to Eq. (18), where $k_{\rm F}(r) = (3\pi^2\rho(r))^{1/3}$. Within this procedure, we employ the gap $\Delta(k)$ calculated for neutron matter in the Bardeen–Cooper–Schrieffer approximation with the $v18$ nucleon–nucleon interaction for the set of $k_{\rm F}$ values required for our purposes. The resulting values of $\Delta(k_{\rm F})$ (gap at the Fermi surface) are given in Fig. 1.

4. COMPUTATIONAL SCHEME

As was indicated above, the application of the variational principle to the generalized energy functional in (2) leads to a set of equations for the Bogolyubov functions $u_i(\mathbf{r})$ and $v_i(\mathbf{r})$ that is equivalent to the set of equations in the Hartree–Fock–Bogolyubov method. In the present study, which includes a massive array of calculations for the structure of the inner crust in a neutron star, we abandoned attempts at directly solving the cumbersome problem of pairing in the coordinate representation; instead, we used the simpler method of expansion in the eigenfunctions $\phi_\lambda(\mathbf{r}) = R_{nlj}(r)\Phi_{jlm}(\mathbf{n})$ of the problem without pairing (the isotopic index τ is suppressed for the sake of brevity), Φ_{jlm} being spin–angular functions. The

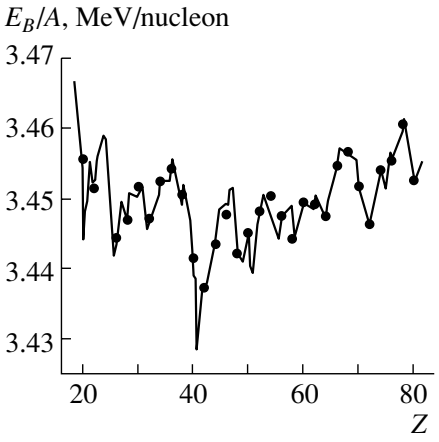

Fig. 2. Comparison of two methods of calculations.

functions $R_{nlj}(r)$ satisfy the radial Schrödinger equation with the potential $U(r)$ and boundary conditions at the boundary of the cell that are typical of the Wigner–Seitz method:

$$R_{nlj}(r = R_c) = 0 \tag{21}$$

for even l and

$$\left(\frac{dR_{nlj}}{dr}\right)_{r=R_c} = 0 \tag{22}$$

for odd l. The radial Schrödinger equation for the functions $y_\lambda(r) = rR_\lambda(r)$ is solved by Numerov's method. The equation for the gap is solved in the model subspace including single-particle states whose energy satisfies the condition $\varepsilon_\lambda - \mu < E_0$, where $E_0 \simeq 35$ MeV.

The calculations in question are strongly simplified in the diagonal approximation for the gap: $\Delta_{\lambda\lambda'} = \Delta_\lambda \delta_{\lambda\lambda'}$. The analysis performed in [19] for ordinary nuclei revealed that, in the case of volume pairing, the errors of the diagonal approximation are insignificant. In the case of surface pairing, the errors caused by the diagonal approximation are greater, but they concern primarily the aforementioned even–odd effects. But for the mean features of the system, which are the subject of our prime interest, the errors of the diagonal approximation are insignificant; therefore, we will use this approximation.

For a given baryon density, the computational scheme was constructed in the following way. At a fixed number Z of protons, one seeks the radius R_c of the Wigner–Seitz cell such that the beta-stability condition (1) is satisfied for it. For this equality to be satisfied to a fairly high precision, it is necessary to perform calculations for various values of R_c that are very close to one another. In this case, it is convenient to employ a coordinate mesh that has the same number N_c of nodes and slightly different integration steps $H = R_c/(N_c - 1)$. Owing to this, a self-consistent solution found for one value of R_c can be used as an initial iteration for a neighboring value. For this value of R_c, the total number of nucleons in the cell is $A = [4/(9\pi)](k_\mathrm{F}R_c)^3$. It is distributed between Z protons and N neutrons ($A = Z + N$) with allowance for the condition in (1). The self-consistent solution to the equations of the generalized-energy-functional method at a fixed value of R_c automatically yields density distributions and other features of the ground state of the system being considered that correspond to the minimum of the total binding energy E_B, which, in addition to the integral in (2) over the cell volume, involves the contribution from electrons, which is calculated by standard methods [5].

A numerical accuracy of solutions to Eq. (1) was prescribed in such a way that the difference of two values of the binding energy per nucleon, E_B/A, that correspond to two values of $\delta\mu > 0$ and $\delta\mu < 0$ closest to zero did not exceed 0.2 keV. Of course, the criterion of accuracy for the convergence of our iterative process in determining a self-consistent solution is still stricter (about 0.02 keV). Among all values of E_B/A that correspond to different values of Z, we then find the absolute minimum, and it is this minimum that determines the equilibrium (Z, R_c) configuration for the density being considered.

We would like to note that, in [13], we employed a computational scheme that involved the constant integration step of $H = 0.05$ fm; in solving Eq. (1) at a fixed value of Z, this prevented the use of arbitrary values of R_c (R_c is an integer in units of H). In view of this, we employed a formal trick, letting the number Z of protons be fractional at a fixed R_c, this making it possible to solve Eq. (1) to the required degree of precision. After that, all points of E_B/A versus R_c and the points of Z versus R_c were connected by continuous lines, on which only the points corresponding to integral values of Z were physically meaningful. That these two methods are nearly equivalent is demonstrated in Fig. 2, where the binding energies per nucleon calculated by the two methods in question are compared for the case of $k_\mathrm{F} = 0.8$ fm^{-1}. The solid line represents the results of the calculation at a constant step, in which case there appear fractional values of Z. The results of the calculation for a variable step and integral values of Z (only even values were taken) are shown by the points. One can see that almost all of the points lie on the curve. The only significant distinction is observed at $Z = 46$, in which case the continuous dependence of E_B/A on Z in the calculation with fractional Z proves to be anomalously sharp. Fortunately, the binding energy attains a minimum at a different value of Z—namely, at $Z = 42$—so that the possible inaccuracy of the

scheme at a single point, $Z = 46$, does not entail any physical consequences.

5. EEFECT OF SUPERFLUIDITY ON THE STRUCTURE OF THE INNER CRUST OF A NEUTRON STAR

The effect of taking into account neutron and proton pairing on the structure of the inner crust of a neutron star—more precisely, on the features of the equilibrium (Z, R_c) configuration—was demonstrated in [13–15]. All of these calculations were performed for a single value of the mean density, that which corresponds to the Fermi momentum of $k_F = 0.7$ fm^{-1} for uniform neutron matter. For the sake of completeness, the results of this analysis are presented in this section. Figure 3 illustrates the effect of pairing for the functional used by Fayans *et al.* [13], the pairing interaction being taken into account in the simplest form (7). Here, use was made of the computational scheme involving the constant integration step of $H = 0.05$ fm, in which case the radius R_c of the Wigner–Seitz cell appears to be a natural independent variable. The quantities under comparison are the binding energies per nucleon E_B/A and the corresponding Z values calculated without pairing ($f^\xi = 0$) and with allowance for pairing (dotted and solid curves, respectively). The coupling constant for the volume pairing interaction was chosen to be $f^\xi = -0.45$ (the average value [19] of those for Ca and Pb). We will first consider the case featuring no pairing. The absolute minimum is attained at $R_c = 25$ fm and $Z = 40$, but it is only 1 keV below the second local minimum at $R_c = 28$ fm and $Z = 50$ and the third minimum at $R_c = 30$ fm and the nonmagic value of $Z = 58$. In addition, there is a pronounced local minimum at $R_c = 21$ fm and the magic value of $Z = 20$, but this minimum is much higher on the energy scale. Thus, this calculation without pairing confirms qualitatively the results reported in [5]. However, the inclusion of pairing changes the pattern drastically: although a very weak local minimum is observed at $R_c = 28$ fm and $Z = 50$, the absolute minimum is shifted to the region of greater values of $R_c \simeq 31$ fm and $Z \simeq 70$.

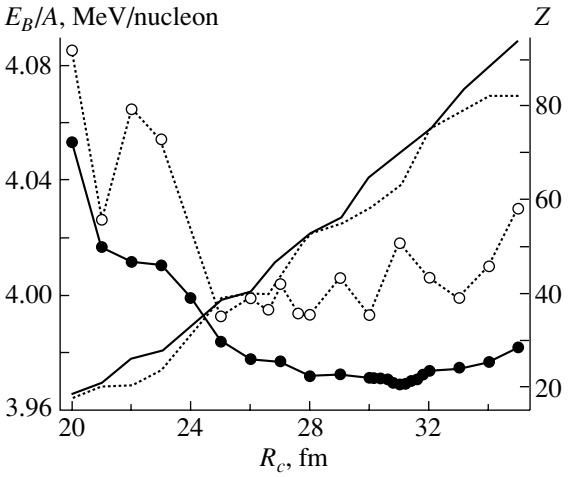

Fig. 3. Effect of pairing on the equilibrium configuration in the case of a phenomenological generalized energy functional.

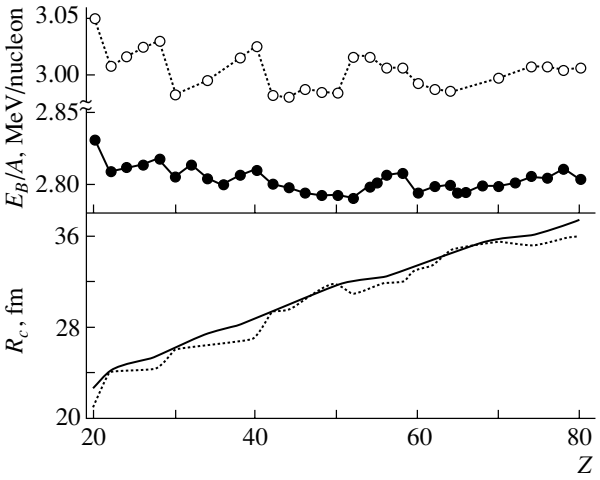

Fig. 4. Effect of pairing on the equilibrium configuration in the case of a semimicroscopic generalized energy functional.

Figure 4 shows the results of a similar calculation for the semimicroscopic generalized energy functional proposed in [15]. Here, we employed the basic computational scheme involving a variable integration step H, in which case the number Z of protons in the cell is a natural independent variable. The binding energy per nucleon E_B/A and the radius R_c of the Wigner–Seitz cell are given separately in individual panels versus Z. In just the same way as in Fig. 3, the results obtained with allowance for pairing and their counterparts that take no account of pairing are represented by, respectively, the solid and the dotted curves. As before, the effect of taking into account pairing is quite sizable, albeit it is not as drastic as in the preceding case. We note that, in the absence of pairing, the binding energy per nucleon reaches a minimum at the nonmagic value of $Z = 44$ (the corresponding radius of the Wigner–Seitz cell is $R_c = 29.473$ fm). Thus, the features of the equilibrium configuration of the Wigner–Seitz cell exhibit a significant dependence on the specific form of energy functional. The optimum configuration that emerges from the calculation that takes into account pairing is characterized by the values of $Z = 52$ and $R_c = 32.02$ fm. This value of Z is close to the cor-

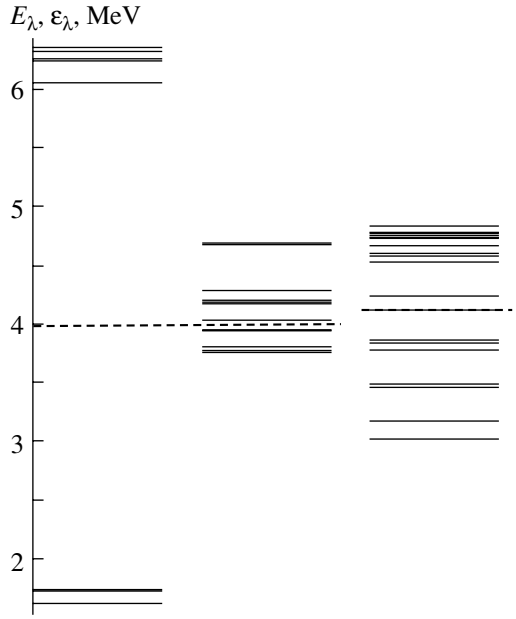

Fig. 5. Effect of pairing on the single-particle spectrum of neutrons.

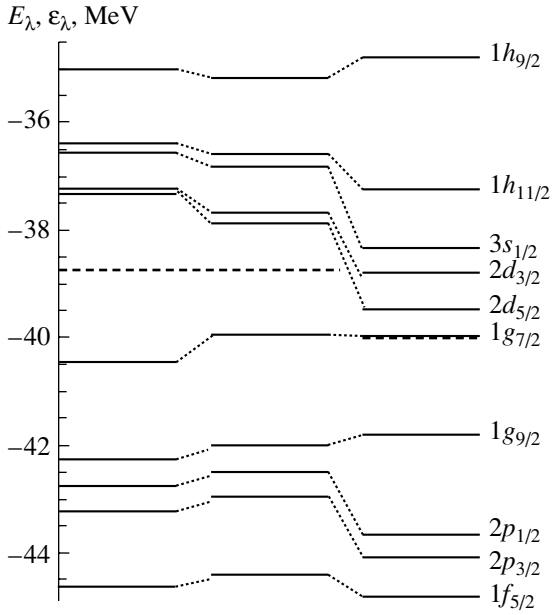

Fig. 6. Effect of pairing on the single-particle spectrum of protons.

responding result of Negele and Vautherin, but this agreement is accidental to some extent. We will see below that, at other density values, the results of our calculations differ significantly from the predictions of those authors in [5].

The fact that pairing effects change strongly the equilibrium features of the Wigner–Seitz cell can be qualitatively illustrated by an analysis of the beta-stability condition (1). Indeed, we can see that, since the electrons in the inner crust of a neutron star are ultrarelativistic, the relation $\mu_e \simeq (9\pi Z/4)^{1/3}/R_c$ holds. Substituting this relation into (1), we obtain

$$Z \simeq \frac{4}{9\pi}(\mu_n - \mu_p)^3 R_c^3. \qquad (23)$$

Upon taking into account pairing, both μ_n and μ_p may change by a value of about the gap, which is $\Delta \simeq 1$–2 MeV. In fact, this change is usually several times smaller, which should not seem significant, at first glance, but the difference $(\mu_n - \mu_p)$ is raised to the third power in (23); therefore, even a small change in it at fixed R_c leads to a sizable change in the equilibrium value of Z. The effect of pairing on the single-particle spectrum is illustrated in Figs. 5 and 6 for neutrons and protons, respectively. As above, the equilibrium configuration characterized by $Z = 52$ and $R_c = 32.02$ fm is considered there for $k_\mathrm{F} = 0.7$ fm^{-1}. For our analysis to be consistent, the system without pairing was considered for the same value of Z. In either figure, the right, the middle, and the left column show, respectively, the single-particle energies ε_λ of the problem without pairing, the single-particle energies ε_λ (basis) of the problem that involves pairing, and the true single-particle energies $E_\lambda = \mu \pm \sqrt{(\varepsilon_\lambda - \mu)^2 + \Delta_\lambda^2}$ with allowance for pairing. The dashed line indicates the position of the chemical potential μ. The effect of pairing is much more pronounced in the spectrum of neutrons. At the same time, the effect of pairing on the chemical potential is stronger in the case of protons.

6. RESULTS OF THE CALCULATIONS FOR THE SEMIMICROSCOPIC FUNCTIONAL

On the basis of the method developed by our group, a number of calculations for the structure of the inner crust in a neutron star were performed for the density range corresponding to Fermi momenta in the range $k_\mathrm{F} = 0.6$–1.2 fm^{-1}, which was covered with a step of $\delta k_\mathrm{F} = 0.1$ fm^{-1}. As was indicated in the Introduction, the intensity of pairing in uniform neutron matter is maximal precisely in this region (see Fig 1). This is the reason why the effect of pairing on the structure of the crust is expected to be maximal there and why this region is the most interesting from the point of view of studying superfluid vortices. As follows from the self-consistent calculations performed in [7] with Skyrme forces, a phase transition to a uniform state occurs at the density of $\rho_c = 0.077$ fm^{-3}, which corresponds to

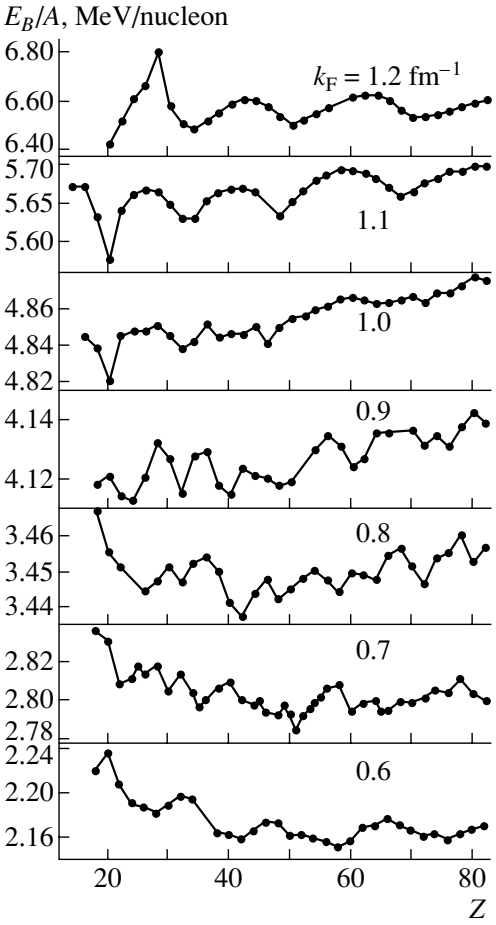

Fig. 7. Binding energy per nucleon for various values of k_F in the case of a semimicroscopic generalized energy functional.

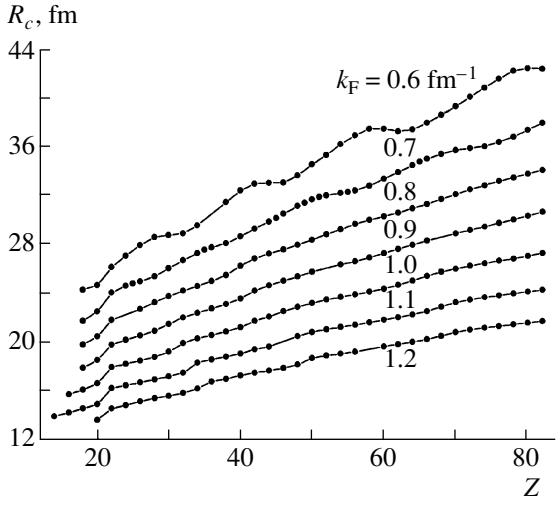

Fig. 8. Radius of the Wigner–Seitz cell as a function of Z for various values of k_F in the case of a semimicroscopic generalized energy functional.

$k_F^c = 1.316$ fm^{-1}. This agrees qualitatively with the estimate presented in [5]: $k_F^c \gtrsim 1.327$ fm^{-1}. At the same time, the calculations performed in [8] predict that, at a density of $\rho_c \simeq 0.05$ fm^{-3}, a spherical shape of the Wigner–Seitz cell becomes energetically unfavorable, giving way to a more complicated deformed configuration resembling a rod ("spaghetti" structure). The last of the densities considered here, $k_F = 1.2$ fm^{-1}, falls within this region; therefore, the corresponding calculation possibly has only a methodological meaning. All of the preceding densities are rather far off the critical region, so that one can dispense with considering the possible existence of exotic structures.

The procedure for determining a self-consistent solution for a given value of k_F was described above. At a fixed number Z of protons, we find the cell radius R_c in such a way as to ensure fulfillment of the beta-stability condition (1). We then compare the values of the energy of the system per nucleon E_B/A that correspond to different values of Z and then find the absolute minimum that realizes the equilibrium (Z, R_c) configuration for a given density. The set of curves E_B/A versus Z for all of the k_F values considered here is given in Fig. 7. Figure 8 shows the radius R_c of the Wigner–Seitz cell as a function of Z.

With the aim of reducing the time of computations, we considered only even values of Z, with the only exception of $Z = 51$ in the case of $k_F = 0.7$ fm^{-1}, where it was difficult to choose between the two neighboring even values of $Z = 50$ and 52 and where it turned out that the actual minimum of the binding energy per nucleon corresponds to precisely the odd number of $Z = 51$. We note that, in considering odd Z, we disregarded the so-called blocking effect, which leads to a regular even–odd pairing effect. It follows that some irregularity in the Z dependence of E_B/A rather than the even–odd pairing effect in the binding energy manifests itself in the present case, the latter usually having an opposite sign. Possibly, the inclusion of the blocking effect would compensate for this irregularity, so that, in fact, it would be more correct to take $Z = 50$–52 for equilibrium Z. An analysis of the data in Fig. 7 reveals that, with increasing mean density, the position of the absolute minimum on the curve $E_B/A(Z)$ is shifted leftward, the minimum remaining fixed at the magic number $Z = 20$ of protons from $k_F = 1.0$ fm^{-1}. We note that, in the case of $k_F = 1.2$ fm^{-1}, the curve is truncated at $Z = 20$, because, for $Z = 18$, we were unable to find a self-consistent solution of the cluster type. In all probability, this is due to the proximity of a phase transition to a uniform state.

Table 1. Properties of equilibrium configurations of the Wigner–Seitz cell at various densities

k_F, fm^{-1}	Z		A	R_c, fm	x		
	our study	[5]			our study	[5]	[23]
0.6	58	50	1612.10	37.505	0.036	0.037	0.0004
0.7	51	50	1573.70	31.890	0.032	0.037	0.0010
0.8	42	50	1409.10	26.895	0.030	0.028	0.0019
0.9	24	50	857.02	20.255	0.028	0.028	0.0034
1.0	20	40	658.07	16.693	0.030	0.027	0.0057
1.1	20	40	634.62	14.993	0.032	0.027	0.0086
1.2	20	40	626.47	13.684	0.032	0.027	0.0125

The properties of the equilibrium configurations are given in Table 1. In the last three columns of the table, we present the ratio $x = Z/A$—that is, the mean concentration of protons in the matter of the inner crust of a neutron star for a given density.

For the sake of comparison, we also present the values of Z and x from the article of Negele and Vautherin [5]. We note that, in fact, those authors performed calculations at somewhat different values of the density (namely, at $k_F = 0.55$, 0.64, 0.85, 1.12, 1.33 fm^{-1}). The Z and x values from [5] that are given in Table 1 correspond to data for the closest value of k_F. One can see that the equilibrium values of Z that were found here differ significantly from the predictions of Negele and Vautherin in [5], while the average concentrations of protons prove to be much closer to each other. At the same time, they are much larger (by an order of magnitude, on average) than the corresponding values for uniform matter, which are presented in the last column. Thus, the crystallization of nuclear matter leads to an increase in the average concentration of protons. With increasing density, the distinction between the proton concentrations in nonuniform and in uniform matter decreases, as it must, as one approaches the critical density ρ_c.

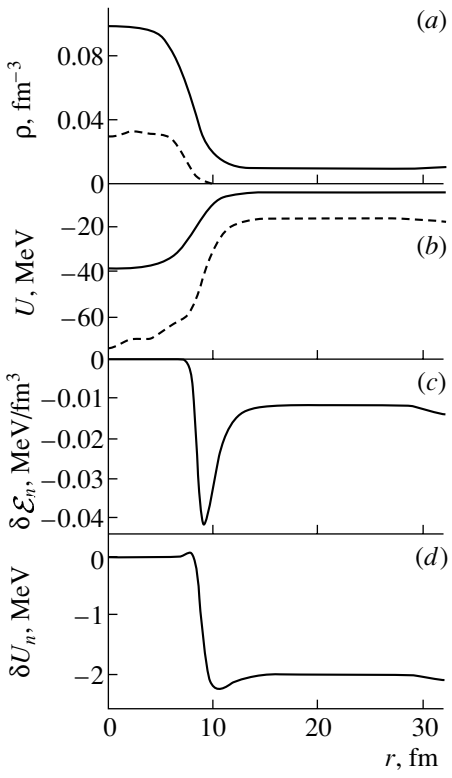

Fig. 9. Basic normal mean-field features of the equilibrium Wigner–Seitz cell for $k_F = 0.7$ fm^{-1} in the case of a semimicroscopic generalized energy functional.

Figure 9 shows basic "normal" ingredients of a self-consistent calculation for the equilibrium configuration corresponding to $k_F = 0.7$ fm^{-1}. Figures 9*a* and 9*b* display the densities and potentials, respectively. The solid and dashed curves were constructed for neutrons and protons, respectively. One can see that a well-defined nuclear-like cluster occurs within the Wigner–Seitz cell, but this cluster features a very large neutron excess: at the cluster center, the neutron density is three times as high as the proton density. At the same time, the proton potential is nearly twice as deep as the neutron potential. Beyond the cluster boundary, the proton density tends to zero very fast, while the neutron density approaches a constant. By and large, Fig. 9 confirms the pattern known from [5] for the structure of the Wigner–Seitz cell in the density region being considered: a compact cluster resembling in properties conventional nuclei, which is surrounded by nearly uniform neutron mat-

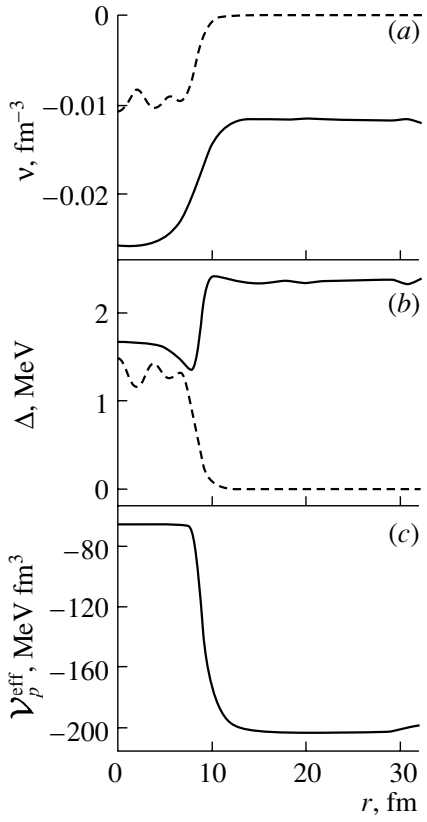

Fig. 10. Basic anomalous mean-field features of the equilibrium Wigner–Seitz cell for $k_{\text{F}} = 0.7$ fm^{-1} in the case of a semimicroscopic generalized energy functional.

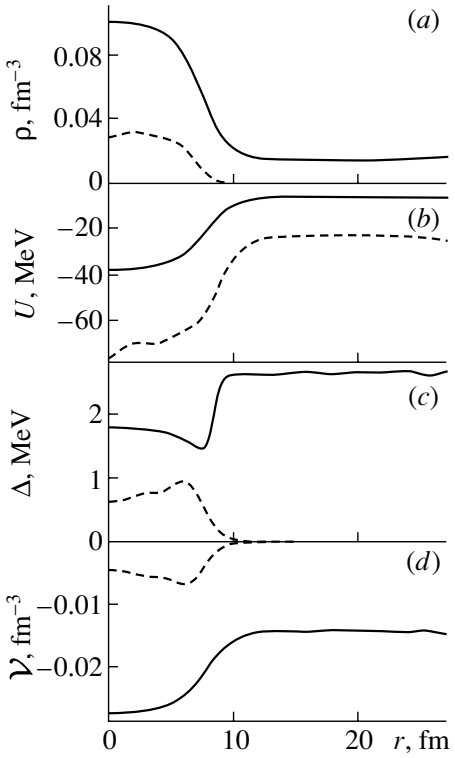

Fig. 11. Basic mean-field features of the equilibrium Wigner–Seitz cell for $k_{\text{F}} = 0.8$ fm^{-1} in the case of a semimicroscopic energy functional.

ter, occurs at the cell center. Separately, we show, in Fig. 9*c*, the microscopic addition to the normal part of the neutron phenomenological generalized energy potential,

$$\delta\mathcal{E}_n(r) = \Big(\mathcal{E}^{\text{mi}}(\rho_n(r)) - \mathcal{E}^{\text{ph}}(\rho_n(r))\Big)\,(1 - F_m(r)), \tag{24}$$

and, in Fig. 9*d*, the analogous addition to the mean neutron potential,

$$\delta U_n(r) = \Big(U_n^{\text{mi}}(\rho_n(r)) - U_n^{\text{ph}}(\rho_n(r))\Big)\,(1 - F_m(r)). \tag{25}$$

It is necessary to discuss an irregular behavior of the quantity $\delta\mathcal{E}_n(r)$ in the vicinity of the point $R_m = 8.51$ fm. The reason for this is that, in this region, the neutron density $\rho_n(r)$ begins growing with increasing r, the microscopic quantity $\mathcal{E}^{\text{mi}}(\rho_n)$ calculated according to (13) growing faster (in absolute value) with increasing ρ_n than its phenomenological analog $\mathcal{E}^{\text{ph}}(\rho_n)$ corresponding to [19]. Therefore, the difference in Eq. (24) increases as $r \to R_m$ from the right up to the point $R_0 \simeq R_m + d$, where the factor $(1 - F_m(r))$ comes into play, suppressing this growth. In the region $r < R_m - d$, the quantity in (24) tends fast to zero. At first glance, the application of the local-density approximation to the normal part of the generalized energy functional is questionable because of this irregular behavior. However, there are two arguments removing these doubts. First, the region of a fast variation in the quantity $\delta\mathcal{E}_n(r)$ is rather narrow and makes but a small contribution to the integral of the quantity in (24) over the volume, this integral yielding a microscopic correction to the binding energy per nucleon E_B/A (in the case being considered, the correction in question is $\delta(E_B/A) = -1.0769$ MeV). Second, the quantity $\mathcal{E}$ is not directly involved in the self-consistent calculation. The quantity $\delta U_n(r)$—that is, its variational derivative with respect to the density—is much more important, but it behaves in a considerably more regular way, its maximum value being approximately 20 times less than the average depth of the neutron potential within the cluster (see Fig. 9*b*). It follows that the application of the local-density approximation to the normal part of the generalized energy functional is quite legitimate.

The anomalous features of the equilibrium configuration of the Wigner–Seitz cell are displayed in

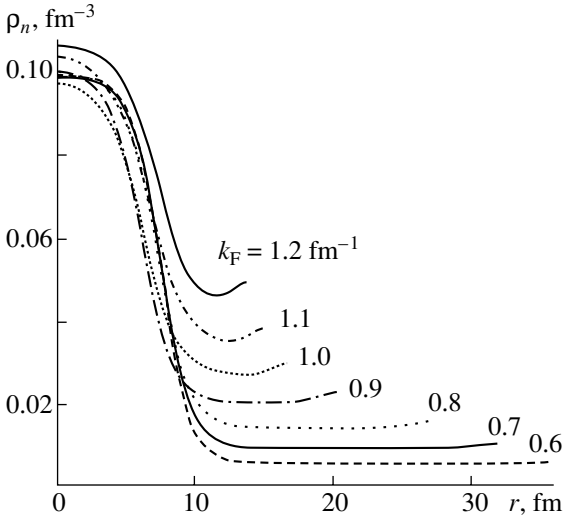

Fig. 12. Neutron-density distribution within the equilibrium Wigner–Seitz cell for various values of k_F in the case of a semimicroscopic generalized energy functional.

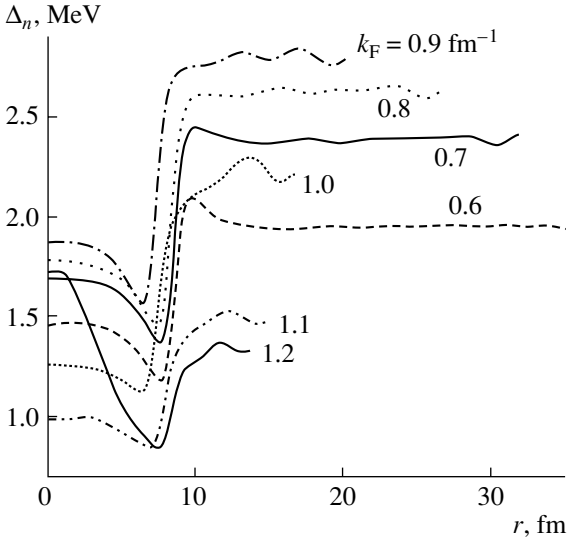

Fig. 14. Pairing gap for neutrons within the equilibrium Wigner–Seitz cell for various values of k_F in the case of a semimicroscopic generalized energy functional.

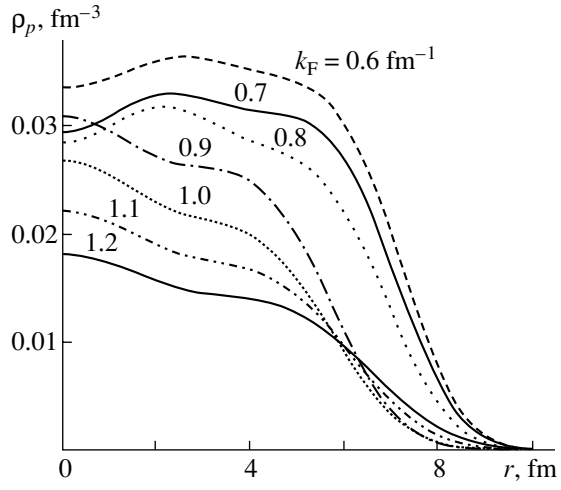

Fig. 13. As in Fig. 12, but for protons.

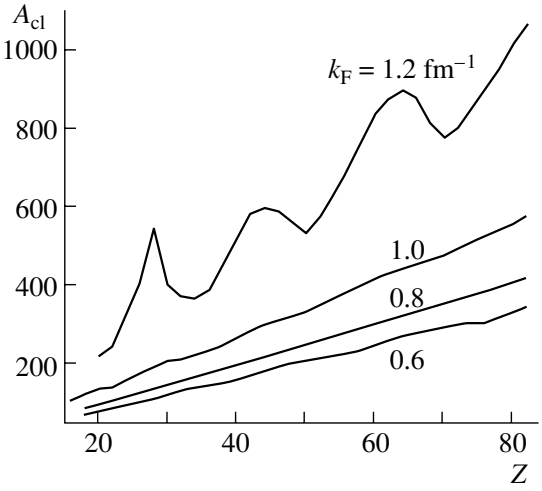

Fig. 15. Number of nucleons in the cluster as a function of Z for various values of k_F in the case of a semimicroscopic generalized energy functional.

Fig. 10 for the same value of $k_F = 0.7$ fm^{-1}. The anomalous densities and the pairing gaps are shown in Figs. 10*a* and 10*b*, respectively. As in Fig. 9, the solid and dashed curves refer to neutrons and protons, respectively. From Fig. 10, one can see that, beyond the cluster, the anomalous proton density tends fast to zero, as its normal analog in Fig. 9 does. It follows that, upon the substitution of the interpolation formula (17) for the effective proton interaction into the equation for the proton gap, the second term is virtually inoperative. Thus, we see that, as was indicated above, the effective pairing interaction of protons can be taken in a purely phenomenological form.

Figure 10*c* shows the effective pairing interaction (17) for neutrons. One can see that, within the cluster, it is much weaker (approximately by a factor of 3) than in neutron matter, while the gap Δ_n is less only by 30%. This is a piece of evidence against a naive application of the local-density approximation in the equation for the gap, in which case we would obtain a much smaller gap within the cluster. At the same time, the application of this approximation to the effective interaction itself is legitimate. The point is that the nonlocality in the equation for the gap stems predominantly from the anomalous den-

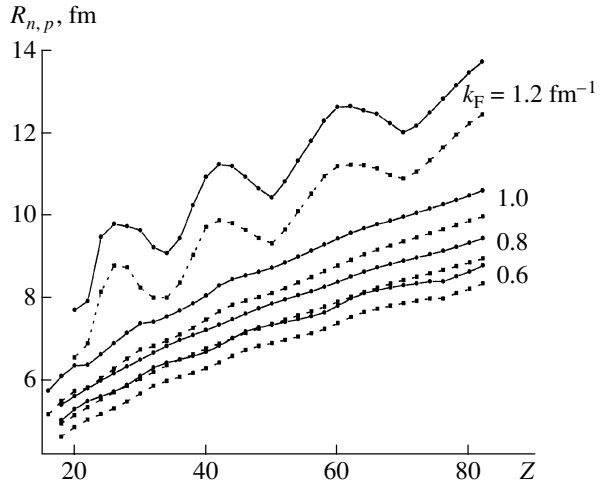

Fig. 16. Radii of neutrons and protons in the cluster versus Z for various values of k_F in the case of a semimicroscopic generalized energy functional.

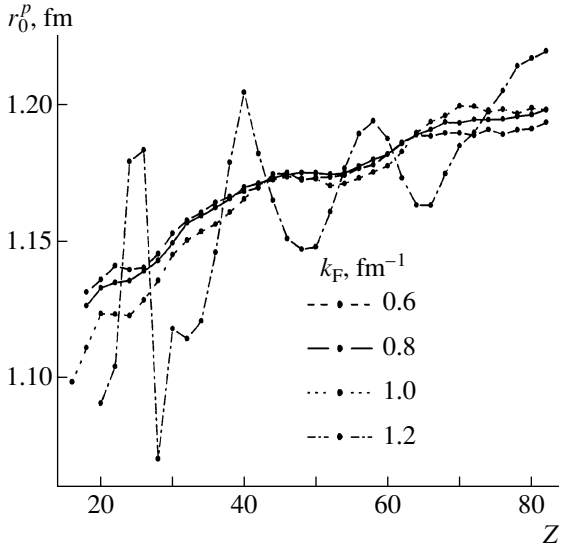

Fig. 18. Parameter r_0^p as a function of Z for various values of k_F.

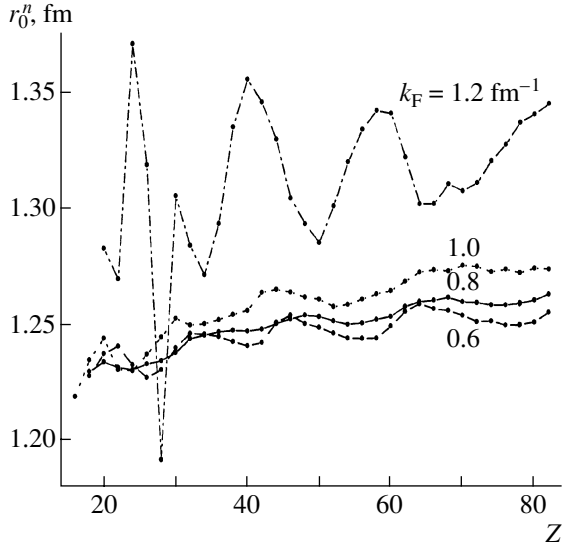

Fig. 17. Parameter r_0^n as a function of Z for various values of k_F.

sity ν, which is calculated in this approach without invoking any form of local approximation. As to the magnitude of the gap in neutron matter, it differs only slightly, in accordance with intuitive ideas, from the corresponding value for infinite neutron matter (see Fig. 1). This conclusion agrees qualitatively with the results of the calculations performed in [3, 4]. We note that, according to the currently prevalent opinion, the Bardeen–Cooper–Schrieffer approximation, which is used here overestimates strongly the gap in neutron matter [14]. Various many-body corrections to this approximation suppress Δ in neutron matter at the densities considered here approximately by a factor of 2. The inclusion of these corrections introduces no changes in the computational scheme that we developed. In order to find a new microscopic effective pairing interaction $\mathcal{V}_p^{\rm mi}(r)$, we must merely substitute, for each value of $k_F(r)$, a new set of $\Delta(k)$ values into Eq. (14). A new result for the gap within neutron matter is again expected to be close to the value of Δ in neutron matter.

Figure 11 displays basic normal and anomalous mean-field features of the Wigner–Seitz cell for the equilibrium configuration ($Z = 42, R_c = 26.895$ fm) corresponding to $k_F = 0.8$ fm^{-1}. The notation there is identical to that in Figs. 9 and 10. All properties of the self-consistent solution are qualitatively very close to the corresponding properties at $k_F = 0.7$ fm^{-1}. We only note that, in the case being considered, the gap $\Delta(r)$ beyond the cluster is somewhat greater than that in Fig. 10. This is because the value of $k_F = 0.8$ fm^{-1} is close to the maximum of the curve $\Delta(k_F)$ (see Fig. 1).

The systematic variations of these quantities versus the density can be traced in Figs. 12–14. The first two of these display the neutron- and proton-density distributions, while the third shows the neutron gap $\Delta_n(r)$. Each curve in Figs. 12 and 14 terminates at the corresponding value of R_c. From these figures, one can see that, with increasing k_F, the cluster at the center of the Wigner–Seitz cell becomes less compact, occupying an ever greater part of the cell. Accordingly, the layer occupied by neutron matter becomes ever narrower, the excess of the density at the

Table 2. Properties of clusters for equilibrium configurations of the inner crust of a neutron star

k_F, fm^{-1}	Z	A	A_{cl}	x_{cl}	y_{cl}	R_m, fm	R_c, fm
0.6	58	1612.10	233.44	0.248	0.145	8.491	37.505
0.7	51	1573.70	224.79	0.227	0.143	8.438	31.890
0.8	42	1409.10	204.64	0.258	0.145	8.214	26.895
0.9	24	857.02	132.32	0.181	0.154	7.221	20.255
1.0	20	658.07	133.81	0.176	0.203	7.311	16.693
1.1	20	634.62	172.39	0.116	0.272	7.800	14.993
1.2	20	626.47	217.57	0.101	0.347	8.291	13.684

Table 3. Energy features of the equilibrium Wigner–Seitz cell

k_F, fm^{-1}	Z	A	E_B/A, MeV/nucleon	μ_n, MeV	μ_p, MeV
0.6	58	1612.10	2.1516	3.2074	−34.4688
0.7	51	1573.70	2.7845	3.9876	−38.6206
0.8	42	1409.10	3.4374	4.8454	−42.7177
0.9	24	857.02	4.1123	5.7340	−46.9934
1.0	20	658.07	4.8210	6.8525	−53.5547
1.1	20	634.62	5.5764	7.4288	−59.9835
1.2	20	626.47	6.4225	8.5814	−65.4208

cluster (cell) center over the density of surrounding neutron matter decreasing monotonically. As to the neutron gap, its asymptotic value beyond the cluster first increases with increasing k_F, reaching a maximum at $k_F = 0.9$ fm^{-1}, and then decreases rather fast. This behavior of the gap agrees qualitatively with the k_F dependence of the gap in uniform neutron matter (see Fig. 1). We note that, in the uniform case, the value of Δ_n at $k_F = 0.9$ fm^{-1} is somewhat less than its counterpart at $k_F = 0.8$ fm^{-1}. Therefore, the fact that, for a nonuniform system, the gap is maximal at $k_F = 0.9$ fm^{-1} is at odds with the naive local-density approximation.

In order to characterize the properties of the clusters quantitatively, we have calculated the following features for them. First, these are the average neutron and proton radii R_n and R_p within the cluster that are defined as the points of the maximum gradient of the corresponding density. In the case where a density distribution is described by a Fermi function, the radius defined in this way coincides with the half-density radius. For protons, R_p is indeed virtually coincident with the respective half-density radius, while, for neutrons, R_n is very close to the position of the point at which the difference $\delta\rho_n(r) = \rho_n(r) - \rho_n^a$ (ρ_n^a is the asymptotic value of the neutron density far off the cluster) is equal to half its value at $r = 0$. The number of neutrons within the cluster is defined as

$$N_{cl} = \int\limits_{r<R_m} d^3r\rho_n(r). \qquad (26)$$

We recall that the matching radius R_m is defined in such a way that the proton density nearly vanishes at this radius value, so that it is natural to identify the cluster radius with it. Accordingly, the total number of nucleons in the cluster is $A_{cl} = N_{cl} + Z$. These "global" features of the clusters versus Z (and, accordingly, versus the size of the Wigner–Seitz cell, since there is an unambiguous relation between Z and this size) are displayed in Figs. 15 and 16 for various densities. In order to avoid encumbering the figures, we restricted ourselves to four values of k_F. As before, the solid and dashed curves in Fig. 16 refer to neutrons and protons, respectively.

For ordinary nuclei, the radii R_n and R_p are close to each other and depend on the mass number A

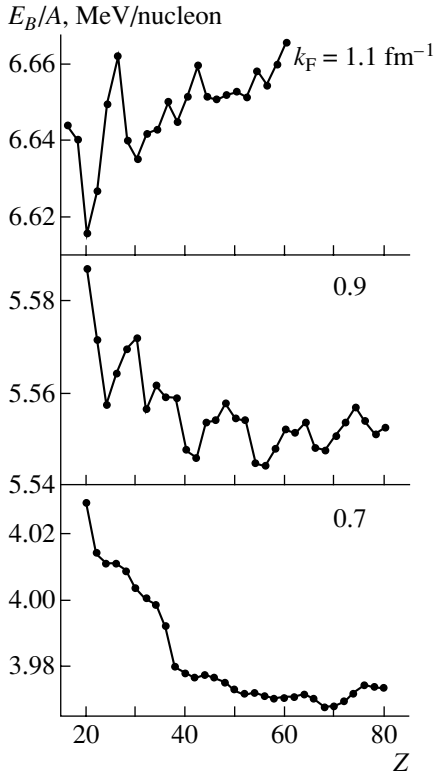

Fig. 19. Binding energy per nucleon for various values of k_F in the case of a phenomenological generalized energy functional.

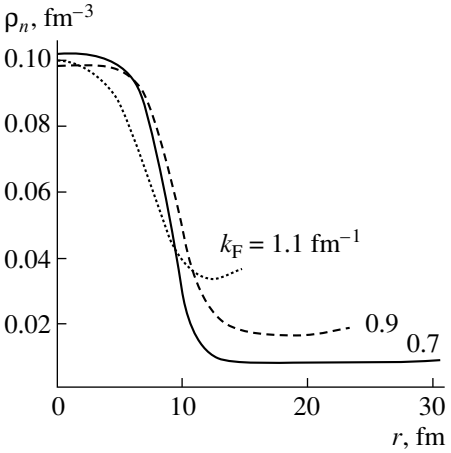

Fig. 20. Neutron-density distribution in the equilibrium Wigner–Seitz cell for various values of k_F in the case of phenomenological generalized energy functional.

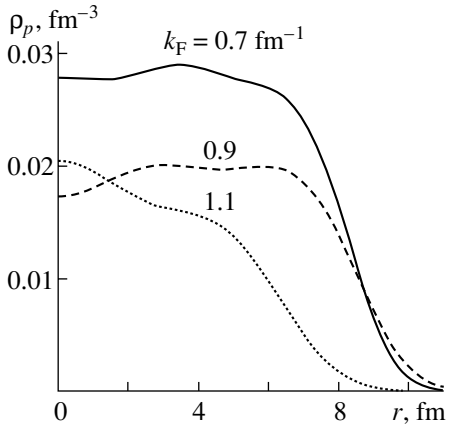

Fig. 21. As in Fig. 20, but for protons.

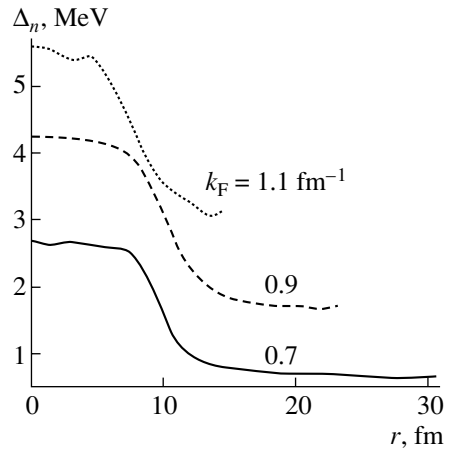

Fig. 22. Pairing gap for neutrons in the equilibrium Wigner–Seitz cell for various values of k_F in the case of a phenomenological generalized energy functional.

according to the law $R = r_0 A^{1/3}$, where the parameter r_0 is virtually constant for all stable nuclei and characterizes the average matter density within a nucleus. It is of interest to calculate the analogous quantities $r_0^n = R_n/A_{cl}^{1/3}$ and $r_0^p = R_p/A_{cl}^{1/3}$ for the clusters being considered and to trace the variations of these parameters versus the average matter density. The graphs representing the dependences of r_0^n and r_0^p on Z are given in Figs. 17 and 18. One can see that the character of the Z dependence of the quantities being studied changes drastically upon going over from $k_F = 1.0$ fm^{-1} to $k_F = 1.2$ fm^{-1}. At three smaller values of k_F, all of the features being considered change very smoothly with increasing Z at a fixed density. Although the parameters r_0^n and r_0^p corresponding to these values grow on average with increasing density (the clusters "swell"), this growth is insignificant. At the same time, all features begin changing sharply (oscillating) with Z (with the cell size) in the case of $k_F = 1.2$ fm^{-1}. Concurrently, the average value of r_0^n increases substantially. In all probability, these oscillations and the increase in the average value of r_0^n is a precursor of a phase transition to a uniform state.

The features of nuclear-like clusters correspond-

Table 4. Features of the equilibrium configurations of the Wigner–Seitz cell for various densities in the case of the generalized energy functional introduced by Fayans and his coauthors

k_F, fm^{-1}	Z	A	R_c, fm	x		
				our study	[5]	[23]
0.7	68	1397.90	30.655	0.049	0.033	0.0010
0.9	56	1323.10	23.410	0.042	0.026	0.0034
1.1	20	601.68	14.729	0.033	0.027	0.0086

Table 5. Features of clusters for equilibrium configurations of the inner crust of a neutron star in the case of the generalized energy functional introduced by Fayans and his coauthors

k_F, fm^{-1}	Z	A	A_{cl}	x_{cl}	y_{cl}	R_c, fm
0.7	68	1397.90	342.99	0.198	0.245	30.655
0.9	56	1323.10	385.57	0.145	0.291	23.410
1.1	20	601.68	180.73	0.111	0.300	14.729

Table 6. Energy features of the equilibrium Wigner–Seitz cell in the case of the generalized energy functional introduced by Fayans and his coauthors

k_F, fm^{-1}	Z	A	E_B/A, MeV/nucleon	μ_n, MeV	μ_p, MeV
0.7	68	1397.2	3.9679	6.0015	−42.7781
0.9	56	1323.3	5.5442	7.5937	−52.6820
1.1	20	601.7	6.6158	8.0193	−60.6237

ing to equilibrium configurations of the inner crust of a neutron star are displayed in Table 2. Given there are the mean proton concentration within the cluster, $x_{cl} = Z/A_{cl}$, and the ratio $y_{cl} = A_{cl}/A$, which indicates the fraction of the total number of nucleons within the cluster. One can see that, with increasing k_F, x_{cl} decreases monotonically, while y_{cl} increases. The last property correlates with a systematic decrease in the difference $(R_c - R_m)$, the thickness of the layer occupied by neutron matter. Finally, Table 3 compiles the energy features of equilibrium configurations for various densities: the binding energies per nucleon E_B/A and the chemical potentials μ_n and μ_p for neutrons and protons, respectively.

7. RESULTS OF THE CALCULATIONS FOR THE PHENOMENOLOGICAL FUNCTIONAL

In this section, we describe the results of the calculations for the purely phenomenological functional proposed by Fayans and his coauthors [19]. In these calculations, the effective pairing interaction was taken in the simplest form (7). Our objective is to make an attempt at understanding which of the results exposed in the preceding section are general and which depend on the specific form of generalized energy functional—that is, on the choice of model. Here, we restricted ourselves to three density values corresponding to $k_F = 0.7$, 0.9, and 1.1 fm^{-1}. The results are given in Figs. 19–22 and in Tables 4–6.

The majority of the properties of the self-consistent solution describing the structure of the inner crust of a neutron star that were indicated in the preceding section for the semimicroscopic functional survive in the case of the phenomenological generalized energy functional. In particular, the equilibrium configuration at $k_F = 1.1$ fm^{-1} corresponds to the magic number $Z = 20$. The neutron gap, whose behavior with increasing k_F is absolutely unrealistic, is an exception. This fact confirms that an extrapolation of the phenomenological nuclear generalized energy functional to neutron matter is very hazardous. In the present case, this is manifested for the anomalous part of the generalized energy functional.

8. CONCLUSION

For a wide interval of densities corresponding to Fermi momenta in the range $k_F = 0.6$–1.2 fm^{-1}, we have performed a self-consistent calculation of the structure of the inner crust in a neutron star, taking into account superfluid correlations of neutrons and protons and relying on the Wigner–Seitz approximation. The approach used here is based on the generalized method of an energy density functional in the form proposed by Fayans and his coauthors [17], this extending the well-known Kohn–Sham method to superfluid systems. For the generalized energy functional describing matter in the inner crust of a star, we have proposed a semimicroscopic model where a nuclear-like cluster occurring at the center of the Wigner–Seitz cell is characterized by a phenomenological nuclear generalized energy functional [17] and where the generalized energy functional for neutron matter surrounding this cluster is calculated microscopically within the Brueckner approach by using the $v18$ nucleon–nucleon potential. The two generalized energy functionals in question have been smoothly matched at the cluster boundary. Systematic calculations have been performed for seven values

of k_F in the above range with a step of 0.1 fm^{-1}. With the aim of studying the dependence of our results on the specific form of generalized energy functional, we have also performed calculations at three values of k_F (0.7, 0.9, 1.1 fm^{-1}) with the purely phenomenological generalized energy functional introduced by Fayans and his coauthors. An analysis of the results and their comparison with the results of the well-known study by Negele and Vautherin [5] have demonstrated a strong effect of pairing on the equilibrium values of the parameters (Z, R_c) of the Wigner–Seitz cell. In particular, the effect of the magic numbers $Z = 40$ and 50 that was predicted by the calculations in [5] (according to [5], only these two values of Z compete over the whole region being considered) is smeared to a considerable extent. For all densities starting from that which corresponds to $k_F = 1.0$ fm^{-1}, our calculations with the semimicroscopic generalized energy functional have also led to the proton magic number $Z = 20$ in the cell. We note that the proton gap Δ_p vanishes concurrently, as it does in ordinary magic nuclei. The magic number $Z = 20$ also arises in the calculations with the phenomenological functional of Fayans and coauthors, but at somewhat higher densities. It is interesting to note that, although the equilibrium values of Z in our calculations and in [5] differ by a factor of 2, the difference in the relative proton concentration x is as small as 10 to 15%. At the same time, the values of x exceed the predictions for uniform matter [23] by an order of magnitude at low k_F and by a factor of 3 at $k_F = 1.2$ fm^{-1}.

For all values of k_F, we have found self-consistently the pairing gap for neutrons, $\Delta_n(r)$; beyond the cluster, it appears to be close to the corresponding value for uniform neutron matter. Thus, we see that, beyond the cluster, the local-density approximation is quite accurate. At the same time, the value of $\Delta_n(r)$ within the cluster exceeds considerably the result obtained by applying the local-density approximation. In solving the equation for the neutron gap, we have employed the interpolated effective pairing interaction (17) whose microscopic part was found in the Bardeen–Cooper–Schrieffer approximation. It is well known that, for neutron matter, this approximation overestimates Δ_n by a factor of about 2. In all probability, the gap $\Delta_n(r)$ found here in the region beyond the cluster is accordingly overestimated by a factor of about 2. The next step that we plan in constructing a realistic generalized energy functional is to improve the microscopic part of the effective pairing interaction with allowance for many-body corrections to the Bardeen–Cooper–Schrieffer approximation.

ACKNOWLEDGMENTS

We are grateful to M.V. Zverev, N.E. Zein, and E.A. Ryabinkin for stimulating discussions.

This work was supported by the Ministry of the Russian Federation for Industry, Science, and Technologies (grant no. 1885.2003.2).

REFERENCES

1. A. B. Migdal, Zh. Éksp. Teor. Fiz. **37**, 249 (1960) [Sov. Phys. JETP **10**, 179 (1960)].
2. C. J. Pethick and D. G. Ravenhall, Annu. Rev. Nucl. Part. Sci. **45**, 429 (1995).
3. F. Barranco, R. A. Broglia, H. Esbensen, and E. Vigezzi, Phys. Lett. B **390**, 13 (1997).
4. F. Barranco, R. A. Broglia, H. Esbensen, and E. Vigezzi, Phys. Rev. C **58**, 1257 (1998).
5. J. W. Negele and D. Vautherin, Nucl. Phys. A **207**, 298 (1973).
6. G. Baym, H. A. Bethe, and C. J. Pethick, Nucl. Phys. A **175**, 225 (1971).
7. F. Douchin and P. Haensel, Phys. Lett. B **485**, 107 (2000).
8. P. Magierski and P.-H. Heenen, Phys. Rev. C **65**, 045 804 (2002).
9. A. Bulgac and P. Magierski, Nucl. Phys. A **683**, 695 (2001).
10. A. Bulgac and P. Magierski, Phys. Scr., T **90**, 150 (2001).
11. A. Bulgac and P. Magierski, Acta Phys. Pol. B **32**, 1099 (2001).
12. P. Magierski, A. Bulgac, and P.-H. Heenen, Nucl. Phys. A **719**, 217 (2003).
13. M. Baldo, U. Lombardo, E. E. Saperstein, and S. V. Tolokonnikov, Pis'ma Zh. Éksp. Teor. Fiz. **80**, 595 (2004) [JETP Lett. **80**, 523 (2004)].
14. M. Baldo, E. E. Saperstein, and S. V. Tolokonnikov, Nucl. Phys. A **749**, 42 (2005).
15. M. Baldo, U. Lombardo, E. E. Saperstein, and S. V. Tolokonnikov, Nucl. Phys. A **750**, 409 (2005).
16. A. V. Smirnov, S. V. Tolokonnikov, and S. A. Fayans, Yad. Fiz. **48**, 1661 (1988) [Sov. J. Nucl. Phys. **48**, 995 (1988)].
17. S. A. Fayans, S. V. Tolokonnikov, E. L. Trykov, and D. Zawischa, JETP Lett. **68**, 276 (1998).
18. W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
19. S. A. Fayans, S. V. Tolokonnikov, E. L. Trykov, and D. Zawischa, Nucl. Phys. A **676**, 49 (2000).
20. A. B. Migdal, *Theory of Finite Fermi Systems and Applications to Atomic Nuclei* (Nauka, Moscow, 1965; Interscience, New York, 1967).
21. M. Baldo, U. Lombardo, E. Saperstein, and M. Zverev, Nucl. Phys. A **628**, 503 (1998).
22. M. Baldo, U. Lombardo, E. E. Saperstein, and M. V. Zverev, Phys. Rep. **391**, 261 (2004).
23. M. Baldo, C. Maieron, P. Schuck, and X. Viñas, Nucl. Phys. A **736**, 241 (2004).

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1830–1841. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1892–1903.

ELEMENTARY PARTICLES AND FIELDS
Theory

Determination of Quark–Antiquark Component of the Photon Wave Function for u, d, s Quarks*

A. V. Anisovich, V. V. Anisovich, **L. G. Dakhno, V. A. Nikonov, and A. V. Sarantsev**

Petersburg Nuclear Physics Institute, Russian Academy of Sciences, Gatchina, 188350 Russia

Received July 9, 2004; in final form, December 9, 2004

Abstract—Based on the data for the transitions $\pi^0, \eta, \eta' \to \gamma\gamma^*(Q^2)$ and reactions of the e^+e^- annihilations $e^+e^- \to \rho^0, \omega, \phi$ and $e^+e^- \to hadrons$ at $1 < E_{e^+e^-} < 3.7$ GeV, we determine the light-quark components of the photon wave function $\gamma^*(Q^2) \to q\bar{q}$ $(q = u, d, s)$ for the region $0 \lesssim Q^2 \lesssim 1$ $(\mathrm{GeV}/c)^2$.

1. INTRODUCTION

In the search for exotic states, one needs to find the quark–gluon content of mesons and establish the meson systematics. The meson radiative decay is a powerful tool for qualitative evaluation of the quark–antiquark components. The study of the two-photon transitions such as $meson \to \gamma\gamma$ and, more generally, $meson \to \gamma^*(Q_1^2)\gamma^*(Q_2^2)$ looks to be a promising way to reveal the quark–antiquark content of mesons.

Experimental data accumulated by the Collaborations L3 [1, 2], ARGUS [3], CELLO [4], TRC/2γ [5], CLEO [6], Mark II [7], Crystal Ball [8], and others make it obvious that the calculation of the processes $meson \to \gamma^*(Q_1^2)\gamma^*(Q_2^2)$ is up to date. To make this reaction informative as concerns the meson quark–gluon content, one needs a reliably determined wave function of the photon at $0 \lesssim Q^2 \lesssim 1$ $(\mathrm{GeV}/c)^2$. Conventionally, one may consider two pieces of the photon wave function: soft and hard ones. The hard component relates to the pointlike vertex $\gamma \to q\bar{q}$; it is responsible for the production of a quark–antiquark pair at high virtuality. At large energy of the e^+e^- system, the ratio of cross sections $R = \sigma(e^+e^- \to hadrons)/\sigma(e^+e^- \to \mu^+\mu^-)$ is determined by the hard component of photon wave function, while the soft component is responsible for the production of low-energy quark–antiquark vector states, such as ρ^0, ω, $\phi(1020)$, and their excitations.

Evaluation of the photon wave function for the $\gamma^*(Q^2) \to u\bar{u}, d\bar{d}, s\bar{s}$ transitions was carried out in [9] on the basis of data of the CLEO Collaboration on the Q^2-dependent transition form factors $\pi^0 \to \gamma\gamma^*(Q^2)$, $\eta \to \gamma\gamma^*(Q^2)$, and $\eta' \to \gamma\gamma^*(Q^2)$ (see [6] and references therein). The goal of the present paper is, by adding information on the process $e^+e^- \to hadrons$, to define more precisely the wave function $\gamma^*(Q^2) \to u\bar{u}, d\bar{d}, s\bar{s}$.

Similarly to what has been done in [9], here we determine the photon wave function working in the approach of the spectral-integration technique. This technique was suggested in [10] for the description of deuteron form factors, the deuteron being treated as a composite two-nucleon system. In [9, 11], the spectral-integration technique was expanded for the composite $q\bar{q}$ systems with wave functions written in terms of the light-cone variables. The wave function depends on the invariant energy squared of the $q\bar{q}$ system as follows:

$$s = \frac{m^2 + k_\perp^2}{x(1-x)}, \tag{1}$$

where m is the quark mass and $\mathbf{k}_\perp$, and x are the light-cone characteristics of quarks (transverse momentum and a part of the longitudinal momentum). In this technique, the quark wave function of the photon, $\gamma^*(Q^2) \to q\bar{q}$, is defined as follows:

$$\Psi_{\gamma^*(Q^2)\to q\bar{q}}(s) = \frac{G_{\gamma\to q\bar{q}}(s)}{s + Q^2} \tag{2}$$

where $G_{\gamma\to q\bar{q}}(s)$ is the vertex for the transition of the photon into the $q\bar{q}$ state and s is the energy squared of the $q\bar{q}$ system. Rather schematically, the vertex function $G_{\gamma\to q\bar{q}}(s)$ may be represented as

$$Ce^{-bs} + \theta(s - s_0), \tag{3}$$

where the first term stands for the soft component which is due to the transition of photons to vector $q\bar{q}$ mesons $\gamma \to V \to q\bar{q}$, while the second one describes the pointlike interaction in the hard domain (here, the

*The text was submitted by the authors in English.

**e-mail: anisovic@thd.pnpi.spb.ru

step function $\theta(s-s_0)=1$ at $s\geq s_0$ and $\theta(s-s_0)=0$ at $s<s_0$). The principal characteristics of the soft component of $G_{\gamma\to q\bar{q}}(s)$ is the threshold value of the vertex, $C\exp(-4m^2b)$, and the rate of its decrease with energy, that is, the slope b. The hard component of the vertex is characterized by the magnitude s_0, which is the quark energy squared when pointlike interaction becomes dominant.

In [9], the photon wave function has been found assuming that the quark relative-momentum dependence is the same for all quark vertices: $g_{\gamma\to u\bar{u}}(k^2)=g_{\gamma\to d\bar{d}}(k^2)=g_{\gamma\to s\bar{s}}(k^2)$, where we redenoted $G_{\gamma\to q\bar{q}}(s)\to g_{\gamma\to q\bar{q}}(k^2)$ with $k^2=s/4-m^2$. The hypothesis of the vertex universality for u and d quarks,

$$G_{\gamma\to u\bar{u}}(s)=G_{\gamma\to d\bar{d}}(s)\equiv G_\gamma(s) \qquad (4)$$

looks rather trustworthy because of the degeneracy of ρ and ω states, though the similarity in the k dependence for nonstrange and strange quarks may be violated. In addition, using experimental data on the transitions $\gamma\gamma^*(Q^2)\to\pi^0,\eta,\eta'$ only, one cannot find the main parameters (C,b,s_0) for both $G_{\gamma\to s\bar{s}}(s)$ and $G_\gamma(s)$. In the present paper, we add the e^+e^--annihilation data for the determination of wave functions, that is, $e^+e^-\to\gamma^*\to\rho^0,\omega,\phi(1020)$, together with the ratio $R(E_{e^+e^-})=\sigma(e^+e^-\to hadrons)/\sigma(e^+e^-\to\mu^+\mu^-)$ at $E_{e^+e^-}$ higher than 1 GeV. The reactions $e^+e^-\to\gamma^*\to\rho^0,\omega,\phi(1020)$ are rather sensitive to the parameters C and b of the soft component of the photon wave function, while the data on $R(E_{e^+e^-})$ allow us to fix the parameter s_0 for the beginning of the pointlike vertex regime [see Eq. (3)].

The paper is organized as follows. In Section 2, which is in fact the introductory one, we present the formulas for the charge form factor of the pseudoscalar meson and transition form factors $\pi^0,\eta,\eta'\to\gamma(Q_1^2)\gamma(Q_2^2)$ in terms of the spectral integration technique. In Section 3, we consider the e^+e^--annihilation processes: the partial decay widths $\omega,\rho^0,\phi\to e^+e^-$ and the ratio $R(E_{e^+e^-})=\sigma(e^+e^-\to hadrons)/\sigma(e^+e^-\to\mu^+\mu^-)$ at $1\leq E_{e^+e^-}\leq 3.7$ GeV. The photon wave function $\gamma\to q\bar{q}$ for the light quarks is determined in Section 4. The results of calculations for the decays $f_0(980),a_0(980)\to\gamma\gamma$ and $f_2(1270),f_2(1525),a_2(1320)\to\gamma\gamma$ carried out with the photon wave function found here are compared with calculations performed with the old photon wave function [9] in Section 5. In the Conclusion, we briefly summarize the results.

2. QUARK–ANTIQUARK-STATE FORM FACTORS IN THE SPECTRAL-INTEGRATION TECHNIQUE

In this section, we recall the main formulas for the calculation of the charge and transition form factors in the spectral-integration technique; these formulas are used for the determination of the photon wave function. First, we present formulas for the pion charge form factor—they are needed to fix up the wave function of the pion and other members of the lowest pseudoscalar nonet, η and η'. The calculation of the charge form factor is based on a fundamental hypothesis of the additive quark model: the mesons consist of a quark and antiquark, and the photon interacts with one of the constituent quarks. Hereafter, the formulas for the transitions $\pi^0,\eta,\eta'\to\gamma(Q_1^2)\gamma(Q_2^2)$ are given; they are written within a similar approach. A more detailed discussion of these formulas and basic assumptions may be found in [12–21].

2.1. Pion Charge Form Factor

Here, we recall the logic of calculation in the spectral-integration technique and write down the formulas for the pion form factor.

The general structure of the amplitude of pion–photon interaction is as follows:

$$A_\mu^{(\pi)}=e(p_\mu+p'_\mu)F_\pi(Q^2), \qquad (5)$$

where e is the absolute value of electron charge, p and p' are the pion incoming and outgoing 4-momenta, and $F_\pi(Q^2)$ is the pion form factor. We are working in the spacelike region of the momentum transfer, so $Q^2=-q^2$, where $q=p-p'$. The amplitude $A_\mu^{(\pi)}$ is the transverse one: $q_\mu A_\mu^{(\pi)}=0$.

In the quark model, the pion form factor is defined as a process shown in Fig. 1*a*: the photon interacts with one of the constituent quarks. In the spectral-integration technique, the method of calculation of the diagram of Fig. 1*a* is as follows: we consider the dispersive integrals over masses of incoming and outgoing $q\bar{q}$ states, and the corresponding cuttings of the triangle diagram are shown in Fig. 1*b*. In this way, we calculate the double discontinuity of the triangle diagram, $\mathrm{disc}_s\mathrm{disc}_{s'}F_\pi(s,s',Q^2)$, where s and s' are the energy squared of the $q\bar{q}$ systems before and after the photon emission, $P^2=s$ and $P'^2=s'$ (recall that, in the dispersion relation technique, the momenta of intermediate particles do not coincide with external momenta, $p\neq P$ and $p'\neq P'$). The double discontinuity is defined by three factors:

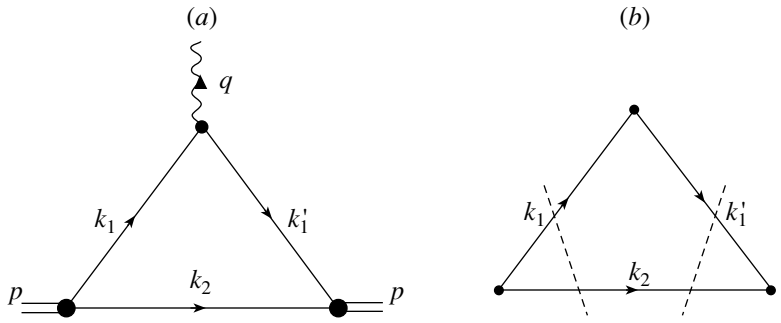

Fig. 1. (*a*) Diagram for the meson charge form factor in the additive quark model. (*b*) Cuts of the triangle diagram in the spectral integral representation.

(i) product of the pion vertex functions and quark charge

$$e_q G_\pi(s) G_\pi(s'), \tag{6}$$

where, due to Eq. (5), e_q is given in units of the charge e;

(ii) phase space of the triangle diagram (Fig. 1*b*) at $s \geq 4m^2$ and $s' \geq 4m^2$:

$$d\Phi_{\rm tr} = d\Phi_2(P; k_1, k_2) d\Phi_2(P'; k_1', k_2')(2\pi)^3 \times 2k_{20}\delta^{(3)}(\mathbf{k}_2 - \mathbf{k}_2'), \tag{7}$$

with the two-particle phase space determined as

$$d\Phi_2(P; k_1, k_2) = \frac{1}{2}\frac{d^3k_1}{(2\pi)^3 \times 2k_{10}} \times \frac{d^3k_2}{(2\pi)^3 \times 2k_{20}}(2\pi)^4\delta^{(4)}(P - k_1 - k_2); \tag{8}$$

(iii) spin factor $S_\pi(s, s', Q^2)$ determined by the trace of the triangle diagram process of Fig. 1*b*:

$$-{\rm Tr}[i\gamma_5(m - \hat{k}_2)i\gamma_5(m + \hat{k}_1')\gamma_\mu^\perp(m + \hat{k}_1)] = (P + P')_\mu^\perp S_\pi(s, s', Q^2). \tag{9}$$

Recall that, in the dispersion integral, we deal with mass-on-shell particles, so $k_1^2 = k_1'^2 = k_2^2 = m^2$. The vertex $i\gamma_5$ corresponds to the transition $\pi \to q\bar{q}$; the photon carries the momentum $\tilde{q} = P - P'$; and the photon momentum squared is fixed, $\tilde{q}^2 = q^2 = -Q^2$. The transversity of the amplitude $A_\mu^{(\pi)}$ is guaranteed by the use of $\gamma_\mu^\perp$ in the trace (9):

$$\gamma_\mu^\perp = g_{\mu\nu}^\perp \gamma_\nu, \tag{10}$$

$$g_{\mu\nu}^\perp = g_{\mu\nu} - \frac{(P_\mu' - P_\mu)(P_\nu' - P_\nu)}{q^2},$$

$$(P + P')_\mu^\perp = \left[P_\mu' + P_\mu - \frac{P_\mu' - P_\mu}{q^2}(s' - s)\right].$$

The spin factor $S_\pi(s, s', Q^2)$ reads

$$S_\pi(s, s', Q^2) = 2\left[(s + s' + Q^2)\alpha(s, s', Q^2) - Q^2\right], \tag{11}$$

$$\alpha(s, s', Q^2) = \frac{s + s' + Q^2}{2(s + s') + (s' - s)^2/Q^2 + Q^2}.$$

As a result, the double discontinuity of the diagram with a photon emitted by a quark is determined as

$$e_q G_\pi(s) G_\pi(s') S_\pi(s, s', Q^2) d\Phi_{\rm tr}. \tag{12}$$

Emission of a photon by an antiquark gives a similar contribution, with the substitution $e_q \to e_{\bar{q}}$, so the total charge factor for the π^+ is unity, $e_u + e_{\bar{d}} = 1$. Then the double discontinuity reads

$${\rm disc}_s{\rm disc}_{s'} F_\pi(s, s', Q^2) = G_\pi(s) G_\pi(s') S_\pi(s, s', Q^2) d\Phi_{\rm tr}. \tag{13}$$

The form factor $F_\pi(Q^2)$ is defined as a double dispersion integral as follows:

$$F_\pi(Q^2) = \int\limits_{4m^2}^{\infty} \frac{ds}{\pi}\frac{ds'}{\pi}\frac{{\rm disc}_s{\rm disc}_{s'} F_\pi(s, s', Q^2)}{(s' - m_\pi^2)(s - m_\pi^2)}. \tag{14}$$

When the form factor calculations are performed, it is suitable to operate with the wave function of the composite system. In the case of a pion, the wave function is defined as follows:

$$\Psi_\pi(s) = \frac{G_\pi(s)}{s - m_\pi^2}. \tag{15}$$

There are different ways to work with formula (14), in accordance with different goals, where the $q\bar{q}$ system is involved. The spectral representation of the form factor appears after the integration in (14) over the momenta of constituents by removing δ functions in the phase space $d\Phi_{\rm tr}$. Then

$$F_\pi(Q^2) = \int\limits_{4m^2}^{\infty} \frac{ds}{\pi}\frac{ds'}{\pi}\Psi_\pi(s)\Psi_\pi(s') \tag{16}$$

$$\times S_\pi(s,s'Q^2)\frac{\theta(s'sQ^2-m^2\lambda(s,s',Q^2))}{16\sqrt{\lambda(s,s',Q^2)}},$$

$$\lambda(s,s',Q^2)=(s'-s)^2+2Q^2(s'+s)+Q^4.$$

Here, the θ function determines the integration region over s and s': $\theta(X)=1$ at $X\geq 0$ and $\theta(X)=0$ at $X<0$.

Another way to present form factor is to remove the integration over the energy squared of the quark–antiquark systems, s and s', by using δ functions entering $d\Phi_{\rm tr}$. Then we have the formula for the pion form factor in the light-cone variables:

$$F_\pi(Q^2)=\frac{1}{16\pi^3}\int\limits_0^1\frac{dx}{x(1-x)^2} \tag{17}$$

$$\times\int d^2k_\perp\Psi_\pi(s)\Psi_\pi(s')S_\pi(s,s',Q^2),$$

$$s=\frac{m^2+k_\perp^2}{x(1-x)},\quad s'=\frac{m^2+(\mathbf{k}_\perp-x\mathbf{Q})^2}{x(1-x)},$$

where $\mathbf{k}_\perp$ and x are the light-cone quark characteristics (transverse momentum of the quark and a part of the momentum along the z axis).

Fitting formula (16) or (17) to data at $0\leq Q^2\leq 1\,(\mathrm{GeV}/c)^2$ with two-exponential parametrization of the wave function Ψ_π,

$$\Psi_\pi(s)=c_\pi\left[\exp(-b_1^\pi s)+\delta_\pi\exp(-b_2^\pi s)\right], \tag{18}$$

we obtain the following values of the pion wave function parameters:

$$c_\pi=209.36\ \mathrm{GeV}^{-2},\quad \delta_\pi=0.01381, \tag{19}$$
$$b_1^\pi=3.57\ \mathrm{GeV}^{-2},\quad b_2^\pi=0.4\ \mathrm{GeV}^{-2}.$$

Figure 2 displays the description of the data by formula (16) [or (17)] with the pion wave function given by (18), (19).

The region $1\leq Q^2\leq 2\ (\mathrm{GeV}/c)^2$ was not used for the determination of parameters of the pion wave function: one may suppose that, at $Q^2\geq 1\,(\mathrm{GeV}/c)^2$, the predictions of the additive quark model fail. However, one can see that the calculated curve fits reasonably well to data in the neighboring region $1\leq Q^2\leq 2\,(\mathrm{GeV}/c)^2$ too (dashed curve in Fig. 2).

The constraint $F_\pi(0)=1$ serves us as a normalization condition for the pion wave function. We have in the low-Q^2 region

$$F_\pi(Q^2)\simeq 1-\frac{1}{6}R_\pi^2Q^2, \tag{20}$$

with $R_\pi^2\simeq 10\ \mathrm{GeV}^{-2}$. The pion radius is just the characteristic which is used later on for comparative estimates of the wave function parameters for other low-lying $q\bar{q}$ states.

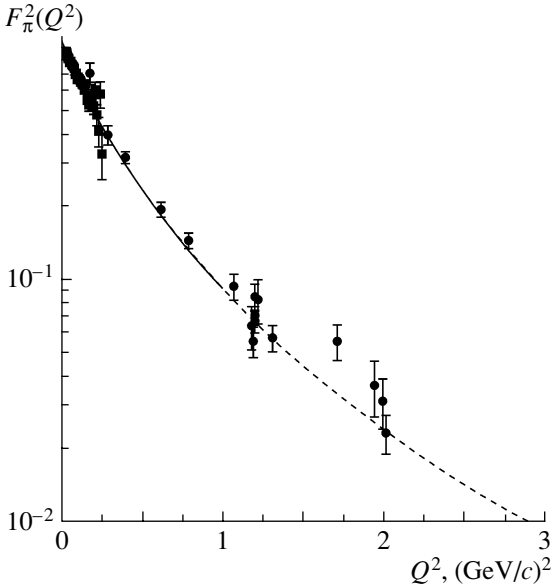

Fig. 2. Description of the experimental data on the pion charge form factor with the pion wave function given by (18), (19).

2.2. Transition Form Factors $\pi^0,\eta,\eta'\to\gamma^*(Q_1^2)\gamma^*(Q_2^2)$

Using the same technique, we can write the formulas for transition form factors of pseudoscalar mesons $\pi^0,\eta,\eta'\to\gamma^*(Q_1^2)\gamma^*(Q_2^2)$; the corresponding diagrams are shown in Fig. 3.

For these processes, the general structure of the amplitude is as follows:

$$A_{\mu\nu}(Q_1^2,Q_2^2)=e^2\epsilon_{\mu\nu\alpha\beta}q_\alpha p_\beta F_{\pi,\eta,\eta'\to\gamma\gamma}(Q_1^2,Q_2^2). \tag{21}$$

In the light-cone variables $(x,\mathbf{k}_\perp)$, the expression for the transition form factor $\pi^0\to\gamma^*(Q_1^2)\gamma^*(Q_2^2)$ determined by two processes of Fig. 3*a* and Fig. 3*b* reads

$$F_{\pi\to\gamma\gamma}(Q_1^2,Q_2^2)=\zeta_{\pi\to\gamma\gamma}\frac{\sqrt{N_c}}{16\pi^3} \tag{22}$$

$$\times\int\limits_0^1\frac{dx}{x(1-x)^2}\int d^2k_\perp\Psi_\pi(s)$$

$$\times\left(S_{\pi\to\gamma\gamma}(s,s_1',Q_1^2)\frac{G_\gamma(s_1')}{s_1'+Q_2^2}\right.$$

$$\left.+\,S_{\pi\to\gamma\gamma}(s,s_2',Q_2^2)\frac{G_\gamma(s_2')}{s_2'+Q_1^2}\right),$$

where

$$s=\frac{m^2+k_\perp^2}{x(1-x)}, \tag{23}$$

$$s_i'=\frac{m^2+(\mathbf{k}_\perp-x\mathbf{Q}_i)^2}{x(1-x)}\qquad(i=1,2).$$

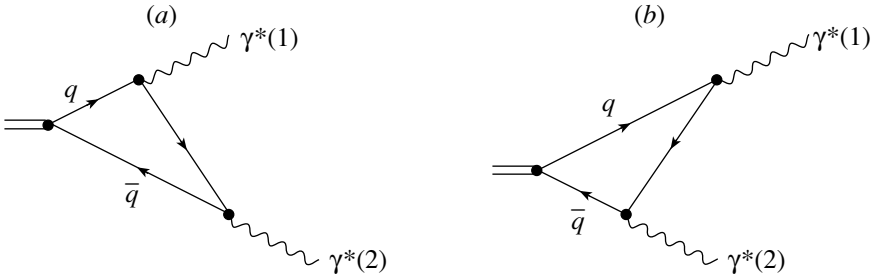

Fig. 3. Diagrams for the two-photon decay of the $q\bar{q}$ state.

The spin factor for pseudoscalar states depends on the quark mass only:

$$S_{\pi\to\gamma\gamma}(s,s_i',Q^2)=4m. \tag{24}$$

The charge factor for the decay $\pi^0\to\gamma\gamma$ is equal to

$$\zeta_{\pi\to\gamma\gamma}=\frac{e_u^2-e_d^2}{\sqrt{2}}=\frac{1}{3\sqrt{2}}. \tag{25}$$

In (22), the ratio $G_\gamma(s_i)/(s_i+Q^2)$ is the photon wave function (recall that we denote $G_{\gamma\to u\bar{u}}(s)=G_{\gamma\to d\bar{d}}(s)\equiv G_\gamma(s)$). The factor $\sqrt{N_c}$ on the right-hand side of (22) appears due to another definition of the color wave function of the photon as compared to pion's one: without $1/\sqrt{N_c}$.

In terms of the spectral integrals over the (s,s') variables, the transition form factor for $\pi^0\to\gamma^*(Q_1^2)\gamma^*(Q_2^2)$ reads

$$F_{\pi\to\gamma\gamma}(Q_1^2,Q_2^2)=\zeta_{\pi\to\gamma\gamma}\frac{\sqrt{N_c}}{16} \tag{26}$$
$$\times\int\limits_{4m^2}^{\infty}\frac{ds}{\pi}\frac{ds'}{\pi}\Psi_\pi(s)\left[\frac{\theta(s'sQ_1^2-m^2\lambda(s,s',Q_1^2))}{\sqrt{\lambda(s,s',Q_1^2)}}\right.$$
$$\times S_\pi(s,s',Q_1^2)\frac{G_\gamma(s')}{s'+Q_2^2}$$
$$\left.+\frac{\theta(s'sQ_2^2-m^2\lambda(s,s',Q_2^2))}{\sqrt{\lambda(s,s',Q_2^2)}}S_\pi(s,s',Q_2^2)\frac{G_\gamma(s')}{s'+Q_1^2}\right],$$

where $\lambda(s,s',Q_i^2)$ is determined in (16).

Similar expressions may be written for the transitions $\eta,\eta'\to\gamma^*(Q_1^2)\gamma^*(Q_2^2)$. One should bear in mind that, because of the presence of two quarkonium components, their flavor wave functions are $s\bar{s}$ and $n\bar{n}=(u\bar{u}+d\bar{d})/\sqrt{2}$; in the η, η' mesons,

$$\eta=n\bar{n}\sin\theta-s\bar{s}\cos\theta,\quad \eta'=n\bar{n}\cos\theta+s\bar{s}\sin\theta,$$

their transition form factors are expressed through mixing angle θ as follows:

$$F_{\eta\to\gamma\gamma}(s)=F_{\eta/\eta'(n\bar{n})\to\gamma\gamma}(s)\sin\theta \tag{27}$$
$$-F_{\eta/\eta'(s\bar{s})\to\gamma\gamma}(s)\cos\theta,$$
$$F_{\eta'\to\gamma\gamma}(s)=F_{\eta/\eta'(n\bar{n})\to\gamma\gamma}(s)\cos\theta$$
$$+F_{\eta/\eta'(s\bar{s})\to\gamma\gamma}(s)\sin\theta.$$

The spin factors for nonstrange components of η and η' are the same as those for the pion, see (24), though with another quark mass entering the strange component: $S_{\eta/\eta'(s\bar{s})\to\gamma\gamma}(s,s',Q^2)=4m_s$.

The charge factors for the $n\bar{n}$ and $s\bar{s}$ components are equal to

$$\zeta_{\eta/\eta'(n\bar{n})\to\gamma\gamma}=\frac{5}{9\sqrt{2}},\quad \zeta_{\eta/\eta'(s\bar{s})\to\gamma\gamma}=\frac{1}{9}. \tag{28}$$

In the calculation of transition form factors of pseudoscalar mesons, the wave function related to nonstrange quarks in η and η' was assumed to be the same as for the pion: $\Psi_{\eta/\eta'(n\bar{n})}(s)=\Psi_\pi(s)$. As to strange components of the wave functions, we suppose the same shape for $n\bar{n}$ and $s\bar{s}$. For $\Psi_{\eta/\eta'(s\bar{s})}(s)=\Psi_\pi(s)$, it results in another normalization only:

$$\Psi_{\eta/\eta'(s\bar{s})}(s)=c_{\eta/\eta'(s\bar{s})}[\exp(-b_1^{\eta/\eta'(s\bar{s})}s) \tag{29}$$
$$+\delta_{\eta/\eta'(s\bar{s})}\exp(-b_2^{\eta/\eta'(s\bar{s})}s)],$$

$$c_{\eta/\eta'(s\bar{s})}=528.78\ \mathrm{GeV}^{3/2},\quad \delta_{\eta/\eta'(s\bar{s})}=\delta_\pi,$$
$$b_1^{\eta/\eta'(s\bar{s})}=b_1^\pi,\quad b_2^{\eta/\eta'(s\bar{s})}=b_2^\pi.$$

3. e^+e^- ANNIHILATION

The e^+e^--annihilation processes provide us with additional information on the photon wave function:

(i) The partial width of the transitions $\omega,\rho^0,\phi\to e^+e^-$ is defined by the quark loop diagrams, which contain the product $G_\gamma(s)\Psi_V(s)$, where $\Psi_V(s)$ is the quark wave function of the vector meson ($V=\omega,\rho^0,\phi$). Supposing that radial wave functions of ω,ρ^0,ϕ coincide with those of the lowest pseudoscalar mesons (this assumption looks reasonable, for these mesons are members of the same lowest 36-plet), we can obtain information about $G_\gamma(s)$ and $G_{\gamma(s\bar{s})}(s)$ from the data on the $\omega,\rho^0,\phi\to e^+e^-$ decays [22].

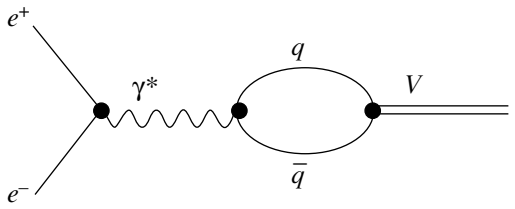

Fig. 4. Production of the vector $q\bar{q}$ state in the e^+e^- annihilation (V stands for a vector meson).

(ii) The ratio $R(s) = \sigma(e^+e^- \to hadrons)/\sigma(e^+e^- \to \mu^+\mu^-)$ at high center-of-mass energies but below the open charm production ($\sqrt{s} \equiv E_{e^+e^-} < 3.7$ GeV) is determined by hard components of the photon vertices $G_\gamma(s)$ and $G_{\gamma(s\bar{s})}(s)$ (transitions $\gamma^* \to u\bar{u}, d\bar{d}, s\bar{s}$), thus giving us a well-known magnitude $R(s) = 2$ (small deviations from $R(s) = 2$ come from corrections related to the gluon emission $\gamma^* \to q\bar{q}g$; see [23] and references therein). Hence, the deviation of the ratio from the value $R(s) = 2$ at decreasing $E_{e^+e^-}$ provides us with information about the energies when the regime changes: hard components in $G_\gamma(s)$ and $G_{\gamma(s\bar{s})}(s)$ stop working, while soft components start to play their role.

3.1. Partial Decay Widths $\omega, \rho^0, \phi \to e^+e^-$

Figure 4 is a diagrammatic representation of the reaction $V \to e^+e^-$: a virtual photon produces the $q\bar{q}$ pair, which turns into a vector meson.

The partial width of the vector meson is determined as follows:

$$m_V \Gamma_{V\to e^+e^-} = \pi\alpha^2 A^2_{e^+e^-\to V} \frac{1}{m_V^4} \times \left(\frac{4}{3}m_V^2 + \frac{8}{3}m_e^2\right)\sqrt{\frac{m_V^2 - 4m_e^2}{m_V^2}}. \quad (30)$$

Here, m_V is the vector meson mass, the factor $1/m_V^2$ is associated with the photon propagator, and $\alpha = e^2/(4\pi)$. In (30), the integration over electron–positron phase space results in $\sqrt{(m_V^2 - 4m_e^2)/m_V^2}/(16\pi)$, while the averaging over vector-meson polarizations and summing over electron–positron spins gives

$$\frac{1}{3}\mathrm{Tr}\left[\gamma_\mu^\perp(\hat{k}_1 + m_e)\gamma_\mu^\perp(-\hat{k}_2 + m_e)\right] = \frac{4}{3}m_V^2 + \frac{8}{3}m_e^2. \quad (31)$$

The amplitude $A_{V\to e^+e^-}$ is determined through the quark–antiquark loop calculations, within the spectral-integration technique. In this way, we get for the decays $\omega, \rho^0 \to e^+e^-$

$$A_{\omega,\rho^0\to e^+e^-} = Z_{\omega,\rho^0}\frac{\sqrt{N_c}}{16\pi} \quad (32)$$

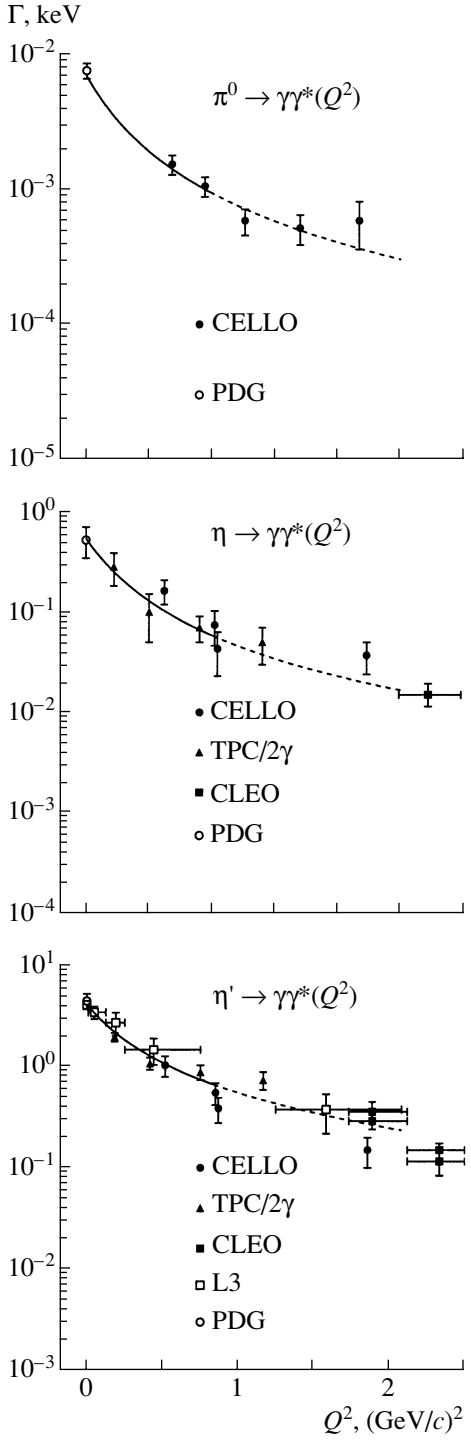

Fig. 5. Data for $\pi^0 \to \gamma\gamma^*$, $\eta \to \gamma\gamma^*$, and $\eta' \to \gamma\gamma^*$ vs. the calculation curves.

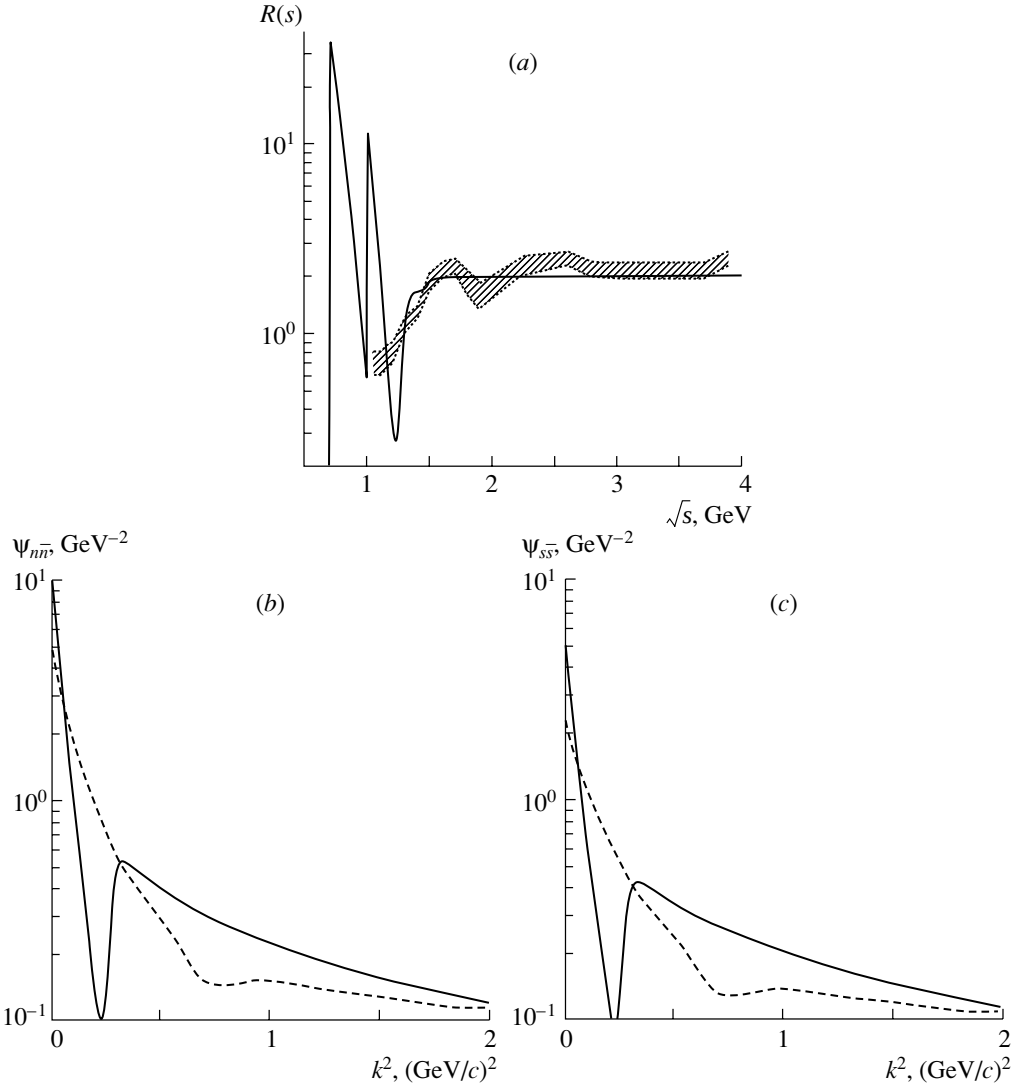

Fig. 6. (*a*) $R_{\rm vert}(s)$ [solid line, Eq. (39)] vs. $R(s) = \sigma(e^+e^- \to hadrons)/\sigma(\mu^+\mu^- \to hadrons)$ (shaded area). The k^2 dependence of the photon wave function (k^2 is the relative quark momentum squared): (*b*) $\Psi_{\gamma\to n\bar{n}}(k^2)$ and (*c*) $\Psi_{\gamma\to s\bar{s}}(k^2)$. Solid curves stand for the new wave functions and dashed curves for the old one.

$$\times \int\limits_{4m^2}^{\infty} \frac{ds}{\pi} G_\gamma(s)\Psi_{\omega,\rho}(s)\sqrt{\frac{s-4m^2}{s}}\left(\frac{8}{3}m^2+\frac{4}{3}s\right),$$

where Z_{ω,ρ^0} is the quark charge factor for vector mesons: $Z_\omega = 1/(3\sqrt{2})$ and $Z_{\rho^0} = 1/\sqrt{2}$. We have a similar expression for the $\phi(1020) \to e^+e^-$ amplitude:

$$A_{\phi\to e^+e^-} = Z_\phi \frac{\sqrt{N_c}}{16\pi} \tag{33}$$
$$\times \int\limits_{4m_s^2}^{\infty} \frac{ds}{\pi} G_{\gamma\to s\bar{s}}(s)\Psi_\phi(s)\sqrt{\frac{s-4m_s^2}{s}}\left(\frac{8}{3}m_s^2+\frac{4}{3}s\right),$$

with $Z_\phi = 1/3$. The normalization condition for the vector-meson wave function reads

$$\frac{1}{16\pi}\int\limits_{4m^2}^{\infty}\frac{ds}{\pi}\Psi_V^2(s)\sqrt{\frac{s-4m^2}{s}}\left(\frac{8}{3}m^2+\frac{4}{3}s\right)=1. \tag{34}$$

The wave function is parametrized in a one-exponential form:

$$\Psi_V(s) = c_V \exp(-b_V s), \tag{35}$$

with

$$b_{\omega,\rho} = 2.2\ \text{GeV}^{-2}, \quad c_{\omega,\rho} = 95.1\ \text{GeV}^{-2} \tag{36}$$

for the nonstrange mesons and

$$b_\phi = 2.5\ \text{GeV}^{-2}, \quad c_\phi(s\bar{s}) = 374.8\ \text{GeV}^{-2} \tag{37}$$

for the $\phi(1020)$. Within the parametrization used, the vector mesons are characterized by the following mean radii squared: $R^2_{\omega,\rho} = 10\ \mathrm{GeV}^{-2}$ and $R^2_\phi = 11\ \mathrm{GeV}^{-2}$.

3.2. The Ratio $R(s) = \sigma(e^+e^- \to hadrons)/\sigma(e^+e^- \to \mu^+\mu^-)$ at Energies below the Open Charm Production

At high energies but below the open charm production, $E_{e^+e^-} = \sqrt{s} < 3.7$ GeV, the ratio $R(s)$ is determined by the sum of quark charges squared in the transition $e^+e^- \to \gamma^* \to u\bar{u} + d\bar{d} + s\bar{s}$ multiplied by the factor $N_c = 3$:

$$R(s) = \frac{\sigma(e^+e^- \to hadrons)}{\sigma(e^+e^- \to \mu^+\mu^-)} = N_c(e_u^2 + e_d^2 + e_s^2) = 2. \quad (38)$$

Since the $G_\gamma(s)$ and $G_{\gamma(s\bar{s})}(s)$ vertices are normalized as $G_\gamma(s) = G_{\gamma(s\bar{s})}(s) = 1$ at $s \to \infty$, at large s, we can relate $R(s)$ and

$$R_{\mathrm{vert}}(s) = 3(e_u^2 + e_d^2)G_\gamma^2(s) + 3e_s^2 G^2_{\gamma(s\bar{s})}(s) = \frac{5}{3}G_\gamma^2(s) + \frac{1}{3}G^2_{\gamma(s\bar{s})}(s) \quad (39)$$

to each other:

$$R(s) \simeq R_{\mathrm{vert}}(s). \quad (40)$$

Following (40), we determine the energy region where the hard components in $G_\gamma(s)$ and $G_{\gamma(s\bar{s})}(s)$ start to dominate.

4. PHOTON WAVE FUNCTION

To determine the photon wave function we use

(i) transition widths $\pi^0, \eta, \eta' \to \gamma\gamma^*(Q^2)$,

(ii) partial decay widths $\omega, \rho^0, \phi \to e^+e^-, \mu^+\mu^-$,

(iii) the ratio $R(s) = \sigma(e^+e^- \to hadrons)/\sigma(e^+e^- \to \mu^+\mu^-)$.

Transition vertices for $u\bar{u}, d\bar{d} \to \gamma$, and $s\bar{s} \to \gamma$ have been chosen in the following form:

$$u\bar{u}, d\bar{d}: \quad G_\gamma(s) = c_\gamma(e^{-b_1^\gamma s} + c_2^\gamma e^{-b_2^\gamma s}) + \frac{1}{1 + e^{-b_0^\gamma(s - s_0^\gamma)}}, \quad (41)$$

$$s\bar{s}: \quad G_{\gamma(s\bar{s})}(s) = c_{\gamma(s\bar{s})}e^{-b_1^{\gamma(s\bar{s})}s} + \frac{1}{1 + e^{-b_0^{\gamma(s\bar{s})}(s - s_0^{\gamma(s\bar{s})})}}.$$

Recall that photon wave function is determined as $\Psi_\gamma(s, Q^2) = G_\gamma(s)/(s + Q^2)$ [see Eq. (2)].

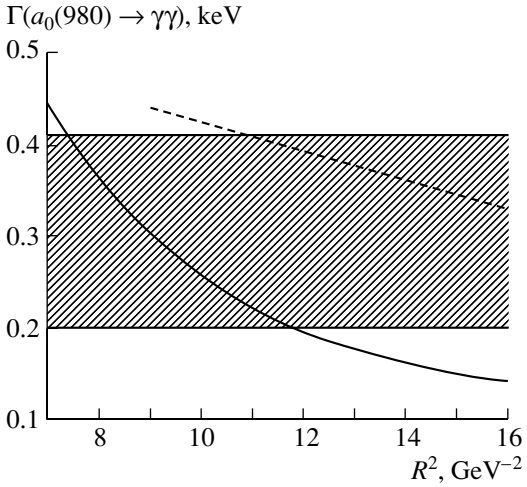

Fig. 7. Partial width $\Gamma(a_0(980) \to \gamma\gamma)$ calculated under the assumption that $a_0(980)$ is the $q\bar{q}$ system, being a function of radius squared of $a_0(980)$. The solid curve stands for the calculation with the new photon wave function, and the dashed curve stands for the old one. The shaded area corresponds to the values allowed by the data [22].

The following parameter values have been found in fitting to data:

$$u\bar{u}, d\bar{d}: \quad c^\gamma = 32.577, \quad c_2^\gamma = -0.0187, \quad b_1^\gamma = 4\ \mathrm{GeV}^{-2}, \quad b_2^\gamma = 0.8\ \mathrm{GeV}^{-2}, \quad b_0^\gamma = 15\ \mathrm{GeV}^{-2}, \quad s_0^\gamma = 1.62\ \mathrm{GeV}^2, \quad (42)$$

$$s\bar{s}: \quad c_{\gamma(s\bar{s})} = 310.55, \quad b_1^{\gamma(s\bar{s})} = 4\ \mathrm{GeV}^{-2}, \quad b_0^{\gamma(s\bar{s})} = 15\ \mathrm{GeV}^{-2}, \quad s_0^{\gamma(s\bar{s})} = 2.15\ \mathrm{GeV}^2. \quad (43)$$

Now, let us present the results of the fit in more detail.

Figure 5 shows the data for $\pi^0 \to \gamma\gamma^*(Q^2)$ [4, 22], $\eta \to \gamma\gamma^*(Q^2)$ [4–6, 22], and $\eta' \to \gamma\gamma^*(Q^2)$ [2, 4–6, 22]. We perform the fitting procedure in the interval $0 \le Q^2 \le 1\ (\mathrm{GeV}/c)^2$; the fitting curves are shown by solid curves. The continuation of the curves into the neighboring region $1 \le Q^2 \le 2\ (\mathrm{GeV}/c)^2$ (dashed curves) demonstrates that there is also a reasonable description of the data.

The calculation results for the $V \to e^+e^-$ decay partial widths vs. data [22] are shown below (in keV):

$$\Gamma^{\mathrm{calc}}_{\rho^0\to e^+e^-} = 7.50, \quad \Gamma^{\mathrm{exp}}_{\rho^0\to e^+e^-} = 6.77 \pm 0.32, \quad (44)$$
$$\Gamma^{\mathrm{calc}}_{\omega\to e^+e^-} = 0.796, \quad \Gamma^{\mathrm{exp}}_{\omega\to e^+e^-} = 0.60 \pm 0.02,$$
$$\Gamma^{\mathrm{calc}}_{\phi\to e^+e^-} = 1.33, \quad \Gamma^{\mathrm{exp}}_{\phi\to e^+e^-} = 1.32 \pm 0.06,$$
$$\Gamma^{\mathrm{calc}}_{\rho^0\to \mu^+\mu^-} = 7.48, \quad \Gamma^{\mathrm{exp}}_{\rho^0\to \mu^+\mu^-} = 6.91 \pm 0.42,$$
$$\Gamma^{\mathrm{calc}}_{\phi\to \mu^+\mu^-} = 1.33, \quad \Gamma^{\mathrm{exp}}_{\phi\to \mu^+\mu^-} = 1.65 \pm 0.22.$$

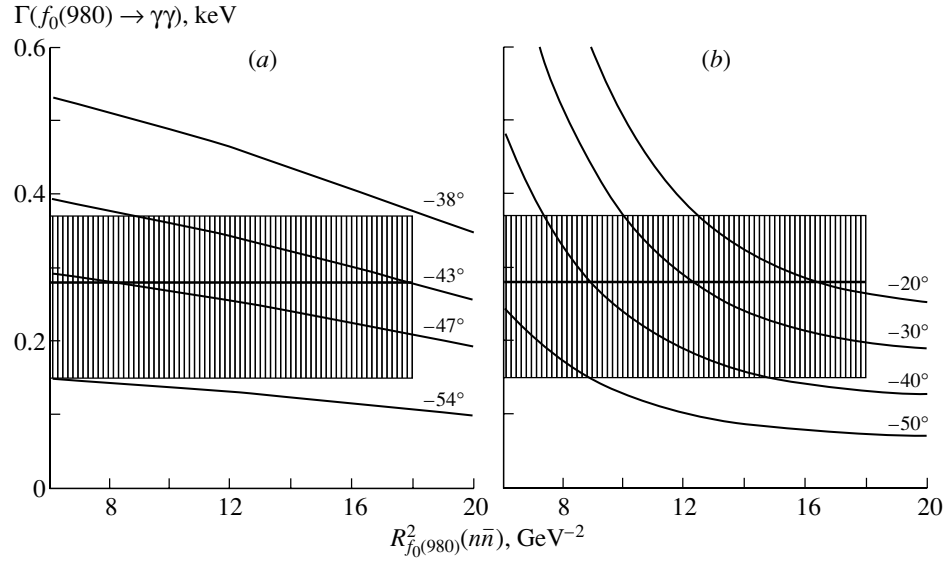

Fig. 8. Partial width $\Gamma(f_0(980) \to \gamma\gamma)$ calculated under the assumption that $f_0(980)$ is the $q\bar{q}$ system, $q\bar{q} = n\bar{n}\cos\varphi + s\bar{s}\sin\varphi$, depending on the radius squared of the $q\bar{q}$ system: (a) with the old photon wave function and (b) with the new one. Calculations were carried out with different values of mixing angle φ in the region $\varphi < 0$. The shaded area shows the allowed experimental values [26].

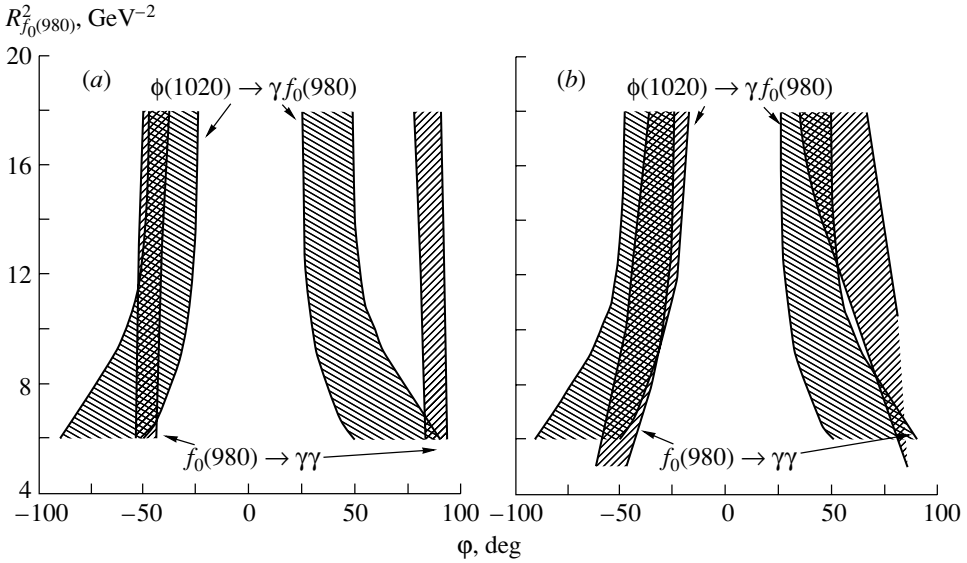

Fig. 9. Combined presentation of the $(R^2_{f_0(980)}, \varphi)$ areas allowed by the experiment for the decays $f_0(980) \to \gamma\gamma$ and $\phi(1020) \to \gamma f_0(980)$ with (a) the old photon wave function and (b) the new one.

Figure 6*a* displays the data for $R(s)$ [23] at $E_{e^+e^-} > 1$ GeV (dashed area) versus $R_{\rm vert}(s)$ given by Eq. (39) with parameters (42) (solid line).

In Figs. 6*b* and 6*c*, one can see the k^2 dependence of photon wave functions on the quark relative-momentum squared k^2 (here, $s = 4m^2 + 4k^2$) for the nonstrange and strange components found in our fit (solid curves) and those found in [9] (dashed curves). One may see that, in the region $0 \le k^2 \le 2\,(\mathrm{GeV}/c)^2$, the scrupulous distinction is rather considerable, though, on average, the old and new wave functions almost coincide. In the next section, we compare the results obtained for the two-photon decays of scalar and tensor mesons, $S \to \gamma\gamma$ and $T \to \gamma\gamma$, calculated with the old and new wave functions.

5. TRANSITIONS $S \to \gamma\gamma$ AND $T \to \gamma\gamma$

As was mentioned above, on average, the old [9] and new photon wave functions coincide, though they

differ in details. So it would be useful to understand to what extent this difference influences the calculation results for the two-photon decays of scalar and tensor mesons.

The calculation of the two-photon decays of scalar mesons $f_0(980) \to \gamma\gamma$ and $a_0(980) \to \gamma\gamma$ has been performed in [12, 13] with the old wave function, under the assumption that $f_0(980)$ and $a_0(980)$ are the $q\bar{q}$ systems. The results of the calculations are shown in Fig. 7 (dashed curve). The solid curve represents the values found with the new photon wave function; for $a_0(980)$, the new wave function reveals a steeper dependence on the radius squared as compared to the old wave function. In the region $R^2_{a_0(980)} \sim R^2_\pi = 10\ \text{GeV}^{-2}$, the value $\Gamma(a_0(980) \to \gamma\gamma)$ calculated with the new wave function becomes 1.5–2 times smaller as compared to the old wave function—this reflects the more precise definition of the photon wave function. In the calculations, the flavor wave function of $f_0(980)$ was defined as follows:

$$n\bar{n}\cos\varphi + s\bar{s}\sin\varphi.$$

In Fig. 8, the calculated areas are shown for the region $\varphi < 0$ that is governed by the K-matrix analysis results of meson spectra [24, 25]. The dashed area corresponds to the experimental data [26].

The $f_0(980)$ being the $q\bar{q}$ system is characterized by two parameters: the mean radius squared of $f_0(980)$ and mixing angle φ. In Fig. 9, the areas allowed for these parameters are shown; they were obtained for the processes $f_0(980) \to \gamma\gamma$ and $\phi(1020) \to \gamma f_0(980)$ with the old photon wave function (Fig. 9*a*) and the new one (Fig. 9*b*). The change in the allowed areas $(R^2_{f_0(980)}, \varphi)$ for the reaction $f_0(980) \to \gamma\gamma$ is rather noticeable, but it should be emphasized that it does not lead to a cardinal alteration of the parameter magnitudes.

Another set of reactions calculated with the photon wave function is the two-photon decay of tensor mesons as follows: $a_2(1320) \to \gamma\gamma$, $f_2(1270) \to \gamma\gamma$, and $f_2(1525) \to \gamma\gamma$. The calculations of $a_2(1320) \to \gamma\gamma$ with the old and new wave functions are shown in Fig. 10 (dashed and solid curves, respectively); experimental data [22, 26] are presented in Fig. 10 too (shaded areas). The description of experimental data has been carried out at $R^2_{a_2(1320)} \sim 8\ \text{GeV}^{-2}$: in this region, the difference between the calculated values of partial widths, which is due to a change in the wave function, is of the order of 10–20%.

The amplitude of the transition $f_2 \to \gamma\gamma$ is determined by four form factors related to the existence of two flavor components and two spin structures (see [12, 14] for the details). The calculations of these four form factors with the old and new wave functions

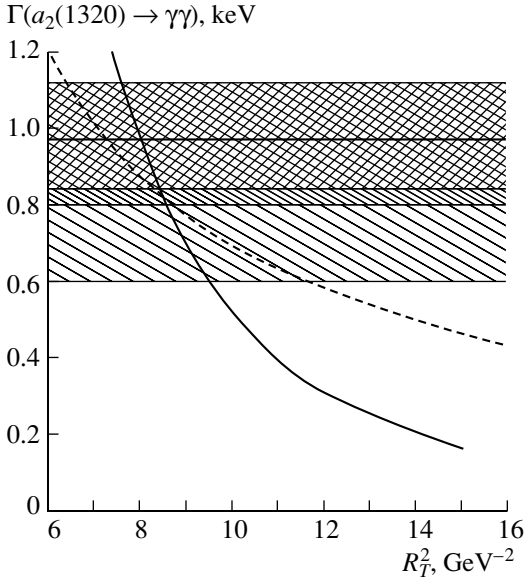

Fig. 10. Calculated curves vs. experimental data (shaded areas) for $\Gamma(a_2(1320) \to \gamma\gamma)$. The solid curve stands for the new photon wave function and the dashed curve stands for the old one.

are shown in Fig. 11—at $R^2_T \sim 8$–$10\ \text{GeV}^{-2}$, the difference is of the order of 10–20%. In Fig. 12, we show the allowed areas (R^2_T, φ_T) obtained in the description of experimental widths $\Gamma(f_2(1270) \to \gamma\gamma)$ and $\Gamma(f_2(1525) \to \gamma\gamma)$ [22] with the old (Fig. 12*a*) and new (Fig. 12*b*) wave functions. The new photon wave function results in a more rigid constraint for the areas (R^2_T, φ_T), though there are no qualitative changes in the description of data. The two-photon-decay data give us two solutions for the (R^2_T, φ_T) parameters:

$$(R^2_T, \varphi_T)_{\text{I}} \simeq \left(8\ \text{GeV}^{-2}, 0\right), \qquad (45)$$
$$(R^2_T, \varphi_T)_{\text{II}} \simeq \left(8\ \text{GeV}^{-2}, 25^\circ\right).$$

The solution with $\varphi \simeq 0$, when $f_2(1270)$ is nearly a pure $n\bar{n}$ state and $f_2(1525)$ is an $s\bar{s}$ system, is more preferable from the point of view of hadronic decays as well as the analysis [27].

6. CONCLUSION

Meson–photon transition form factors have been discussed within various approaches such as perturbative QCD formalism [28, 29], QCD sum rules [30–32], and variants of the light-cone quark model [9, 33–37]. A distinctive feature of the quark-model approach [9] consists in taking into account soft interaction of quarks in the subprocess $\gamma \to q\bar{q}$, that is, taking into account the production of vector mesons in the intermediate state: $\gamma \to V \to q\bar{q}$.

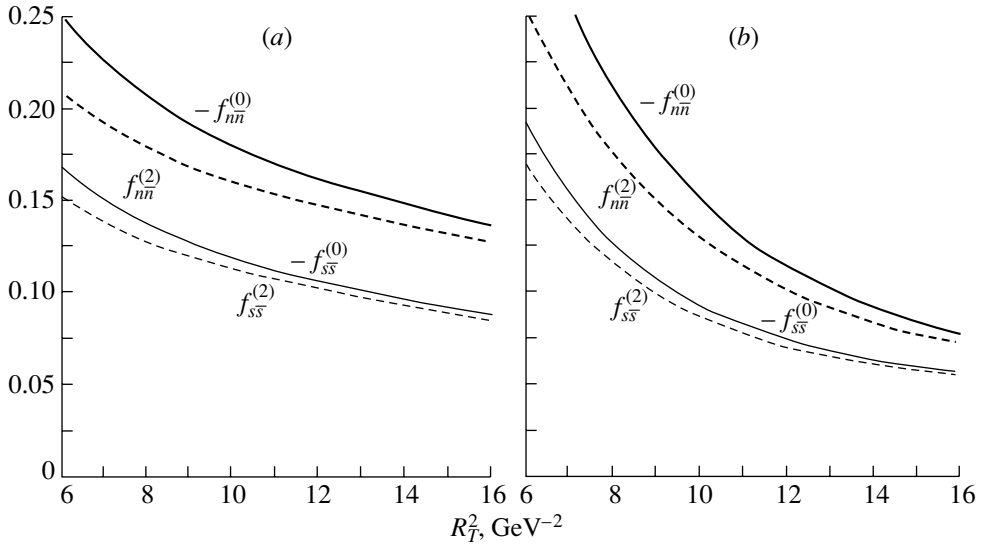

Fig. 11. Transition form factors in the decay of tensor quark–antiquark states $1^3P_2 n\bar{n} \to \gamma\gamma$ and $1^3P_2 s\bar{s} \to \gamma\gamma$ as functions of radius squared of the $q\bar{q}$ system calculated with (a) the old and (b) the new photon wave functions.

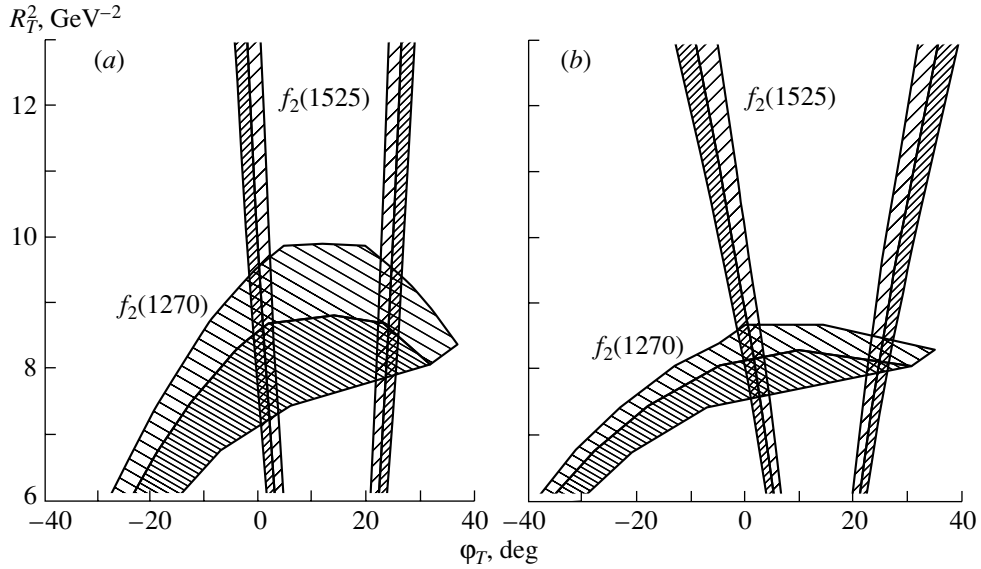

Fig. 12. Allowed areas $(R^2_{f_0(980)}, \varphi)$ for partial widths $\Gamma(f_2(1270) \to \gamma\gamma)$ and $\Gamma(f_2(1525) \to \gamma\gamma)$ calculated with (a) the old and (b) the new photon wave functions. Mixing angle φ_T defines the flavor content of mesons as follows: $f_2(1270) = n\bar{n}\cos\varphi_T + s\bar{s}\sin\varphi_T$ and $f_2(1525) = -n\bar{n}\sin\varphi_T + s\bar{s}\cos\varphi_T$.

In the present paper, we have reanalyzed the quark components of the photon wave function (the $\gamma^*(Q^2) \to u\bar{u}, d\bar{d}, s\bar{s}$ transitions) on the basis of data on the reactions $\pi^0, \eta, \eta' \to \gamma\gamma^*(Q^2)$, $e^+e^- \to \rho^0, \omega, \phi$, and $e^+e^- \to hadrons$. On the qualitative level, the wave function obtained here coincides with that defined before [9] by using the transitions $\pi^0, \eta, \eta' \to \gamma\gamma^*(Q^2)$ only. The data on the reactions $e^+e^- \to \rho^0, \omega, \phi$ and $e^+e^- \to hadrons$ allowed us to get a more precise wave function structure, in particular, in the region of the relative quark momenta $k \sim 0.4$–1.0 GeV/c.

Such an improvement in our knowledge of the photon wave function does not lead to a cardinal change in the description of two-photon decays or basis scalar and tensor mesons obtained before [12–14]. Still, a more detailed definition of the photon wave function is important for the calculations of the decays of a loosely bound $q\bar{q}$ state, such as the ra-

dial excitation state or reactions with virtual photons, $q\bar{q} \to \gamma^*(Q_1^2)\gamma^*(Q_2^2)$.

ACKNOWLEDGMENTS

We thank M.G. Ryskin for useful discussions.

This work was supported by the Russian Foundation for Basic Research, project no. 04-02-17091.

REFERENCES

1. M. N. Kinzle-Focacci, in *Proceedings of the VIII Blois Workshop, Protvino, Russia, 1999*, Ed. by V. A. Petrov and A. V. Prokudin (World Sci., Singapore, 2000); V. A. Schegelsky, Talk given at *Open Session of HEP Division of PNPI "On the Eve of the XXI Century," 2000*.
2. M. Acciarri *et al.* (L3 Collab.), Phys. Lett. B **501**, 1 (2001); **418**, 389 (1998); L. Vodopyanov (L3 Collab.), Nucl. Phys. B (Proc. Suppl.) **82**, 327 (2000).
3. H. Albrecht *et al.* (ARGUS Collab.), Z. Phys. C **74**, 469 (1997); **65**, 619 (1995); Phys. Lett. B **332**, 451 (1994); **267**, 535 (1991).
4. H. J. Behrend *et al.* (CELLO Collab.), Z. Phys. C **49**, 401 (1991).
5. H. Aihara *et al.* (TRC/2γ Collab.), Phys. Rev. D **38**, 1 (1988).
6. R. Briere *et al.* (CLEO Collab.), Phys. Rev. Lett. **84**, 26 (2000).
7. F. Butler *et al.* (Mark II Collab.), Phys. Rev. D **42**, 1368 (1990).
8. K. Karch *et al.* (Crystal Ball Collab.), Z. Phys. C **54**, 33 (1992).
9. V. V. Anisovich, D. I. Melikhov, and V. A. Nikonov, Phys. Rev. D **55**, 2918 (1997).
10. V. V. Anisovich, M. N. Kobrinsky, D. I. Melikhov, and A. V. Sarantsev, Nucl. Phys. A **544**, 747 (1992); A. V. Anisovich and V. A. Sadovnikova, Yad. Fiz. **55**, 2657 (1992) [Sov. J. Nucl. Phys. **55**, 1483 (1992)]; **57**, 1393 (1994); Eur. Phys. J. A **2**, 199 (1998).
11. V. V. Anisovich, D. I. Melikhov, and V. A. Nikonov, Phys. Rev. D **52**, 5295 (1995).
12. A. V. Anisovich, V. V. Anisovich, and V. A. Nikonov, Eur. Phys. J. A **12**, 103 (2001).
13. A. V. Anisovich, V. V. Anisovich, V. N. Markov, and V. A. Nikonov, Yad. Fiz. **65**, 523 (2002) [Phys. At. Nucl. **65**, 497 (2002)].
14. A. V. Anisovich, V. V. Anisovich, M. A. Matveev, and V. A. Nikonov, Yad. Fiz. **66**, 946 (2003) [Phys. At. Nucl. **66**, 914 (2003)].
15. A. V. Anisovich, V. V. Anisovich, and V. A. Nikonov, hep-ph/0305216 v2.
16. A. V. Anisovich, V. V. Anisovich, V. N. Markov, *et al.*, Yad. Fiz. **68**, 1614 (2005) [Phys. At. Nucl. **68**, 1554 (2005)]; hep-ph/0403123.
17. D. I. Melikhov, Phys. Rev. D **56**, 7089 (1997); Phys. Lett. B **516**, 61 (2001); M. Beyer, D. I. Melikhov, N. Nikitin, and B. Stech, Phys. Rev. D **64**, 094 006 (2001).
18. V. V. Anisovich, L. G. Dakhno, and V. A. Nikonov, Yad. Fiz. **67**, 1593 (2004) [Phys. At. Nucl. **67**, 1571 (2004)].
19. A. V. Anisovich, V. V. Anisovich, V. N. Markov, *et al.*, J. Phys. G **28**, 15 (2002).
20. V. V. Anisovich and M. A. Matveev, Yad. Fiz. **67**, 634 (2004) [Phys. At. Nucl. **67**, 614 (2004)].
21. A. V. Anisovich, V. V. Anisovich, L. Montanet, and V. A. Nikonov, Eur. Phys. J. A **6**, 247 (1999).
22. K. Hagiwara *et al.* (Particle Data Group), Phys. Rev. D **66**, 010001 (2002).
23. M. G. Ryskin, A. Martin, and J. Outhwaite, Phys. Lett. B **492**, 67 (2000).
24. V. V. Anisovich and A. V. Sarantsev, Eur. Phys. J. A **16**, 229 (2003).
25. V. V. Anisovich, Usp. Fiz. Nauk **174**, 49 (2004) [Phys. Usp. **47**, 45 (2004)].
26. M. Boglione and M. Pennington, Eur. Phys. J. C **9**, 11 (1999).
27. V. A. Schegelsky *et al.*, hep-ph/0404226.
28. G. P. Lepage and S. J. Brodsky, Phys. Rev. D **22**, 2157 (1980).
29. F.-G. Cao, T. Huang, and B.-Q. Ma, Phys. Rev. D **53**, 6582 (1996).
30. A. V. Radyushkin and R. Ruskov, Phys. Lett. B **374**, 173 (1996).
31. A. Schmedding and O. Yakovlev, Phys. Rev. D **62**, 116 002 (2000).
32. A. P. Bakulev, S. V. Mikhailov, and N. Stefanis, Phys. Rev. D **67**, 074 012 (2003).
33. C.-W. Hwang, Eur. Phys. J. C **19**, 105 (2001).
34. H. M. Choi and C. R. Ji, Nucl. Phys. A **618**, 291 (1997).
35. P. Kroll and M. Raulfus, Phys. Lett. B **387**, 848 (1996).
36. B.-W. Xiao and B.-Q. Ma, Phys. Rev. D **68**, 034 020 (2003).
37. M. A. DeWitt, H. M. Choi, and C. R. Ji, Phys. Rev. D **68**, 054 026 (2003).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1842–1850. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1904–1912.

ELEMENTARY PARTICLES AND FIELDS
Theory

Rare Radiative Leptonic Decays $B^0_{d,s} \to \ell^+\ell^-\gamma$

D. I. Melikhov*, N. V. Nikitin, and K. S. Toms*****

Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia

Received August 30, 2004

Abstract—The rare radiative leptonic decays $\bar{B}^0_{d,s} \to \ell^+\ell^-\gamma$ are studied. The contributions to the respective amplitude from the emission of photons from the quark loop, bremsstrahlung from leptons, and weak-annihilation effects are taken into account in relevant calculations. Results are presented for the partial widths and distributions of leptons in the final state. It is shown that the previously disregarded contributions of vector resonances associated with virtual-photon emission from the light valence quark of the B meson have a significant effect on the mass spectra of dileptons. © *2005 Pleiades Publishing, Inc.*

INTRODUCTION

The rare radiative leptonic decays $\bar{B}^0_{d,s} \to \ell^+\ell^-\gamma$ (for the sake of brevity, they are referred to below as merely rare radiative decays) are associated with $b \to s, d$ transitions, which are caused by a flavor-changing weak neutral current. Within the Standard Model, such a current arises in the lowest order of perturbation theory only beginning from loop diagrams of the penguin and box types. It follows that, within the Standard Model, the partial widths with respect to the radiative decays being considered range between 10^{-8} and 10^{-15}. The partial widths with respect to the rare radiative leptonic decays $B^0_{d,s} \to \ell^+\ell^-$ are on the same order of magnitude. Decays characterized by partial widths as small as those quoted above can be detected in none of the currently operating experiments, including those at the Tevatron proton–antiproton collider and those at the BaBar and Belle B-meson factories. However, it is planned to record the decays $B^0_{d,s} \to \mu^+\mu^-$ and $B^0_{d,s} \to \mu^+\mu^-\gamma$ at the ATLAS, CMS, and LHCb facilities, which will operate at the LHC proton–proton collider presently under construction at CERN. The respective partial widths are about 10^{-10} and 10^{-9} for the former and, as will be shown in the present study below, 10^{-10} and 10^{-8} for the latter.

The smallness of the partial widths with respect to the decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ opens the possibility for studying, at a high level of precision, the predictions of the Standard Model in higher orders of perturbation theory and for seeking new-physics effects (that is, effects beyond the Standard Model). Moreover, such rare radiative decays can make a nontrivial background contribution to experimental measurements of the partial width for the decay $B^0_d \to \ell^+\ell^-$. Knowledge of this contribution is of importance for the problem of new-physics searches in the rare leptonic decays of B mesons at future accelerators.

It is convenient to study rare radiative decays theoretically by using the effective Hamiltonian for $b \to q$ transitions that is represented in the form of Wilson's operator-product expansion [1]:

$$\mathcal{H}^{b\to q}_{\rm eff} = \frac{G_{\rm F}}{\sqrt{2}} V_{tb}V^*_{tq} \sum_i C_i(\mu) O_i(\mu), \qquad (1)$$

where $q = \{d, s\}$; $G_{\rm F}$ is the Fermi constant; C_i are Wilson coefficients, which involve all high-energy singularities of the Standard Model or its extensions; and O_i is a set of basis operators. The scale parameter μ separates the soft and the hard contribution from strong interactions. In describing B-meson decays, one usually sets $\mu \simeq 5$ GeV.

The hard contribution is contained in Wilson coefficients, which are calculated by means of perturbative QCD. The soft contribution is contained in invariant amplitudes—that is, in form factors parametrizing the matrix elements of basis operators between the initial and the final hadron state. A few types of basis operators are involved in describing rare radiative decays of B mesons. Accordingly, it is necessary to calculate form factors of a few types.

The decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ have been explored in a number of studies. In [2–6], the radiative-decay partial widths, the dilepton mass spectra, the photon energy spectra, and the charge asymmetries were obtained by means of various nonperturbative methods for taking into account the contribution of strong

*e-mail: melikhov@sinp.msu.ru
**e-mail: nik679@monet.npi.msu.su
***e-mail: ktoms@mail.cern.ch

interactions. The authors of those studies considered primarily the contribution to the amplitude of the rare radiative decays from the transition $B_q^0 \to \gamma$ induced by the Hamiltonian for the $b \to q\ell^+\ell^-$ transition. This is accompanied by the emission of a real photon by the valence quarks of the B_q^0 meson. We discuss the transitions $B_q^0 \to \gamma$ in Section 1, following the results reported in [2].

As contrasted to what was done in [2], we consider here three additional contributions to the amplitudes for the rare radiative decays.

In Section 2, we analyze the contribution of the process in which the emission of a virtual photon by the valence quarks of the B_q^0 meson ($B_q^0 \to \gamma^*$ transition) is followed by the conversion of this photon into a lepton pair. In describing this process, it is necessary to take into account the resonance contribution of vector mesons in the physical region of the decay process. It is shown how the form factors for the transitions $B_q^0 \to \gamma^*$ can be obtained from the form factors for the decays $B_q^0 \to V\gamma$ and $B_q^0 \to \gamma\gamma$.

The weak-annihilation process is considered in Section 3. It is well known that, against photon emission from the B_q^0-meson loop, this process is suppressed in proportion to $1/m_b$ [7]. Nonetheless, its contribution is quite sizable in the region of small values of the spectrum with respect to the invariant dilepton mass; therefore, it is included in our consideration.

Lepton bremsstrahlung in the final state is considered in Section 4. The contribution of this process to the cross section is proportional to $(m_\ell/M_{B_q^0})^2$. Therefore, it is dominant only in the case where τ leptons are produced in the final state. For some problems, however, the inclusion of bremsstrahlung is of importance even in the case where, in the final state, the bremsstrahlung photon is emitted by muons. For example, this is so in studying the decays $B_{d,s}^0 \to \mu^+\mu^-\gamma$ as background processes to the decays $B_{d,s}^0 \to \mu^+\mu^-$ at the LHC energies.

In Section 5, numerical estimates for the partial widths with respect to the decays $B_{d,s}^0 \to \ell^+\ell^-\gamma$ and for the relevant dilepton mass spectra are given with allowance for all of the aforementioned contributions.[1)]

[1)]In the present study, we do not consider asymmetries and various polarization effects, since a small expected number of detected $B_{d,s}^0 \to \ell^+\ell^-\gamma$ events prevents the determination of these quantities at LHC to a reasonable degree of accuracy [8].

1. TRANSITIONS INVOLVING REAL-PHOTON EMISSION BY THE VALENCE QUARKS OF THE $\bar{B}_q^0$ MESON

The $\bar{B}_q^0 \to \gamma$ transition in which the valence quarks of the $\bar{B}_q^0$ meson emit a real photon are described by the respective matrix element of the effective Hamiltonian

$$H_{\rm eff}^{b\to q\ell^+\ell^-} = \frac{G_{\rm F}}{\sqrt{2}}\frac{\alpha_{\rm em}}{2\pi}V_{tb}V_{tq}^*$$
$$\times\left[-2\frac{C_{7\gamma}(\mu)}{q^2}m_b\left(\bar{q}i\sigma_{\mu\nu}\left(1+\gamma_5\right)q^\nu b\right)\left(\bar{\ell}\gamma^\mu\ell\right)\right.$$
$$+C_{9V}^{\rm eff}(\mu,q^2)\left(\bar{q}O_\mu b\right)\left(\bar{\ell}\gamma^\mu\ell\right)$$
$$\left.+C_{10A}(\mu)\left(\bar{q}O_\mu b\right)\left(\bar{\ell}\gamma^\mu\gamma_5\ell\right)\right]+\ \text{h.c.},$$

where $\alpha_{\rm em} = e^2/4\pi$ is the fine-structure constant, $e = |e|$, and m_b is the b-quark mass and we have adopted the following conventions: $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, $\sigma_{\mu\nu} = \frac{i}{2}[\gamma_\mu,\gamma_\nu]$, and $O_\mu = \gamma_\mu(1-\gamma_5)$. The amplitude of this transition is given by the two diagrams in Fig. 1*a*. The coefficient $C_{9V}^{\rm eff}(\mu,q^2)$ is the sum of the coefficient $C_{9V}(\mu)$ [1] and the resonance contribution described in [9].

By using the form factors for the $\bar{B}_q^0 \to \gamma$, we can represent the required amplitude in the form [2]

$$\langle\gamma(k,\epsilon),\ell^+(p_1),\ell^-(p_2)\left|H_{\rm eff}^{b\to q\ell^+\ell^-}\right|\bar{B}_q^0(p,M_1)\rangle \quad (2)$$
$$=\frac{G_{\rm F}}{\sqrt{2}}V_{tb}V_{tq}^*\frac{\alpha_{em}}{2\pi}e\epsilon_\alpha^*\left[\frac{2C_{7\gamma}(\mu)}{q^2}m_b\right.$$
$$\times\left(\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta F_{TV}(q^2,0)-i\big(g_{\mu\alpha}(pk)\right.$$
$$\left.-p_\alpha k_\mu\big)F_{TA}(q^2,0)\big)\bar{\ell}(p_2)\gamma_\mu\ell(-p_1)$$
$$\times C_{9V}^{\rm eff}(\mu,q^2)\left(\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta\frac{F_V(q^2)}{M_1}-i\big(g_{\mu\alpha}(pk)\right.$$
$$\left.-p_\alpha k_\mu\big)\frac{F_A(q^2)}{M_1}\right)\bar{\ell}(p_2)\gamma_\mu\ell(-p_1)$$
$$+C_{10A}(\mu)\left(\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta\frac{F_V(q^2)}{M_1}-i\big(g_{\mu\alpha}(pk)\right.$$
$$\left.\left.-p_\alpha k_\mu\big)\frac{F_A(q^2)}{M_1}\right)\bar{\ell}(p_2)\gamma_\mu\gamma_5\ell(-p_1)\right],$$

where p_1 and p_2 are, respectively, the antilepton and lepton momenta; k is the real-photon momentum; p

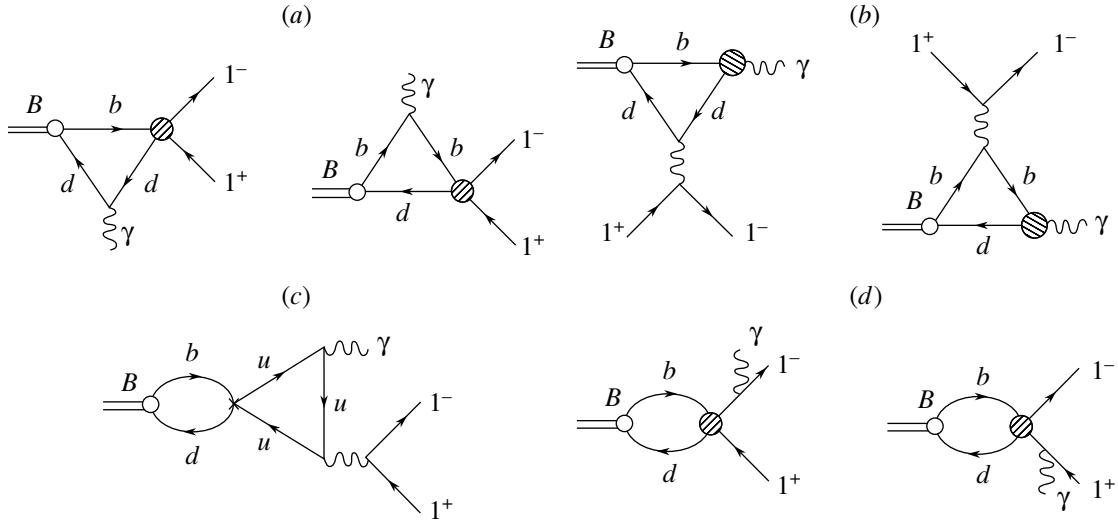

Fig. 1. Examples of diagrams contributing to the decays $B_d^0 \to \ell^+\ell^-\gamma$. Shaded circles in the diagrams represent vertices corresponding to the effective Hamiltonians for the $(a, d)\, b \to d\ell^+\ell^-$ and $(b)\, b \to d\gamma$ transitions.

$(p = k + p_1 + p_2 = k + q)$ and M_1 are the $\bar{B}_q^0$-meson momentum and mass, respectively; and $\varepsilon^{0123} = -1$.

The form factors $F_A(q^2)$, $F_V(q^2)$, $F_{TA}(q^2, 0)$, and $F_{TV}(q^2, 0)$ are defined as follows [2]:[2)]

$$\langle\gamma(k,\epsilon)|\bar{q}\gamma_\mu\gamma_5 b|\bar{B}_q^0(p, M_1)\rangle = ie\epsilon_\alpha^*\left(g_{\mu\alpha}(pk) - p_\alpha k_\mu\right)\frac{F_A(q^2)}{M_1}, \tag{3}$$

$$\langle\gamma(k,\epsilon)|\bar{q}\gamma_\mu b|\bar{B}_q^0(p, M_1)\rangle = e\epsilon_\alpha^*\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta\frac{F_V(q^2)}{M_1},$$

$$\langle\gamma(k,\epsilon)|\bar{q}\sigma_{\mu\nu}\gamma_5 b|\bar{B}_q^0(p, M_1)\rangle(p-k)_\nu = e\epsilon_\alpha^*\left[g_{\mu\alpha}(pk) - p_\alpha k_\mu\right]F_{TA}(q^2, 0),$$

$$\langle\gamma(k,\epsilon)|\bar{q}\sigma_{\mu\nu} b|\bar{B}_q^0(p, M_1)\rangle(p-k)_\nu = ie\epsilon_\alpha^*\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta F_{TV}(q^2, 0).$$

In [2], Krüger and Melikhov, who relied on the relativistic dispersion approach [10], studied these form factors as functions of the variable $(p-k)^2 = q^2$ and derived for them convenient parametrizations that satisfy all known rigorous constraints on the behavior of the form factors (3) in the limit $m_b \to \infty$. The parametrizations from [2] are used in Section 5 to obtain numerical estimates. For further applications, it is convenient to introduce the dimensionless Mandelstam invariant quantities

$$\hat{s} = \frac{(p-k)^2}{M_1^2}, \qquad \hat{t} = \frac{(p-p_1)^2}{M_1^2}, \qquad \hat{u} = \frac{(p-p_2)^2}{M_1^2}, \tag{4}$$

which are related by the equation $\hat{s} + \hat{t} + \hat{u} = 1 + 2\hat{m}_\ell^2$, where $\hat{m}_\ell^2 = m_\ell^2/M_1^2$, with m_ℓ being the lepton mass; we also have $\hat{m}_b = m_b/M_1$. The variables x and $\cos\theta$ are defined as follows [2]:

$$x = 1 - \hat{s}, \qquad \cos\theta = \frac{\xi\left(\hat{s}, \hat{t}\right)}{x\sqrt{1 - 4\hat{m}_l^2/\hat{s}}}, \qquad \xi\left(\hat{s}, \hat{t}\right) = \hat{u} - \hat{t}. \tag{5}$$

Equation (34) from [2] can then be recast into the form

$$\frac{d^2\Gamma^{(1)}}{d\hat{s}d\hat{t}} = \frac{G_{\rm F}^2\alpha_{\rm em}^3 M_1^5}{2^{10}\pi^4}\left|V_{tb}V_{tq}^*\right|^2 \times\Big[x^2 B_0(\hat{s},\hat{t}) + x\xi(\hat{s},\hat{t})\tilde{B}_1(\hat{s},\hat{t}) + \xi^2(\hat{s},\hat{t})\tilde{B}_2(\hat{s},\hat{t})\Big], \tag{6}$$

where

$$B_0(\hat{s},\hat{t}) = \left(\hat{s} + 4\hat{m}_\ell^2\right)\left(F_1(\hat{s}) + F_2(\hat{s})\right) - 8\hat{m}_\ell^2\left|C_{10A}(\mu)\right|^2\left(F_V^2(q^2) + F_A^2(q^2)\right), \tag{7}$$

[2)] We note that, in [2], the form factors $F_{TV}(q^2, 0)$ and $F_{TA}(q^2, 0)$ were denoted by $F_{TV}(q^2)$ and $F_{TA}(q^2)$, respectively. In the present article, the notation is changed in order to highlight the different dependences of the tensor form factors on the momenta of the two photons, a real and a virtual one, the momentum squared of the photon emitted from the effective vertex described by the Hamiltonian $H_{\rm eff}^{b\to q\ell^+\ell^-}(x)$ or the Hamiltonian $H_{\rm eff}^{b\to q\gamma}(x)$ (Section 2) always appearing as the first argument.

$$\tilde{B}_1(\hat{s},\hat{t}) = 8\Big[\hat{s}F_V(q^2)F_A(q^2)\mathrm{Re}\left(C_{9V}^{\mathrm{eff}*}(\mu,q^2)C_{10A}(\mu)\right) + \hat{m}_b F_V(q^2)\mathrm{Re}\left(C_{7\gamma}^*(\mu)F_{TA}^*(q^2,0)C_{10A}(\mu)\right) + \hat{m}_b F_A(q^2)\mathrm{Re}\left(C_{7\gamma}^*(\mu)F_{TV}^*(q^2,0)C_{10A}(\mu)\right)\Big],$$

$$\tilde{B}_2(\hat{s},\hat{t}) = \hat{s}\left(F_1(\hat{s}) + F_2(\hat{s})\right),$$

$$F_1(\hat{s}) = \left(\left|C_{9V}^{\mathrm{eff}}(\mu,q^2)\right|^2 + |C_{10A}(\mu)|^2\right)F_V^2(q^2) + \left(\frac{2\hat{m}_b}{\hat{s}}\right)^2\left|C_{7\gamma}(\mu)F_{TV}(q^2,0)\right|^2 + \frac{4\hat{m}_b}{\hat{s}}F_V(q^2)\mathrm{Re}\left(C_{7\gamma}(\mu)F_{TV}(q^2,0)C_{9V}^{\mathrm{eff}*}(\mu,q^2)\right),$$

$$F_2(\hat{s}) = \left(\left|C_{9V}^{\mathrm{eff}}(q^2,\mu)\right|^2 + |C_{10A}(\mu)|^2\right)F_A^2(q^2) + \left(\frac{2\hat{m}_b}{\hat{s}}\right)^2\left|C_{7\gamma}(\mu)F_{TA}(q^2,0)\right|^2 + \frac{4\hat{m}_b}{\hat{s}}F_A(q^2)\mathrm{Re}\left(C_{7\gamma}(\mu)F_{TA}(q^2,0)C_{9V}^{\mathrm{eff}*}(\mu,q^2)\right).$$

In the ensuing analysis, it is convenient to treat the form factors $F_{TV}(q^2,0)$ and $F_{TA}(q^2,0)$ as complex-valued quantities.

2. TRANSITIONS INVOLVING VIRTUAL-PHOTON EMISSION BY THE VALENCE QUARKS OF THE $\bar{B}_q^0$ MESON

In Section 1, we have studied diagrams in which the valence quarks of the $\bar{B}_q^0$ meson emit a real photon. In addition to these diagrams, there exist diagrams describing the process in which the electromagnetic-interaction-induced emission of a virtual photon from valence quarks in the $\bar{B}_q^0$-meson loop is followed by virtual-photon transition to a lepton pair ($\bar{B}_q^0 \to \gamma\gamma^* \to \gamma\ell^+\ell^-$). The relevant diagrams are given in Fig. 1*b*. The analogous electroweak process featuring the emission of a virtual Z^0 boson can be disregarded. If m_q is neglected against m_b, real-photon emission in such processes is described by the effective Hamiltonian for $b \to q\gamma$ transitions,

$$H_{\mathrm{eff}}^{b\to q\gamma} = \frac{G_{\mathrm{F}}}{\sqrt{2}}V_{tb}V_{tq}^*C_{7\gamma}(\mu)\frac{e}{8\pi^2} \times m_b\left(\bar{q}\sigma_{\mu\nu}\left(1+\gamma_5\right)b\right)F^{\mu\nu} + \text{ h.c.} \quad (8)$$

The calculation of the respective matrix elements of the Hamiltonian in (8) leads to the appearance of Lorentz-invariant structures that are coincident with the structures in front of the form factors $F_{TV}(q^2,0)$ and $F_{TA}(q^2,0)$ (see Section 1). The form factors accompanying these structures are denoted by $F_{TV}(0,q^2)$ and $F_{TA}(0,q^2)$. In order to take into account the diagrams in Figs. 1*a* and 1*b*, it is then necessary to make the following substitutions in Eq. (6):

$$F_{TV}(q^2,0) \to F_{TV}(q^2,0) + F_{TV}(0,q^2), \quad (9)$$
$$F_{TA}(q^2,0) \to F_{TA}(q^2,0) + F_{TA}(0,q^2).$$

In order to calculate the form factors $F_{TV}(0,q^2)$ and $F_{TA}(0,q^2)$, one can invoke the vector-meson-dominance (VMD) model [11, 12]. The decay $\bar{B}_d^0 \to \ell^+\ell^-\gamma$ involving virtual-photon emission by the $\bar{d}$ antiquark can then be considered as the transition of the $\bar{B}_d^0$ meson to ρ^0 and ω mesons. In describing virtual-photon emission by the $\bar{s}$ antiquark in the decays $\bar{B}_s^0 \to \ell^+\ell^-\gamma$, it is necessary to take into account only the ϕ-resonance contribution. Vector mesons in turn go over to a virtual photon, which undergoes conversion to a lepton pair. We can disregard the Υ, Υ', ... contributions that arise in describing virtual-photon emission by the b quark, since the main contribution from these narrow resonances lies beyond the physical region of the decays $\bar{B}_{d,s}^0 \to \ell^+\ell^-\gamma$.

For $q^2 \to 0$, the amplitude for the transition $\bar{B}_q^0 \to \gamma\gamma^*$ must reduce to the amplitude for the decay $\bar{B}_q^0 \to \gamma\gamma$. Owing to this, the form factors $F_{TV}(0,q^2)$ and $F_{TA}(0,q^2)$ can be written as

$$F_{TV}(0,q^2) = F_{TV}(0,0) - \sum_i \frac{2g_+^i(0)}{f_i}\frac{q^2}{q^2 - M_i^2 + iM_i\Gamma_i}, \quad (10)$$

$$F_{TA}(0,q^2) = F_{TA}(0,0) - \sum_i \frac{2g_+^i(0)}{f_i}\frac{q^2}{q^2 - M_i^2 + iM_i\Gamma_i},$$

where M_i and Γ_i are, respectively, the masses and total decay widths of the vector mesons V_i. The constants f_i are calculated on the basis of the known widths with respect to the decays $V_i \to e^+e^-(\mu^+\mu^-)$ according to the formula

$$\Gamma\left(V_i \to e^+e^-(\mu^+\mu^-)\right) = \frac{4\pi}{3}\frac{\alpha_{\mathrm{em}}^2}{f_i^2}M_i. \quad (11)$$

The form factors $g_+^i(k^2)$ are defined as follows [13]:

$$\langle V_i(q,M_i,\epsilon)|\bar{d}\sigma_{\mu\nu}b|\bar{B}_d^0(p,M_1)\rangle \quad (12)$$

$$= i\epsilon^{*\alpha}\varepsilon_{\mu\nu\beta\gamma}\left[g_+^i(k^2)g_{\alpha\beta}(p+q)^\gamma + g_-^i(k^2)g_{\alpha\beta}k^\gamma + g_0^i(k^2)p_\alpha p^\beta q^\gamma\right].$$

The numerical values of the dimensionless quantities f_i and $g_+^i(0)$ with allowance for the $d\bar{d}$- or $s\bar{s}$-pair contributions to the ρ^0-, ω-, and ϕ-resonance wave functions are given in Table 1.

We note that Dincer and Sehgal [3] also considered the contribution of the diagrams in Fig. 1*b*, but they set $F_{TV,TA}(q^2,0) = F_{TV,TA}(0,q^2)$, which is not a physically justified approximation.

The resonance contribution calculated in this section is analogous to the contribution of vector resonances to the Wilson coefficient $C_{9V}^{\rm eff}(\mu,q^2)$, which was considered in [9]. The resonance contribution to the coefficient $C_{9V}^{\rm eff}(\mu,q^2)$ arises owing to the chain of transitions

$$b \to (d,s)(\bar{c}c,\bar{u}u) \to (d,s)\gamma^* \to (d,s)\ell^+\ell^-, \quad (13)$$

whose origin is associated with the presence of the four-quark operators O_1 and O_2 in Wilson's expansion (1). The structure of the vector-meson contributions to the decay chain (13) is similar to the structure of the contribution from the operator O_{9V}; that is, the resonance contribution can be described as an addition to the Wilson coefficient for the corresponding operator. From (13), it follows that, in calculating $C_{9V}^{\rm eff}(\mu,q^2)$, it is necessary to take into account the contribution of only those resonances whose quark structure involves $u\bar{u}$ and $c\bar{c}$ pairs (ρ^0, ω, J/ψ, ψ', etc.).

3. WEAK ANNIHILATION

Let us consider yet another class of diagrams, that which represents the weak-annihilation process in the decays $B^{0,\pm} \to \rho^{0,\pm}(\gamma,\ell^+\ell^-)$ [14]. The contribution of these diagrams, which are depicted in Fig. 1*c*, is proportional to the divergence of the axial current and is the well-known axial anomaly. The effective weak-annihilation Hamiltonian for $\bar{B}_d^0$-meson decay has the form

$$H_{\rm eff}^{\bar{B}_d^0\to\bar{u}u}(x) = -\frac{G_{\rm F}}{\sqrt{2}}V_{ub}V_{ud}^* a_1 \times \left(\bar{d}(x)\gamma_\mu(1-\gamma_5)b(x)\right)\left(\bar{u}(x)\gamma_\mu(1-\gamma_5)u(x)\right),$$

where $a_1 = C_1 + C_2/N_c$, N_c being the number of colors. At $N_c = 3$, we obtain $a_1 = -0.13$; that is, weak annihilation is dynamically suppressed. For the matrix element associated with the axial anomaly, we have

$$-\frac{G_{\rm F}}{\sqrt{2}}V_{td}^*V_{tb}\frac{\alpha_{\rm em}}{\pi}e\epsilon_\alpha^*\frac{C_{7\gamma}(\mu)}{q^2}m_b\varepsilon_{\mu\alpha\xi\eta}p_\xi k_\eta \times 4\frac{V_{ub}V_{ud}^*}{V_{tb}V_{td}^*}\frac{a_1N_cQ_u^2}{C_{7\gamma}(\mu)}\frac{f_{B_d}}{m_b}\bar{\ell}(p_2)\gamma_\mu\ell(-p_1), \quad (14)$$

where the leptonic $\bar{B}_q^0$-meson decay constant $f_{B_q} > 0$ is defined according to the condition

$$\langle 0|\bar{q}(0)\gamma_\mu\gamma_5 b(0)|\,\bar{B}_q^0(p,M_1)\rangle = if_{B_q}p_\mu.$$

Comparing (14) with (2), we find that, for the decay $\bar{B}_d^0 \to \ell^+\ell^-\gamma$, formula (6) undergoes no changes if one makes the substitution

$$F_{TV}(q^2,0) \to F_{TV}(q^2,0) + F_{TV}(0,q^2) - 4\frac{V_{ub}V_{ud}^*}{V_{tb}V_{td}^*}\frac{a_1N_cQ_u^2}{C_{7\gamma}(\mu)}\frac{f_{B_d}}{m_b}, \quad (15)$$

$$F_{TA}(q^2,0) \to F_{TA}(q^2,0) + F_{TA}(0,q^2).$$

In the approximation of heavy-quark effective theory, the asymptotic behavior of the form factors (3) at photon energies E^γ in the $\bar{B}_q^0$-meson rest frame that are much higher than the characteristic confinement scale is given by [2, 15]

$$F_V(q^2) \approx F_A(q^2) \approx F_{TV}(q^2,0) \approx F_{TA}(q^2,0) \sim \frac{f_{B_q}M_1}{E^\gamma},$$

while the weak-annihilation contribution is proportional to the factor f_{B_d}/m_b. Thus, we see that, even at the maximum photon energy $E_{\rm max}^\gamma \sim M_1/2$, the weak-annihilation correction to $F_{TV}(q^2,0)$ is suppressed in proportion to $1/m_b$. For the radiative decays of the $\bar{B}_s^0$ meson, this contribution is additionally Cabibbo-suppressed. Therefore, it can be disregarded in the last case.

Table 1. Numerical values of the vector-meson decay constants f_i and form factors $g_+^i(0)$ with allowance for additional isotopic coefficients with which the $d\bar{d}$ and $s\bar{s}$ states appear in the ρ^0-, ω-, and ϕ-meson wave functions

Transition	$B_d^0 \to \rho^0$	$B_d^0 \to \omega$	$B_s^0 \to \phi$
f_i	5.04	17.1	−13.2
$g_+^i(0)$	$0.27/\sqrt{2}$	$-0.27/\sqrt{2}$	−0.38

4. BREMSSTRAHLUNG

Let us consider lepton bremsstrahlung in the decays $\bar{B}_{d,s}^0 \to \ell^+\ell^-\gamma$. The corresponding diagrams are shown in Fig. 1*d*. The bremsstrahlung amplitude has the form

$$ie\frac{G_{\rm F}}{\sqrt{2}}\frac{\alpha_{\rm em}}{\pi}V_{td}^*V_{tb}\frac{f_{B_q}}{M_1}\hat{m}_\ell C_{10A}(\mu)\bar{\ell}(p_2) \times \left[\frac{(\gamma\epsilon^*)(\gamma p)}{\hat{t}-\hat{m}_\ell^2} - \frac{(\gamma p)(\gamma\epsilon^*)}{\hat{u}-\hat{m}_\ell^2}\right]\gamma_5\ell(-p_1). \quad (16)$$

Table 2. Partial widths with respect to the decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ versus the cutoff in the minimum photon energy $E^\gamma_{\min}$ in the B^0_q-meson rest frame [only the contributions of the $B^0_q \to \gamma$ transition (Section 1) and bremsstrahlung (Section 4) are taken into account]

m_ℓ	m_e			m_μ			m_τ		
$E^\gamma_{\min}$ [GeV]	0.02	0.05	0.08	0.02	0.05	0.08	0.02	0.05	0.08
$\mathrm{Br}(B^0_d \to \ell^+\ell^-\gamma) \times 10^{10}$	1.20	1.20	1.20	0.78	0.74	0.72	4.01	2.81	2.22
$\mathrm{Br}(B^0_s \to \ell^+\ell^-\gamma) \times 10^{9}$	2.23	2.22	2.22	1.46	1.39	1.35	7.80	5.50	4.37

Table 3. Partial widths with respect to the decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ versus the cutoff in the minimum photon energy $E^\gamma_{\min}$ in the B^0_q-meson rest frame (the contributions of all processes considered in Sections 1–4 are taken into account; the region of J/ψ and ψ' resonances is eliminated in integration with respect to $\hat{s}$)

m_ℓ	m_e			m_μ			m_τ		
$E^\gamma_{\min}$ [GeV]	0.02	0.05	0.08	0.02	0.05	0.08	0.02	0.05	0.08
$\mathrm{Br}(B^0_d \to \ell^+\ell^-\gamma) \times 10^{10}$	4.67	4.67	4.66	1.58	1.56	1.55	4.00	2.80	2.21
$\mathrm{Br}(B^0_s \to \ell^+\ell^-\gamma) \times 10^{9}$	16.6	16.6	16.6	12.8	12.7	12.7	7.77	5.47	4.33

The bremsstrahlung contribution to the differential width with respect to the decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ has a simple form,

$$\frac{d^2\Gamma^{(2)}}{d\hat{s}d\hat{t}} = \frac{G_F^2\alpha_{\mathrm{em}}^3 M_1^5}{2^4\pi^4} \times \left|V_{tb}V_{tq}^*\right|^2 \left(\frac{f_{B_q}}{M_1}\right)^2 \hat{m}_\ell^2 \left|C_{10A}(\mu)\right|^2 \times \left[\frac{\hat{s}+x^2/2}{(\hat{u}-\hat{m}_\ell^2)(\hat{t}-\hat{m}_\ell^2)} - \left(\frac{x\hat{m}_\ell}{(\hat{u}-\hat{m}_\ell^2)(\hat{t}-\hat{m}_\ell^2)}\right)^2\right]. \quad (17)$$

The term that describes the interference between the amplitudes in (2) and (16) is given by

$$\frac{d^2\Gamma^{(12)}}{d\hat{s}d\hat{t}} = \frac{G_F^2\alpha_{\mathrm{em}}^3 M_1^5}{2^6\pi^4}\left|V_{tb}V_{tq}^*\right|^2 \frac{f_{B_q}}{M_1} \times \hat{m}_\ell^2 \frac{x^2}{(\hat{u}-\hat{m}_\ell^2)(\hat{t}-\hat{m}_\ell^2)} \times \left[\frac{2x\hat{m}_b}{\hat{s}}\mathrm{Re}\left(C_{10A}^*(\mu)C_{7\gamma}(\mu)F_{TV}(q^2,0)\right) + xF_V(q^2)\mathrm{Re}\left(C_{10A}^*(\mu)C_{9V}^{\mathrm{eff}}(\mu,q^2)\right) + \xi(\hat{s},\hat{t})F_A(q^2)\left|C_{10A}(\mu)\right|^2\right], \quad (18)$$

where the form factor $F_{TV}(q^2,0)$ implies the substitution in (15). In (17) and (18), there is an infrared pole, which requires introducing a cutoff in the emitted-photon energy. Obviously, the higher the lepton mass and the lower the minimum photon energy, the greater the enhancement associated with this pole. The dependence of the decay width on the lepton mass and on the minimum photon energy is studied in the next section.

5. NUMERICAL RESULTS

In order to obtain numerical estimates, it is necessary to know respective Wilson coefficients, elements of the Cabibbo–Maskawa matrix, lifetimes of neutral B mesons, and form factors. The Wilson coefficients were calculated by the formulas from [1]. Taking into account the condition $C_2(M_W) = -1$, one can find that $C_1(5\ \mathrm{GeV}) = 0.241$, $C_2(5\ \mathrm{GeV}) = -1.1$, $C_{7\gamma}(5\ \mathrm{GeV}) = 0.312$, $C_{9V}(5\ \mathrm{GeV}) = -4.21$, and $C_{10A}(5\ \mathrm{GeV}) = 4.64$. The elements of the Cabibbo–Maskawa matrix are $|V_{td}^*V_{tb}|^2 = 8.1 \times 10^{-5}$ and $|V_{ts}^*V_{tb}|^2 = 1.5 \times 10^{-3}$. The B^0_q-meson lifetimes are taken to be $\tau(B^0_d) = 1.55 \times 10^{-12}$ s and $\tau(B^0_s) = 1.49 \times 10^{-12}$ s. The $B^0_d \to \gamma$ form factors are parametrized as in [2]. The form factors for the transitions $B^0_d \to \rho^0$, $B^0_d \to \omega$, and $B^0_s \to \phi$ are given in Table 1. It is assumed that the form factors for the transition $B^0_s \to \gamma$ differ insignificantly from the form factors for the transition $B^0_d \to \gamma$.

Table 2 displays the partial widths with respect to the decays $B^0_{d,s} \to \ell^+\ell^-\gamma$ versus the cutoff in the minimum photon energy ("mass") $E^\gamma_{\min}$ in the

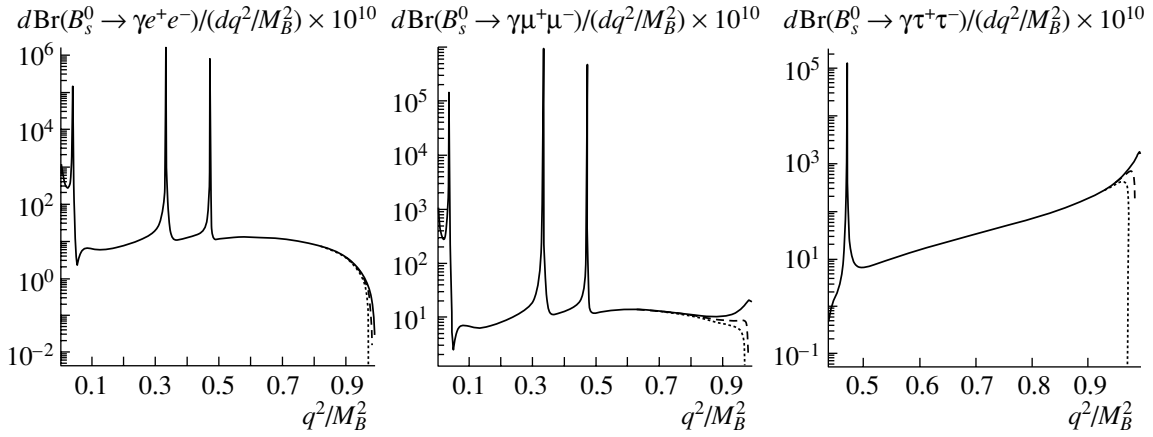

Fig. 2. Invariant-mass distributions of dileptons in the decays $B_s^0 \to \ell^+\ell^-\gamma$ for $E_{\mathrm{min}}^{\gamma}$ = (solid curves) 0.02, (dashed curves) 0.05, and (dotted curves) 0.08 GeV. The contribution of all processes considered in Sections 1–4 is taken into account.

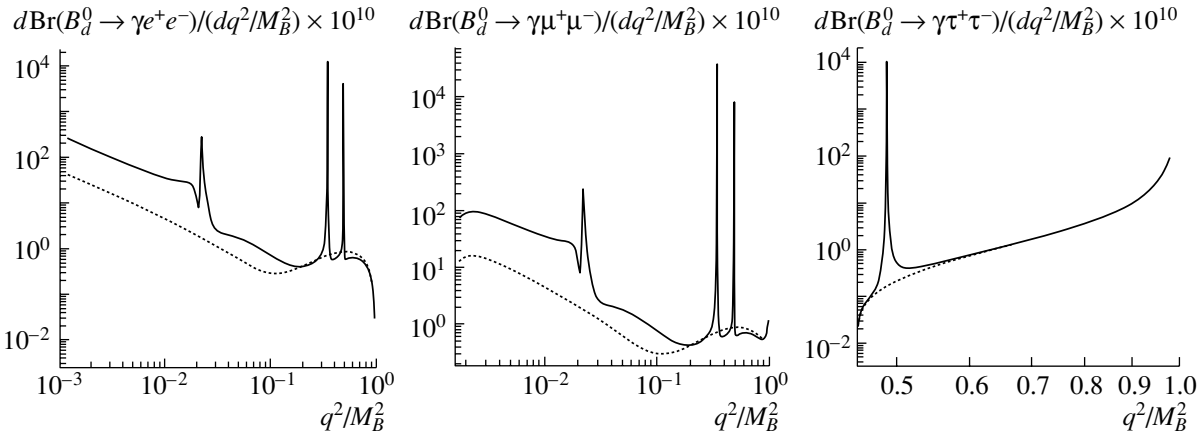

Fig. 3. Invariant-mass distributions of dileptons in the decays $B_d^0 \to \ell^+\ell^-\gamma$ at the minimum photon energy ("mass") of $E_{\mathrm{min}}^{\gamma} = 0.02$ GeV. The dotted curves represent the distributions including only the contributions of the transition $B_q^0 \to \gamma$ (Section 1) and photon bremsstrahlung (Section 4), this corresponding to the analysis in [4]; the solid curves were calculated with allowance for all contributions considered in Sections 1–4. In the case of electrons, the variable $\hat{s} = q^2/M_B^2$ is restricted in the graph by the value of 10^{-3}.

B_q^0-meson rest frame. The data in this table correspond to the version where the partial widths include the contributions from only the transition $B_q^0 \to \gamma$ and lepton bremsstrahlung [Eqs. (6), (17), and (18)]. In this case, the substitutions specified by (15) are not made for the form factors $F_{TV}(q^2, 0)$ and $F_{T\mathrm{A}}(q^2, 0)$ and only the nonresonance part $C_{9V}(\mu)$ is taken into account in the Wilson coefficient $C_{9V}^{\mathrm{eff}}(\mu, q^2)$. Table 2 is presented here in order to compare the numerical results of our study with their counterparts from earlier articles (see, for example, [4, 6]) and to demonstrate compellingly the importance of taking into account the transitions $B_q^0 \to \gamma^*$ and weak annihilation in the partial widths and dilepton mass spectra. The choice of $E_{\mathrm{min}}^{\gamma}$ value is dictated by the energy resolution of a specific experimental facility. The maximum (minimum) value of $E_{\mathrm{min}}^{\gamma}$ in Table 2 approximately corresponds to the planned resolution in the B-meson mass at the ATLAS (LHCb) facility.

The results in Table 2 are easily understandable. For electrons, the pole of the photon propagator in the vicinity of $\hat{s}_{\mathrm{min}} = 4\hat{m}_e^2$ makes a dominant contribution to the partial decay width. Since $m_e \ll E_{\mathrm{min}}^{\gamma}$, then the partial widths with respect to the decays $B_{d,s}^0 \to e^+e^-\gamma$ must depend only slightly on $E_{\mathrm{min}}^{\gamma}$.

For the τ lepton, the situation is diametrically opposite. Since the τ-particle mass is large, the lepton pole at $\hat{s}_{\mathrm{min}} = 4\hat{m}_\tau^2$ is absent. On the contrary, $E_{\mathrm{min}}^{\gamma} \ll m_\tau$; therefore, the decay width depends

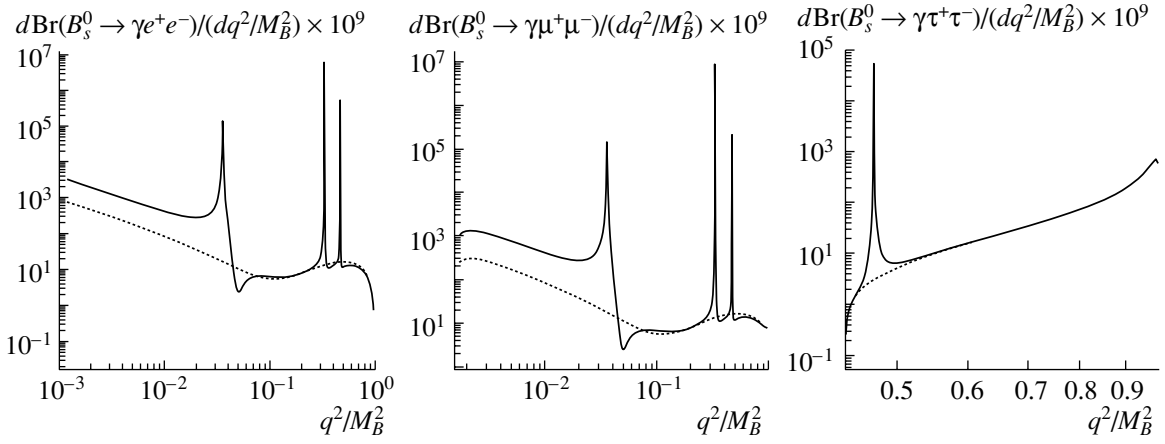

Fig. 4. Invariant-mass distributions of dileptons in the decays $B_s^0 \to \ell^+\ell^-\gamma$ at the minimum photon energy ("mass") of $E_{\min}^{\gamma} = 0.05$ GeV. The dotted curves represent the distributions including only the contributions of the transition $B_q^0 \to \gamma$ (Section 1) and photon bremsstrahlung (Section 4), this corresponding to the analysis in [4]; the solid curves were calculated with allowance for all contributions considered in Sections 1–4. In the case of electrons, the variable $\hat{s} = q^2/M_B^2$ is restricted in the graph by the value of 10^{-3}.

greatly on the cutoff in the photon "mass," in perfect agreement with expressions (17) and (18).

In the case of $\ell \equiv \mu$, which is of importance for LHC, there is also a pole at $\hat{s}_{\min} = 4\hat{m}_\mu^2$. However, its contribution to the partial width with respect to radiative decay is much less significant than in the case of electrons. Moreover, $E_{\min}^{\gamma} \sim m_\mu$, this leading to a weak dependence of the widths with respect to the decays $B_{d,s}^0 \to \mu^+\mu^-\gamma$ on the choice of $E_{\min}^{\gamma}$.

Upon rescaling the respective partial widths to the case of identical values of the lifetimes and the elements of the Cabibbo–Maskawa matrix, a comparison of the results in Table 2 with the numerical results reported in [4] reveals that the two calculations in question are quite consistent. However, it should be emphasized once again that, in view of the disregard of the contribution from $B_q^0 \to \gamma^*$ transitions in [4], this numerical agreement is nothing but an independent test of correctness of the calculations in Sections 1 and 4.

Table 3 displays the $E_{\min}^{\gamma}$ dependence of the partial widths calculated for the decays $B_{d,s}^0 \to \ell^+\ell^-\gamma$ with allowance for all of the contributions described in Sections 1–4. In order to simplify the weak-annihilation contribution, we set $V_{ub}V_{ud}^*/V_{tb}V_{td}^* = 1$ in Eq. (15). In performing integration with respect to $\hat{s}$, we eliminated the region of J/ψ and ψ' resonances (that is, the region where $\hat{s} \in [0.33, 0.55]$), this corresponding to the experimental procedure of searches for rare semileptonic and radiative decays of B mesons at LHC [8, 16]. Figure 2 shows the invariant-mass distributions $d\mathrm{Br}\left(B_s^0 \to \ell^+\ell^-\gamma\right)/d\hat{s}$ of dileptons for various values of the parameter $E_{\min}^{\gamma}$. One can see that, for decays involving electrons and muons, a variation of $E_{\min}^{\gamma}$ over a region of a few tens of megaelectronvolts has virtually no effect on the partial decay width, while, for τ leptons, the choice of $E_{\min}^{\gamma}$ is crucial in calculating the total width.

From a comparison of the numerical results in Tables 2 and 3, it follows that, in order to calculate the widths with respect to the decays $B_{d,s}^0 \to e^+e^-\gamma$ and $B_{d,s}^0 \to \mu^+\mu^-\gamma$ correctly, it is of paramount importance to take into account the contributions described in Sections 2 and 3. For the radiative decays of the B_d^0 meson, the increase in the partial widths stems, to a greater extent, from the appearance of the terms $F_{TV,TA}(0,0)$ in (15) and from weak annihilation and, to a less extent, from the contribution of ρ^0 and ω resonances. For the radiative decays of the B_s^0 meson to light leptons, the ϕ-resonance contribution to the increase in the partial width is dominant.

In addition, one can see from Figs. 3 and 4 that the presence of narrow and high resonance peaks in the region of low $\hat{s}$ (solid curves) changes rather strongly the shape of the distributions in the invariant mass of dileptons and in E^{γ} in relation to the distributions from [4, 6] (dotted curves). For the decays of $B_{d,s}^0$ mesons to light leptons and a photon, the solid and the dotted curve are parallel at low $\hat{s}$. This is due to the presence of the terms $F_{TV,TA}(0,0)$ and an axial anomaly in expression (15). In Fig. 4, the choice of a logarithmic scale highlights the importance of the ϕ-meson contribution to the partial widths with respect to the decays $B_s^0 \to e^+e^-\gamma$ and $B_s^0 \to \mu^+\mu^-\gamma$.

A comparison of the numerical results presented in Table 3 and in [3] shows good agreement for

the partial widths with respect to the decays $B_s^0 \to e^+e^-(\mu^+\mu^-)\gamma$. However, there cannot be perfect agreement since no account of the resonance contribution and the weak-annihilation contribution was taken in [3] and since the q^2 dependence of the tensor form factors was defined differently in [3] and in this study (see Section 2).

We note that the region of light resonances cannot be eliminated from an experimental analysis by analogy with the region of J/ψ and ψ' resonances, since, at low $\hat{s}$, the energy of the photons is sufficient for their detection in an electromagnetic calorimeter, and this is precisely the region where experimentalists plan to collect the main signal for identifying the rare radiative decays $B_{d,s}^0 \to \mu^+\mu^-\gamma$ at LHC.

In the decays $B_{d,s}^0 \to \tau^+\tau^-\gamma$, the partial widths undergo virtually no changes upon taking into account the results obtained in Sections 2 and 3. Indeed, the peaks of the ρ^0, ω, ϕ and J/ψ resonances lie in the region below the minimum kinematical boundary $\hat{s}_{\min} = 4\hat{m}_\tau^2$ for the decays being considered. The region around the ψ'-resonance peak is eliminated in performing integration with respect to $\hat{s}$. Thus, the partial widths calculated for the decays $B_{d,s}^0 \to \tau^+\tau^-\gamma$ with allowance for resonances prove to be even somewhat smaller than their counterparts not including resonance contributions (see Tables 2 and 3).

CONCLUSION

We have calculated the partial widths with respect to the decays $B_{d,s}^0 \to \ell^+\ell^-\gamma$, taking into account $B_q^0 \to \gamma$ and $B_q^0 \to \gamma^*$ transitions, lepton bremsstrahlung, and weak-annihilation effects (see Table 3). By employing the hypothesis of vector-meson dominance, we have found the $B_q^0 \to \gamma^*$ form factors, which differ radically from their counterparts in [3]. It has been shown that the contribution of these form factors affects substantially the partial widths with respect to rare radiative decays and the respective dilepton mass spectra if light leptons appear in the final state and is not significant if the final state involves a $\tau^+\tau^-$ pair. In view of this, it is necessary to introduce corrections in the earlier studies reported in [2–6].

ACKNOWLEDGMENTS

We are grateful to L.N. Smirnov and B. Stech for stimulating discussions.

The work of N.V. Nikitin was supported by INTAS (grant no. YSF2001/2-122) and the work of D.I. Melikhov was supported by the Alexander von Humboldt Foundation.

REFERENCES

1. T. Inami and C. S. Lim, Prog. Theor. Phys. **65**, 297 (1981); B. Grinstein, M. B. Wise, and M. J. Savage, Nucl. Phys. B **319**, 271 (1989); A. Buras and M. Münz, Phys. Rev. D **52**, 186 (1995).
2. F. Krüger and D. Melikhov, Phys. Rev. D **67**, 034 002 (2003).
3. Yu. Dincer and L. M. Sehgal, Phys. Lett. B **521**, 7 (2001).
4. C. Q. Geng, C. C. Lih, and W. M. Zhang, Phys. Rev. D **62**, 074 017 (2000).
5. S. Descotes-Genon and C. T. Sachrajda, Phys. Lett. B **557**, 213 (2003).
6. T. M. Aliev, A. Ozpineci, and M. Savci, Phys. Rev. D **55**, 7059 (1997).
7. S. Bosch and G. Buchalla, J. High Energy Phys. **0208**, 054 (2002).
8. P. Ball *et al.*, in *Proceedings of the Workshop on Standard Model Physics (and More) at the LHC*, CERN 2000-004 (2000), p. 305.
9. F. Krüger and L. M. Sehgal, Phys. Rev. D **55**, 2799 (1997); D. Melikhov, N. Nikitin, and S. Simula, Phys. Lett. B **430**, 332 (1998); Phys. Rev. D **57**, 6814 (1998); D. Melikhov, Phys. Lett. B **516**, 61 (2001).
10. D. Melikhov, Phys. Rev. D **53**, 2460 (1996); **56**, 7089 (1997).
11. J. J. Sakurai, Ann. Phys. (N.Y.) **11**, 1 (1960); M. Gell-Mann and F. Zachariasen, Phys. Rev. **124**, 953 (1961); G. J. Gounaris and J. J. Sakurai, Phys. Rev. Lett. **21**, 244 (1968).
12. D. Melikhov, O. Nachtmann, V. Nikonov, and T. Paulus, Eur. Phys. J. C **34**, 345 (2004).
13. D. Melikhov and B. Stech, Phys. Rev. D **62**, 014 006 (2000).
14. A. Khodjamirian, G. Stoll, and D. Wyler, Phys. Lett. B **358**, 129 (1995); A. Ali and V. M. Braun, Phys. Lett. B **359**, 223 (1995); M. Beyer, D. Melikhov, N. Nikitin, and B. Stech, Phys. Rev. D **64**, 094 006 (2001).
15. G. P. Korchemsky, D. Pirjol, and T.-M. Yan, Phys. Rev. D **61**, 114 510 (2000).
16. N. V. Nikitin, F. K. Rizatdinova, and L. N. Smirnova, Yad. Fiz. **62**, 1823 (1999) [Phys. At. Nucl. **62**, 1697 (1999)].

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1851–1865. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1913–1927.

ELEMENTARY PARTICLES AND FIELDS
Theory

Higgs Bosons in the Two-Doublet Model Involving CP Violation

É. N. Akhmetzyanova[1)*], M. V. Dolgopolov[1)], and M. N. Dubinin[2)***]**

Received April 1, 2004; in final form, February 10, 2005

Abstract—An effective two-doublet Higgs potential that involves complex-valued parameters and whose CP invariance is violated both explicitly and spontaneously is considered. The problem of diagonalizing the mass term of this potential at a local minimum is solved. The eigenstates of Higgs bosons and their mass spectrum are obtained for the special case of the two-doublet Higgs sector of the minimal supersymmetric model, where the CP invariance of the effective potential is violated owing to the interactions of the Higgs fields with the third-generation scalar quarks. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

It is well known that the origin of the Cabibbo–Kobayashi–Maskawa (CKM) mixing matrix [1] is associated with the Lagrangian terms of the Standard Model that describe Higgs boson interactions with quarks (Yukawa terms),

$$L = -g^u_{ij}\bar{\psi}'^i_L H u'^j_R - g^d_{ij}\bar{\psi}'^i_L \tilde{H} d'^j_R + \text{h.c.}, \qquad (1)$$

where $\bar{\psi}^{1\prime}_L = (\bar{u}', \bar{d}')_L$, $\bar{\psi}^{2\prime}_L = (\bar{c}', \bar{s}')_L$, $\bar{\psi}^{3\prime}_L = (\bar{t}', \bar{b}')_L$; $u^{1\prime}_R = u'_R$, $u^{2\prime}_R = c'_R$, $u^{3\prime}_R = t'_R$; $d^{1\prime}_R = d'_R$, $d^{2\prime}_R = s'_R$, $d^{3\prime}_R = b'_R$; H denotes the doublet of complex scalar fields; $\tilde{H}_k = \epsilon_{kl}H^*_l$; and g^u_{ij} and g^d_{ij} are 3×3 matrices. Their matrix elements are in general complex numbers that can be determined, apart from the phases of the CP transformation[3)] of the spinor fields of quarks and the scalar fields of the Higgs boson. In order to diagonalize the mass term of the quarks after the spontaneous symmetry breaking $H \to (0, v/\sqrt{2})$, one must perform unitary transformations of the fields of the up and down quarks $u^{i\prime}$ and $d^{i\prime}$, defining the "rotated" fields $u^i_{L,R} = U_{L,R}u^{i\prime}_{L,R}$ and $d^i_{L,R} = D_{L,R}d^{i\prime}_{L,R}$. After this diagonalization, the unitary-transformation matrices U_L and D_L do not appear either in the terms of the Lagrangian in (1) or in the interactions of neutral vector quark currents, but they arise explicitly in the Lagrangian terms that describe the interactions of the charged currents of $u^{i\prime}$ and $d^{i\prime}$: $g\bar{u}'_L\gamma_\mu d'_L = g\bar{u}_L\gamma_\mu U_L D^+_L d_L$. The product $V_{\text{CKM}} = U_L D^+_L$ determines the complex Cabibbo–Kobayashi–Maskawa matrix, which describes CP-violation effects. Within the Standard Model pattern, there arises CP violation because one cannot obtain in general purely real matrix elements V_{CKM} by employing CP transformations for six up and down quarks. In other words, CP violation in the Standard Model is due to the fact that there are precisely three generations of fundamental fermions.

At the same time, one can introduce, for a system of several scalar fields, Hermitian Lagrangians that explicitly violate CP invariance [2]. By way of example, we indicate that, for three complex fields φ_1, φ_2, and φ_3, this may be

$$L = \lambda\varphi_1\varphi^*_2\varphi^*_3 + \lambda^*\varphi^*_1\varphi_2\varphi_3, \quad CPLP^+C^+$$
$$= L^{CP} = \lambda e^{i\alpha}\varphi^*_1\varphi_2\varphi_3 + \lambda^* e^{-i\alpha}\varphi_1\varphi^*_2\varphi^*_3,$$

where λ is a complex-valued parameter and α is the CP-transformation phase, which is immaterial in the case being considered. It can be removed by means of a phase transformation of the fields that is associated with the conservation of the corresponding charge. After that, one can see that L and L^{CP} differ only by the sign of the imaginary part of λ. In the extremely simple example being considered, this will not lead to observable consequences, because the phase of the parameter λ can also be removed by means of a $U(1)$ rotation. However, this procedure cannot be implemented in general for a system involving a trilinear interaction $\lambda_i\varphi_j\varphi_k\varphi_l$ of four scalar fields. It can easily be found that the Lagrangian of such a system is CP-invariant only in the case where the phases of the four parameters λ_i satisfy specific conditions that

[1)]Samara State University, ul. Akademika Pavlova 1, Samara, 443011 Russia.
[2)]Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia.
[*]e-mail: elza@ssu.samara.ru
[**]e-mail: dolg@ssu.samara.ru
[***]e-mail: dubinin@theory.sinp.msu.ru

[3)]We recall that, from the definition concerning, for example, P of the form $Pa^+_\sigma(\mathbf{p})P^+ = \eta_\sigma a^+_\sigma(-\mathbf{p})$, where the complex factor containing the phase of the P transformation satisfies the condition $|\eta_\sigma| = 1$ and where $\sigma = 0$ or $1/2$, it follows that $P\phi(x)P^+ = \eta^*_0\phi(x')$ and $P\psi(x)P^+ = \eta^*_{1/2}\gamma_0\psi(x')$, where $x' = Px$.

1063-7788/05/6811-1851$26.00

ensure the possibility of removing them by means of $U(1)$ rotations of the fields φ_i. Models that contain an extended Higgs sector and where the CP invariance of the Higgs potential involving complex-valued parameters is explicitly violated are of great interest from this point of view. The two-doublet effective potential of the minimal supersymmetric model—if one does not consider the possibility of a spontaneous violation of CP invariance [3], this potential contains ten parameters, four of which can be complex-valued [4, 5]—is the simplest of such cases.

2. VIOLATION OF THE CP INVARIANCE OF THE EFFECTIVE TWO-DOUBLET POTENTIAL

In the general two-doublet model (two Higgs doublet model, also known as THDM), there are two $SU(2)$ doublets of complex scalar fields,

$$\Phi_1 = \begin{pmatrix} \phi_1^+(x) \\ \phi_1^0(x) \end{pmatrix} = \begin{pmatrix} -i\omega_1^+ \\ \dfrac{1}{\sqrt{2}}(v_1 + \eta_1 + i\chi_1) \end{pmatrix}, \quad (2)$$

$$\Phi_2 = e^{i\xi}\begin{pmatrix} \phi_2^+(x) \\ \phi_2^0(x) \end{pmatrix} = e^{i\xi}\begin{pmatrix} -i\omega_2^+ \\ \dfrac{1}{\sqrt{2}}(v_2 e^{i\zeta} + \eta_2 + i\chi_2) \end{pmatrix}, \quad (3)$$

their vacuum expectation values being nonzero,

$$\langle\Phi_1\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \quad (4)$$

$$\langle\Phi_2\rangle = \frac{e^{i\xi}}{\sqrt{2}}\begin{pmatrix} 0 \\ v_2 e^{i\zeta} \end{pmatrix} \equiv \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v_2 e^{i\theta} \end{pmatrix}.$$

Here, the quantities v_1 and v_2 are real. The phases ζ and ξ reflect the possible arbitrariness in the choice of the relative rotation of the vacuum expectation values and the relative rotation of the doublets of complex scalar fields. The sum θ of the phases will appear as a parameter in the conditions of the existence of a local minimum of the effective Higgs potential.

It is convenient to single out the phase ξ in order to analyze the possible relation between this phase and the phase of the elements of the Cabibbo–Kobayashi–Maskawa matrix. In the following, it will be shown that these phases are different. A representation of vacuum expectation values that is similar to (4) was studied in detail, for example, in [6] for a simplified THDM potential not containing terms of dimension $2\Phi_1^\dagger\Phi_2 + \Phi_2^\dagger\Phi_1$ and featuring real-valued parameters μ_{12}^2 and $\lambda_{5,6,7}$ and spontaneous CP violation (that is, the angle of the relative rotation of the doublets is $\xi = 0$, but $\zeta \neq 0$). In the present study, we consider the most general case of the THDM, that which involves complex-valued parameters μ_{12}^2 and $\lambda_{5,6,7}$ in the scalar two-doublet potential and nonzero phases ζ and ξ.

The most general Hermitian form of the renormalizable $SU(2) \times U(1)$-invariant Lagrangian for the system of fields specified by Eqs. (2) and (3) is

$$\mathcal{L}_\mathrm{H} = (\mathcal{D}_\nu\Phi_1)^\dagger\mathcal{D}^\nu\Phi_1 + (\mathcal{D}_\nu\Phi_2)^\dagger\mathcal{D}^\nu\Phi_2 \quad (5)$$
$$+ \kappa(\mathcal{D}_\nu\Phi_1)^\dagger\mathcal{D}^\nu\Phi_2 + \kappa^*(\mathcal{D}_\nu\Phi_2)^\dagger\mathcal{D}^\nu\Phi_1 - U(\Phi_1, \Phi_2),$$

where the potential may contain the following invariant terms:

$$U(\Phi_1, \Phi_2) = -\mu_1^2(\Phi_1^\dagger\Phi_1) - \mu_2^2(\Phi_2^\dagger\Phi_2) \quad (6)$$
$$- \mu_{12}^2(\Phi_1^\dagger\Phi_2) - (\mu_{12}^2)^*(\Phi_2^\dagger\Phi_1) + \lambda_1(\Phi_1^\dagger\Phi_1)^2$$
$$+ \lambda_2(\Phi_2^\dagger\Phi_2)^2 + \lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2)$$
$$+ \lambda_4(\Phi_1^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1) + \frac{\lambda_5}{2}(\Phi_1^\dagger\Phi_2)(\Phi_1^\dagger\Phi_2)$$
$$+ \frac{\lambda_5^*}{2}(\Phi_2^\dagger\Phi_1)(\Phi_2^\dagger\Phi_1) + \lambda_6(\Phi_1^\dagger\Phi_1)(\Phi_1^\dagger\Phi_2)$$
$$+ \lambda_6^*(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_1) + \lambda_7(\Phi_2^\dagger\Phi_2)(\Phi_1^\dagger\Phi_2)$$
$$+ \lambda_7^*(\Phi_2^\dagger\Phi_2)(\Phi_2^\dagger\Phi_1).$$

The potential parameters μ_{12}^2, λ_5, λ_6, and λ_7 are complex-valued. The complex-valued parameter κ describes the mixed kinetic term of the Lagrangian (for details, see [7]). Precision experimental data on the gauge-boson masses $m_{W,Z}$ impose stringent constraints on its real part. Moreover, the presence of the mixed kinetic term prevents a simultaneous construction of a diagonal 4×4 matrix of kinetic terms for Higgs bosons and a diagonal matrix for their mass term.[4)] In the following, we therefore set $\kappa = 0$.

The potential of the Higgs sector in the minimal supersymmetric model (MSSM) is a special case of the two-doublet potential (6). Within this model, the parameters $\lambda_{1,\ldots,7}$ are real-valued in the tree approximation at the energy scale of M_SUSY (that is, at energies on the order of the sparticle masses); they are expressed in terms of the coupling constants g_1 and g_2 of the electroweak gauge-symmetry group $SU(2) \times U(1)$ as [9]

$$\lambda_1(M_\mathrm{SUSY}) = \lambda_2(M_\mathrm{SUSY}) = \frac{1}{8}(g_2^2(M_\mathrm{SUSY}) \quad (7)$$

[4)] The corresponding conditions can be written as a set of ten linear equations, which, in the cases of practical importance, has a solution only at $\kappa = 0$. We also note that, in order to ensure renormalizability, it is not necessary to introduce a mixed kinetic term. It is shown below that the corresponding contributions from the renormalization of the scalar-quark fields to the effective parameters $\lambda_{5,6,7}$ vanish (see also [8]).

$$+ g_1^2(M_{\rm SUSY})),$$
$$\lambda_3(M_{\rm SUSY}) = \frac{1}{4}(g_2^2(M_{\rm SUSY}) - g_1^2(M_{\rm SUSY})),$$
$$\lambda_4(M_{\rm SUSY}) = -\frac{1}{2}g_2^2(M_{\rm SUSY}),$$
$$\lambda_5(M_{\rm SUSY}) = \lambda_6(M_{\rm SUSY}) = \lambda_7(M_{\rm SUSY}) = 0.$$

The CP-invariance of the potential is not violated at the scale of $M_{\rm SUSY}$. However, the parameters of any model depend on the energy scale at which they are measured or fixed. This dependence is described by renormalization-group equations. The conditions in (7) play the role of boundary conditions for the renormalization-group equations. Below the scale of $M_{\rm SUSY}$, the supersymmetric relations (7) are violated by a finite quantum effect caused by Higgs boson interaction with the third-generation squarks (the interaction with the first- and second-generation squarks is strongly suppressed). The potential for this interaction can be written as [8]

$$\mathcal{V}^0 = \mathcal{V}_M + \mathcal{V}_\Gamma + \mathcal{V}_\Lambda + \mathcal{V}_{\widetilde{Q}}, \tag{8}$$

where

$$\mathcal{V}_M = (-1)^{i+j} m_{ij}^2 \Phi_i^\dagger \Phi_j + M_{\widetilde{Q}}^2 \left(\widetilde{Q}^\dagger \widetilde{Q}\right) \tag{9}$$
$$+ M_{\widetilde{U}}^2 \widetilde{U}^* \widetilde{U} + M_{\widetilde{D}}^2 \widetilde{D}^* \widetilde{D},$$

$$\mathcal{V}_\Gamma = \Gamma_i^D \left(\Phi_i^\dagger \widetilde{Q}\right) \widetilde{D} + \Gamma_i^U \left(i\Phi_i^T \sigma_2 \widetilde{Q}\right) \widetilde{U} \tag{10}$$
$$+ (\Gamma_i^D)^* \left(\widetilde{Q}^\dagger \Phi_i\right) \widetilde{D}^* - (\Gamma_i^U)^* \left(i\widetilde{Q}^\dagger \sigma_2 \Phi_i^*\right) \widetilde{U}^*,$$

$$\mathcal{V}_\Lambda = \Lambda_{ik}^{jl} \left(\Phi_i^\dagger \Phi_j\right) \left(\Phi_k^\dagger \Phi_l\right) + \left(\Phi_i^\dagger \Phi_j\right) \tag{11}$$
$$\times \left[\Lambda_{ij}^Q \left(\widetilde{Q}^\dagger \widetilde{Q}\right) + \Lambda_{ij}^U \widetilde{U}^* \widetilde{U} + \Lambda_{ij}^D \widetilde{D}^* \widetilde{D}\right]$$
$$+ \overline{\Lambda}_{ij}^Q \left(\Phi_i^\dagger \widetilde{Q}\right) \left(\widetilde{Q}^\dagger \Phi_j\right)$$
$$+ \frac{1}{2}\left[\Lambda \epsilon_{ij} \left(i\Phi_i^T \sigma_2 \Phi_j\right) \widetilde{D}^* \widetilde{U} + \text{h.c.}\right], \quad i,j,k,l = 1,2,$$

$\mathcal{V}_{\widetilde{Q}}$ stands for terms associated with the interaction of four scalar quarks, and σ_2 is the Pauli matrix

$$\sigma_2 \equiv \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

The Yukawa coupling constants for the third-generation scalar quarks are defined in a standard way as

$$h_t = \frac{\sqrt{2} m_t}{v \sin\beta}, \quad h_b = \frac{\sqrt{2} m_b}{v \cos\beta},$$

where $v^2 \equiv v_1^2 + v_2^2$, and $\tan\beta \equiv v_2/v_1$. Following [10], we obtain[5)]

$$\Gamma_{\{1;2\}}^U = h_U\{-\mu^*; A_U\}, \quad \Gamma_{\{1;2\}}^D = h_D\{A_D; -\mu^*\}. \tag{12}$$

In the case considered here, these parameters are complex-valued. It is obvious that, in the sector of scalar-quark interaction with Higgs bosons, there appear CP-odd structures of the type in (1) and that it is natural to expect the emergence of mixing matrices that are 3×3 analogs of the Cabibbo–Kobayashi–Maskawa matrix. The trilinear parameters A_t and A_b and the higgsino–neutralino mass parameter μ must be taken to be complex-valued. Their imaginary parts may be quite large.

Thus, we see that, at the energy scale of m_Z or m_t, which is much less than $M_{\rm SUSY}$, the effective potential of the minimal supersymmetric model [11] is the general two-doublet potential (6), whose parameters λ_i are expressed in terms of the coupling constants for Higgs boson interaction with the scalar quarks and the scalar-quark masses, which play the role of Pauli–Villars regulators. For the renormalization-group equations that determine the evolution of the parameters λ_i, as well as the evolution of the Yukawa coupling constants $h_{t,b}$ and the gauge-interaction coupling constants $g_{1,2}$, there are the boundary conditions in (7), which are formulated at the scale $M_{\rm SUSY}$ and which are modified with allowance for the Higgs boson interaction (8) with third-generation scalar quarks. The radiative corrections $\Delta\lambda_i^{\text{eff.pot.}}$ to the boundary conditions in (7) for the parameters λ_i at the scale m_t were calculated within the effective-potential method. Taking also into account the one-loop contributions $\Delta\lambda_i^{\text{field}}$ from field renormalization (which are not included within the effective-potential method—for details, see [12]), we obtain

$$\lambda_1 = \frac{g_2^2 + g_1^2}{8} + \frac{3}{32\pi^2}\left[h_b^4 \frac{|A_b|^2}{M_{\rm SUSY}^2}\right. \tag{13}$$
$$\times \left(2 - \frac{|A_b|^2}{6M_{\rm SUSY}^2}\right) - h_t^4 \frac{|\mu|^4}{6M_{\rm SUSY}^4} + 2h_b^4 l$$
$$\left. + \frac{g_2^2 + g_1^2}{4M_{\rm SUSY}^2}(h_t^2|\mu|^2 - h_b^2|A_b|^2)\right] + \Delta\lambda_1^{\text{field}}$$
$$+ \frac{1}{768\pi^2}\left(11g_1^4 + 9g_2^4 - 36(g_1^2 + g_2^2)h_b^2\right) l,$$
$$\lambda_2 = \lambda_1(t \longleftrightarrow b, \Delta\lambda_1^{\text{field}} \longrightarrow \Delta\lambda_2^{\text{field}}),$$

[5)] For the CP-conservation case, which was considered in [8], the trilinear parameters in (10) are real-valued. For them, use was made of the notation $\Gamma_{\{1;2\}}^U \equiv h_U\{-\mu; A_U\}$ and $\Gamma_{\{1;2\}}^D \equiv h_D\{A_D; -\mu\}$.

$$\lambda_3 = \frac{g_2^2 - g_1^2}{4}\left[1 - \frac{3}{16\pi^2}(h_t^2 + h_b^2)l\right] \quad (14)$$
$$+ \frac{3}{8\pi^2}h_t^2 h_b^2\left[l + \frac{1}{2}X_{tb}\right] + \frac{3}{96\pi^2}\frac{|\mu|^2}{M_{\rm SUSY}^2}$$
$$\times\left[h_t^4\left(3 - \frac{|A_t|^2}{M_{\rm SUSY}^2}\right) + h_b^4\left(3 - \frac{|A_b|^2}{M_{\rm SUSY}^2}\right)\right]$$
$$+ \frac{3(g_2^2 - g_1^2)\left[h_b^2(|\mu|^2 - |A_b|^2) + h_t^2(|\mu|^2 - |A_t|^2)\right]}{128\pi^2 M_{\rm SUSY}^2}$$
$$+ \Delta\lambda_3^{\rm field} + \frac{9g_2^4 - 11g_1^4}{384\pi^2}l,$$

$$\lambda_4 = -\frac{g_2^2}{2}\left[1 - \frac{3}{16\pi^2}(h_t^2 + h_b^2)l\right] \quad (15)$$
$$- \frac{3}{8\pi^2}h_t^2 h_b^2\left[l + \frac{1}{2}X_{tb}\right] + \frac{3}{96\pi^2}\frac{|\mu|^2}{M_{\rm SUSY}^2}$$
$$\times\left[h_t^4\left(3 - \frac{|A_t|^2}{M_{\rm SUSY}^2}\right) + h_b^4\left(3 - \frac{|A_b|^2}{M_{\rm SUSY}^2}\right)\right]$$
$$- \frac{3g_2^2\left[h_b^2(|\mu|^2 - |A_b|^2) + h_t^2(|\mu|^2 - |A_t|^2)\right]}{64\pi^2 M_{\rm SUSY}^2}$$
$$+ \Delta\lambda_4^{\rm field} - \frac{3g_2^4}{64\pi^2}l,$$

where

$$X_{tb} \equiv \frac{|A_t|^2 + |A_b|^2 + 2{\rm Re}e(A_b^* A_t)}{2M_{\rm SUSY}^2} \quad (16)$$
$$- \frac{|\mu|^2}{M_{\rm SUSY}^2} - \frac{||\mu|^2 - A_b^* A_t|^2}{6M_{\rm SUSY}^4}.$$

Here and in other formulas, $l \equiv \ln\left(M_{\rm SUSY}^2/\sigma^2\right)$, where σ is the renormalization scale. The effective complex-valued parameters $\lambda_{5,6,7}$ are given by

$$\lambda_5 = -\Delta\lambda_5 = -\frac{3}{96\pi^2} \quad (17)$$
$$\times\left(h_t^4\left(\frac{\mu A_t}{M_{\rm SUSY}^2}\right)^2 + h_b^4\left(\frac{\mu A_b}{M_{\rm SUSY}^2}\right)^2\right),$$

$$\lambda_6 = -\Delta\lambda_6 = \frac{3}{96\pi^2} \quad (18)$$
$$\times\left[h_t^4\frac{|\mu|^2\mu A_t}{M_{\rm SUSY}^4} - h_b^4\frac{\mu A_b}{M_{\rm SUSY}^2}\left(6 - \frac{|A_b|^2}{M_{\rm SUSY}^2}\right)\right.$$
$$\left. + (h_b^2 A_b - h_t^2 A_t)\frac{3\mu}{M_{\rm SUSY}^2}\frac{g_2^2 + g_1^2}{4}\right],$$

$$\lambda_7 = -\Delta\lambda_7 = \frac{3}{96\pi^2} \quad (19)$$
$$\times\left[h_b^4\frac{|\mu|^2\mu A_b}{M_{\rm SUSY}^4} - h_t^4\frac{\mu A_t}{M_{\rm SUSY}^2}\left(6 - \frac{|A_t|^2}{M_{\rm SUSY}^2}\right)\right.$$
$$\left. + (h_t^2 A_t - h_b^2 A_b)\frac{3\mu}{M_{\rm SUSY}^2}\frac{g_2^2 + g_1^2}{4}\right].$$

The one-loop contributions $\lambda_i^{\rm field}$ from field renormalization to the aforementioned parameters λ_i have the form

$$\Delta\lambda_1^{\rm field} = \frac{1}{2}(g_1^2 + g_2^2)A'_{11}, \quad (20)$$
$$\Delta\lambda_2^{\rm field} = \frac{1}{2}(g_1^2 + g_2^2)A'_{22},$$

$$\Delta\lambda_3^{\rm field} = -\frac{1}{4}(g_1^2 - g_2^2)(A'_{11} + A'_{22}),$$
$$\Delta\lambda_4^{\rm field} = -\frac{1}{2}g_2^2(A'_{11} + A'_{22}), \quad \Delta\lambda_5^{\rm field} = 0,$$
$$\Delta\lambda_6^{\rm field} = \frac{1}{8}(g_1^2 + g_2^2)(A'_{12} - A'^*_{21}) = 0,$$
$$\Delta\lambda_7^{\rm field} = \frac{1}{8}(g_1^2 + g_2^2)(A'_{21} - A'^*_{12}) = 0.$$

They differ from the formulas given in [8] for the case of CP conservation by the presence of the logarithmic term l [see Eq. (12) and footnote 5]. In our case, the matrices $[A]$ have the form

$$A'_{ij} = -\frac{3\left(1 - \frac{1}{2}l\right)}{96\pi^2 M_{\rm SUSY}^2} \quad (21)$$
$$\times\left(h_U^2\begin{bmatrix} |\mu|^2 & -\mu^* A_U^* \\ -\mu A_U & |A_U|^2 \end{bmatrix}\right.$$
$$\left. + h_D^2\begin{bmatrix} |A_D|^2 & -\mu^* A_D^* \\ -\mu A_D & |\mu|^2 \end{bmatrix}\right).$$

The one-loop renormalization of the wave function does not make a CP-violating contribution to λ_i. In the following, the notation for the deviations of the effective parameters λ_i from $\lambda_i^{\rm SUSY} = \lambda_i(M_{\rm SUSY})$ is identical to that in [5]; that is,

$$\lambda_{1,2} \equiv \lambda_{1,2}^{\rm SUSY} - \Delta\lambda_{1,2}/2, \quad (22)$$
$$\lambda_{3,4} \equiv \lambda_{3,4}^{\rm SUSY} - \Delta\lambda_{3,4}, \quad \lambda_{5,6,7} \equiv -\Delta\lambda_{5,6,7},$$

where[6)]

$$\Delta\lambda_i \equiv \Delta\lambda_i^{\rm eff.\ pot} - \Delta\lambda_i^{\rm field}, \quad (23)$$
$$\Delta\lambda_i^{\{\rm eff.\ pot;\ field\}} \equiv \Delta\lambda_i^{\rm log} + \Delta\lambda_i^{\rm finite}.$$

It should be noted that

$$\Delta\lambda_{5,6,7}^{\rm log} = 0, \quad \Delta\lambda_{5,6,7}^{\rm field} = 0. \quad (24)$$

[6)] On the left-hand side of the first equation in (23), it is necessary to introduce an additional factor of $1/2$ for $\Delta\lambda_{1,2}$ because of the notation $\Delta\lambda_{1,2}/2$ in (22).

Table 1. Comparison of corrections to the parameters λ_i at the scale of m_t for $m_Z = 91.19$ GeV, $m_b = 3$ GeV, $m_t = 175$ GeV, $m_W = 79.96$ GeV, $g_2 = 0.6517$, $g_1 = 0.3573$, $v = 245.4$ GeV, $G_F = 1.174 \times 10^{-5}$ GeV^{-2}, $\alpha_s(m_t) = 0.1072$, $\tan\beta = 5$, $M_{SUSY} = 500$ GeV, $\sigma = m_t$, $m_{H^\pm} = 300$ GeV, $|A_t| = |A_b| = A = 1000$ GeV, $|\mu| = 2000$ GeV, $\varphi \equiv \arg(\mu A_{t,b}) = 0$, these values being coincident with those in [15]

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
$\mathcal{O}(h_t^4)$ approximation	0.907	−0.203	0.057	0.057	0.227	−0.453	0.057
$\Delta\lambda_i$	0.860	−0.182	0.054	0.072	0.227	−0.442	0.046
One-loop approximation [4]	0.907	−0.191	0.064	0.043	0.227	−0.453	0.057
One- + two-loop approximation [4]	0.761	−0.152	0.052	0.032	0.135	−0.371	0.044
Two-loop approximation [4]	−0.146	0.039	−0.012	−0.011	−0.092	0.082	−0.013
One-loop(D+ field) approximation	−0.047	0.009	−0.010	0.028	0	0.011	−0.011
$\dfrac{\Delta\lambda(D+\text{field}) - \Delta\lambda(\text{two-loop})}{\Delta\lambda(D+\text{field})} \times 100\%$	−211	−311	−20	139	−	−647	−15
$\dfrac{\lvert\Delta\lambda(D+\text{field})\rvert - \lvert\Delta\lambda(\text{two-loop})\rvert}{\lvert\Delta\lambda(D+\text{field})\rvert} \times 100\%$	−211	−311	−20	61	−	−647	−15
One-loop + two-loop + one-loop(D + field) approximation	0.715	−0.143	0.042	0.061	0.135	−0.360	0.033

Note: The one-loop(D + field) approximation includes one-loop corrections only from the D terms and from field renormalization.

Here, $\Delta\lambda_i^{\log}$ are logarithmic corrections, while $\Delta\lambda_i^{\text{finite}}$ are finite corrections.

To conclude this section, we will make several general comments concerning the results of other authors. A feature that distinguishes the present analysis from the standard scheme of summation of leading logarithms by means of renormalization-group equations is that, in the boundary conditions, we take into account effects of Higgs boson interaction with third-generation scalar quarks. In other words, we consider a two-doublet Higgs sector model that is an effective theory for minimal supersymmetry at the energy scale of m_t. The "running" one-loop effective parameters (13)–(19) satisfy boundary conditions that are specified by Eqs. (7) and finite power-law terms that are determined by the soft-supersymmetry-breaking Higgs boson interaction (8) with scalar quarks. The logarithmic term l describes the evolution of the parameters as the scale decreases from M_{SUSY} to $\sigma = m_t$. Finite power-law corrections to the parameters $\lambda_{6,7}$ appear owing to so-called F terms [terms of the trilinear interaction in (10)] and D terms [which are contained in (11)]. The corrections to λ_5 appear only owing to the F terms. The parameters λ_i of the effective two-doublet potential in the minimal supersymmetric model were considered previously in [4] for the case of CP violation and in [8, 13] for the case of CP conservation. For a phenomenological analysis, one usually sets $A_t = A_b \equiv A$ and introduces a universal phase μA, so that $\lambda_5 = |\lambda_5|\exp[i \cdot 2\arg(\mu A)]$, $\lambda_6 = |\lambda_6|\exp[i\arg(\mu A)]$, and $\lambda_7 = |\lambda_7|\exp[i\arg(\mu A)]$.

In [4, 13], the contributions of the D terms were taken into account only partly. The additional terms involving gauge-interaction coupling constants g_2^2 and g_1^2 and appearing in expressions (13)–(15), which we obtained for the effective parameters, correspond to their total contributions. The one-loop contributions $\Delta\lambda_{1,\ldots,4}^{\text{field}}$ from field renormalization were also disregarded in [4, 13]. Phases do not appear in expressions (13)–(15) up to the two-loop approximation inclusive, and these expressions coincide, apart from the contributions of the D terms and apart from the terms $\Delta\lambda_{1,\ldots,4}^{\text{field}}$, with the results reported in [4, 13]. At real-valued μ and A, expressions (13)–(15) correspond precisely to the results of Haber and Hempfling [8], who took into account the contributions of the D terms. We note that it would be illegitimate to generalize the results for $\lambda_{5,6,7}$ to the case of CP nonconservation by directly replacing real-valued μ and A by their complex-valued counterparts—this would lead to an erroneous result.

If one disregards the contribution of the D terms, the field-renormalization contributions $\Delta\lambda_{1,\ldots,4}^{\text{field}}$, and terms of order h_b^2 for the Yukawa coupling of the b quark, there arise parameters in the one-loop approximation, which contain only $\mathcal{O}(h_t^4)$ corrections. For the case of real-valued μ and A, the respective results are presented in [14]. By way of example, we indicate that the result for λ_2 is

$$\lambda_2 \approx \frac{g_2^2 + g_1^2}{8} + \frac{3}{32\pi^2} \times \left[h_t^4 \frac{|A|^2}{M_{SUSY}^2}\left(2 - \frac{|A|^2}{6M_{SUSY}^2}\right) + 2h_t^4 l\right]. \quad (25)$$

It is interesting to note that only for λ_2 does the renormalization-group beta function involve a constant (λ_i-independent) negative contribution, $-6h_t^4$, this being coincident with the respective result in [8].

A detailed numerical comparison of the correction to λ_i can be found in Table 1, where, in our case (see the second row),

$$\Delta\lambda_i = \{\text{one-loop contribution}\} + \{\text{one-loop contribution } (D \text{ terms} + \text{field renormalization})\}.$$

We conclude that the one-loop corrections from field renormalization and from the D terms must be taken into account since they can be on the same order as or even greater than the leading two-loop corrections.

3. DIAGONALIZATION OF THE MASS TERM IN THE EFFECTIVE POTENTIAL AT A LOCAL MINIMUM

3.1. Complex-Valued Parameters μ_{12}^2 and $\lambda_{5,6,7}$; $\theta = 0$

The components ω_i, η_i, and χ_i of the $SU(2)$ doublets in (2) and (3) are not mass eigenstates. In order to obtain the masses of the Higgs bosons and their interactions determined by the effective potential (6), it is necessary to diagonalize its mass term at a local minimum. For the case of complex-valued parameters μ_{12}^2 and $\lambda_{5,6,7}$ and zero phase $\theta = 0$ of the vacuum expectation value, this problem was considered in [5]. The diagonalization at a minimum is performed in two steps. First, the CP-even fields h and H, the CP-odd mass eigenstate A[7] (so-called pseudoscalar), and the Goldstone field G^0 are determined by means of a linear transformation,[8]

$$h = -\eta_1 \sin\alpha + \eta_2 \cos\alpha, \tag{26}$$

$$H = \eta_1 \cos\alpha + \eta_2 \sin\alpha, \tag{27}$$

$$A = -\chi_1 \sin\beta + \chi_2 \cos\beta, \tag{28}$$

$$G^0 = \chi_1 \cos\beta + \chi_2 \sin\beta, \tag{29}$$

$$\tan(2\alpha) = \frac{s_{2\beta}(m_A^2 + m_Z^2) + v^2((\Delta\lambda_3 + \Delta\lambda_4)s_{2\beta} + 2c_\beta^2 \mathrm{Re}\Delta\lambda_6 + 2s_\beta^2 \mathrm{Re}\Delta\lambda_7)}{c_{2\beta}(m_A^2 - m_Z^2) + v^2(\Delta\lambda_1 c_\beta^2 - \Delta\lambda_2 s_\beta^2 - \mathrm{Re}\Delta\lambda_5 c_{2\beta} + (\mathrm{Re}\Delta\lambda_6 - \mathrm{Re}\Delta\lambda_7)s_{2\beta})}, \tag{30}$$

where we have used the relations $g_1^2 + g_2^2 = g_2^2 m_Z^2/m_W^2$, $g_2^2 - g_1^2 = g_2^2(2 - m_Z^2/m_W^2)$. Further, we substitute into the effective potential the real-valued parameters $\mu_{1,2}$ and $\lambda_{1,2,3,4}$ and the real parts $\mathrm{Re}\mu_{12}^2$ and $\mathrm{Re}\lambda_{5,6,7}$, which are related by linear equations,

$$\lambda_1 = \frac{1}{2v^2}\left[\left(\frac{s_\alpha}{c_\beta}\right)^2 m_h^2 + \left(\frac{c_\alpha}{c_\beta}\right)^2 m_H^2 - \frac{s_\beta}{c_\beta^3}\mathrm{Re}\mu_{12}^2\right] + \frac{1}{4}(\mathrm{Re}\lambda_7 \tan^3\beta - 3\mathrm{Re}\lambda_6 \tan\beta), \tag{31}$$

$$\lambda_2 = \frac{1}{2v^2}\left[\left(\frac{c_\alpha}{s_\beta}\right)^2 m_h^2 + \left(\frac{s_\alpha}{s_\beta}\right)^2 m_H^2 - \frac{c_\beta}{s_\beta^3}\mathrm{Re}\mu_{12}^2\right] + \frac{1}{4}(\mathrm{Re}\lambda_6 \cot^3\beta - 3\mathrm{Re}\lambda_7 \cot\beta), \tag{32}$$

$$\lambda_3 = \frac{1}{v^2}\left[2m_{H^\pm}^2 - \frac{\mathrm{Re}\mu_{12}^2}{s_\beta c_\beta} + \frac{s_{2\alpha}}{s_{2\beta}}(m_H^2 - m_h^2)\right] - \frac{\mathrm{Re}\lambda_6}{2}\cot\beta - \frac{\mathrm{Re}\lambda_7}{2}\tan\beta, \tag{33}$$

$$\lambda_4 = \frac{1}{v^2}\left(\frac{\mathrm{Re}\mu_{12}^2}{s_\beta c_\beta} + m_A^2 - 2m_{H^\pm}^2\right) - \frac{\mathrm{Re}\lambda_6}{2}\cot\beta - \frac{\mathrm{Re}\lambda_7}{2}\tan\beta, \tag{34}$$

$$\mathrm{Re}\lambda_5 = \frac{1}{v^2}\left(\frac{\mathrm{Re}\mu_{12}^2}{s_\beta c_\beta} - m_A^2\right) - \frac{\mathrm{Re}\lambda_6}{2}\cot\beta - \frac{\mathrm{Re}\lambda_7}{2}\tan\beta, \tag{35}$$

$$\mu_1^2 = \lambda_1 v_1^2 + (\lambda_3 + \lambda_4 + \mathrm{Re}\lambda_5)\frac{v_2^2}{2} - \mathrm{Re}\mu_{12}^2 \tan\beta + \frac{v^2 s_\beta^2}{2}(3\mathrm{Re}\lambda_6 \cot\beta + \mathrm{Re}\lambda_7 \tan\beta), \tag{36}$$

$$\mu_2^2 = \lambda_2 v_2^2 + (\lambda_3 + \lambda_4 + \mathrm{Re}\lambda_5)\frac{v_1^2}{2} - \mathrm{Re}\mu_{12}^2 \cot\beta + \frac{v^2 c_\beta^2}{2}(\mathrm{Re}\lambda_6 \cot\beta + 3\mathrm{Re}\lambda_7 \tan\beta). \tag{37}$$

At real-valued parameters (in the following, we will also refer to this case as that of the CP-conserving

[7] The fields h, H, and A are mass eigenstates at $\phi = 0$.

[8] In the following, we employ a condensed notation; that is, $s_\beta \equiv \sin\beta$, $c_\beta \equiv \cos\beta$, $s_\alpha \equiv \sin\alpha$, $c_\alpha \equiv \cos\alpha$, $c_{2\beta} \equiv \cos 2\beta$, and so on.

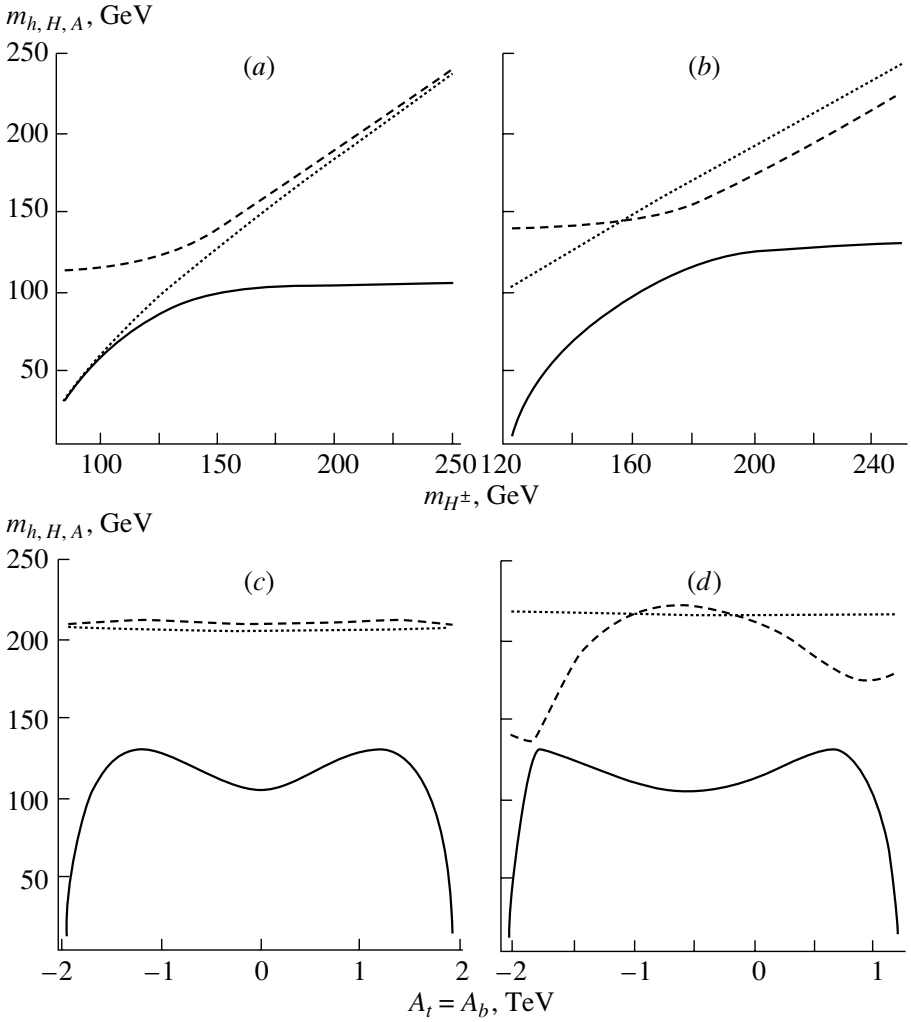

Fig. 1. Masses of the neutral Higgs bosons h, H, and A versus $m_{H^\pm}$ and versus the trilinear parameters A_t and A_b ($A_t = A_b$) in the limit of CP conservation at $\tan\beta = 5$ and $M_{\mathrm{SUSY}} = 0.5$ TeV: (solid curves) m_h, (dotted curves) m_A, and (dashed curves) m_H. The results are given here for the cases of (*a*) $A_t = A_b = \mu = 0$, (*b*) $A_t = A_b = 0.9$ TeV and $\mu = -1.5$ TeV, (*c*) $A_t = A_b$, $m_{H^\pm} = 220$ GeV and $\mu = 0$, and (*d*) $A_t = A_b$, $m_{H^\pm} = 220$ GeV and $\mu = -2$ TeV.

limit at $\varphi = 0$, $\mathrm{Re}\lambda_i = |\lambda_i|$, and $\mathrm{Re}\Delta\lambda_i = |\Delta\lambda_i|$), relations (36) and (37) ensure the vanishing of the terms in the potential that are linear in the fields and are therefore minimization conditions. From (31)–(35), it follows that, in the CP-conserving limit, the masses of the CP-even Higgs bosons and the real part of the parameter μ_{12}^2 are given by

$$m_H^2 = c_{\alpha+\beta}^2 m_Z^2 + s_{\alpha-\beta}^2 m_A^2 - v^2(\Delta\lambda_1 c_\alpha^2 c_\beta^2 + \Delta\lambda_2 s_\alpha^2 s_\beta^2 + 2(\Delta\lambda_3 + \Delta\lambda_4)c_\alpha c_\beta s_\alpha s_\beta + \mathrm{Re}\Delta\lambda_5(c_\alpha^2 s_\beta^2 + s_\alpha^2 c_\beta^2) + 2s_{\alpha+\beta}(\mathrm{Re}\Delta\lambda_6 \cdot c_\alpha c_\beta + \mathrm{Re}\Delta\lambda_7 \cdot s_\alpha s_\beta)), \quad (38)$$

$$m_h^2 = s_{\alpha+\beta}^2 m_Z^2 + c_{\alpha-\beta}^2 m_A^2 - v^2(\Delta\lambda_1 s_\alpha^2 c_\beta^2 + \Delta\lambda_2 c_\alpha^2 s_\beta^2 - 2(\Delta\lambda_3 + \Delta\lambda_4)c_\alpha c_\beta s_\alpha s_\beta + \mathrm{Re}\Delta\lambda_5(s_\alpha^2 s_\beta^2 + c_\alpha^2 c_\beta^2) - 2c_{\alpha+\beta}(\mathrm{Re}\Delta\lambda_6 \cdot s_\alpha c_\beta - \mathrm{Re}\Delta\lambda_7 \cdot c_\alpha s_\beta)), \quad (39)$$

$$m_{H^\pm}^2 = m_W^2 + m_A^2 - \frac{v^2}{2}(\mathrm{Re}\Delta\lambda_5 - \Delta\lambda_4), \quad (40)$$

$$\mathrm{Re}\mu_{12}^2 = s_\beta c_\beta\left[m_A^2 - \frac{v^2}{2}(2\mathrm{Re}\Delta\lambda_5 + \mathrm{Re}\Delta\lambda_6 \cdot \cot\beta + \mathrm{Re}\Delta\lambda_7 \cdot \tan\beta)\right].$$

Upon the substitution of (31)–(37) into (6), we obtain the mass term of the effective potential in the form

$$U_{\mathrm{mass}}(h,H,A) = c_0 A + c_1 hA + c_2 HA + \frac{m_h^2}{2}h^2 + \frac{m_H^2}{2}H^2 + \frac{m_A^2}{2}A^2 + m_{H^\pm}^2 H^+ H^-. \quad (41)$$

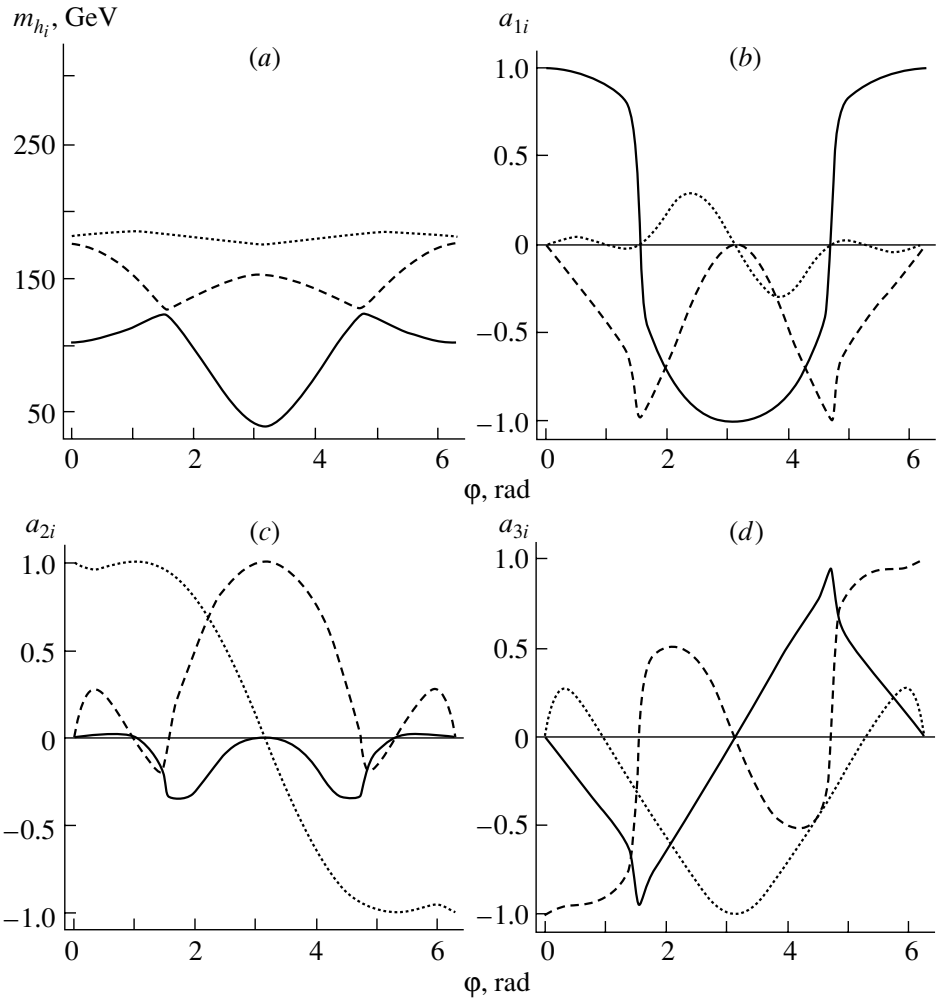

Fig. 2. (*a*) Masses of the neutral Higgs bosons h_i and (*b*–*d*) matrix elements a_{ij} versus the phase $\varphi = \arg(\mu A_{t,b})$ at $\tan\beta = 5$, $m_{H^\pm} = 180$ GeV, $M_{\rm SUSY} = 0.5$ TeV, $A_t = A_b = 1$ TeV, and $\mu = 2$ TeV: (solid curves) $i = 1$, (dashed curves) $i = 2$, and (dotted curves) $i = 3$.

The minimization condition $c_0 = 0$ fixes the imaginary part of the parameter μ_{12}^2:

$$\mathrm{Im}\mu_{12}^2 = \frac{v^2}{2}(s_\beta c_\beta \mathrm{Im}\lambda_5 + c_\beta^2 \mathrm{Im}\lambda_6 + s_\beta^2 \mathrm{Im}\lambda_7). \quad (42)$$

After that, the coefficients of the off-diagonal terms terms hA and HA at the minimum corresponding to $c_0 = 0$ can be represented in the form

$$\begin{aligned} c_1 &= \frac{v^2}{2}(s_\alpha s_\beta - c_\alpha c_\beta)\mathrm{Im}\lambda_5 \\ &+ v^2(s_\alpha c_\beta \mathrm{Im}\lambda_6 - c_\alpha s_\beta \mathrm{Im}\lambda_7), \\ c_2 &= -\frac{v^2}{2}(s_\alpha c_\beta + c_\alpha s_\beta)\mathrm{Im}\lambda_5 \\ &- v^2(c_\alpha c_\beta \mathrm{Im}\lambda_6 + s_\alpha s_\beta \mathrm{Im}\lambda_7). \end{aligned} \quad (43)$$

They involve only the imaginary parts $\mathrm{Im}\mu_{12}^2$ and $\mathrm{Im}\lambda_{5,6,7}$. Obviously, there appears no off-diagonal term of the form hH in (41), so that, below, we have $M_{12} = M_{21} = 0$ in the mixing matrix (45).

The second step consists in removing off-diagonal terms of the form hA and HA by means of an orthogonal transformation in the sector spanned by h, H, and A; that is,

$$\begin{aligned} &(h, H, A)M^2 \begin{pmatrix} h \\ H \\ A \end{pmatrix} \\ &= (h_1, h_2, h_3)a_{ik}^T M_{kl}^2 a_{lj} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}, \end{aligned} \quad (44)$$

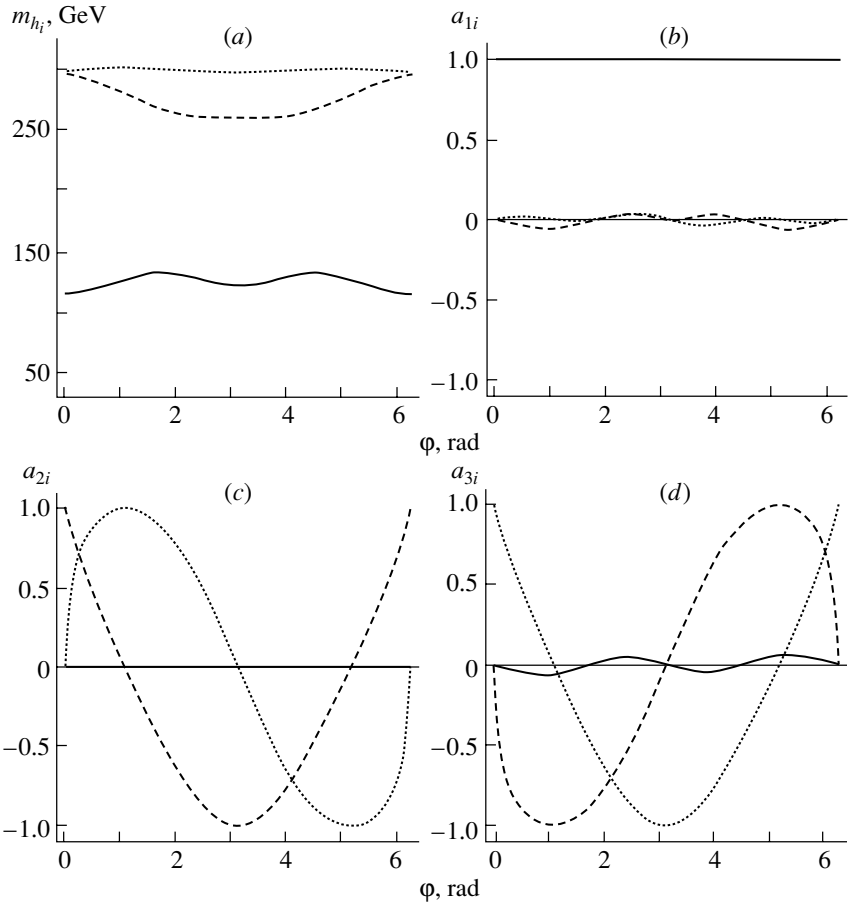

Fig. 3. As in Fig. 2, but for $m_{H^\pm} = 300$ GeV.

where the mass matrix has the form

$$M^2 = \frac{1}{2}\begin{pmatrix} m_h^2 & 0 & c_1 \\ 0 & m_H^2 & c_2 \\ c_1 & c_2 & m_A^2 \end{pmatrix}. \qquad (45)$$

As a result, there appear Higgs boson mass eigenstates h_1, h_2, and h_3, which are not CP eigenstates. The eigenvalues of the matrix M^2 determine the squares of their masses, while its normalized eigenvectors determine the rows in the mixing matrix a_{ij}. The squares of the masses of the Higgs boson mass eigenstates ($m_{h_1}^2 \le m_{h_2}^2 \le m_{h_3}^2$) are given by

$$m_{h_1}^2 = 2\sqrt{(-q)}\cos\left(\frac{\Theta + 2\pi}{3}\right) - \frac{a_2}{3}, \qquad (46)$$

$$m_{h_2}^2 = 2\sqrt{(-q)}\cos\left(\frac{\Theta + 4\pi}{3}\right) - \frac{a_2}{3},$$

$$m_{h_3}^2 = 2\sqrt{(-q)}\cos\left(\frac{\Theta}{3}\right) - \frac{a_2}{3},$$

where

$$\Theta = \arccos\frac{r}{\sqrt{(-q^3)}},$$

$$r = \frac{1}{54}(9a_1a_2 - 27a_0 - 2a_2^3),$$

$$q = \frac{1}{9}(3a_1 - a_2^2),$$

$$a_1 = m_h^2m_H^2 + m_h^2m_A^2 + m_H^2m_A^2 - c_1^2 - c_2^2,$$

$$a_2 = -m_h^2 - m_H^2 - m_A^2,$$

$$a_0 = c_1^2m_H^2 + c_2^2m_h^2 - m_h^2m_H^2m_A^2,$$

while the components of the eigenvectors $\begin{pmatrix} h \\ H \\ A \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \\ h_3 \end{pmatrix}$, $a_{ij} = a'_{ij}/n_j$, are

$$a'_{11} = ((m_H^2 - m_{h_1}^2)(m_A^2 - m_{h_1}^2) - c_2^2),$$

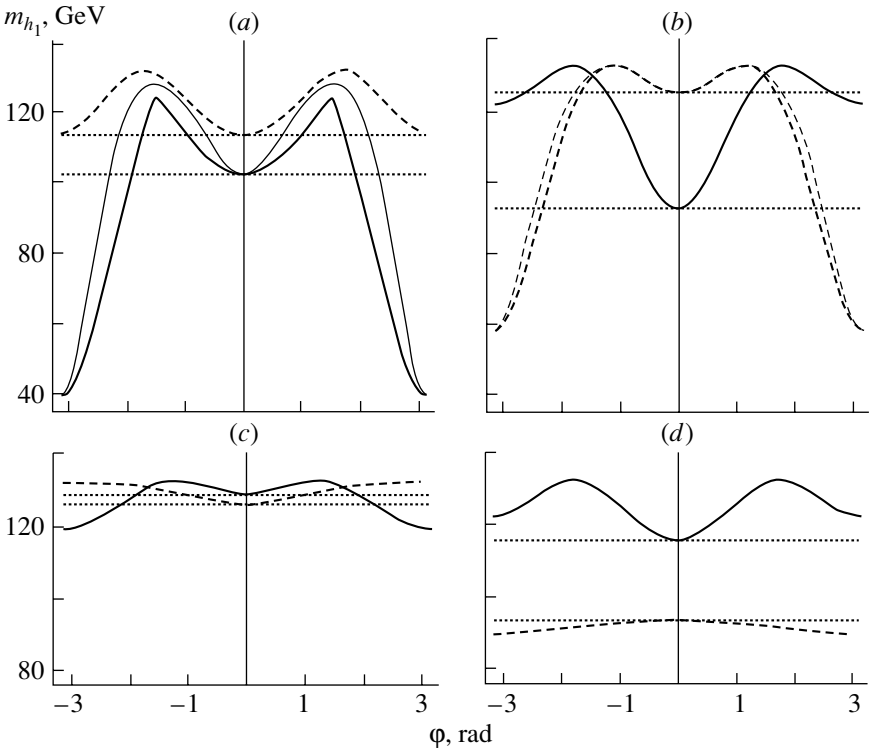

Fig. 4. Mass of the Higgs boson h_1 versus the phase φ at $M_{\rm SUSY} = 0.5$ TeV. The dotted straight lines correspond to $m_h(\varphi = 0)$. In Fig. 4*a*, for which we choose the values of $\tan\beta = 5$, $A_t = A_b = 1000$ GeV, and $\mu = 2000$ GeV, the thick and thin solid curves represent, respectively, m_{h_1} and $m_h(\varphi)$ at $m_{H^\pm} = 180$ GeV, while the dashed curve corresponds to m_{h_1} at $m_{H^\pm} = 250$ GeV; in Fig. 4*b*, for which we used the values of $\tan\beta = 5$, $m_{H^\pm} = 300$ GeV, and $\mu = 2000$ GeV, the solid curve represents m_{h_1} at $A_t = A_b = -1200$ GeV, while the thick and thin dashed curves show, respectively, m_{h_1} and $m_h(\varphi)$ for $A_t = A_b = 1300$ GeV; in Fig. 4*c*, the solid (dashed) curve corresponds to $\mu = -1600$ GeV ($\mu = 700$ GeV) at $\tan\beta = 5$, $m_{H^\pm} = 300$ GeV, and $A_t = A_b = 1000$ GeV; and, in Fig. 4*d*, the solid (dashed) curve corresponds to $\tan\beta = 5$ ($\tan\beta = 40$) at $\mu = 2000$ GeV, $m_{H^\pm} = 300$ GeV, and $A_t = A_b = 1000$ GeV.

$$a'_{21} = c_1c_2, a'_{31} = -c_1(m_H^2 - m_{h_1}^2),$$
$$a'_{12} = -c_1c_2,$$
$$a'_{22} = -((m_h^2 - m_{h_2}^2)(m_A^2 - m_{h_2}^2) - c_1^2),$$
$$a'_{32} = c_2(m_h^2 - m_{h_2}^2),$$
$$a'_{13} = -c_1(m_H^2 - m_{h_3}^2),$$
$$a'_{23} = -c_2(m_h^2 - m_{h_3}^2),$$
$$a'_{33} = (m_h^2 - m_{h_3}^2)(m_H^2 - m_{h_3}^2),$$

$n_i = \pm\sqrt{(a_{1i}'^2 + a_{2i}^2 + a_{3i}^{f\prime 2})}$. The dependences of the Higgs boson masses on the parameters $A_{t,b}$ and μ of the minimal supersymmetric model and on the universal phase $\varphi = \arg(\mu A_{t,b})$ are illustrated in Figs. 1–4, along with the behavior of the matrix elements a_{ij}, which characterize mixed states. In contrast to what was done in [5], we took $m_{H^\pm}$ and $\tan\beta$ for the independent parameters of the two-doublet sector instead of m_A and $\tan\beta$, our choice being more convenient for a comparison of the results with those reported in [4, 15]. A change in the phase φ may lead to the reversal of the sign of the parameters c_1 and c_2, the regions of positive and negative definite c_1 and c_2 depending functionally on the choice of values for the parameters m_A, $\tan\beta$, A, μ, and $M_{\rm SUSY}$ in the CP-conserving limit. Upon the passage of the zeros of c_1 and c_2, the matrix elements a_{ij} must change sign according to the requirement of a left-hand orthonormalized basis of eigenvectors. It is the most important that, in the mass matrix, m_{h_1}, m_{h_2}, and m_{h_3} lie on the diagonal from top to bottom and that, in the limiting case of $c_1 = c_2 = 0$, we have $m_{h_1} = \min(m_h, m_H, m_A)$ and $m_{h_3} = \max(m_h, m_H, m_A)$ ("mass ordering"). We note that, with increasing $\Delta\lambda_i$, the denominator on the right-hand side of (30) may change sign, with the result that the mass ordering requires determining the angle $\alpha(\varphi)$ consistently with the boundary condition at the scale of $M_{\rm SUSY}$: $m_A^2 + m_Z^2 = (-\sin(2\alpha)/\sin(2\beta))(m_H^2 - m_h^2)$, this boundary condition following from (31)–(35) and (7). To conclude this section, we present in Table 2 some numerical values obtained for the Higgs boson masses within our present approach by using various

Table 2. Higgs boson masses (in GeV) and decay widths (in GeV) in our case and in the case where use is made of the CPsuperH package [15] (in either case, the parameter values are $\alpha_{\rm EM}(m_Z) = 0.7812$, $\alpha_s(m_Z) = 0.1172$, $G_{\rm F} = 1.174 \times 10^{-5}$ GeV^{-2}, $\tan\beta = 5$, $M_{\rm SUSY} = 500$ GeV, $|A_t| = |A_b| = A$, $|\mu| = 2000$ GeV, $A = 1000$ GeV, and $m_{H^\pm} = 300$ GeV)

φ	m_{h_1}		m_{h_2}		m_{h_3}		$\Gamma_{h_1\to gg} \times 10^4$		$\Gamma_{h_1\to\gamma\gamma} \times 10^6$	
	our study	[15]	our study	[15]	our study	[15]	our study	[15]	our study	[15]
0	115.4	106.8	295.5	302.2	297.1	302.3	1.378	1.878	7.703	5.796
$\pi/6$	118.7	109.0	289.6	297.8	299.5	304.4	1.529	1.964	8.593	6.287
$\pi/3$	125.9	113.9	279.7	290.9	300.4	305.0	1.907	2.107	10.981	7.605
$\pi/2$	131.4	117.4	269.3	282.2	299.9	304.5	2.220	1.961	13.313	8.996
$2\pi/3$	130.7	114.9	262.2	273.9	298.8	303.5	2.101	1.262	12.953	8.969
$5\pi/6$	125.2	105.7	259.8	268.3	297.6	302.4	1.707	0.503	10.645	7.223
π	122.0	99.4	259.6	264.4	297.1	302.0	1.516	0.263	9.508	6.101

values of the phase φ. In this table, one can also find the results derived on the basis of the aforementioned CPsuperH package [15]. The authors of CPsuperH employed a different method for diagonalizing the mass term [in particular, they did not introduce at all the mixing angle for the CP-even states h and H, with the result that M_{12} and M_{21} did not vanish in the analog of the mass matrices (45)] for the Higgs potential, where the radiation-induced parameters $\lambda_{6,7}$ are absent; also, they took into account two-loop corrections to the Yukawa coupling constants $h_{t,b}$. The results are seen to be qualitatively consistent, but a detailed numerical comparison is complicated by the difference in the approaches used.

3.2. Real-Valued Parameters μ_{12}^2 and $\lambda_{5,6,7}$; $\theta \neq 0$

If the parameters μ_{12}^2 and $\lambda_{5,6,7}$ are real-valued, the effective potential (6) is CP-invariant. It can easily be shown [4, 5, 16] that the phases of the complex-valued parameters μ_{12}^2 and $\lambda_{5,6,7}$ can be removed by means of a $U(1)_Y$ rotation in the hypercharge space under the conditions

$$\mathrm{Im}(\mu_{12}^4\lambda_5^*) = 0, \quad \mathrm{Im}(\mu_{12}^2\lambda_6^*) = 0, \quad \mathrm{Im}(\mu_{12}^2\lambda_7^*) = 0. \tag{47}$$

Since there are no physical grounds that would substantiate these conditions of the fine tuning of the phases, it does not seem reasonable to restrict ourselves to a CP-invariant potential featuring real-valued parameters. If the phase of the vacuum expectation value is nonzero, $\theta \neq 0$, CP invariance is violated spontaneously. A local minimum of the effective potential (6) is attained for $\lambda_5 > 0$ [this is the case of a pure imaginary value of μA—see Eq. (17)], and

$$\cos\theta = \frac{\mu_{12}^2 - \dfrac{v_1^2}{2}\lambda_6 - \dfrac{v_2^2}{2}\lambda_7}{\lambda_5 v_1 v_2}. \tag{48}$$

Upon taking into account the diagonalization condition (35), we obtain

$$\cos\theta = \frac{m_A^2}{\lambda_5 v^2} + 1; \tag{49}$$

that is, there is no minimum for $m_A^2 > 0$. At $\lambda_5 = 0$, the condition in (48) does not have solutions for θ. For $\lambda_5 < 0$, Eq. (48) is a condition of maximum, while the absolute minimum (the lowest value) is attained at the ends of the interval $\cos\theta = \pm 1$. By way of example, we indicate that, with allowance for the diagonalization condition (35), there is no absolute minimum at $\theta = 0$ if

$$m_A^2 > 2|\lambda_5|v^2. \tag{50}$$

Thus, we see that, in the case of real-valued μ_{12}^2 and $\lambda_{5,6,7}$ and a spontaneous violation of CP invariance, it is not possible at all to diagonalize the mass term at a minimum of the effective potential, at least at not small masses m_A.

3.3. Complex-Valued Parameters μ_{12}^2 and $\lambda_{5,6,7}$; $\theta \neq 0$

In the case of complex-valued parameters and a nonzero phase of the vacuum expectation value for the doublet Φ_2 (3), in which case the CP invariance of the potential is violated both explicitly and spontaneously, the condition that the derivative of the potential with

respect to θ vanishes depends on the real and imaginary parts of μ_{12}^2 and $\lambda_{5,6,7}$; that is,

$$\cos\theta(2\mathrm{Im}\mu_{12}^2 - v_1^2\mathrm{Im}\lambda_6 - v_2^2\mathrm{Im}\lambda_7) \qquad (51)$$
$$- v_1v_2\mathrm{Im}\lambda_5\cos(2\theta) + \sin\theta(2\mathrm{Re}\mu_{12}^2 - v_1^2\mathrm{Re}\lambda_6$$
$$- v_2^2\mathrm{Re}\lambda_7) - v_1v_2\mathrm{Re}\lambda_5\sin(2\theta) = 0.$$

In this case, the condition of extremum for $\mathrm{Im}\mu_{12}^2$ depends on the relative phase θ of the vacuum expectation values; for our choice of the phases, the diagonalization condition for $\mathrm{Re}\mu_{12}^2$ additionally depends on the phase ξ of the relative rotation of the doublets [see Eqs. (3) and (4)]. At real-valued μ_{12}^2 and $\lambda_{5,6,7}$ and nonzero θ, relation (51) leads to (48).

The conditions of extremum (the vanishing of the derivatives of the potential with respect to the lower, neutral, real-valued, and imaginary components of the doublet fields) for a general and a particular ($\theta = 0$) case are given in the form of the coefficients of the Higgs potential parameters in Table 3 (conditions of extremum for μ_1^2 and μ_2^2 in the general case and at $\theta = 0$) and in Table 4 (conditions of extremum for $\mathrm{Re}\mu_{12}^2$ at $\theta = 0$ and for $\mathrm{Im}\mu_{12}^2$ in the general case and at $\theta = 0$). Since the additional condition that specifies, in the general case, the choice of the parameter m_A for $\mathrm{Re}\mu_{12}^2$ is rather cumbersome, it is given separately below:

$$\mathrm{Re}\mu_{12}^2 = -\lambda_2\frac{v^2\cos\theta\sin^3(2\beta)\sin^2(\theta+\xi)}{3 + (1-\cos\theta\cos\xi)\left(\cos^4\beta - \frac{3}{2}\sin^2(2\beta)\right) + \sin^4\beta + \cos\theta\cos\xi(1-\sin^4\beta)} \qquad (52)$$
$$+ \mathrm{Re}\lambda_5\frac{v^2(\cos^4\beta\cos^2\xi + \cos^2\theta\sin^4\beta + \cos\beta\cos(\theta-\xi)\sin\beta\sin(2\beta))}{\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta}$$
$$- \mathrm{Im}\lambda_5\frac{v^2(\sin^2(2\beta)\sin(\theta-\xi) + \sin^4\beta(\sin(2\theta) + \tan\theta) + \cos^4\beta(\tan\theta - \sin(2\xi)))}{2(\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta)}$$
$$+ \mathrm{Re}\lambda_6\frac{1}{2}v^2\cos^2\beta + \mathrm{Im}\lambda_6\frac{v^2\cos^3\beta\sin\beta\sin\xi}{\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta}$$
$$+ \mathrm{Re}\lambda_7\left(\frac{v^2\cos^4\beta(4\cos(\theta+2\xi) - 2\cos(2\theta)\sec\theta)\tan\beta}{4(\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta)}\right.$$
$$\left. + \frac{v^2(2\sin^2(2\beta)\cos\xi + 2\sec\theta\sin^4\beta - \cos(2\theta+\xi)\sin^2(2\beta))\tan\beta}{4(\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta)}\right)$$
$$+ \mathrm{Im}\lambda_7\frac{v^2\sin(2\beta)(2\cos^2\beta\cos\xi\sin(\theta+\xi) + \sin^2\beta(2\sin\xi + \sin(2\theta+\xi)))}{2(\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta)}$$
$$- \mathrm{Im}\mu_{12}^2\frac{\sin(2\beta)\sin\xi}{\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\cot\beta}$$
$$+ m_A^2\frac{1}{\cos^2\beta\cot\beta\sec\theta + \cos\xi\sin(2\beta) + \sec\theta\sin^2\beta\tan\beta}.$$

If one sets $\theta = 0$ and $\xi = 0$, there arise expressions coincident with those in the particular case of explicit CP violation [see Eqs. (35) and (42)].

It is noteworthy that, in the general case, the substitution of the conditions of extremum from Tables 3 and 4 into (51) leads to an identity [irrespective of expression (52) for $\mathrm{Re}\mu_{12}^2$].

The requirement that the second derivative of the potential with respect to θ is positive is the condition that an extremum is a minimum:

$$-\sin\theta(2\mathrm{Im}\mu_{12}^2 - v_1^2\mathrm{Im}\lambda_6 - v_2^2\mathrm{Im}\lambda_7) \qquad (53)$$
$$+ 2v_1v_2\mathrm{Im}\lambda_5\sin(2\theta) + \cos\theta(2\mathrm{Re}\mu_{12}^2 - v_1^2\mathrm{Re}\lambda_6$$
$$- v_2^2\mathrm{Re}\lambda_7) - 2v_1v_2\mathrm{Re}\lambda_5\cos(2\theta) > 0.$$

In the general case of $\theta \neq 0$ and $\xi \neq 0$, the diagonalization of the mass term in the effective potential at a local minimum can be performed on the basis of a procedure similar to that which was considered above:

Table 3. Conditions of extremum for μ_1^2 and μ_2^2 in the general case and at $\theta = 0$

	μ_1^2		μ_2^2	
	$\theta \neq 0$	$\theta = 0$	$\theta \neq 0$	$\theta = 0$
λ_1	v_1^2	v_1^2	0	0
λ_2	0	0	v_2^2	v_2^2
λ_3	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_1^2$	$\frac{1}{2}v_1^2$
λ_4	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_1^2$	$\frac{1}{2}v_1^2$
$\mathrm{Re}\lambda_5$	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_2^2$	$\frac{1}{2}v_1^2$	$\frac{1}{2}v_1^2$
$\mathrm{Im}\lambda_5$	$-\frac{1}{2}v_2^2\tan\theta$	0	$-\frac{1}{2}v_1^2\tan\theta$	0
$\mathrm{Re}\lambda_6$	$\frac{1}{2}v_1v_2(2+\cos(2\theta))\sec\theta$	$\frac{3}{2}v_1v_2$	$\frac{1}{2}v_1^2\sec\theta\cot\beta$	$\frac{1}{2}v_1^2\cot\beta$
$\mathrm{Im}\lambda_6$	$-v_1v_2\sin\theta$	0	0	0
$\mathrm{Re}\lambda_7$	$\frac{1}{2}v_2^2\sec\theta\tan\beta$	$\frac{1}{2}v_2^2\tan\beta$	$\frac{1}{2}v_1v_2(2+\cos(2\theta))\sec\theta$	$\frac{3}{2}v_1v_2$
$\mathrm{Im}\lambda_7$	0	0	$-v_1v_2\sin\theta$	0
$\mathrm{Re}\mu_{12}^2$	$-\tan\beta\sec\theta$	$-\tan\beta$	$-\cot\beta\sec\theta$	$-\cot\beta$

(i) We define $\widetilde{h}$, $\widetilde{H}$, $\widetilde{A}$, and $\widetilde{G}^0$ as four linear combinations of the independent fields η_1, η_2, χ_1, and χ_2 in the system specified by Eqs. (2) and (3), the zero row and the zero column in the symmetric 4×4 matrix of the masses squared corresponding to the Goldstone field $\widetilde{G}^0$. In other words, the Goldstone mode is a linear combination that is orthogonal to the plane spanned by those directions in the space of complex fields that are parallel to the vacuum expectation values v_1 and $v_2\exp\{i(\xi+\zeta)\}$. The 4×4 mass matrix of scalar fields then contains a 3×3 minor, while its remaining matrix elements vanish upon taking into account the conditions of extremum from Tables 3 and 4.

(ii) We specify an orthogonal transformation for the respective 3×3 matrix by choosing the rotation angle $\widetilde{\alpha}$ in the $\widetilde{h}$–$\widetilde{H}$ sector in such a way as to annihilate the crossed term $\widetilde{h}\widetilde{H}$. Concurrently, one can readily trace the limiting transitions from the fields and variables for the general case of nonzero phases ξ and θ (tilde-labeled fields and variables) to the fields and variables for the particular case of $\xi=\theta=0$ (non-tilde-labeled fields and variables), as well as to the limit of CP conservation in the basis of the mass eigenstates h, H, and A. We will consider that the effective-potential parameters are real or complex-valued in diagonalizing the matrix of the masses squared for a local minimum of the effective potential.

Thus, we find for the CP states of the Higgs fields in the case of $\xi=0$ and $\theta\neq0$ that

$$\widetilde{h} = -\eta_1\sin\widetilde{\alpha} + (\chi_2\sin\theta + \eta_2\cos\theta)\cos\widetilde{\alpha}, \quad (54)$$

$$\widetilde{H} = \eta_1\cos\widetilde{\alpha} + (\chi_2\sin\theta + \eta_2\cos\theta)\sin\widetilde{\alpha}, \quad (55)$$

$$\widetilde{A} = -\chi_1\sin\beta + (\chi_2\cos\theta - \eta_2\sin\theta)\cos\beta, \quad (56)$$

$$\widetilde{G}^0 = \chi_1\cos\beta + (\chi_2\cos\theta - \chi_2\sin\theta)\sin\beta. \quad (57)$$

A direct substitution of these fields into the potential (6) results in that, in the symmetric 4×4 matrix of the masses squared, the row and the column corresponding to the Goldstone mode have zero elements; that is, the 4×4 mass matrix degenerates into a symmetric 3×3 matrix upon taking into account the conditions of extremum from Tables 3 and 4. At a local minimum, the matrix elements $\widetilde{M}_{13}$ and $\widetilde{M}_{23}$,

Table 4. Conditions of extremum for $\mathrm{Re}\mu_{12}^2$ at $\theta = 0$ and for $\mathrm{Im}\mu_{12}^2$ in the general case and at $\theta = 0$

	$\mathrm{Re}\mu_{12}^2$	$\mathrm{Im}\mu_{12}^2$	
	$\theta = 0$ and $\xi = 0$	$\theta \neq 0$	$\theta = 0$
λ_1	0	0	0
λ_2	0	0	0
λ_3	0	0	0
λ_4	0	0	0
$\mathrm{Re}\lambda_5$	$v_1 v_2$	$v_1 v_2 \sin\theta$	0
$\mathrm{Im}\lambda_5$	0	$\frac{1}{2} v_1 v_2 \cos(2\theta)\sec\theta$	$\frac{1}{2} v_1 v_2$
$\mathrm{Re}\lambda_6$	$\frac{1}{2} v_1^2$	$\frac{1}{2} v_1^2 \tan\theta$	0
$\mathrm{Im}\lambda_6$	0	$\frac{1}{2} v_1^2$	$\frac{1}{2} v_1^2$
$\mathrm{Re}\lambda_7$	$\frac{1}{2} v_2^2$	$\frac{1}{2} v_2^2 \tan\theta$	0
$\mathrm{Im}\lambda_7$	0	$\frac{1}{2} v_2^2$	$\frac{1}{2} v_2^2$
m_A^2	$\sin\beta\cos\beta$	0	0
$\mathrm{Re}\mu_{12}^2$	–	$-\tan\theta$	0

which correspond to the off-diagonal terms $\widetilde{h}\widetilde{A}$ and $\widetilde{H}\widetilde{A}$, have the form

$$\begin{aligned}\widetilde{c}_1 = -\frac{v^2}{2}(&\cos(\widetilde{\alpha}+\beta)\cos(2\theta)\mathrm{Im}\lambda_5 \\ &- 2\sin\widetilde{\alpha}\cos\beta\cos\theta\mathrm{Im}\lambda_6 \\ &+ 2\cos\widetilde{\alpha}\sin\beta\cos\theta\mathrm{Im}\lambda_7 \\ &- \cos(\widetilde{\alpha}+\beta)\sin(2\theta)\mathrm{Re}\lambda_5 \\ &- 2\sin\widetilde{\alpha}\cos\beta\sin\theta\mathrm{Re}\lambda_6 \\ &+ 2\cos\widetilde{\alpha}\sin\beta\sin\theta\mathrm{Re}\lambda_7),\end{aligned} \tag{58}$$

$$\begin{aligned}\widetilde{c}_2 = -\frac{v^2}{2}(&\sin(\widetilde{\alpha}+\beta)\cos(2\theta)\mathrm{Im}\lambda_5 \\ &- 2\cos\widetilde{\alpha}\cos\beta\cos\theta\mathrm{Im}\lambda_6 \\ &+ 2\sin\widetilde{\alpha}\sin\beta\cos\theta\mathrm{Im}\lambda_7 \\ &+ \cos(\widetilde{\alpha}+\beta)\sin(2\theta)\mathrm{Re}\lambda_5 \\ &- 2\cos\widetilde{\alpha}\cos\beta\sin\theta\mathrm{Re}\lambda_6 \\ &+ 2\sin\widetilde{\alpha}\sin\beta\sin\theta\mathrm{Re}\lambda_7).\end{aligned} \tag{59}$$

In the case of $\theta = 0$, they coincide with (43).

The case of $\xi \neq 0$ and $\theta \neq 0$ is analogous to that considered above, this being associated with arbitrariness in choosing the phase ξ of the relative rotation of the scalar-field doublets. Field mass eigenstates are obtained by means of the substitution $\theta \to \theta - \xi$:

$$\begin{aligned}\widetilde{h} = -\eta_1\sin\widetilde{\alpha} + (&\chi_2\sin(\theta-\xi) \\ &+ \eta_2\cos(\theta-\xi))\cos\widetilde{\alpha},\end{aligned} \tag{60}$$

$$\begin{aligned}\widetilde{H} = \eta_1\cos\widetilde{\alpha} + (&\chi_2\sin(\theta-\xi) \\ &+ \eta_2\cos(\theta-\xi))\sin\widetilde{\alpha},\end{aligned} \tag{61}$$

$$\begin{aligned}\widetilde{A} = -\chi_1\sin\beta + (&\chi_2\cos(\theta-\xi) \\ &- \eta_2\sin(\theta-\xi))\cos\beta,\end{aligned} \tag{62}$$

$$\begin{aligned}\widetilde{G}^0 = \chi_1\cos\beta + (&\chi_2\cos(\theta-\xi) \\ &- \chi_2\sin(\theta-\xi))\sin\beta.\end{aligned} \tag{63}$$

4. CONCLUDING REMARKS

In conclusion, we emphasize once again that the potential of the general two-doublet model is not CP-invariant and that the parameters μ_{12}^2 and $\lambda_{5,6,7}$ of the two-doublet effective potential in the Higgs sector of the minimal supersymmetric model must be taken to be complex-valued. The choice of purely real parameters of the general two-doublet model is not quite compelling because this implies that the conditions in (47) are additionally imposed, but there is no physical validation of these conditions. Within the minimal supersymmetric model, complex-valued parameters of the effective Higgs potential arise in quite a natural way if one assumes that, in the sector of scalar quarks, there is mixing similar to the Cabibbo–Kobayashi–Maskawa mixing for the three generations of Standard Model quarks. If this mixing leads to a strong CP violation (for a discussion on a weak CP violation, see [17]) and if the scalar sector of the minimal supersymmetric model features a fairly strong coupling (that is, there arise large imaginary parts of the parameters μ_{12}^2 and $\lambda_{5,6,7}$), the deviations of observable effects in a model featuring CP violation from the respective Standard Model predictions for the production of Higgs bosons at new-generation colliders may be so large that they would require revising experimental priorities [18] for the possibilities of observing Higgs bosons in known channels such as $\gamma\gamma$, ttH, and bbH.

The phase ζ of the relative rotation of the scalar doublets and the phase ξ of the relative rotation of the vacuum expectation values [see (4)] can be constrained by the conditions of diagonalization of the mass term at a minimum [for example, the condition in (52)], and this would make it possible to establish nontrivial relations between ζ and ξ, on one hand, and the variables of the parameter space of the minimal supersymmetric model. However, we cannot see a

direct relationship between the phase of the matrix elements of the Cabibbo–Kobayashi–Maskawa matrix and the phases ζ and ξ, which characterize CP violation in the two-doublet potential. On the basis of the notation used above in the Introduction, one can establish that the Lagrangian for the Yukawa interaction of quarks in the type II THDM (the masses of the fermions belonging to the upper and lower sectors are generated to the different doublets Φ_1 and Φ_2) is given by an expression of the form (1). Upon going over to the basis formed by the mass eigenstates of the quark fields, a unitary mixing matrix V_{u^i,d^j} appears in the terms representing the interaction with a charged Higgs boson:

$$\frac{M_d \tan\beta}{\sqrt{2}v}\overline{u}^i_L V_{u^i,d^j} d^j_R H^+ + \frac{M_u}{\sqrt{2}v\tan\beta}\overline{d}^i_L V^\dagger_{u^i,d^j} u^j_R H^-. \quad (64)$$

If, in the mixing-matrix elements, one singles out a common phase, $V_{u^i,d^j} \to e^{i\varphi}\left|V_{u^i,d^j}\right|$ and $V^\dagger_{u^i,d^j} \to e^{-i\varphi}\left|V_{u^i,d^j}\right|$, then the terms of the Yukawa interaction will take the form

$$\frac{M_d \tan\beta}{\sqrt{2}v}\overline{u}^i_L e^{i\varphi}\left|V_{u^i,d^j}\right| d^j_R H^+ + \frac{M_u}{\sqrt{2}v\tan\beta}\overline{d}^i_L e^{-i\varphi}\left|V_{u^i,d^j}\right| u^j_R H^-; \quad (65)$$

that is, the common phase φ can be identified with the phase ξ of the relative rotation of the doublets. However, a structure of this type does not correspond to mixing in the sector of charged weak currents, because the Cabibbo–Kobayashi–Maskawa matrix, which determines this mixing, does not feature a universal complex factor.

ACKNOWLEDGMENTS

The work of M.N. Dubinin and M.V. Dolgopolov was supported by the Russian Foundation for Basic Research (project no. 04-02-17448). The work of M.N. Dubinin was also supported by INTAS (grant no. 03-51-4007). É.N. Akhmetzyanova gratefully acknowledges the financial support of the Dynasty Foundation and the International Center for Fundamental Physics in Moscow.

REFERENCES

1. M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
2. H. Georgi, Hadronic J. **1**, 155 (1978).
3. T. D. Lee, Phys. Rev. D **8**, 1226 (1973).
4. A. Pilaftsis and C. E. M. Wagner, Nucl. Phys. B **553**, 3 (1999).
5. M. N. Dubinin and A. V. Semenov, Eur. Phys. J. C **28**, 223 (2003).
6. J. Liu and L. Wolfenstein, Nucl. Phys. B **289**, 1 (1987); Y. L. Wu and L. Wolfenstein, Phys. Rev. Lett. **73**, 1762 (1994); **73**, 2809 (1994).
7. I. Ginzburg and M. Krawczyk, hep-ph/0408011.
8. H. E. Haber and R. Hempfling, Phys. Rev. D **48**, 4280 (1993).
9. K. Inoue, A. Kakuto, H. Komatsu, and S. Takeshita, Prog. Theor. Phys. **67**, 1889 (1982); R. A. Flores and M. Sher, Ann. Phys. (N. Y.) **148**, 95 (1983).
10. M. Carena, J. R. Ellis, A. Pilaftsis, and C. E. M. Wagner, Nucl. Phys. B **586**, 92 (2000).
11. Y. Okada, M. Yamaguchi, and T. Yanagida, Phys. Lett. B **262**, 54 (1991); J. Ellis, G. Ridolfi, and F. Zwirner, Phys. Lett. B **257**, 83 (1991); H. E. Haber and R. Hempfling, Phys. Rev. Lett. **66**, 1815 (1991); R. Barbieri, M. Frigeni, and F. Caravaglios, Phys. Lett. B **258**, 167 (1991).
12. E. Akhmetzyanova, M. Dolgopolov, and M. Dubinin, in *Proceedings of the International Workshop "Supersymmetries and Quantum Symmetries—SQS'03"* (Dubna, 2004), p. 265; *Proceedings of the XVII International Workshop on High-Energy Physics and Quantum Field Theory (QFTHEP 2003), Samara–Saratov, 2003* (Moscow, 2004), p. 273.
13. M. Carena, J. R. Espinosa, M. Quiros, and C. E. M. Wagner, Phys. Lett. B **355**, 209 (1995).
14. M. Quiros, in *Perspectives on Higgs Physics II*, Ed. by G. L. Kane (World Sci., Singapore, 1998); hep-ph/9703412.
15. J. S. Lee, A. Pilaftsis, M. Carena, *et al.*, Comput. Phys. Commun. **156**, 283 (2004).
16. J. F. Gunion and H. E. Haber, Phys. Rev. D **67**, 075 019 (2003).
17. I. F. Ginzburg and M. V. Vychugin, in *Proceedings of the XVI International Workshop on High-Energy Physics and Quantum Field Theory (QFTHEP 2001)* (Moscow, 2002), p. 64; hep-ph/0201117.
18. S. Abdullin *et al.*, CMS Note 2003/033.

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1866–1872. Translated from Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1928–1934.

ELEMENTARY PARTICLES AND FIELDS
Theory

Color Flows for the Process $gg \to B_c + c + \bar{b}$

A. V. Berezhnoy*

Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119899 Russia

Received July 13, 2004; in final form, February 1, 2005

Abstract—Contributions of various color flows to the cross section for the gluonic production of B_c mesons are calculated. This is essential for simulating events involving B_c mesons with the aid of the PYTHIA package because the method used in PYTHIA to perform the hadronization of final partons and hadron remnants depends on the type of color flow. A modified method of partition into color flows is proposed.
© *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Particle-production processes in present-day hadronic experiments can hardly be performed without employing Monte Carlo methods. This is because integro-differential equations that describe the evolution of initial partons and the hadronization of parton-interaction products are very complicated. At the present time, PYTHIA [1] is considered to be the most reliable package for simulating particle-production processes at high energies by Monte Carlo methods. This package describes all three stages into which one can break down, by convention, the production process—these are the evolution of initial partons, the hard subprocess of initial-parton interaction, and the hadronization of hard-interaction products and the hadron remnant. The hard subprocess is considered within perturbative QCD, and the subsequent hadronization is described on the basis of the color-string model [2], where it is assumed that the color charges of hard-interaction products are connected by color-field strings, which are eventually ruptured, producing color-singlet final states. It should be noted that strings between the color triplet and antitriplet are considered in this model. The color octet (for example, a gluon) is treated approximately either as a weakly bound triplet–antitriplet state or as an excitation of the string stretched between the triplet and the antitriplet. The technique that makes it possible to break down the matrix element into components corresponding to specific color flows was given in [2], and it was shown there that, in the limit of an infinite number of colors ($N_c \to \infty$), the interference between color flows becomes negligible.

A rather wide set of hard processes can be simulated within the PYTHIA package; however, the entire variety of processes that can be studied at high interaction energies are not exhausted by this set. In view of this, recent versions of the PYTHIA package provide the possibility of including an appropriate matrix element that corresponds to the hard part of the process. In this case, there arises the problem of breaking down the matrix element into color flows such that a specific way of stretching strings between final partons and hadron remnants corresponds to each color flow in the PYTHIA package.

In the present study, we discuss color flows for the production of B_c mesons in gluon interaction and give methods that can be used to perform a partition into flows and which differ slightly from standard methods.

It should be noted that it is rather difficult to calculate the cross section for the gluonic production of B_c mesons, since B_c-meson production does not proceed via hard $b\bar{b}$-pair production followed by the hadronization of the b quark into a B_c meson. Therefore, the respective production process cannot be described in terms of the $b \to B_c$ fragmentation function (this fragmentation function was calculated in [3]). Under conditions of planned experiments and those that are currently operating, B_c mesons are predominantly produced via the recombination mechanism. In order to estimate the contribution of the recombination mechanism, it is necessary to calculate 36 tree Feynman diagrams of order α_s^4. These calculations were performed independently by several research groups [4–9]. The results obtained in [4–7] are in good agreement with one another.

The approaching commissioning of LHC has rekindled interest in the problem of studying B_c-meson production in hadron interactions. For example, the latest versions of the SIMUB package [10], which is intended for a complete simulation of processes involving heavy quarks under LHC conditions, make it possible to generate B_c-meson events. The authors of the SIMUB package employed

*e-mail: aber@ttk.ru

the codes reported in [4, 5]. Close scientific contacts between the present author and the developers of SIMUB have given impetus to this study.

2. CALCULATION OF COLOR FLOWS AT THE LEVEL OF AMPLITUDES

The Feynman diagrams describing the process $gg \to B_c + X$ are given in Fig. 1. Since the diagram involving the four-gluon vertex contains three different color structures, we will treat it as three different diagrams. The color components of the diagrams can be represented as

$$T_1 = f^{n_1 g_2 n_2} f^{n_3 n_2 g_1} t^{n_1}_{c\bar{c}} t^{n_3}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_2 = f^{n_1 g_2 n_2} f^{n_3 n_2 g_1} t^{n_3}_{c\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_3 = f^{n_1 g_1 g_2} f^{n_2 n_1 n_3} t^{n_2}_{b\bar{b}} t^{n_3}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_4 = f^{n_1 g_2 n_2} f^{n_2 n_3 g_1} t^{n_3}_{b\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_5 = f^{g_2 n_1 n_2} f^{n_2 n_3 g_1} t^{n_1}_{b\bar{b}} t^{n_3}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_6 = f^{n_1 n_2 n_3} f^{n_3 g_2 g_1} t^{n_1}_{b\bar{b}} t^{n_2}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_7 = i f^{n_1 n_2 g_2} t^{n_1}_{c\bar{c}} t^{g_1}_{bl_1} t^{n_2}_{l_1\bar{b}} \delta_{b\bar{c}},$$

$$T_8 = i f^{n_1 n_2 g_2} t^{n_1}_{c\bar{c}} t^{n_2}_{bl_1} t^{g_1}_{l_1\bar{b}} \delta_{b\bar{c}},$$

$$T_9 = i f^{n_1 n_2 g_2} t^{n_1}_{b\bar{b}} t^{g_1}_{cl_1} t^{n_2}_{l_1\bar{c}} \delta_{b\bar{c}},$$

$$T_{10} = i f^{n_1 n_2 g_2} t^{n_1}_{b\bar{b}} t^{n_2}_{cl_1} t^{g_1}_{l_1\bar{c}} \delta_{b\bar{c}},$$

$$T_{11} = i f^{n_1 n_2 g_1} t^{n_1}_{b\bar{b}} t^{n_2}_{cl_1} t^{g_2}_{l_1\bar{c}} \delta_{b\bar{c}},$$

$$T_{12} = i f^{n_1 n_2 g_1} t^{n_1}_{b\bar{b}} t^{g_2}_{cl_1} t^{n_2}_{l_1\bar{c}} \delta_{b\bar{c}},$$

$$T_{13} = i f^{n_1 n_2 g_1} t^{n_1}_{c\bar{c}} t^{n_2}_{bl_1} t^{g_2}_{l_1\bar{b}} \delta_{b\bar{c}},$$

$$T_{14} = i f^{n_1 n_2 g_1} t^{n_1}_{c\bar{c}} t^{g_2}_{bl_1} t^{n_2}_{l_1\bar{b}} \delta_{b\bar{c}},$$

$$T_{15} = t^{g_1}_{bl_1} t^{n_1}_{l_1\bar{b}} t^{n_1}_{cl_2} t^{g_2}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{16} = t^{n_1}_{bl_1} t^{g_1}_{l_1\bar{b}} t^{n_1}_{cl_2} t^{g_2}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{17} = t^{g_1}_{bl_1} t^{n_1}_{l_1\bar{b}} t^{g_2}_{cl_2} t^{n_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{18} = t^{g_1}_{bl_1} t^{n_1}_{l_1\bar{b}} t^{g_2}_{cl_2} t^{n_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{19} = t^{n_1}_{bl_1} t^{g_2}_{l_1\bar{b}} t^{g_1}_{cl_2} t^{n_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{20} = t^{n_1}_{bl_1} t^{g_2}_{l_1\bar{b}} t^{n_1}_{cl_2} t^{g_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{21} = t^{n_1}_{bl_1} t^{g_2}_{l_1\bar{b}} t^{n_1}_{cl_2} t^{g_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{22} = t^{g_2}_{bl_1} t^{n_1}_{l_1\bar{b}} t^{n_1}_{cl_2} t^{g_1}_{l_2\bar{c}} \delta_{b\bar{c}},$$

$$T_{23} = t^{n_1}_{bl_1} t^{g_1}_{l_1l_2} t^{g_2}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{24} = t^{g_1}_{bl_1} t^{n_1}_{l_1l_2} t^{g_2}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{25} = t^{g_1}_{bl_1} t^{g_2}_{l_1l_2} t^{n_1}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{26} = t^{n_1}_{bl_1} t^{g_2}_{l_1l_2} t^{g_1}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{27} = t^{g_2}_{bl_1} t^{n_1}_{l_1l_2} t^{g_1}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{28} = t^{g_2}_{bl_1} t^{g_1}_{l_1l_2} t^{n_1}_{l_2\bar{b}} t^{n_1}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{29} = t^{n_1}_{cl_1} t^{g_1}_{l_1l_2} t^{g_2}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{30} = t^{g_1}_{cl_1} t^{n_1}_{l_1l_2} t^{g_2}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{31} = t^{g_1}_{cl_1} t^{g_2}_{l_1l_2} t^{n_1}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{32} = t^{n_1}_{cl_1} t^{g_2}_{l_1l_2} t^{g_1}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{33} = t^{g_2}_{cl_1} t^{n_1}_{l_1l_2} t^{g_1}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{34} = t^{g_2}_{cl_1} t^{g_1}_{l_1l_2} t^{n_1}_{l_2\bar{c}} t^{n_1}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{35} = i f^{n_1 g_1 g_2} t^{n_2}_{bl_1} t^{n_1}_{l_1\bar{b}} t^{n_2}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{36} = i f^{n_1 g_1 g_2} t^{n_1}_{bl_1} t^{n_2}_{l_1\bar{b}} t^{n_2}_{c\bar{c}} \delta_{b\bar{c}},$$

$$T_{37} = i f^{n_1 g_1 g_2} t^{n_2}_{cl_1} t^{n_1}_{l_1\bar{c}} t^{n_2}_{b\bar{b}} \delta_{b\bar{c}},$$

$$T_{38} = i f^{n_1 g_1 g_2} t^{n_1}_{cl_1} t^{n_2}_{l_1\bar{c}} t^{n_2}_{b\bar{b}} \delta_{b\bar{c}},$$

where the superscripts g_1 and g_2 indicate the color states of initial gluons; the subscripts b, $\bar{b}$, c, and $\bar{c}$ stand for the color states of the b, $\bar{b}$, c, and $\bar{c}$ quarks, respectively; and $\delta_{b\bar{c}}$ is the color component of the B_c-meson wave function (the normalization factor $1/\sqrt{3}$ is omitted to avoid encumbering the presentation).

By way of example, we now consider the color part of diagram *1* in Fig. 1 (in this diagram, the initial gluons exchange a gluon in the t channel, whereupon they split into quark–antiquark pairs). We have

$$T_1 = f^{n_1 g_2 n_2} f^{n_3 n_2 g_1} t^{n_1}_{c\bar{c}} t^{n_3}_{b\bar{b}} \delta_{b\bar{c}} = f^{n_1 g_2 n_2} f^{n_3 n_2 g_1} (t^{n_1} t^{n_3})_{\bar{b}c}.$$

Using the equality $t^a t^b - t^b t^a = i f^{abc} t^c$, we arrive at

$$f^{n_1 g_2 n_2} f^{n_3 n_2 g_1} t^{n_1} t^{n_3} = -(t^{g_2} t^{n_2} - t^{n_2} t^{g_2})$$

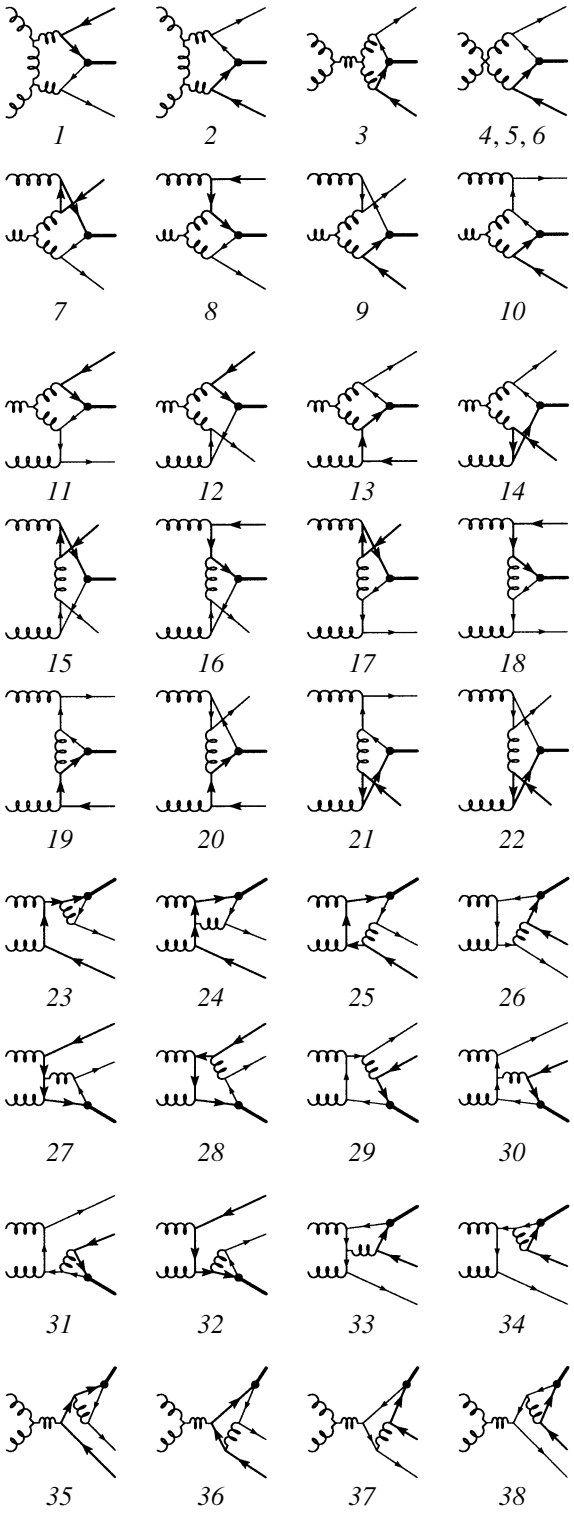

Fig. 1. Feynman diagrams for the gluonic production of B_c mesons.

$$\times (t^{n_2}t^{g_1} - t^{g_1}t^{n_1}) = -t^{g_2}t^{n_2}t^{n_2}t^{g_1} + t^{n_2}t^{g_2}t^{n_2}t^{g_1}$$
$$+ t^{g_2}t^{n_2}t^{g_1}t^{n_1} - t^{n_2}t^{g_2}t^{g_1}t^{n_1} = -\frac{4}{3}t^{g_2}t^{g_1}$$

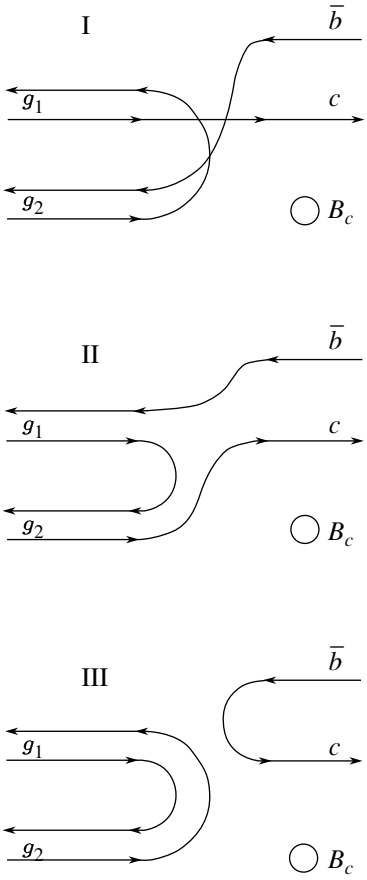

Fig. 2. Color flows for the process $gg \to B_c + c + \bar{b}$. The color strings are stretched as follows: c quark $\longrightarrow$ remnant of the hadron that contained the gluon g_1 $\longrightarrow$ remnant of the hadron that contained the gluon g_2 $\longrightarrow$ $\bar{b}$ quark for flow I; c quark $\longrightarrow$ remnant of the hadron that contained the gluon g_2 $\longrightarrow$ remnant of the hadron that contained the gluon g_1 $\longrightarrow$ $\bar{b}$ quark for flow II; and c quark $\longrightarrow$ $\bar{b}$ quark and remnant of the hadron that contained the gluon $g_1 \rightleftarrows$ remnant of the hadron that contained the gluon g_2 for flow III.

$$-\frac{1}{6}t^{g_2}t^{g_1} - \frac{1}{6}t^{g_2}t^{g_1} - \left(\frac{1}{4}\delta^{g_1g_2} - \frac{1}{6}t^{g_2}t^{g_1}\right)$$
$$= -\frac{1}{4}\delta^{g_1g_2} - \frac{3}{2}t^{g_2}t^{g_1}.$$

Finally, we have

$$T_1 = -\frac{3}{2}t^{g_2}_{ck}t^{g_1}_{k\bar{b}} - \frac{1}{4}\delta^{g_1g_2}\delta_{c\bar{b}}.$$

The first term (see scheme II in Fig. 2) corresponds to the color flow where the color of the second gluon (g_2) flows to the c quark, the anticolor of the first gluon (g_1) flows to the $\bar{b}$ quark, and the color of the first gluon (g_2) and the anticolor of the second gluon undergo annihilaton (sum over k). The second term corresponds to the color flow where the color and anticolor of the initial gluons are annihilated, while

the color and anticolor of the c and $\bar{b}$ quarks are produced from a vacuum (see scheme III in Fig. 2). In the other diagrams being considered, there is also a contribution of the form $t^{g_1}_{ck} t^{g_2}_{k\bar{b}}$. It corresponds to the case where the color of the first gluon (g_1) flows to the c quark, the anticolor of the second gluon (g_2) flows to the $\bar{b}$ quark, and the anticolor of the first gluon and the color of the second gluon undergo annihilation (see scheme I in Fig. 2). There are no color flows in the process in question other than those mentioned above. This color-flow partition corresponds to the recipe given in [2].

However, it would be desirable to rely on more fundamental QCD principles. For example, the term $\delta^{g_1 g_2}\delta_{c\bar{b}}$ corresponds to the case where the c and $\bar{b}$ quarks are created in a singlet state; therefore, the string is naturally stretched between them. We note that the term $t^{g_2}_{ck} t^{g_1}_{k\bar{b}}$ also features a singlet contribution:

$$t^a t^b = \frac{1}{6}\delta^{ab} + \frac{1}{2}\Big(d^{abc} + if^{abc}\Big)t^c. \qquad (1)$$

It turns out that one part of the singlet is hadronized according to one law, while its other part is hadronized according to a different law. Therefore, it may be more correct to treat the whole-color singlet contribution as a individual color flow. In this case, the remaining two flows are composed from two octet states d and f.

Thus, we define three flows as follows:

(i) The color of the first gluon flows to the c quark, the anticolor of the second gluon flows to the $\bar{b}$ quark, and the anticolor of the first gluon and the color of the second gluon undergo annihilation:

$$\frac{1}{2}\big(d^{g_1 g_2 k} + if^{g_1 g_2 k}\big)t^k_{c\bar{b}}.$$

(ii) The color of the second gluon flows to the c quark, the anticolor of the first gluon flows to the $\bar{b}$ quark, and the color of the first gluon and the anticolor of the second gluon undergo annihilation:

$$\frac{1}{2}\big(d^{g_1 g_2 k} - if^{g_1 g_2 k}\big)t^k_{c\bar{b}}.$$

(iii) The color and the anticolor of the initial gluons are annihilated, while the color and the anticolor of the c and $\bar{b}$ quarks are produced from a vacuum:

$$\delta^{g_1 g_2}\delta_{c\bar{b}}.$$

In our opinion, this partition is more physically justified because the singlet-state contribution is treated as an individual flow that does not interfere with the other flows.

The partition method described here differs slightly from the standard method where $t^{g_1}_{ck} t^{g_2}_{k\bar{b}}$, $t^{g_2}_{ck} t^{g_1}_{k\bar{b}}$, and $\delta^{g_1 g_2}\delta_{c\bar{b}}$ are taken to be color flows, but where the partition is performed to terms of order $1/N_c$ inclusive, N_c being the number of colors. Formula (1) then assumes the form

$$t^a t^b = \frac{1}{2N_c}\delta^{ab} + \frac{1}{2}\Big(d^{abc} + if^{abc}\Big)t^c, \qquad (2)$$

whence it follows that, in the limit $N_c \to \infty$, the two partition methods lead to the same result.

Vectors A_n, B_n, and C_n (see schemes I, II, and III in Fig. 2) characterizing the partition of the matrix element in flows

n	A_n	B_n	C_n	n	A_n	B_n	C_n
1	0	−3/2	−1/2	20	−1/6	0	−1/36
2	−3/2	0	−1/2	21	−1/6	0	−1/36
3	3/2	−3/2	0	22	−1/6	0	2/9
4	0	3/2	1/2	23	4/3	0	2/9
5	−3/2	0	−1/2	24	−1/6	0	1/36
6	3/2	−3/2	0	25	−1/6	0	2/9
7	0	0	1/4	26	0	4/3	2/9
8	0	−3/2	−1/4	27	0	−1/6	−1/36
9	3/2	0	1/4	28	0	−1/6	2/9
10	0	0	−1/4	29	−1/6	0	2/9
11	0	0	−1/4	30	−1/6	0	−1/36
12	0	3/2	1/4	31	4/3	0	2/9
13	−3/2	0	−1/4	32	0	−1/6	2/9
14	0	0	1/4	33	0	−1/6	−1/36
15	0	−1/6	2/9	34	0	4/3	2/9
16	0	−1/6	−1/36	35	4/3	−4/3	0
17	0	−1/6	−1/36	36	−1/6	1/6	0
18	0	4/3	2/9	37	−1/6	1/6	0
19	4/3	0	2/9	38	4/3	−4/3	0

Thus, the color component of the matrix element (n) can be represented in the form

$$T_n = \frac{1}{2}\Big(d^{g_1 g_2 k} + if^{g_1 g_2 k}\Big)t^k_{c\bar{b}} \cdot A_n \qquad (3)$$
$$+ \frac{1}{2}\Big(d^{g_1 g_2 k} - if^{g_1 g_2 k}\Big)t^k_{c\bar{b}} \cdot B_n + \delta^{g_1 g_2}\delta_{c\bar{b}} \cdot C_n.$$

Upon averaging over the color states of initial particles and summation over the color states of final particles, the color matrix takes the form

$$M_{mn} = \frac{1}{64}\Big((D+F)\cdot A_m A_n + (D+F)\cdot B_m B_n \qquad (4)$$

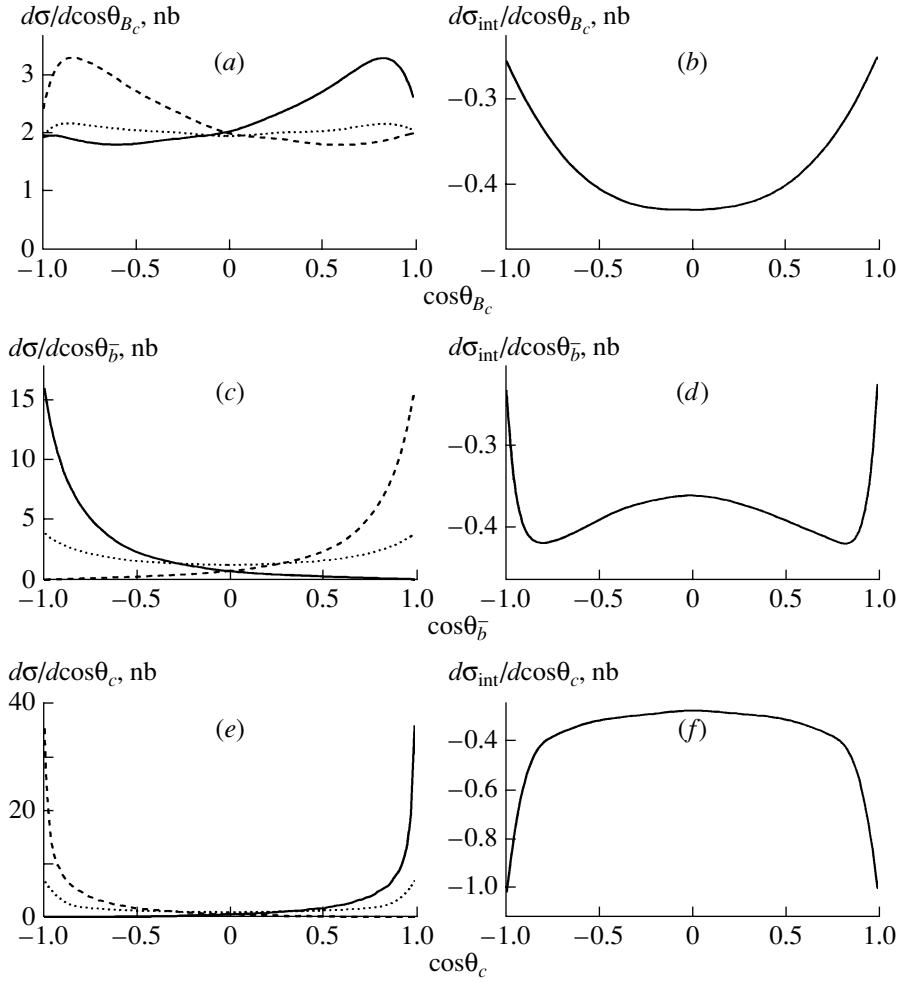

Fig. 3. Differential distributions of the cross sections with respect to the cosine of the emission angles of final particles for various color flows and analogous distribution of the interference term at the gluon-interaction energy of 25 GeV for the gluonic production of a pseudoscalar B_c meson ($gg \to B_c + \bar{b} + c$): (a, b) angular distributions for the B_c meson, (c, d) angular distributions for the $\bar{b}$ quark, and (e, f) angular distributions for the c quark. In Figs. 3a, 3c, and 3e, the solid, dashed, and dotted curves correspond to flows I, II, and III, respectively. The interference contribution is shown in Figs. 3b, 3d, and 3f.

$$+ (D-F)\cdot(A_mB_n + B_mA_n) + S\cdot C_mC_n\Big),$$

where $D = 5/3$, $F = 3$, and $S = 24$. This matrix is determined by three vectors A_n, B_n, and C_n, which are given in the table.

In our approach, the color matrix corresponding to the interference between the color flows, $(d^{g_1g_2k} + if^{g_1g_2k})t^k_{c\bar{b}}$ and $(d^{g_1g_2k} - if^{g_1g_2k})t^k_{c\bar{b}}$, has a simple form,

$$M^{\rm int}_{mn} = \frac{1}{64}\Big((D-F)\cdot(A_mB_n + B_mA_n)\Big). \quad (5)$$

It should be noted that formula (3) can easily be rewritten in terms of standard color flows:

$$T_n = t^{g_1}_{ck}t^{g_2}_{k\bar{b}}\cdot A_n + t^{g_2}_{ck}t^{g_1}_{k\bar{b}}\cdot B_n + \delta^{g_1g_2}\delta_{c\bar{b}}\cdot\Big(C_n - \frac{A_n+B_n}{6}\Big). \quad (6)$$

From a comparison of Eqs. (3) and (6), one can see that the expressions for flow I differ only in the value of the common factor. It can readily be shown that, for flow I, the squares of the matrix elements in these two partition schemes are in the ratio 7 : 8, the same being valid for flow II.

3. RESULTS OF THE CALCULATIONS

Figures 3 and 4 present the differential distributions of the cross sections with respect to the cosine of the emission angle of final particles (B_c, $\bar{b}$, and c) for various color flows and the analogous distributions of the interference term at the gluon-interaction energy of 25 GeV, which is characteristic of B_c-meson production at LHC. We note that our intuitive ideas of color flows are in accord with the results of accurate calculations. From the distributions presented

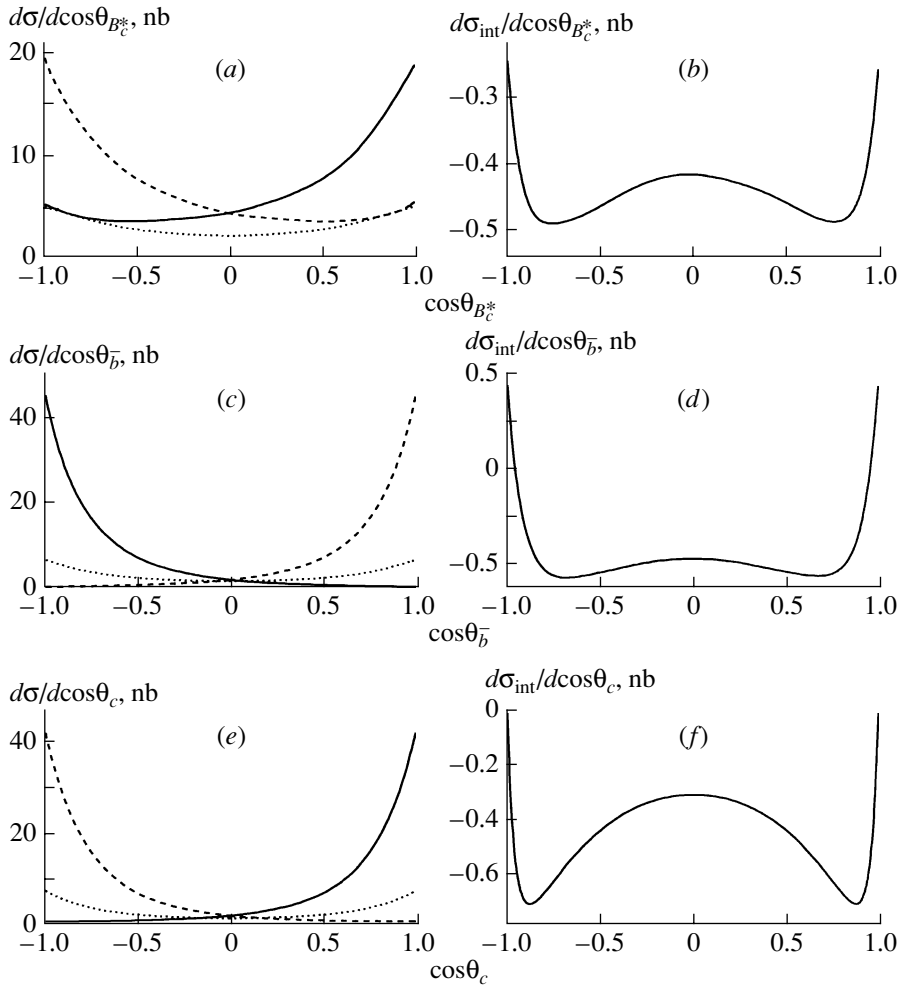

Fig. 4. As in Fig. 3 but for the gluonic production of a vector B_c meson (B_c^*).

here for flows I and II, one can indeed see that the $\bar{b}$ quark moves predominantly in the direction of the gluon from which the anticolor flew to it (Figs. 3*c* and 4*c*), while the c quark moves in the direction of the gluon from which the color flew to it (Fig. 3*e* and 4*e*). As might have been expected, flow III, which corresponds to the color singlet, is represented by distributions that are symmetric in angles.

The interference between flows I and III is small (Figs. 3*b*, 3*d*, 3*f*, 4*b*, 4*d*, and 4*f*). In the case of the production of a pseudoscalar B_c meson at the interaction energy of 25 GeV, it is destructive over the entire region of emission angles of final particles. In the case of the production of a vector B_c meson, the interference term is also negative everywhere, with the exception of peripheral regions in the distributions with respect to the cosine of the emission angle of the $\bar{b}$ and c quarks. For both pseudoscalar- and vector-meson production, the total contribution of the interference to the cross section is negative. The calculations show that, at low energies, the interference contribution becomes positive for the vector B_c meson, but this is not so for the pseudoscalar B_c meson. The absolute value of the interference term is small over the entire range of gluon energies.

As was mentioned above, the contributions of flows I and II within the present and within the traditional scheme of partition into flows differ only in normalization. As to the color-singlet flows obtained in these two approaches, they differ in form as well; therefore, it is of interest to compare the contributions of this flow within the different approaches. In Fig. 5, the distributions of the cross sections with respect to the B_c-meson emission angle in our and in the traditional approach are given for flow III. One can see that the contribution of the singlet flow is greater in our approach than in the standard scheme. This enhancement can be explained by the fact that, in our approach, there is no interference between flows I and III and between flows II and III; that part of the cross section which corresponded to the respective

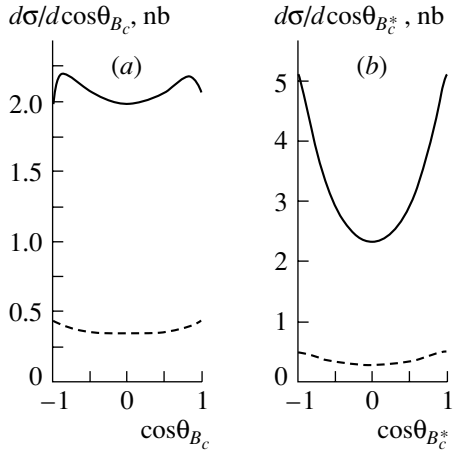

Fig. 5. Differential distributions of the cross sections with respect to the cosine of the emission angles of the (a) pseudoscalar and (b) vector B_c mesons for flow III according to calculations by (solid curves) the method proposed in the present study and (dashed curves) the standard method.

interference terms transformed into the contribution of flow III.

4. CONCLUSION

The proposed method of partition into color flows for the process $gg \to B_c + \bar{b} + c \to B_c + X$ makes it possible to reduce the interference between these flows to a minimum level; therefore, one can neglect the interference in simulating events involving a B_c meson. In addition, this method enables one to establish a clearer correspondence between the color states of final partons and color flows. The redistribution of the interference terms leads to a significant enhancement of flow III. In particular, the contribution of this flow in the central region becomes commensurate with the contribution of flows I and II, and this may lead to a different pattern of the hadronization of final partons $\bar{b}$ and c.

ACKNOWLEDGMENTS

I am grateful to A.K. Likhoded for a discussion on the results of this study and valuable comments and to Yu.P. Gouz and S.G. Shulga for a collaboration in simulating B_c-meson production under conditions of experiments planned for LHC.

This work was supported in part by the Ministry of Education of the Russian Federation (grant no. E02-3.1-96), the CRDF (grant no. M0-011-0), and the Russian Foundation for Basic Research (project no. 04-02-17530). It was also funded within the Program for Support of Leading Scientific Schools (grant no. 1303.2003.2).

REFERENCES

1. T. Sjöstrand, P. Edèn, C. Friberg, *et al.*, Comput. Phys. Commun. **135**, 238 (2001).
2. B. Anderson, G. Gustafson, G. Ingelman, and T. Sjöstrand, Phys. Rep. **97**, 31 (1983); Z. Phys. C **20**, 317 (1983); H. U. Bengtsson and G. Ingelman, Comput. Phys. Commun. **34**, 251 (1985).
3. L. Clavelli, Phys. Rev. D **26**, 1610 (1982); C.-R. Ji and R. Amiri, Phys. Rev. D **35**, 3318 (1987); Phys. Lett. B **195**, 593 (1987); C.-H. Chang and Y.-Q. Chen, Phys. Lett. B **284**, 127 (1992); Phys. Rev. D **46**, 3845 (1992); E. Braaten, K. Cheung, and T.-C. Yuan, Phys. Rev. D **48**, 4230 (1993); **48**, 5049 (1993); V. V. Kiselev, A. K. Likhoded, and M. V. Shevlyagin, Z. Phys. C **63**, 77 (1994); T.-C. Yuan, Phys. Rev. D **50**, 5664 (1994); K. Cheung, Phys. Rev. Lett. **71**, 3413 (1993); K. Cheung and T.-C. Yuan, Phys. Rev. D **53**, 3591 (1996).
4. A. V. Berezhnoy, A. K. Likhoded, and M. V. Shevlyagin, Yad. Fiz. **58**, 730 (1995) [Phys. At. Nucl. **58**, 672 (1995)]; A. V. Berezhnoy, A. K. Likhoded, and O. P. Yushchenko, Yad. Fiz. **59**, 742 (1996) [Phys. At. Nucl. **59**, 709 (1996)]; A. V. Berezhnoy, V. V. Kiselev, and A. K. Likhoded, Yad. Fiz. **60**, 108 (1997) [Phys. At. Nucl. **60**, 100 (1997)]; Z. Phys. A **356**, 79 (1996); A. V. Berezhnoy, V. V. Kiselev, A. K. Likhoded, and A. I. Onishchenko, Yad. Fiz. **60**, 1889 (1997) [Phys. At. Nucl. **60**, 1875 (1997)].
5. C.-H. Chang, Y.-Q. Chen, G.-P. Han, and H.-Q. Jiang, Phys. Lett. B **364**, 78 (1995); C.-H. Chang and X.-G. Wu, hep-ph/0309121; C.-H. Chang, C. Driouichi, P. Eerola, and X.-G. Wu, Comput. Phys. Commun. **159**, 192 (2004).
6. K. Kołodziej, A. Leike, and R. Rückl, Phys. Lett. B **355**, 337 (1995); Acta Phys. Pol. B **27**, 2591 (1996).
7. S. P. Baranov, Phys. Rev. D **56**, 3046 (1997); Nucl. Phys. B (Proc. Suppl.) **55A**, 33 (1997); Yad. Fiz. **60**, 1459 (1997) [Phys. At. Nucl. **60**, 1322 (1997)].
8. S. R. Slabospitsky, Yad. Fiz. **58**, 1061 (1995) [Phys. At. Nucl. **58**, 988 (1995)].
9. F. Sartogo and M. Masetti, Phys. Lett. B **357**, 659 (1995).
10. A. A. Belkov and S. G. Shulga, hep-ph/0201283.

Translated by A. Isaakyan

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1873–1903. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1935–1965.

ELEMENTARY PARTICLES AND FIELDS
Theory

A Systematic Study of QCD Coupling Constant from Deep-Inelastic Measurements*

V. G. Krivokhizhin and A. V. Kotikov

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

Received November 11, 2004

Abstract—We reanalyze deep-inelastic scattering data of the BCDMS Collaboration by including proper cuts of ranges with large systematic errors. We also perform the fits of high-statistic deep-inelastic scattering data of the BCDMS, SLAC, NM, and BFP Collaborations taking the data separately and in a combined way and find good agreement between these analyses. We extract the values of both the QCD coupling constant $\alpha_s(M_Z^2)$ up to the NLO level and of the power corrections to the structure function F_2. The fits of the combined data for the nonsinglet part of the structure function F_2 predict the coupling constant value $\alpha_s(M_Z^2) = 0.1174 \pm 0.0007\,(\text{stat.}) \pm 0.0019\,(\text{syst.}) \pm 0.0010\,(\text{norm.})$ (or QCD parameter $\Lambda^{(5)}_{\overline{MS}} = 204 \pm 25\,(\text{total experimental error})\,\text{MeV}$). The fits of the combined data for both the nonsinglet part and the singlet part lead to the values $\alpha_s(M_Z^2) = 0.1177 \pm 0.0007\,(\text{stat.}) \pm 0.0021\,(\text{syst.}) \pm 0.0009\,(\text{norm.})$ (or QCD parameter $\Lambda^{(5)}_{\overline{MS}} = (208 \pm 27\,(\text{total experimental error})\,\text{MeV}$). The above values are in very good agreement with each other. We estimate theoretical uncertainties for $\alpha_s(M_Z^2)$ at $+0.0047$ and -0.0057 from fits of the combined data when complete singlet and nonsinglet Q^2 evolution is taken into account.

1. INTRODUCTION

The deep-inelastic scattering (DIS) of leptons on hadrons is the basic process to study the values of the parton distribution functions (PDFs) which are universal (after choosing factorization and renormalization schemes) and can be used in other processes. The accuracy of the present data for deep-inelastic structure functions (SFs) has reached the level at which the Q^2 dependence of logarithmic QCD-motivated and powerlike ones may be studied separately (for a review, see the recent papers [1, 2] and references therein).

In the present article, we analyze at the next-to-leading (NLO) order[1)] of perturbative QCD the most known DIS SF $F_2(x, Q^2)$ taking into account SLAC, NM, BCDMS, and BFP experimental data [4–10]. We stress the powerlike effects, so-called twist-4 (i.e., $\sim 1/Q^2$) contributions. For our purposes, we represent the SF $F_2(x, Q^2)$ as the contribution of the leading-twist part $F_2^{\text{pQCD}}(x, Q^2)$ described by perturbative QCD and the nonperturbative part (twist-4 terms $\sim 1/Q^2$):

$$F_2(x, Q^2) \equiv F_2^{\text{full}}(x, Q^2) = F_2^{\text{pQCD}}(x, Q^2)\left(1 + \frac{\tilde{h}_4(x)}{Q^2}\right). \quad (1)$$

The SF $F_2^{\text{pQCD}}(x, Q^2)$ obeys the (leading-twist) perturbative QCD dynamics including the target mass corrections (TMC) (and coincides with $F_2^{\text{tw2}}(x, Q^2)$ when the TMC are withdrawn).

Equation (1) allows us to separate pure kinematical power corrections, i.e., TMC, so that the function $\tilde{h}_4(x)$ corresponds to the "dynamical" contribution of the twist-4 operators. The parametrization (1) implies[2)] that the anomalous dimensions of the twist-2 and twist-4 operators are equal to each other, which is not correct in principle. Moreover, there are estimations of these anomalous dimensions (see [11]). Meanwhile, in view of limited precision of the data, the approximation (1) and the one in the footnote 2

*The text was submitted by the authors in English.

[1)] The evaluation of $\alpha_s^3(Q^2)$ corrections to anomalous dimensions of Wilson operators, which will be done in near future by J.A.M. Vermaseren and his coauthors (see discussions in [3]), makes it possible to apply many modern programs to perform fits of data at the next-next-to-leading order (NNLO) of perturbative theory (see detailed discussions in Summary).

[2)] The right-hand side of Eq. (1) is represented sometimes as $F_2^{\text{pQCD}}(x, Q^2) + \overline{h}_4(x)/Q^2$. It implies that the anomalous dimensions of the twist-4 operators are equal to zero.

1063-7788/05/6811-1873$26.00

give rather good predictions (see discussions in [12, 13]).

Contrary to standard fits (see, for example, [14, 15]) when the direct numerical calculations based on the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equation [16] are used to evaluate SFs, we use the exact solution of the DGLAP equation for the Mellin moments $M_n^k(Q^2)$ of SFs $F_2^k(x,Q^2)$,

$$M_n^k(Q^2) = \int_0^1 x^{n-2} F_2^k(x,Q^2)dx \qquad (2)$$

(hereafter, $k =$ full, pQCD, tw2, …),

and the subsequent reproduction of $F_2^{\rm full}(x,Q^2)$, $F_2^{\rm pQCD}(x,Q^2)$, and/or $F_2^{\rm tw2}(x,Q^2)$ at every needed Q^2 value with help of the Jacobi polynomial expansion method [17–19] (see similar analyses at the NLO level [18–21] and at the NNLO level and above [22–27]).

The method of the Jacobi polynomial expansion was developed in [17, 18] and described in detail in [19]. Here, we consider only some basic definitions in Section 3.

The paper has the following structure: In Section 2, we present basic formulas which are needed in our analysis: we consider different types of Q^2 dependence of SF moments, effects of nuclear corrections and heavy-quark thresholds, and the structure of normalization of parton densities in singlet and nonsinglet channels. In Section 3, we introduce the basic elements of our fits. Sections 4 and 5 contain conditions and results of several types of fits with the nonsinglet and singlet evolutions for different sets of data. In Section 6, we study the dependence of the results on choice of factorization and renormalization scales. In Section 7, we summarize the basic observations following from the fits and discuss possible future extensions of the analysis.

2. Q^2 DEPENDENCE OF SFs AND THEIR MOMENTS

In this section, we analyze Eq. (1) in detail, considering separately different types of Q^2 dependence of structure function F_2.

2.1. The Leading-Twist Q^2 Dependence

To study the Q^2 dependence of the SF $F_2^{\rm tw2}(x,Q^2) = F_2^{\rm NS}(x,Q^2) + F_2^S(x,Q^2)$, which splits explicitly into the nonsinglet (NS) part and the singlet (S) part, it is very useful to introduce PDFs:[3] a gluon one $f_G(x,Q^2)$ and singlet and nonsinglet quarks ones $f_S(x,Q^2)$ and $f_{\rm NS}(x,Q^2)$.

The moments $M_n^{\rm NS}(Q^2)$ and $M_n^S(Q^2)$ of nonsinglet and singlet parts of SF F_2 (see Eq. (2) for definition) are connected with the corresponding moments of PDFs $\tilde{f}_i(x,Q^2)$ (hereafter, $i = {\rm NS}, S, G$)

$$f_i(n,Q^2) = \int_0^1 x^{n-2}\tilde{f}_i(x,Q^2)dx$$

in the following way (see [28], for example):

$$M_n^{\rm NS}(Q^2) = K_{\rm NS}(f)C_{\rm NS}^{\rm tw2}(n,\overline{a}_s(Q^2))f_{\rm NS}(n,Q^2), \qquad (3)$$

$$M_n^S(Q^2) = K_S(f)[C_S^{\rm tw2}(n,\overline{a}_s(Q^2))f_S(n,Q^2) + C_G^{\rm tw2}(n,\overline{a}_s(Q^2))f_G(n,Q^2)],$$

where[4]

$$\overline{a}_s(Q^2) = \frac{\alpha_s(Q^2)}{4\pi}, \qquad (4)$$

and $C_i^{\rm tw2}(n,\overline{a}_s(Q^2))$ are so-called Wilson coefficient functions. We have also introduced here the coefficients

$$K_S(f) = \sum_{m=1}^{f} e_m^2/f, \qquad (5)$$

$$K_{\rm NS}(f) = e_u^2 - K_S(f),$$

which come from definition of SF F_2 (see, for example, [28]). Here, f is the number of active quarks and e_m^2 is charge square of the active quark of m flavor.

1. The dependence of $\overline{a}_s(Q^2)$ is given by the renormalization group equation, which in the NLO QCD approximation reads

$$\frac{1}{\overline{a}_s(Q^2)} - \frac{1}{\overline{a}_s(M_Z^2)} + \frac{\beta_1}{\beta_0}\ln\left[\frac{\overline{a}_s(Q^2)}{\overline{a}_s(M_Z^2)}\frac{(\beta_0+\beta_1\overline{a}_s(M_Z^2))}{(\beta_0+\beta_1\overline{a}_s(Q^2))}\right] = \beta_0\ln\left(\frac{Q^2}{M_Z^2}\right), \qquad (6)$$

where $\overline{a}_s(M_Z^2)$ serves as a normalization. Here and below, we use β_0 and β_1 for the first and second terms with respect to $\overline{a}_s$ of the QCD β function:

$$\beta(\overline{a}_s) = -\beta_0\overline{a}_s^2 - \beta_1\overline{a}_s^3 + \ldots.$$

[3] Our PDFs are multiplied by x to compare with the standard definition.

[4] Sometimes, we will call the last term, $K_S(f)C_G^{\rm tw2}(n,\overline{a}_s(Q^2))f_G(n,Q^2)$, the gluon part of the singlet moment and denote it as $M_n^G(Q^2)$.

Equation (6) allows us to eliminate the QCD parameter $\Lambda_{\rm QCD}$ from our analysis. However, sometimes we will present it in our discussions, essentially to compare it with the results of old fits. The coupling constant $\overline{a}_s(Q^2)$ is expressed through $\Lambda_{\rm QCD}$ (in the $\overline{MS}$ scheme, where $\Lambda_{\rm QCD} = \Lambda_{\overline{MS}}$) as

$$\frac{1}{\overline{a}_s(Q^2)} + \frac{\beta_1}{\beta_0}\ln\left[\frac{\beta_0^2\overline{a}_s(Q^2)}{(\beta_0+\beta_1\overline{a}_s(Q^2))}\right] = \beta_0\ln\left(\frac{Q^2}{\Lambda^2_{\overline{MS}}}\right). \quad (7)$$

The relation between the normalization $\overline{a}_s(M_Z^2)$ and the QCD parameter $\Lambda_{\rm QCD}$ can be obtained from Eq. (7) with the replacement $Q^2 \to M_Z^2$.

We would like to note that approximations of Eq. (7), based on the expansion of inverse powers of $\ln\left(Q^2/\Lambda^2_{\overline{MS}}\right)$, are very popular. The accuracy of these expansions for evolution of $\overline{a}_s$ from $O(\mathrm{GeV}^2)$ to M_Z^2 may be as large as 0.001 [13], which is comparable with the experimental uncertainties of the $\alpha_s(M_Z^2)$ value extracted from the data (see our analyses in Sections 4 and 5).

Note also that sometimes (see, for example, [22]) the equation

$$\frac{1}{\overline{a}_s(Q^2)} + \frac{\beta_1}{\beta_0}\ln\left(\beta_0\overline{a}_s(Q^2)\right) = \beta_0\ln\left(\frac{Q^2}{\Lambda^2_{\overline{MS}}}\right) \quad (8)$$

is used in the analyses. This equation can be obtained from the basic equation

$$\ln\left(\frac{Q^2}{\Lambda^2_{\overline{MS}}}\right) = \int^{\overline{a}_s(Q^2)} \frac{db}{\beta(b)} \quad (9)$$

by expansion of the inverse QCD β function $1/\beta(\overline{a}_s)$ in powers of $\overline{a}_s$. The difference between Eqs. (8) and (7) may be as large as 0.001 at $O(\mathrm{GeV}^2)$ range. In order to escape the above uncertainties, we use in the analyses the exact numerical solution (with accuracy about 10^{-5}) of Eq. (6) instead. For recalculation of the QCD parameter $\Lambda_{\overline{MS}}$ from $\Lambda^{(f)}_{\overline{MS}}$ to $\Lambda^{(f\pm1)}_{\overline{MS}}$ (i.e., from $\Lambda_{\overline{MS}}$ at f active-quark flavors to $\Lambda_{\overline{MS}}$ at $f\pm1$ active-quark flavors), because β_0 and β_1 are f-dependent functions, we use formulas in the NLO approximation from [29] (see discussions in the Section 2.4).

2. The coefficient functions $C_i^{\rm tw2}(n,\overline{a}_s(Q^2))$ ($i = \mathrm{NS}, S, G$) have the following form:

$$C_i^{\rm tw2}(n,\overline{a}_s(Q^2)) = 1 - \delta_i^G + \overline{a}_s B_i(n) + O(\overline{a}_s^2) \quad (10)$$

(δ_i^m is the Kronecker symbol),

where the NLO coefficients $B_i(n)$ are known exactly (see, for example, [28]).

The Q^2 evolution of the moments $f_i(n,Q^2)$ is given by the well-known perturbative QCD [28, 30] formulas:

$$\frac{f_{\rm NS}(n,Q^2)}{f_{\rm NS}(n,Q_0^2)} = \left[\frac{\overline{a}_s(Q_0^2)}{\overline{a}_s(Q^2)}\right]^{\gamma^{(0)}_{\rm NS}(n)/(2\beta_0)} H^{\rm NS}(n,Q^2,Q_0^2), \quad (11)$$

$$f_j(n,Q^2) = f_j^+(n,Q^2) + f_j^-(n,Q^2) \quad (j = S,G),$$

$$\frac{f_j^\pm(n,Q^2)}{f_j^\pm(n,Q_0^2)} = \left[\frac{\overline{a}_s(Q_0^2)}{\overline{a}_s(Q^2)}\right]^{\gamma^{(0)}_\pm(n)/(2\beta_0)} H_j^\pm(n,Q^2,Q_0^2),$$

where[5)]

$$f_j^\pm(n,Q^2) = \varepsilon_{jl}^\pm(n) f_l(n,Q^2) \quad (j,l = S,G), \quad (12)$$

$$\gamma^{(0)}_\pm(n) = \frac{1}{2}\left[\left(\gamma^{(0)}_{GG}(n) + \gamma^{(0)}_{SS}(n)\right) \pm \sqrt{(\gamma^{(0)}_{SS}(n) - \gamma^{(0)}_{GG}(n)) + 4\gamma^{(0)}_{GS}(n)\gamma^{(0)}_{SG}}\right],$$

$$\varepsilon^\pm_{qq}(n) = \varepsilon^\mp_{gg}(n) = \frac{1}{2}\left(1 + \frac{\gamma^{(0)}_{SS}(n) - \gamma^{(0)}_{GG}(n)}{\gamma^{(0)}_\pm(n) - \gamma^{(0)}_\mp(n)}\right),$$

$$\varepsilon^\pm_{jl}(n) = \frac{\gamma^{(0)}_{jl}(n)}{\gamma^{(0)}_\pm(n) - \gamma^{(0)}_\mp(n)} \quad (j \neq l).$$

The functions $H^{\rm NS}(n,Q^2,Q_0^2)$ and $H_j^\pm(n,Q^2,Q_0^2)$ are nonzero above the leading-order (LO) approximation and may be represented as

$$H^{\rm NS}(n,Q^2,Q_0^2) = 1 + (\overline{a}_s(Q^2) - \overline{a}_s(Q_0^2))Z_{\rm NS}(n) + O(\overline{a}_s^2), \quad (13)$$

$$H_j^\pm(n,Q^2,Q_0^2) = 1 + (\overline{a}_s(Q^2) - \overline{a}_s(Q_0^2))Z_{\pm\pm}(n) + \left(\overline{a}_s(Q_0^2)\left[\frac{\overline{a}_s(Q_0^2)}{\overline{a}_s(Q^2)}\right]^{(\gamma^{(0)}_\mp(n) - \gamma^{(0)}_\pm(n))/(2\beta_0)} - \overline{a}_s(Q^2)\right) Z^j_{\pm\mp}(n) + O(\overline{a}_s^2),$$

where

$$Z_{\rm NS}(n) = \frac{1}{2\beta_0}\left(\gamma^{(1)}_{\rm NS}(n) - \gamma^{(0)}_{\rm NS}(n)\frac{\beta_1}{\beta_0}\right), \quad (14)$$

[5)] We use a nonstandard definition (see [31]) of the projectors $\varepsilon^\pm_{jl}(n)$, which is very convenient beyond the leading order (see Eq. (17) and [32, 33]). The connection with the more usual definition α, $\tilde{\alpha}$, and ε in [34, 28] is given by $\varepsilon^-_{SS}(n) = \alpha(n)$, $\varepsilon^-_{SG}(n) = \tilde{\alpha}(n)$, and $\varepsilon^-_{GS}(n) = \varepsilon(n)$.

$$Z_{\pm\pm}(n) = \frac{1}{2\beta_0}\left(\gamma^{(1)}_{\pm\pm}(n) - \gamma^{(0)}_{\pm}(n)\frac{\beta_1}{\beta_0}\right), \quad (15)$$

$$Z^S_{\pm\mp}(n) = \frac{\gamma^{(1)}_{\pm\mp}(n)}{2\beta_0 + \gamma^{(0)}_{\pm}(n) - \gamma^{(0)}_{\mp}(n)}, \quad (16)$$

$$Z^G_{\pm\mp}(n) = Z^S_{\pm\mp}(n)\frac{\varepsilon^{\mp}_{GG}}{\varepsilon^{\mp}_{SS}}$$

and

$$\gamma^{(1)}_{\pm\pm}(n) = \sum_{j,l=S,G}\varepsilon^{\pm}_{lj}\gamma^{(1)}_{jl}, \quad (17)$$

$$\gamma^{(1)}_{\pm\mp}(n) = \sum_{j=S,G}\varepsilon^{\pm}_{jG}\gamma^{(1)}_{Gj} - \varepsilon^{\mp}_{SS}\gamma^{(1)}_{SS} + (\varepsilon^{\pm}_{GS} - \varepsilon^{\pm}_{GG}/\varepsilon^{\pm}_{SG})\gamma^{(1)}_{GG}.$$

As usual, here we use $\gamma^{(0)}_{\rm NS}(n)$, $\gamma^{(0)}_{jl}(n)$ $(j,l = S,G)$ and $\gamma^{(1)}_{\rm NS}(n)$, $\gamma^{(1)}_{jl}(n)$ as the first and second terms with respect to $\overline{a}_s$ of anomalous dimensions $\gamma_{\rm NS}(n,\overline{a}_s)$ and $\gamma_{jl}(n,\overline{a}_s)$ (see, for example, [35]).

3. In this subsection, we would like to discuss a possible dependence of our results on the factorization scale μ_F and the renormalization scale μ_R, which appear (see, for example, [14, 36]) because perturbative series are truncated. These scales $\mu_F^2 = k_F Q^2$ and $\mu_R^2 = k_R\mu_F^2 = k_R k_F Q^2$ can be added to the right-hand side of Eqs. (3) and (11), respectively.

Then, Eqs. (3) are replaced by

$$M^{\rm NS}_n(Q^2) = K_{\rm NS}(f)\hat{C}^{\rm tw2}_{\rm NS}(n,\overline{a}_s(k_F Q^2)) \times f_{\rm NS}(n,k_F Q^2), \quad (18)$$

$$M^S_n(Q^2) = K_S(f)[\hat{C}^{\rm tw2}_S(n,\overline{a}_s(k_F Q^2)) \times f_S(n,k_F Q^2) + \hat{C}^{\rm tw2}_G(n,\overline{a}_s(k_F Q^2))f_G(n,k_F Q^2)].$$

Equations (11) are replaced, correspondingly, by

$$\frac{f_{\rm NS}(n,k_F Q^2)}{f_{\rm NS}(n,k_F Q_0^2)} = \left[\frac{\overline{a}_s(k_F k_R Q_0^2)}{\overline{a}_s(k_F k_R Q^2)}\right]^{\gamma^{(0)}_{\rm NS}(n)/(2\beta_0)} \times \hat{H}^{\rm NS}(n,k_F k_R Q^2, k_F k_R Q_0^2), \quad (19)$$

$$\frac{f^{\pm}_j(n,k_F Q^2)}{f^{\pm}_j(n,k_F Q_0^2)} = \left[\frac{\overline{a}_s(k_F k_R Q_0^2)}{\overline{a}_s(k_F k_R Q^2)}\right]^{\gamma^{(0)}_{\pm}(n)/(2\beta_0)} \times \hat{H}^{\pm}_j(n,k_F k_R Q^2, k_F k_R Q_0^2).$$

The coefficients $\hat{C}_{\rm NS}$, $\hat{C}_S$, $\hat{C}_G$, $\hat{H}^{\rm NS}$, and $\hat{H}^{\pm}_j$ can be obtained from $C_{\rm NS}$, C_S, C_G, $H^{\rm NS}$, and $H^{\pm}_j$ by modification on the right-hand side of Eqs. (10), (14), and (15) as follows:

$$\overline{a}_s(Q^2) \to \overline{a}_s(k_F Q^2), \quad (20)$$

$$B_{\rm NS}(n) \to B_{\rm NS}(n) + \frac{1}{2}\gamma^{(0)}_{\rm NS}(n)\ln k_F,$$

$$B_j(n) \to B_j(n) + \frac{1}{2}\gamma^{(0)}_{jS}(n)\ln k_F \quad (j = S,G) \quad (21)$$

in Eq. (10);

$$\overline{a}_s(Q^2) \to \overline{a}_s(k_F k_R Q^2), \quad (22)$$

$$Z_{\rm NS}(n) \to Z_{\rm NS}(n) + \frac{1}{2}\gamma^{(0)}_{\rm NS}(n)\ln k_R,$$

$$Z_{\pm\pm}(n) \to Z_{\pm\pm}(n) + \frac{1}{2}\gamma^{(0)}_{\pm}(n)\ln k_R \quad (23)$$

in Eqs. (14) and (15).

Equations (21) can be obtained easily using, for example, the results of [37]. Equations (23) can be found from the expansion of the coupling constant $\overline{a}_s(k_F k_R Q^2)$ around the one $\overline{a}_s(Q^2)$ on the right-hand side of the exact solution of DGLAP equations [see Eqs. (19) and (13)].

The changes (21) and (23) of the results for the Q^2 dependence under variation of k_F and k_R (usually[6] from 1/2 to 2) give an estimation of the errors due to factorization and renormalization scale uncertainties. Evidently, by definition, these uncertainties are connected with the impact of unaccounted terms of the perturbative series and can represent theoretical uncertainties in values of fitted variables. Indeed, incorporation of NNLO corrections to the analysis strongly suppress these uncertainties (see [39, 40]).

We study exactly the μ_F and μ_R dependences here for fitted values of the coupling constant. The results of the study are given in Section 6.

As one can see in Eqs. (20) and (22), the coupling constant $\overline{a}_s$ has different arguments in the NLO corrections of coefficient functions $\hat{C}_{\rm NS}$ and $\hat{C}_j$ $(j = S,G)$ and in the NLO corrections $\hat{H}^{\rm NS}$ and $\hat{H}^{\pm}_j$ of the Q^2 evolution of parton distributions. We would like to note that the difference between the corresponding coupling constants $\overline{a}_s(k_F Q^2)$ and $\overline{a}_s(k_F k_R Q^2)$ is proportional to $\overline{a}_s^2$ and, thus, mathematically negligible in our NLO approximation.

Then, we can use the replacement (22) in coefficient functions too, as has been done in previous studies [38–40, 27]. We note that the replacement $\overline{a}_s(k_F Q^2) \to \overline{a}_s(k_F k_R Q^2)$ in Eq. (20) increases slightly the factorization-scheme dependence of the

[6] In the recent articles [38–40, 26], the variation from 1/4 to 4 has been used. In our opinion, the case $k_F = k_R = 4$ leads to a very small scale of coupling constant $Q^2/16$, which requires us to reject many experimental points of data, because we have the general cut $Q^2 > 1\ {\rm GeV}^2$. So, we prefer to use the variation of scales from 1/2 to 2.

results for coupling constant (see analyses based on nonsinglet evolution and discussions in Section 6).

2.2. Normalization of Parton Distributions

The moments $f_i(n,Q^2)$ at some Q_0^2 are theoretical input of our analysis which is fixed as follows.

For fits of data at $x \geq 0.25$, we can work only with the nonsinglet parton density and use directly its normalization $\tilde{f}_{\rm NS}(x,Q_0^2)$ (see, for example, [22–26]):

$$f_{\rm NS}(n,Q_0^2) = \int_0^1 dx x^{n-2} \tilde{f}_{\rm NS}(x,Q_0^2), \qquad (24)$$

$$\tilde{f}_{\rm NS}(x,Q_0^2) = A_{\rm NS}(Q_0^2)(1-x)^{b_{\rm NS}(Q_0^2)} \times (1 + d_{\rm NS}(Q_0^2)x),$$

where $A_{\rm NS}(Q_0^2)$, $b_{\rm NS}(Q_0^2)$, and $d_{\rm NS}(Q_0^2)$ are some coefficients.[7)]

In the analyses, at arbitrary values of x, we should introduce the normalizations for densities of individual quarks ($q = u,d,s,\dots$) and antiquarks ($\overline{q} = \overline{u},\overline{d},\overline{s},\dots$) $\tilde{f}_q(x,Q_0^2)$ and $\tilde{f}_{\overline{q}}(x,Q_0^2)$ having the moments

$$f_i(n,Q_0^2) = \int_0^1 dx x^{n-2} \tilde{f}_i(x,Q_0^2). \qquad (25)$$

The distributions of u and d quarks $\tilde{f}_u(x,Q_0^2) \equiv u(x,Q_0^2)$ and $\tilde{f}_d(x,Q_0^2) \equiv d(x,Q_0^2)$ are split into two components: the valent ones $u_v(x,Q_0^2)$ and $d_v(x,Q_0^2)$ and the sea ones $u_{\rm sea}(x,Q_0^2)$ and $d_{\rm sea}(x,Q_0^2)$. For other quark distributions and antiquark densities, we keep only sea parts. Moreover, following [28, 41], we assume equality of all sea parts and indicate their sum as $S(x,Q_0^2)$.

We use the following parametrizations for densities $u_v(x,Q_0^2)$, $d_v(x,Q_0^2)$, $S(x,Q_0^2)$, and $\tilde{f}_G(x,Q_0^2)$:

$$u_v(x,Q_0^2) = \frac{2}{B(a_u(Q_0^2), b_u(Q_0^2)+1)} x^{a_u(Q_0^2)} \times (1-x)^{b_u(Q_0^2)}, \qquad (26)$$

$$d_v(x,Q_0^2) = \frac{1}{B(a_d(Q_0^2), b_d(Q_0^2)+1)} x^{a_d(Q_0^2)} \times (1-x)^{b_d(Q_0^2)},$$

$$S(x,Q_0^2) = C_S(Q_0^2) x^{a_{\rm sea}(Q_0^2)} (1-x)^{b_{\rm sea}(Q_0^2)}, \qquad (27)$$

$$\tilde{f}_G(x,Q_0^2) = C_G(Q_0^2) x^{a_G(Q_0^2)} (1-x)^{b_G(Q_0^2)},$$

where $B(a,b)$ is the Euler beta function. The parametrizations (26) have been chosen to satisfy (at the normalization point Q_0^2) the known rule

$$\int_0^1 dx V(x,Q^2) = 3,$$

where $V(x,Q^2) = u_v(x,Q^2) + d_v(x,Q^2)$ is the distribution of valent quarks.

We note that the nonsinglet and singlet parts of quark distributions, $\tilde{f}_{\rm NS}(x,Q_0^2)$ and $\tilde{f}_S(x,Q_0^2)$, can be represented as combination of quark ones:

$$\tilde{f}_S(x,Q_0^2) \equiv \sum_q^f \tilde{f}_q(x,Q_0^2) \qquad (28)$$

$$= V(x,Q_0^2) + S(x,Q_0^2),$$

$$\tilde{f}_{\rm NS}(x,Q_0^2) = u_v(x,Q_0^2) - d_v(x,Q_0^2), \qquad (29)$$

where the right-hand side of Eq. (29) is correct only in the framework of our supposition about equality of antiquarks distributions and sea components of quark ones.

In principle, following the PDF models used in [15, 12] and Eq. (24) above, one can add in Eq. (27) terms proportional to $\sqrt{x}$ and x. However, the terms $\sim\sqrt{x}$ are important only in the region of rather small x (see discussion in [12]). The terms $\sim x$ lead only to replacement of C_i, a_i, and b_i values (see, for example, [42]). Thus, we neglect these terms in our analysis.

In most of our fits, we also do not take into account the terms $\sim x^{a_G(Q_0^2)}$ and $\sim x^{a_{\rm sea}(Q_0^2)}$ in the gluon and sea quark distributions, because we do not consider experimental data at small values of Bjorken variable x.[8)] We hope to include H1 and ZEUS data [43, 44] in our future investigations [45] and then to study the Q^2 dependence of the coefficients $a_G(Q^2)$ and $a_{\rm sea}(Q^2)$, which could be very nontrivial (see, for example, [46–48, 43, 32] and references therein).

We also impose the condition for full-momentum conservation in the form

$$1 = P_G(Q^2) + P_q(Q^2), \qquad (30)$$

[7)] We do not consider here the term $\sim x^{a_{\rm NS}(Q_0^2)}$ in the normalization $\tilde{f}_{\rm NS}(x,Q_0^2)$, because $x \geq 0.25$. The correct small-x asymptotics of nonsinglet distributions will be obtained by Eq. (29) from the corresponding parameters of the valent-quark distributions (26) fitted with complete singlet and nonsinglet evolution in Section 5.

[8)] However, we have performed several fits with nonzero a_G and $a_{\rm sea}$ values taken into account (see Section 5). We have found a negative value for them: $a_G = a_{\rm sea} \sim -0.18$ (which is in agreement with [12]), but these results cannot be considered seriously without taking into account H1 and ZEUS data [43, 44] (see, however, discussions in Section 5.3.4).

where

$$P_G(Q^2) = \int_0^1 dx \tilde{f}_G(x, Q^2), \qquad (31)$$

$$P_q(Q^2) = \int_0^1 dx (\tilde{f}_{\rm NS}(x, Q^2) + \tilde{f}_S(x, Q^2)).$$

The coefficients $C_i(Q_0^2)$, $a_i(Q_0^2)$, $b_i(Q_0^2)$, $c_i(Q_0^2)$, and $d_i(Q_0^2)$ should be found together with $\tilde{h}_4(x)$ (see Section 2.6) and the normalization $\alpha_s(M_Z^2)$ of the QCD coupling constant (or QCD parameter Λ) by the fits of experimental data.

2.3. Target Mass Corrections

TMC modify the SF $F_2^{\rm pQCD}(x, Q^2)$ in the following way (see [28, 49])

$$F_2^{\rm pQCD}(x, Q^2) = \frac{1}{r^3}\frac{x^2}{\xi^2} F_2^{\rm tw2}(\xi, Q^2) \qquad (32)$$

$$+ 6\frac{M_{\rm nucl}^2}{Q^2}\frac{x^3}{r^6}\int_\xi^1 \frac{d\xi'}{(\xi')^2} F_2^{\rm tw2}(\xi', Q^2)$$

$$+ 12\frac{M_{\rm nucl}^4}{Q^4}\frac{x^4}{r^5}\int_\xi^1 d\xi' \int_{\xi'}^1 \frac{d\xi''}{(\xi'')^2} F_2^{\rm tw2}(\xi'', Q^2),$$

where $M_{\rm nucl}$ is the mass of the nucleon, $r = \sqrt{1 + x^2 M_{\rm nucl}^2/Q^2}$, and the Nachtmann variable $\xi = 2x/(1+r)$.

In our analyses below, we will use this representation (32).[9)] We would like to keep the full value of kinematic power corrections, given by nonzero nucleon mass. Then, the excess of the $1/Q^2$ dependence encoded in experimental data will give the magnitude of twist-4 corrections, which is the most important part of dynamical power corrections.

2.4. Thresholds of Heavy Quarks

Modern estimates performed in [51, 52] have revealed a quite significant role of threshold effects in the $\alpha_s(Q^2)$ evolution when the DIS data lie close to threshold points $Q^2 = M_{f+1}^2 \sim m_{f+1}^2$ (to the position of so-called Euclidean-reflected threshold of heavy particles). The corresponding corrections to the normalization $\alpha_s(M_Z^2)$ can reach several percent; i.e., they are of the order of other uncertainties which should be under control at our analysis.

An appropriate procedure for the inclusion of threshold effects in the Q^2 dependence of $\alpha_s(Q^2)$ in the framework of the massless $\overline{MS}$ scheme was proposed more than 20 years ago [53, 54]: transition from the region with a given number of flavors f described by massless $\alpha_s(Q^2; f)$[10)] to the next one with $f+1$ (transition across the M_{f+1} threshold) is realized here with the use of the so-called matching relation for $\alpha_s(Q^2)$ [54]. The latter may be considered as the continuity condition for $\alpha_s(Q^2)$ on (every) heavy-quark mass m_{f+1}:

$$\alpha_s(Q^2 = M_{f+1}^2; f) = \alpha_s(Q^2 = M_{f+1}^2; f+1) \qquad (33)$$

and

$$M_{f+1} = m_{f+1}, \qquad (34)$$

which provides an accurate $\alpha_s(Q^2)$-evolution description for Q^2 values not close to the threshold region (see [55] and references therein).

In the analyses based on nonsinglet evolution, the additional f dependence comes only from the NLO correction of nonsinglet anomalous dimensions (see [35]).[11)] In Section 3, we check numerically the dependence of the results from the matching point. We use two matching points, (34) and

$$M_{f+1} = 2m_{f+1}, \qquad (35)$$

and demonstrate very little variations of the results[12)] (see Section 4 and discussion there).

As we know, for the singlet part of evolution, no simple recipe exists for the exact value of the matching point M_{f+1}. On one hand, as in the nonsinglet case, there is Q^2 evolution of the SF moments which leads to above condition (34). But here we also have the generation of heavy quarks (even in lowest nontrivial order, in the framework of the photon–gluon fusion process), which gives contributions to gluon

[9)] It is contrary to [13], where only the term $\sim M_{\rm nucl}^2/Q^2$ has been used. We note that the appearance of the terms $\sim M_{\rm nucl}^2/Q^2$ at $x = 1$ (see, for example, [50]), i.e., the absence of the equality $F_2^{\rm pQCD}(1, Q^2) = 0$, is not important in our analyses because we do not use experimental data at very large x values: $x \leq 1$.

[10)] Following [55], in this subsection, we use the form $\alpha_s(Q^2; f)$ for the coupling constant with the purpose to demonstrate its f dependence through the ones of β_0 and β_1 coefficients.

[11)] The corresponding moments at any Q^2 value are proportional to the same coefficient $K_{\rm NS}(f)$. Thus, the coefficient can be always taken up by the normalization $M_n^{\rm NS}(Q_0^2)$.

[12)] We will not take into account a small variation (see [56]) of the continuity condition (33) because of the matching point (35).

part of the singlet coefficient function. The photon–gluon fusion needs the matching point at the value of Q^2, when $W^2 = 4m_{f+1}^2$, i.e.,

$$M_{f+1}^2 \frac{1-x}{x} + M_{\rm nucl}^2 = 4m_{f+1}^2. \qquad (36)$$

At small x values, the condition (36) is quite close to (34) (for example, at $x = 0.2$ $M_{f+1}^2 = m_{f+1}^2 - M_{\rm nucl}^2/4$), but in the range of large and intermediate values of x, the value of M_{f+1} is essentially large compared with the one of Eq. (34). For example, at $x = 0.5$ $M_{f+1}^2 = 4m_{f+1}^2 - M_{\rm nucl}^2$, it is very close to the matching point (35). At larger x values, the value of M_{f+1}^2 will be close to those in [38, 39].

We note that the difference between nonsinglet and singlet Q^2 dependences comes from contribution of the gluon distribution. The contribution is negligible at $x > 0.3$, which supports qualitatively the choice (34) as the matching point.

We would like to note also that, in the NLO approximation and above, the situation is even more difficult in the singlet case, because every subprocess itself generates matching point M_{f+1} to coefficient functions. To estimate the possible effect of the dependence on the matching point, we will fit data (in Section 3) with two different matching points: (34) and (35). Surprisingly, in the singlet case, where all functions coming to Q^2 evolution are f-dependent, we do not find a strong f dependence of our results (see Section 5 and discussions there).

2.5. Nuclear Effects

Starting with EMC discovery in [57], the difference between PDFs in free hadrons and ones in hadrons in nuclei is well known. We incorporate the difference in our analyses.

In the nonsinglet case, we parametrize the initial PDF in the form (24) for every type of target. We have

$$f_{\rm NS}^A(n, Q_0^2) = \int_0^1 dx x^{n-2} \tilde{f}_{\rm NS}^A(x, Q_0^2), \qquad (37)$$

where

$$\tilde{f}_{\rm NS}^A(x, Q_0^2) \qquad (38)$$
$$= A_{\rm NS}^A(Q_0^2)(1-x)^{b_{\rm NS}^A(Q_0^2)}(1 + d_{\rm NS}^A(Q_0^2)x)$$

and $A = H$, D, C, and F in the case of H_2, D_2, ^{12}C, and ^{56}Fe targets, respectively.

In the singlet case, we have many parameters in our fits, which should be fitted very carefully. The representations similar to (38) for gluon and sea-quark PDFs should complicate our analyses. To overcome the problem, we apply Eqs. (37) and (38) only to the H_2 and D_2 cases. For heavier targets, we apply simpler representations for F_2^A SFs in the form

$$F_2^A(x, Q^2) = F_2^D(x, Q^2) K_1^A (1 - K_2^A x + K_3^A x^2) \qquad (39)$$
$$(A = C, F),$$

where we use experimental observation[13] (see [60] and references therein) about the approximate Q^2 independence of the EMC ratio $r^A = F_2^A/F_2^D$.

2.6. Higher Twist Corrections

For n space, Eq. (1) transforms to

$$M_n^{\rm full}(Q^2) = M_n^{\rm pQCD}(Q^2) + \frac{h_4(n)}{Q^2}, \qquad (40)$$

where $h_4(n, Q^2)$ are the moments of the function $\tilde{h}_4(x, Q^2)$:

$$h_4(n) = \int_0^1 x^{n-2} \tilde{h}_4(x) F_2^{\rm pQCD}(x) dx. \qquad (41)$$

The shape $\tilde{h}_4(x)$ (or coefficients $h_4(n)$) of the twist-4 corrections are of primary consideration in our analysis. They can be chosen in several different forms:

(i) The twist-4 terms (and twist-6 ones) are fixed in agreement with the infrared renormalon (IRR) model (see [61, 62, 1, 2] and references therein).

(ii) The twist-4 term in the form $\tilde{h}_4(x) \sim d \ln F_2^{\rm NS}(x, Q^2)/dx \sim 1/(1-x)$ (see [30] and references therein). This behavior matches the fact that higher twist effects are usually important only at higher x. The twist-4 coefficient function has the form $C_4^{\rm der}(n) = (n-1)A_4^{\rm der}$.

(iii) The twist-4 term $\tilde{h}_4(x)$ is considered as a set of free parameters at each x_i bin. The set has the form $\tilde{h}_4^{\rm free}(x) = \sum_{i=1}^I \tilde{h}_4(x_i)$, where I is the number of bins. The constants $\tilde{h}_4(x_i)$ (one per x bin) parametrize the x dependence of $\tilde{h}_4^{\rm free}(x)$.

The first two cases have already been considered in [21] and will be studied carefully later [45]. Here, we will follow the last possibility.[14]

[13] The small Q^2 dependence of the EMC ratio has also been observed in theoretical studies. For example, in the framework of the rescaling model [58], the Q^2 dependence is very small (see [59]). It has double-logarithmic form and is found only in the argument of the Euler Ψ function.

[14] In the Summary, we present, however, several comments about an application of higher twist corrections in the form of IRR model.

3. FITS OF F_2: PROCEDURE

To clear up the importance of higher twist terms, we fit SLAC, NM, BCDMS, and BFP experimental data [4–10] (including the systematic errors), keeping identical the form of the perturbative part in the NLO approximation. In this section, we demonstrate the basic ingredients of the analyses.

As has been already discussed in the Introduction, we use the exact solution of the DGLAP equation for the Mellin moments $M_n^{\rm tw2}(Q^2)$ (2) of SF $F_2^{\rm tw2}(x,Q^2)$ and the subsequent reproduction of $F_2^{\rm full}(x,Q^2)$, $F_2^{\rm pQCD}(x,Q^2)$, and/or $F_2^{\rm tw2}(x,Q^2)$ at every needed Q^2 value with help of the Jacobi polynomial expansion method. The method of the Jacobi polynomial expansion was developed in [17, 18] and described in detail in [19]. Here, we consider only some basic definitions.

Having the QCD expressions for the Mellin moments $M_n^k(Q^2)$, we can reconstruct the SFs $F_2^k(x,Q^2)$ as

$$F_2^{k,N_{\max}}(x,Q^2) = x^a(1-x)^b \times \sum_{n=0}^{N_{\max}} \Theta_n^{a,b}(x) \sum_{j=0}^{n} c_j^{(n)}(a,b) M_{j+2}^k(Q^2), \quad (42)$$

where $\Theta_n^{a,b}$ are the Jacobi polynomials[15] and a, b are the parameters fitted by the requirement of minimization of the error of reconstruction of SFs[16] (see [18] for details).

First of all, we choose the cut $Q^2 \geq 1$ GeV^2 in all our studies. For $Q^2 < 1$ GeV^2, the applicability of twist expansion is very questionable.

Secondly, we choose quite large values of the normalization point Q_0^2. There are several reasons for this choice:

(i) Our above perturbative formulas should be applicable at the value of Q_0^2. Moreover, the higher order corrections $\sim\alpha_s^n(Q_0^2)$ $(n \geq 2)$, coming from normalization conditions of PDFs, are less important at higher Q_0^2 values.

(ii) It is necessary to cross heavy-quark thresholds a fewer number of times to reach $Q^2 = M_Z^2$, the point of QCD coupling constant normalization.

(iii) It is better to have the value of Q_0^2 around the middle point of the logarithmic range of considered Q^2 values. Then, the higher order corrections $\sim(\alpha_s(Q^2) - \alpha_s(Q_0^2))^n$ $(n \geq 2)$ are less important.

Basic characteristics of the quality of the fits are χ^2/d.o.f. for SF F_2 and for its slope $d\ln F_2/d\ln Q^2$, which has very sensitive perturbative properties (see [28]).

We note that the characteristics $\chi^2(F_2)$ and χ^2(slope) which come in our fits are not independent, because they come from the same data. We do not have, however, a "double counting" problem. The $\chi^2(F_2)$ value is the basic characteristic in our fits: it determines all needed parameters, i.e., ones in parametrizations of parton distributions and the coupling constant normalization $\alpha_s(M_Z^2)$.

The χ^2(slope) value gives additional information, which demonstrates clearly the effect of agreement and/or disagreement between theory and experiment.

Indeed, as these fits involve many free parameters independent of perturbative QCD, it is important to check whether, in the results of the fits, the features most specific to perturbative QCD are in good agreement with the data. The slope $d\ln F_2/d\ln Q^2$ has really very sensitive perturbative properties and will be used (see Figs. 4–8 and 10–12 below) to check properties of fits. Indeed, the DGLAP equations predict that, to a quite good approximation, the logarithmic derivations of SF and PDF logarithms are proportional to coupling constant $\alpha_s(Q^2)$ with an x-dependent proportionality coefficient that depends (at $x > 0.2$) only weakly on the x dependence of the SF and PDF. Thus, the study of the Q^2 dependence of the slope $d\ln F_2/d\ln Q^2$ leads to obtaining direct information about the corresponding Q^2 dependence of the QCD coupling constant and to verifying the range of accuracy for formulas of perturbative QCD.

We use the MINUIT program [68] for minimization of two χ^2 values:

$$\chi^2(F_2) = \left|\frac{F_2^{\rm exp} - F_2^{\rm theor}}{\Delta F_2^{\rm exp}}\right|^2 \text{ and}$$

$$\chi^2(\text{slope}) = \left|\frac{D^{\rm exp} - D^{\rm theor}}{\Delta D^{\rm exp}}\right|^2 \quad \left(D = \frac{d(\ln F_2)}{d(\ln Q^2)}\right).$$

We would like to apply the following procedure: we will study the dependence of the χ^2/d.o.f. value on value of Q^2 cuts for various sets of experimental

[15] We would like to note here that there is a similar method [63], based on Bernstein polynomials. The method has been used in the analyses at the NLO level in [64, 37] and at the NNLO level in [65, 27].

[16] There is another possibility to fit data. It is possible to transfer experimental information about SFs to their moments and to analyze these moments directly. This approach was very popular in the past (see, for example, [66]), but it is used very rarely at present (see, however, [67] and references therein) because a transformation of experimental information to the SF moments is quite a difficult procedure.

data. The study will be done for both cases: including higher twists corrections (HTC) and without them.

We use free normalizations of data for different experiments. For the reference, we use the most stable deuterium BCDMS data at the value of energy $E_0 = 200$ GeV.[17] Using other types of data as reference gives negligible changes in our results. The usage of fixed normalization for all data leads to fits with a bit worse χ^2.

4. RESULTS OF FITS OF F_2: THE NONSINGLET EVOLUTION PART

Firstly, we will consider the Q^2 evolution of the SF F_2 in the nonsinglet case, where there are only contributions of quark densities and, thus, the corresponding fits are essentially simpler. The consideration of the nonsinglet part limits the range of data by the cut $x \geq 0.25$. At smaller x values, the contributions of the gluon distribution are no longer negligible.

Hereafter, in the nonsinglet case of evolution, we choose $Q_0^2 = 90$ GeV^2 for the BCDMS data and all data combined and $Q_0^2 = 20$ GeV^2 for the combined SLAC, NM, BFP data, respectively. The choice of Q_0^2 values is in good agreement with above conditions (see the previous section). We use also $N_{\max} = 8$, the cut $0.25 \leq x \leq 0.8$. The $N_{\max}$ dependence of the results has been studied carefully in [18] (see also Table 3 below).

Except the studies done in Section 4.3.2, the effects of the thresholds of heavy quarks are taken into account in agreement with Eqs. (33) and (34).

4.1. BCDMS $^{12}C + H_2 + D_2$ Data

We start our analysis with the most precise experimental data [7–9] obtained by BCDMS muon scattering experiment at high Q^2 values. The full set of data is 607 points (when $x \geq 0.25$). The starting point of QCD evolution is $Q_0^2 = 90$ GeV^2.

It is well known that the original analyses given by the BCDMS Collaboration itself (see also [14]) lead to quite small values of $\alpha_s(M_Z^2)$: for example, $\alpha_s(M_Z^2) = 0.113$ has been obtained in [14].[18] Although in some recent papers (see, for example, [12, 13, 69]) higher values of $\alpha_s(M_Z^2)$ have been observed, we think that an additional reanalysis of BCDMS data should be very useful.

[17] E_0 is the initial energy of the lepton beam.

[18] We would like to note that the paper [14] has a quite strange result. The authors of the article have obtained the value $\Lambda^{(4)}_{\overline{MS}} = 263$ MeV, which should lead to the value of coupling constant $\alpha_s(M_Z^2)$ which is equal to 0.1157.

Table 1. The values of Y_{cut3}, Y_{cut4}, and Y_{cut5}

N	0	1	2	3	4	5	6
Y_{cut3}	0	0.14	0.16	0.16	0.18	0.22	0.23
Y_{cut4}	0	0.16	0.18	0.20	0.20	0.23	0.24
Y_{cut5}	0	0.20	0.20	0.22	0.22	0.24	0.25

Based on study [70] (see also [71, 69]), we propose that the reason for small values of $\alpha_s(M_Z^2)$ coming from BCDMS data is the existence of a subset of data having large systematic errors. Indeed, the original analyses of H_2, D_2, and ^{12}C data performed by the BCDMS Collaboration lead to the following value of the QCD mass parameter (see [7–9]):

$$\Lambda^{(4)}_{\overline{MS}} = 220 \pm 13\,(\text{stat.}) \pm 50\,(\text{syst.})\,\text{MeV}; \quad (43)$$

i.e., the systematic error is four times bigger than the statistical one. (Hereafter, "stat." and "syst." mark the statistical and systematic errors, respectively).

We study this subject by introducing several so-called Y cuts[19] (see [70] and Sections 4.1.1 and 5.1). Excluding this set of data with large systematic errors leads to substantially larger values of $\alpha_s(M_Z^2)$ and very slow dependence of the values on the concrete choice of the Y cut (see below).

4.1.1. The study of systematics. The correlated systematic errors of the data have been studied in [70], together with the other parameters. Regions of data have been identified in which the fits cause large systematic shifts of the data points. We would like to exclude these regions from our analyses.

We have studied the influence of the experimental systematic errors on the results of the QCD analysis as a function of Y_{cut3}, Y_{cut4}, and Y_{cut5} applied to the data. We use the following x-dependent Y cuts:

$$\begin{aligned} &Y \geq 0.14 \text{ when } 0.3 < x \leq 0.4, \quad (44)\\ &Y \geq 0.16 \text{ when } 0.4 < x \leq 0.5,\\ &Y \geq Y_{\text{cut3}} \text{ when } 0.5 < x \leq 0.6,\\ &Y \geq Y_{\text{cut4}} \text{ when } 0.6 < x \leq 0.7,\\ &Y \geq Y_{\text{cut5}} \text{ when } 0.7 < x \leq 0.8. \end{aligned}$$

We use several sets N of values for the cuts at $0.5 < x \leq 0.8$, which are given in Table 1.

The systematic errors for BCDMS data are given [7–9] as multiplicative factors to be applied to $F_2(x, Q^2)$: f_r, f_b, f_s, f_d, and f_h are the uncertainties due to spectrometer resolution, beam momentum

[19] Hereafter, we use the kinematical variable $Y = (E_0 - E)/E_0$, where E_0 and E are the initial and scattering energies of lepton, respectively.

Table 2. The values of $\alpha_s(M_Z^2)$ at different values of N

N	Number of points	$\chi^2(F_2)$/d.o.f.	$\alpha_s(90\ \mathrm{GeV}^2) \pm$ stat.	Full syst. error
0	607	1.03	0.1590 ± 0.0020	0.0090
1	511	0.97	0.1711 ± 0.0027	0.0075
2	502	0.97	0.1720 ± 0.0027	0.0071
3	495	0.97	0.1723 ± 0.0027	0.0063
4	489	0.94	0.1741 ± 0.0027	0.0061
5	458	0.95	0.1730 ± 0.0028	0.0052
6	452	0.95	0.1737 ± 0.0029	0.0050

calibration, spectrometer magnetic field calibration, detector inefficiencies, and energy normalization, respectively.

For this study each experimental point of the undistorted set was multiplied by a factor characterizing a given type of uncertainties and a new (distorted) data set was fitted again in agreement with our procedure considered in the previous section. The factors (f_r, f_b, f_s, f_d, f_h) were taken from CERN preprint versions in [7–9]. The absolute differences between the values of α_s for the distorted and undistorted sets of data are given in Table 2 and Fig. 1 as the total systematic error of α_s estimated in quadratures. The number of experimental points and the value of α_s for the undistorted set of F_2 are also given in Table 2 and Fig. 1.

From Table 2 and Fig. 1, we can see that the α_s values obtained for $N = 1-6$ of Y_{cut3}, Y_{cut4}, and Y_{cut5} are very stable and statistically consistent. The case $N = 6$ reduces the systematic error in α_s by a factor of 1.8 and increases the value of α_s, while increasing the statistical error by 30%.

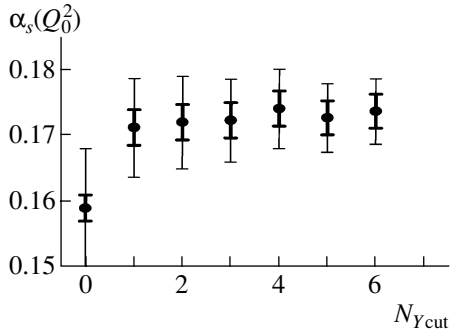

Fig. 1. The study of systematics at different Y-cut values in the fits based on nonsinglet evolution. The QCD analysis of BCDMS ^{12}C, H_2, D_2 data (nonsinglet case): $x_{\mathrm{cut}} = 0.25$, $Q_0^2 = 90\ \mathrm{GeV}^2$. Thresholds of c and b quarks are chosen at $Q^2 = 9\ \mathrm{GeV}^2$ and $Q^2 = 80\ \mathrm{GeV}^2$, respectively. The inner (outer) error bars show statistical (systematic) errors.

After the cuts have been implemented (in this section below, we use the set $N = 6$), we have 452 points in the analysis. Fitting them in agreement with the same procedure considered in the previous section, we obtain the following results:

$$\begin{aligned}\alpha_s(90\ \mathrm{GeV}^2) &= 0.1737 \pm 0.0029\ (\mathrm{stat.}) \\ &\pm 0.0050\ (\mathrm{syst.}) \pm 0.0025\ (\mathrm{norm.}), \\ \alpha_s(M_Z^2) &= 0.1153 \pm 0.0013\ (\mathrm{stat.}) \\ &\pm 0.0022\ (\mathrm{syst.}) \pm 0.0012\ (\mathrm{norm.}),\end{aligned} \tag{45}$$

where hereafter "norm." indicates the error of normalization of experimental data. Thus, the last error (± 0.0012 to $\alpha_s(M_Z^2)$) comes from the difference in fits with free and fixed normalizations of BCDMS data [7–9] having different values of energy.

So, for the fits with nonsinglet evolution of BCDMS data [7–9] with minimization of systematic errors, we have the following results:

$$\begin{aligned}\alpha_s(M_Z^2) &= 0.1153 \\ &\pm 0.0028\ \ (\text{total experimental error}).\end{aligned} \tag{46}$$

Here, total experimental error is the square root of the sum of squares of statistical error, systematic one, and error of normalization.

The value of $\alpha_s(M_Z^2)$ corresponds to the following value of the QCD mass parameter:

$$\begin{aligned}\Lambda^{(5)}_{\overline{MS}} &= 181 \pm 32\ (\text{total experimental error})\ \mathrm{MeV}, \\ \Lambda^{(4)}_{\overline{MS}} &= 257 \pm 40\ (\text{total experimental error})\ \mathrm{MeV}.\end{aligned} \tag{47}$$

4.1.2. The study of $N_{\max}$ dependence. Following [18, 19], we study the dependence of our results on the $N_{\max}$ value. The full set of data is 452 points. The Q^2 evolution starts at $Q_0^2 = 90\ \mathrm{GeV}^2$.

As can be seen in Table 3, our results are very stable, which is in very good agreement with [18].

Starting with $N_{\max} = 5$, where our results are already very stable, we put the results together and can calculate the average value of $\alpha_s(M_Z^2) = 0.1152$ and estimate the average deflection. The deflection is 0.0002 and can be considered as the error of the Jacobi polynomial expansion method, i.e., method error.

4.2. SLAC and NM $H_2 + D_2$ Data and BFP Fe Data

We continue our nonsinglet evolution analyses by fits of experimental data [4–6, 10] obtained by the SLAC, NM, and BFP Collaborations. The full set of data is 345 points (when $x \geq 0.25$): 238 from SLAC,

Table 3. The values of $\alpha_s(M_Z^2)$ at different values of $N_{\max}$

$N_{\max}$	$\chi^2(F_2)$/d.o.f.	χ^2(slope) for 6 points	$\alpha_s(90\ \mathrm{GeV}^2) \pm$ stat. (stat. = 0.0038)	$\alpha_s(M_Z^2) \pm$ stat. (stat. = 0.0013)
3	1.08	7.3	0.1720	0.1155
4	0.97	11.3	0.1715	0.1143
5	1.11	6.9	0.1729	0.1144
6	0.95	3.6	0.1747	0.1157
7	0.94	5.4	0.1740	0.1154
8	0.94	6.8	0.1738	0.1153
9	0.94	7.6	0.1735	0.1152
10	1.07	7.7	0.1735	0.1152
11	1.08	7.2	0.1726	0.1149
12	1.04	7.1	0.1731	0.1152
13	1.11	7.1	0.1725	0.1149

Table 4. The values of $\alpha_s(M_Z^2)$ and χ^2 at different regimes of fits

Fit	TMC	HTC	Syst. error	$\chi^2(F_2)$/d.o.f.	χ^2(slope) for 6 points	$\alpha_s(20\ \mathrm{GeV}^2) \pm$ stat.	$\alpha_s(M_Z^2)$
1	No	No	Yes	6.0	1050	0.2131 ± 0.0012	0.1167
2	Yes	No	Yes	2.3	224	0.2017 ± 0.0013	0.1133
3	Yes	Yes	No	1.8	12.0	0.2230 ± 0.0030	0.1195
4	Yes	Yes	Yes	0.8	6.1	0.2231 ± 0.0060	0.1195

66 from NM, and 41 from BFP. The starting point of QCD evolution is $Q_0^2 = 20\ \mathrm{GeV}^2$; the Q^2 cut is $Q^2 > 1\ \mathrm{GeV}^2$.

For illustration of importance of $1/Q^2$ corrections, we fit the data in the following way. First of all, we analyze the data applying only the perturbative QCD part of SF F_2, i.e., F_2^{tw2}. Later, we add $1/Q^2$ corrections: firstly, target mass and, later, twist-4 ones. As one can see in Table 4, we have a very bad fit when we work only with the twist-2 part F_2^{tw2}. The agreement with the data is improved substantially when TMC are added. The incorporation of twist-4 corrections leads to very good fit of the data. Neglect of systematic errors deteriorates twice our results. We combine the statistical and systematic errors in quadrature.

We have obtained the following values for parameters in parametrizations (24) of parton distributions (at $Q_0^2 = 20\ \mathrm{GeV}^2$):

$$A_{\mathrm{NS}}^P = 1.44, \quad A_{\mathrm{NS}}^D = 2.06, \quad A_{\mathrm{NS}}^F = 1.87, \quad (48)$$
$$b_{\mathrm{NS}}^P = 3.88, \quad b_{\mathrm{NS}}^D = 3.84, \quad b_{\mathrm{NS}}^F = 4.23,$$
$$d_{\mathrm{NS}}^P = 10.9, \quad d_{\mathrm{NS}}^D = 4.04, \quad d_{\mathrm{NS}}^F = 5.03,$$

where the symbols P, D, and F denote the parameters of parametrizations for proton, deuteron, and iron data, respectively.

We note that the values of the coefficients are close to those obtained in other numerical analyses (see [12, 13, 26, 27] and references therein). The values of b_{NS}^l ($l = P, D, F$) are in a quite good agreement with quark-counting rules of [72]. There is also good agreement with theoretical studies [73, 42].

The values of parameters of twist-4 terms are given in Table 5. We would like to note that the twist-4 terms for H_2 and D_2 data coincide with each other within the errors taken into account. It is in full agreement with analogous analysis [14].

We obtain the following results (at $\chi^2(F_2) = 250$, $\chi^2(\mathrm{slope}) = 6.1$ on six points):

$$\alpha_s(20\ \mathrm{GeV}^2) = 0.2231 \pm 0.0060\ (\mathrm{stat.}) \quad (49)$$
$$\pm 0.0075\ (\mathrm{syst.}) + 0.0030\ (\mathrm{norm.}),$$
$$\alpha_s(M_Z^2) = 0.1195 \pm 0.0017\ (\mathrm{stat.})$$
$$\pm 0.0022\ (\mathrm{syst.}) + 0.0010\ (\mathrm{norm.}).$$

Table 5. The values of the twist-4 terms

x_i	$\tilde{h}_4(x_i) \pm$ stat.	
	H_2	D_2
0.25	-0.149 ± 0.015	-0.176 ± 0.014
0.35	-0.151 ± 0.013	-0.178 ± 0.012
0.45	-0.214 ± 0.012	-0.147 ± 0.022
0.55	-0.228 ± 0.022	-0.065 ± 0.037
0.65	0.024 ± 0.070	0.053 ± 0.080
0.75	0.227 ± 0.154	0.130 ± 0.131

The last error (± 0.0010 to $\alpha_s(M_Z^2)$) comes from fits with free and fixed normalizations between different data of the SLAC, NM, and BFP Collaborations.

So, the fits of SLAC, NM, and BFP data based on the nonsinglet evolution give for coupling constant

$$\alpha_s(M_Z^2) = 0.1195 \pm 0.0030 \text{ (total experimental error)}, \quad (50)$$

which corresponds to the following value of the QCD mass parameter:

$$\Lambda^{(5)}_{\overline{MS}} = 231 \pm 37 \text{ (total experimental error) MeV},$$
$$\Lambda^{(4)}_{\overline{MS}} = 321 \pm 44 \text{ (total experimental error) MeV}, \quad (51)$$

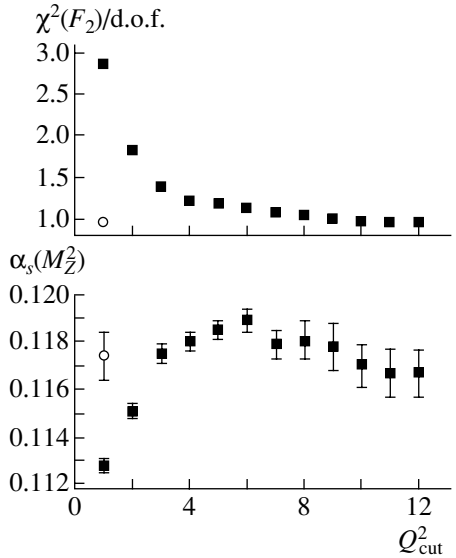

Fig. 2. The values of $\alpha_s(M_Z^2)$ and χ^2 at different Q^2 values of data cuts in the fits based on nonsinglet evolution. The black squares show the analyses of data without twist-4 contributions; the circles correspond to the case where twist-4 contributions were added. Only statistical errors are shown.

where the errors connected with the type of data normalization are already included in the systematic error.

Looking at the results obtained in two previous subsections, we see good agreement (within existing errors) between the values of the coupling constant $\alpha_s(M_Z^2)$ obtained separately in the fits of BCDMS data and in the fits of combined SLAC, NM, and BFP data [see Eqs. (45)–(47) and (49)–(51)]. Thus, we have the possibility to fit together all the data that are the subject of the following subsection.

4.3. SLAC, BCDMS, NM, and BFP Data

We use the following common x cut, $x \geq 0.25$, and Y cut with $N = 6$ (see Table 1) for the BCDMS data. After these cuts have been incorporated, the full set of data is 797 points. The starting point of QCD evolution is $Q_0^2 = 90\ \text{GeV}^2$.

4.3.1. The results of fits. We verify here the range of applicability of perturbative QCD. To do it, we analyze firstly the data without a contribution of twist-4 terms, i.e., when $F_2 = F_2^{\text{pQCD}}$. We do several fits using the cut $Q^2 \geq Q^2_{\text{cut}}$ and increase the value of Q^2_{cut} step by step. We observe good agreement of the fits with the data when $Q^2_{\text{cut}} \geq 10\ \text{GeV}^2$ (see Table 6).

Later, we add the twist-4 corrections and fit the data with the usual cut $Q^2 \geq 1\ \text{GeV}^2$. We have found very good agreement with the data. Moreover the predictions for $\alpha_s(M_Z^2)$ in both above procedures are very similar (see Table 6 and Fig. 2).

We have obtained the following values for parameters in parametrizations of parton distributions (at $Q_0^2 = 90\ \text{GeV}^2$):

$$A^P_{\text{NS}} = 2.40, \quad A^D_{\text{NS}} = 2.46, \quad A^C_{\text{NS}} = 2.46, \quad A^F_{\text{NS}} = 1.65,$$
$$b^P_{\text{NS}} = 3.98, \quad b^D_{\text{NS}} = 3.94, \quad b^C_{\text{NS}} = 4.08, \quad b^F_{\text{NS}} = 4.72,$$
$$d^P_{\text{NS}} = 4.85, \quad d^D_{\text{NS}} = 2.38, \quad d^C_{\text{NS}} = 1.55, \quad d^F_{\text{NS}} = 7.97. \quad (52)$$

The values are in good agreement with those presented in the previous subsection. Then all discussions given there can be applied here.

Table 7 contains the value of parameters of the twist-4 terms. As in the previous subsection, the twist-4 terms for H_2 and D_2 data coincide with each other within the errors taken into account, which is in agreement with [14].

So, the analysis of combined SLAC, NM, BCDMS, and BFP data gives the following results:

Table 6. The values of $\alpha_s(M_Z^2)$ and χ^2 for different regimes of fits

Fit	$Q^2_{\rm cut}$, GeV^2	Number of points	HTC	$\chi^2(F_2)$/d.o.f.	$\alpha_s(90\ \text{GeV}^2) \pm$ stat.	$\alpha_s(M_Z^2)$
1	1.0	797	No	2.87	0.1679 ± 0.0007	0.1128
2	2.0	772	No	1.82	0.1733 ± 0.0007	0.1151
3	3.0	745	No	1.38	0.1789 ± 0.0009	0.1175
4	4.0	723	No	1.23	0.1802 ± 0.0009	0.1180
5	5.0	703	No	1.19	0.1813 ± 0.0011	0.1185
6	6.0	677	No	1.13	0.1803 ± 0.0013	0.1189
7	7.0	650	No	1.09	0.1799 ± 0.0016	0.1179
8	8.0	632	No	1.06	0.1803 ± 0.0019	0.1181
9	9.0	613	No	1.01	0.1797 ± 0.0023	0.1178
10	10.0	602	No	0.98	0.1776 ± 0.0022	0.1170
11	11.0	688	No	0.97	0.1770 ± 0.0024	0.1167
12	12.0	574	No	0.97	0.1768 ± 0.0025	0.1167
13	1.0	797	Yes	0.97	0.1785 ± 0.0025	0.1174

Table 7. The values of the twist-4 terms $\tilde{h}_4(x_i) \pm$ stat. for H_2, D_2, C, and Fe data

x_i	H_2	D_2	x_i	C and Fe
0.275	-0.221 ± 0.010	-0.226 ± 0.010	0.250	-0.118 ± 0.187
0.350	-0.252 ± 0.010	-0.214 ± 0.010	0.350	-0.415 ± 0.233
0.450	-0.232 ± 0.019	-0.159 ± 0.020	0.450	-0.656 ± 0.494
0.550	-0.122 ± 0.360	-0.058 ± 0.300		
0.650	-0.159 ± 0.031	-0.057 ± 0.031		
0.750	0.040 ± 0.050	0.020 ± 0.049		

When twist-4 corrections are not included and the cut of Q^2 is 10 GeV^2 at the free normalization,

$$\chi^2/\text{d.o.f.} = 0.98 \text{ and } \alpha_s(90\ \text{GeV}^2) = 0.1776 \pm 0.0022\ (\text{stat.}), \quad (53)$$
$$\alpha_s(M_Z^2) = 0.1170 \pm 0.0009\ (\text{stat.}).$$

When twist-4 corrections are included and the cut of Q^2 is 1 GeV^2

$$\chi^2/\text{d.o.f.} = 0.97 \text{ and } \alpha_s(90\ \text{GeV}^2) = 0.1785 \pm 0.0025\ (\text{stat.}), \quad (54)$$
$$\alpha_s(M_Z^2) = 0.1174 \pm 0.0010\ (\text{stat.}).$$

Thus, as follows from nonsinglet fits of experimental data, perturbative QCD works rather well at $Q^2 \geq 10$ GeV^2.

4.3.2. The study of threshold effects. Here, we would like to study threshold effects in Q^2 evolution of SF F_2. Note that, at the NLO level in the nonsinglet case, the coefficient function of F_2 and anomalous dimension do not depend on the number f of active quarks. Then, our study of the threshold effects in Q^2 evolution of SF F_2 is exactly equal to the investigation of the role of threshold effects in the QCD coupling constant $\alpha_s(Q^2)$.

To study the threshold effects (with TMC and HTC included), we consider two types of possible thresholds of heavy quarks: $Q_f^2 = 4m_f^2$ and $Q_f^2 = m_f^2$. The first type of threshold appears when a heavy quark with mass m_f takes a possibility to be born. The second one lies close to the position of the Euclidean-reflected threshold of heavy quarks. It should play a significant role (see [55]) in the $\alpha_s(Q^2)$ evolution.

(i) Let thresholds appear at $Q_f^2 = 4m_f^2$. Then we split the range of data into three separate ranges:

Table 8. The values of $\alpha_s(M_Z^2)$ and χ^2 at different regimes of fits

Fit	Q^2, GeV^2	f	Q_0^2, GeV^2	Number of points	$\chi^2(F_2)$	$\Lambda\frac{(3)}{MS} \pm$ stat., MeV	$\Lambda\frac{(4)}{MS} \pm$ stat., MeV	$\Lambda\frac{(5)}{MS} \pm$ stat., MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	1–10	3	5	195	124	400 ± 30	308 ± 26	220 ± 23	0.1187 ± 0.0020
2	10–80	4	20	455	471		291 ± 17	208 ± 13	0.1177 ± 0.0012
3	80–300	5	90	190	143			199 ± 54	0.1169 ± 0.0040

Table 9. The values of $\alpha_s(M_Z^2)$ and χ^2 at different regimes of fits

Fit	Q^2, GeV^2	f	Q_0^2, GeV^2	Number of points	$\chi^2(F_2)$	$\Lambda\frac{(4)}{MS} \pm$ stat., MeV	$\Lambda\frac{(5)}{MS} \pm$ stat., MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	2.5–20.5	4	10	241	197	298 ± 10	213 ± 8	0.1181 ± 0.0007
2	20.5–300	5	90	558	533		187 ± 16	0.1159 ± 0.0014

Table 10. The values of Y_{cut3}, Y_{cut4}, and Y_{cut5}

N	0	1	2	3	4	5
Y_{cut3}	0	0.14	0.16	0.18	0.22	0.23
Y_{cut4}	0	0.16	0.18	0.20	0.23	0.24
Y_{cut5}	0	0.20	0.20	0.22	0.24	0.25

Table 11. The values of $\alpha_s(M_Z^2)$ at different values of N

N	Number of points	$\chi^2(F_2)$/d.o.f.	$\alpha_s(20\ \text{GeV}^2) \pm$ stat.	Total syst. error
0	762	1.22	0.1992 ± 0.0034	0.0122
1	649	1.06	0.2116 ± 0.0042	0.0096
2	640	1.07	0.2126 ± 0.0044	0.0088
3	627	1.05	0.2152 ± 0.0045	0.0080
4	596	1.04	0.2172 ± 0.0047	0.0076
5	590	1.04	0.2160 ± 0.0047	0.0068

The Q^2 values are between 1 and 10 GeV^2, where the number f of active quarks is 3.

The Q^2 values are between 10 and 80 GeV^2, where the number f of active quarks is 4.

The Q^2 values are above 80 GeV^2, where the number f of active quarks is 5.

The results are shown in Table 8. The average $\alpha_s(M_Z^2)$ value can be calculated and it has the following value:

$$\alpha_s(M_Z^2) = 0.1178 \pm 0.0010\,(\text{stat.}). \qquad (55)$$

(ii) Let thresholds appear at $Q_f^2 = m_f^2$. Then we split the range of data into two separate ranges:

The Q^2 values are between 2.5 and 20.5 GeV^2, where the number f of active quarks is 4.

The Q^2 values are above 20.5 GeV^2, where the number f of active quarks is 5.

The results are shown in Table 9. The average $\alpha_s(M_Z^2)$ value can be calculated and it has the following value:

$$\alpha_s(M_Z^2) = 0.1176 \pm 0.0006\,(\text{stat.}). \qquad (56)$$

Thus, we do not find a strong dependence on the exact value of thresholds of heavy quarks, i.e., between Eqs. (55) and (56). However, there is a difference between these results and results (54), which shows that the theoretical uncertainties due to threshold effects can be estimated for $\alpha_s(M_Z^2)$ at 0.0004.

4.4. The Results of the Analysis Based on Nonsinglet Evolution

Thus, using the analyses based on nonsinglet evolution of the SLAC, NM, BCDMS, and BFP experimental data for SF F_2, we obtain for $\alpha_s(M_Z^2)$ the following expressions:

(i) When we switch off the twist-4 corrections and set the cut $Q^2 > 10$ GeV^2, we get at $\chi^2/\text{d.o.f.} = 0.98$

$$\alpha_s(M_Z^2) = 0.1170 \pm 0.0009\,(\text{stat.}) \qquad (57)$$
$$\pm 0.0019\,(\text{syst.}) \pm 0.0010\,(\text{norm.})$$

or

$$\alpha_s(M_Z^2) = 0.1170 \qquad (58)$$
$$\pm 0.0023\ \ (\text{total experimental error}).$$

Table 12. The values of $\alpha_s(M_Z^2)$ and χ^2 for different regimes of fits

Fit	TMC	HTC	Syst. error	$\chi^2(F_2)$/d.o.f.	χ^2(slope) for 23 points	$\alpha_s(20\ \mathrm{GeV}^2) \pm$ stat.	$\alpha_s(M_Z^2)$
1	No	No	Yes	5.5	800	0.2400 ± 0.0017	0.1241
2	Yes	No	Yes	2.2	179	0.2153 ± 0.0018	0.1174
3	Yes	Yes	Yes	0.85	21	0.2138 ± 0.0058	0.1170

Table 13. The values of $\alpha_s(M_Z^2)$ and χ^2 in fits with different values of W^2 cut

Fit	W^2 cut, GeV^2	$\chi^2(F_2)$/d.o.f.	$\alpha_s(20\ \mathrm{GeV}^2) \pm$ stat.	$\Lambda\frac{(4)}{MS} \pm$ stat., MeV	$\Lambda\frac{(5)}{MS} \pm$ stat., MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	2.0	1.30	0.2407 ± 0.0013	400 ± 6	296 ± 4	0.1243 ± 0.0004
2	4.0	1.00	0.2135 ± 0.0018	280 ± 7	194 ± 5	0.1169 ± 0.0004
3	6.0	1.00	0.2070 ± 0.0023	253 ± 9	178 ± 7	0.1150 ± 0.0007
4	8.0	0.91	0.2128 ± 0.0043	277 ± 18	197 ± 14	0.1167 ± 0.0012
5	10	0.91	0.2107 ± 0.0053	268 ± 22	190 ± 18	0.1162 ± 0.0015

(ii) When we add the twist-4 corrections and set the cut $Q^2 > 1\ \mathrm{GeV}^2$, we get at $\chi^2/\mathrm{d.o.f.} = 0.97$

$$\alpha_s(M_Z^2) = 0.1174 \pm 0.0007\ (\mathrm{stat.}) \pm 0.0021\ (\mathrm{syst.}) \pm 0.0005\ (\mathrm{norm.}) \quad (59)$$

or

$$\alpha_s(M_Z^2) = 0.1174 \pm 0.0022\ \ (\mathrm{total\ experimental\ error}). \quad (60)$$

Looking at the results obtained in this section, we see that the central value of the coupling constant $\alpha_s(M_Z^2)$ obtained in the fits (based on nonsinglet evolution) of combined SLAC, BCDMS, NM, and BFP data lies between the central values of the coupling constants obtained separately in the fits of BCDMS data and in the fits of SLAC, BCDMS, NM, and BFP data. All obtained values of $\alpha_s(M_Z^2)$ are in good agreement within existing statistical errors.

5. RESULTS OF FITS OF F_2: THE COMBINED NONSINGLET AND SINGLET EVOLUTION

For this case, the quite low x experimental data lie in the low-Q^2 range and we choose $Q_0^2 = 20\ \mathrm{GeV}^2$. We use also $N_{\max} = 8$.

The study of the $N_{\max}$ dependence of the results in the combined nonsinglet and singlet case of evolution has been found in [19]. Note here only that the analysis in [19] shows the $N_{\max}$ independence of the obtained results starting already with $N_{\max} = 7$.

Except the studies done in Section 5.3.2, the effects of the thresholds of heavy quarks are taken into account in agreement with Eqs. (33) and (34).

5.1. BCDMS $^{12}C + H_2 + D_2$ Data

As in the previous section, we start our analyses with the experimental data [7–9] obtained by the BCDMS muon scattering experiment. The full set of data is 762 points. The starting point of QCD evolution is $Q_0^2 = 20\ \mathrm{GeV}^2$.

As in the nonsinglet evolution case, we have studied the influence of the experimental systematic errors on the results of the QCD analysis as a function of Y_{cut3}, Y_{cut4}, and Y_{cut5} applied to the data. Here we also use several sets N of the values for the cuts at $0.5 < x \leq 0.8$, which are given in Table 10.

The absolute differences between the values of α_s for the distorted and undistorted sets of data are given in Table 11 and Fig. 3 as the total systematic

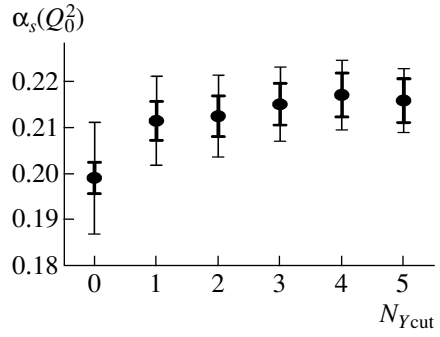

Fig. 3. The study of systematics at different Y-cut values in the fits based on combined singlet and nonsinglet evolution. The QCD analysis of BCDMS ^{12}C, H_2, D_2 data (the case of combined evolution): no x_{cut}, $Q_0^2 = 20\ \mathrm{GeV}^2$. Thresholds of c and b quarks are chosen at $Q^2 = 9\ \mathrm{GeV}^2$ and $Q^2 = 80\ \mathrm{GeV}^2$, respectively. The inner (outer) error bars show statistical (systematic) errors.

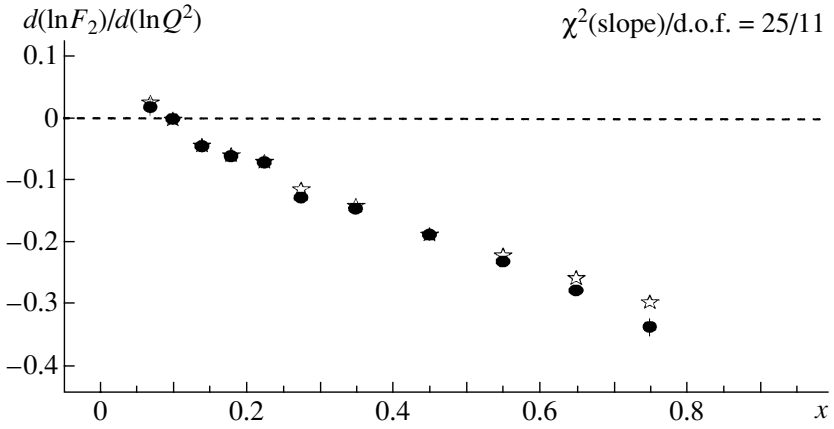

Fig. 4. The values of the slope $d(\ln F_2)/d(\ln Q^2)$ at $Q^2 = 20$ GeV^2. The stars correspond to the theoretical predictions based on combined singlet and nonsinglet evolution; the closed circles show BCDMS ^{12}C, H_2, and D_2 data without a Y cut.

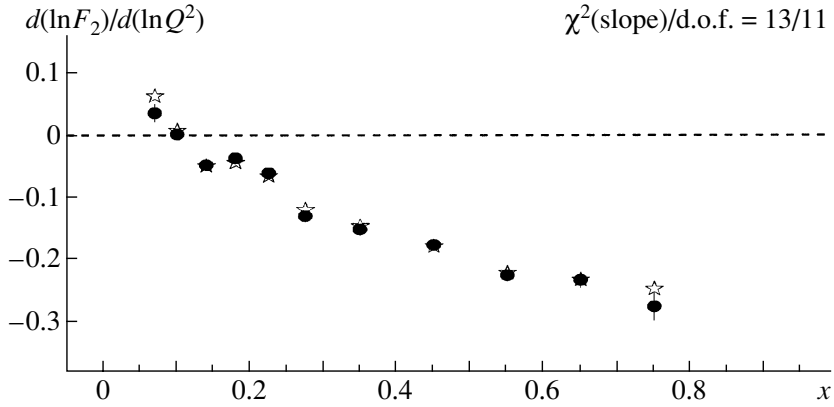

Fig. 5. The same as in Fig. 4 with one exception: the Y cut with $N = 5$ is taken into account.

error of α_s estimated in quadratures. The number of the experimental points and the value of α_s for the undistorted set of F_2 are also given in Table 11 and Fig. 3.

From Table 11 and Fig. 3, we can see that the α_s values obtained for $N = 1$–5 of Y_{cut3}, Y_{cut4}, and Y_{cut5} are very stable and statistically consistent. The case $N = 5$ reduces the systematic error in α_s by a factor of 1.8 and increases the value of α_s while increasing the statistical error by 27%.

The importance of the Y cut can be shown also in Figs. 4 and 5, where the slope $d(\ln F_2)/d(\ln Q^2)$ has been shown at $Q^2 = 20$ GeV^2. As we can see, there is a significant improvement (the corresponding χ^2(slope) decreases by half) when the Y cut has been taken into account.

After the cuts have been implemented (in this section below, we use the set $N = 5$), we have 590 points in the analysis. Fitting them in agreement with the same procedure considered in Section 3, we obtain the following results:

$$\alpha_s(20\ \text{GeV}^2) = 0.2160 \pm 0.0047\ (\text{stat.}) \pm 0.0068\ (\text{syst.}) \pm 0.0031\ (\text{norm.}), \quad (61)$$

$$\alpha_s(M_Z^2) = 0.1175 \pm 0.0014\ (\text{stat.}) \pm 0.0020\ (\text{syst.}) \pm 0.0011\ (\text{norm.}).$$

As in the nonsinglet case, the last error (± 0.0011 to $\alpha_s(M_Z^2)$) comes from the difference in fits with free and fixed normalizations of BCDMS data [7–9] having different values of energy.

So, for the fits of BCDMS data [7–9] based on complete singlet and nonsinglet evolution with minimization of systematic errors, we have the following results (total experimental error is the square root of the sum of squares of statistical error, systematic error, and error of normalization):

$$\alpha_s(M_Z^2) = 0.1175 \pm 0.0026\ (\text{total experimental error}). \quad (62)$$

The value of $\alpha_s(M_Z^2)$ corresponds to the following value of the QCD mass parameter:

$$\Lambda^{(4)}_{\overline{MS}} = 290 \pm 20\,(\text{stat.}) \pm 29\,(\text{syst.})\,\text{MeV}, \quad (63)$$
$$\Lambda^{(5)}_{\overline{MS}} = 206 \pm 17\,(\text{stat.}) \pm 24\,(\text{syst.})\,\text{MeV},$$

where the errors connected with the type of normalization of data are already included in the systematic error.

5.2. SLAC and NM $H_2 + D_2$ Data and BFP Fe Data

We continue our analyses with the experimental data [4–10] obtained by the SLAC, NM, and BFP Collaborations. The full set of data is 719 points (with the cut $Q^2 > 1$ GeV^2): 364 from SLAC, 300 from NM, and 55 from BFP. The starting point of QCD evolution is $Q_0^2 = 20$ GeV^2.

As in the previous section, we give an illustration of the importance of $1/Q^2$ corrections. First of all, we analyze the data applying only the perturbative QCD part of SF F_2, i.e., F_2^{tw2}. Later, we add the $1/Q^2$ corrections: firstly, TMC and, later, twist-4 corrections. As one can see in Table 12 and Figs. 6–8, we have a very bad fit (χ^2(slope)/d.o.f. = 40) when we work only with twist-2 part F_2^{tw2}. The agreement with the data is significantly better (χ^2(slope)/d.o.f. ≈ 9) when TMC have been added. The incorporation of twist-4 corrections leads to a very good fit of the data: χ^2(slope)/d.o.f. ≈ 1.05 (see Table 12 and Fig. 8). We note that the statistical and systematic errors are combined in quadratures.

Thus, we see that χ^2(slope)/d.o.f. decreases by 38 times when the $1/Q^2$ corrections have been taken into account.

Looking at the results in Table 12, we see the following results for coupling constants:

$$\alpha_s(20\text{ GeV}^2) = 0.2138 \pm 0.0058\,(\text{stat.}) \quad (64)$$
$$\pm 0.0075\,(\text{syst.}) + 0.0030\,(\text{norm.}),$$
$$\alpha_s(M_Z^2) = 0.1170 \pm 0.0016\,(\text{stat.})$$
$$\pm 0.0021\,(\text{syst.}) + 0.0011\,(\text{norm.}).$$

As in the nonsinglet evolution fits, the last error ± 0.0011 to $\alpha_s(M_Z^2)$ comes from fits with free and fixed normalizations between different data of the SLAC, NM, and BFP Collaborations.

We would like to compare the results in Table 12 with the results of the analyses of the data when an additional W^2 cut is taken into account. The inclusion of the W^2 cut is very popular (see [74] and references therein) and we fit the data considered with variation of the W^2 cut (and with the standard cut $Q^2 > 1$ GeV^2). The results of the fits (without twist-4 correction) are presented in Table 13 (the systematic errors of the data are included in the fits).

As we can see from Tables 12 (the last line, where twist-4 corrections are incorporated) and 13, the results for $\alpha_s(M_Z^2)$ are in very good agreement for values of W^2 cut larger than 4 GeV^2. So, the W^2-cut procedure can be used successfully to switch off the range of experimental data where HTC are required.

We would like to note that the results obtained in two previous subsections show a very good agreement between the values of coupling constant $\alpha_s(M_Z^2)$ obtained separately in the fits of BCDMS data and in the fits of combined SLAC, NM, and BFP data [see Eqs. (61), (63) and (64)]. Thus, as in the case of nonsinglet evolution, we have the possibility to fit together all the data. This is the subject of the following subsection.

5.3. SLAC, BCDMS, NM, and BFP Data

Here, we start to analyze the maximal number of experimental points which have been produced in the considered experiments. The full set of data is 1309 points.

5.3.1. The study of Q^2 range where $1/Q^2$ corrections are important. Here, we would like to repeat our analysis given in Section 4.3. Firstly, we fit the data without a contribution of twist-4 terms. We use the cut $Q^2 \geq Q^2_{\text{cut}}$ and increase the value of Q^2_{cut} step by step. Later, we do fits including the twist-4 corrections and the cut $Q^2 \geq 1$ GeV^2.

As in the nonsinglet case, we observe a very good agreement for the first type of fits starting with $Q^2_{\text{cut}} \geq 15$ GeV^2 (see Table 14 and Fig. 9). For the second type of fits, the agreement is already good at $Q^2 \geq 1$ GeV^2. Both types of fits give very similar results. Moreover, the results are very close to those obtained earlier in the nonsinglet case (see Table 6 and Fig. 2).

We obtain the following results:

When twist-4 corrections are not included and the cut of Q^2 is 15 GeV^2,

$$\chi^2/\text{d.o.f.} = 1.14 \text{ and } \alpha_s(20\text{ GeV}^2) \quad (65)$$
$$= 0.2177 \pm 0.0042\,(\text{stat.}),$$
$$\alpha_s(M_Z^2) = 0.1180 \pm 0.0013\,(\text{stat.}).$$

When twist-4 corrections are included and the cut of Q^2 is 1 GeV^2,

$$\chi^2/\text{d.o.f.} = 1.11 \text{ and } \alpha_s(20\text{ GeV}^2) \quad (66)$$
$$= 0.2167 \pm 0.0024\,(\text{stat.}),$$
$$\alpha_s(M_Z^2) = 0.1177 \pm 0.0007\,(\text{stat.}).$$

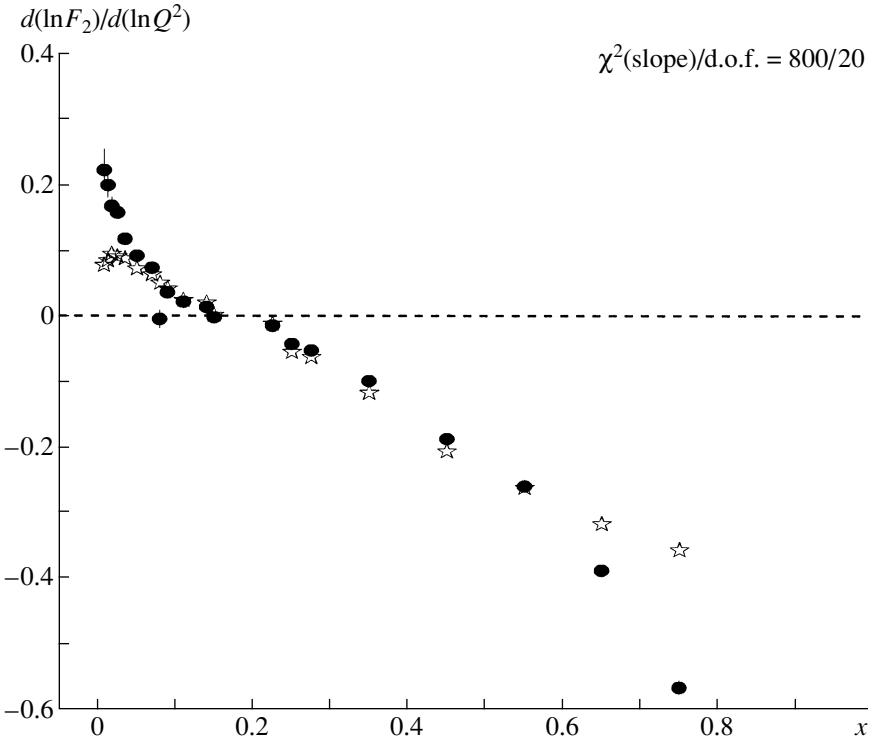

Fig. 6. The values of the slope $d(\ln F_2)/d(\ln Q^2)$ at $Q^2 = 20$ GeV^2. The stars correspond to the theoretical predictions based on the twist-2 approximation of perturbative QCD and combined singlet and nonsinglet evolution; the closed circles show SLAC, NM, and BFP experimental data.

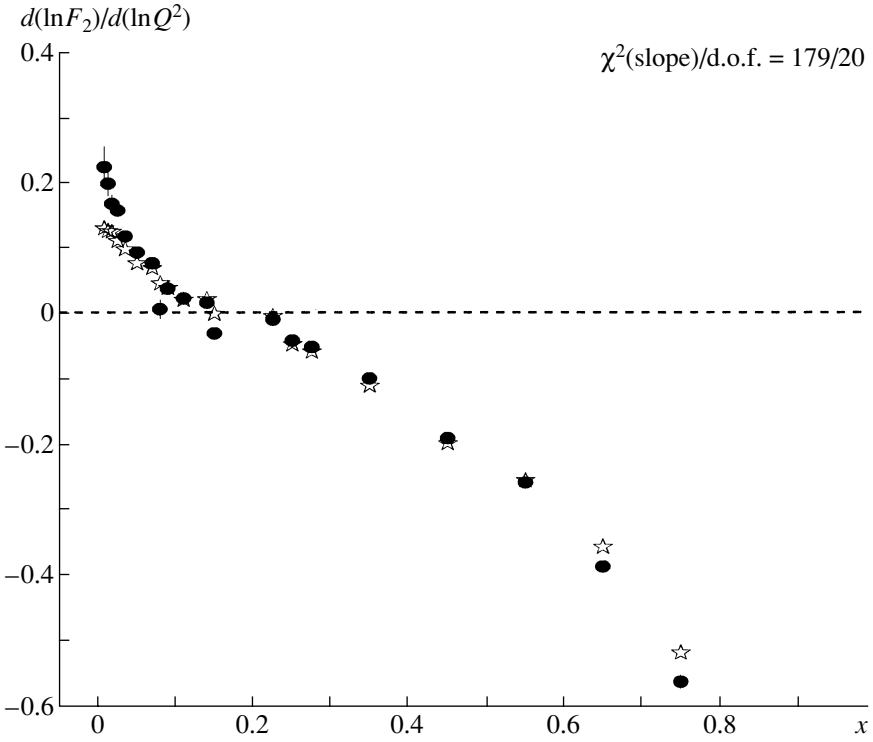

Fig. 7. The same as in Fig. 6 with one exception: TMC are taken into account for theoretical predictions.

Table 14. The values of $\alpha_s(M_Z^2)$ and χ^2 for different regimes of fits

Fit	$Q^2_{\rm cut}$, GeV^2	Number of points	HTC	$\chi^2(F_2)$/d.o.f.	$\alpha_s(20\ \text{GeV}^2) \pm$ stat.	$\Lambda\frac{(4)}{MS}$, MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	1.0	1309	No	1.55	0.2258 ± 0.0011	333	0.1203 ± 0.0004
2	4.0	1051	No	1.27	0.2364 ± 0.0017	380	0.1232 ± 0.0005
3	6.0	942	No	1.30	0.2385 ± 0.0022	390	0.1237 ± 0.0005
4	8.0	870	No	1.32	0.2232 ± 0.0035	321	0.1196 ± 0.0010
5	10.0	817	No	1.27	0.2226 ± 0.0035	318	0.1194 ± 0.0011
6	11.0	793	No	1.21	0.2187 ± 0.0038	301	0.1183 ± 0.0011
7	12.0	758	No	1.18	0.2192 ± 0.0039	304	0.1185 ± 0.0011
8	13.0	754	No	1.17	0.2180 ± 0.0039	297	0.1181 ± 0.0012
9	14.0	740	No	1.17	0.2169 ± 0.0041	294	0.1178 ± 0.0013
10	15.0	714	No	1.14	0.2177 ± 0.0042	297	0.1180 ± 0.0013
11	1.0	1309	Yes	1.11	0.2167 ± 0.0024	293	0.1177 ± 0.0007

Table 15. The values of $\alpha_s(M_Z^2)$ and χ^2 for different regimes of fits

Fit	Q^2, GeV^2	f	Q_0^2, GeV^2	Number of points	χ^2	$\Lambda\frac{(3)}{MS} \pm$ stat., MeV	$\Lambda\frac{(4)}{MS} \pm$ stat., MeV	$\Lambda\frac{(5)}{MS} \pm$ stat., MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	1–10	3	3.0	467	290	331 ± 24	250 ± 20	176 ± 16	0.1148 ± 0.0015
2	10–80	4	20	627	595		274 ± 21	194 ± 17	0.1165 ± 0.0014
3	80–300	5	90	190	156			220 ± 70	0.1187 ± 0.0050

Table 16. The values of $\alpha_s(M_Z^2)$ and χ^2 for different regimes of fits

Fit	Q^2, GeV^2	f	Q_0^2, GeV^2	Number of points	χ^2	$\Lambda\frac{(4)}{MS} \pm$ stat., MeV	$\Lambda\frac{(5)}{MS} \pm$ stat., MeV	$\alpha_s(M_Z^2) \pm$ stat.
1	2.5–20.5	4	10	519	396	230 ± 21	160 ± 16	0.1132 ± 0.0016
2	20.5–300	5	90	631	670		205 ± 15	0.1174 ± 0.0013

For additional illustration of the importance of $1/Q^2$ corrections at nonlarge Q^2 values, we study the slope $d(\ln F_2)/d(\ln Q^2)$ as has been done in Section 5.2. First of all, we analyze the data applying only the perturbative QCD approximation of SF F_2 (with TMC taken into account), i.e., $F_2^{\rm pQCD}$. Later, we add the cut $Q^2 \geq 15$ GeV^2. As can be seen in Figs. 10 and 11, we have a bad fit (χ^2(slope)/d.o.f. ≈ 7.78) in the case without a Q^2 cut. The agreement with the data is much better when this Q^2 cut has been added: χ^2(slope)/d.o.f. ≈ 1.26 in the case.

As in the previous subsection, the incorporation of twist-4 corrections also leads to a very good fit of the data (without a Q^2 cut): χ^2(slope)/d.o.f. ≈ 1.09 (see Fig. 12). These results demonstrate the importance of twist-4 corrections at nonlarge Q^2 values.

Thus, as follows from the fits of experimental data based on combined singlet and nonsinglet evolution, perturbative QCD works well at $Q^2 \geq 15$ GeV^2.

5.3.2. The study of threshold effects. Here, we continue our study of threshold effects in Q^2 evolution of SF F_2. Note that, at LO and NLO levels in the singlet case of evolution, the coefficient functions of F_2 and anomalous dimensions depend on the number f of active quarks.

By analogy with the nonsinglet case of evolution (see Section 4.3.2), to study the threshold effects (with TMC and HTC included), we consider two types of possible thresholds of heavy quarks: $Q_f^2 = 4m_f^2$ and $Q_f^2 = m_f^2$. The first type of threshold appears

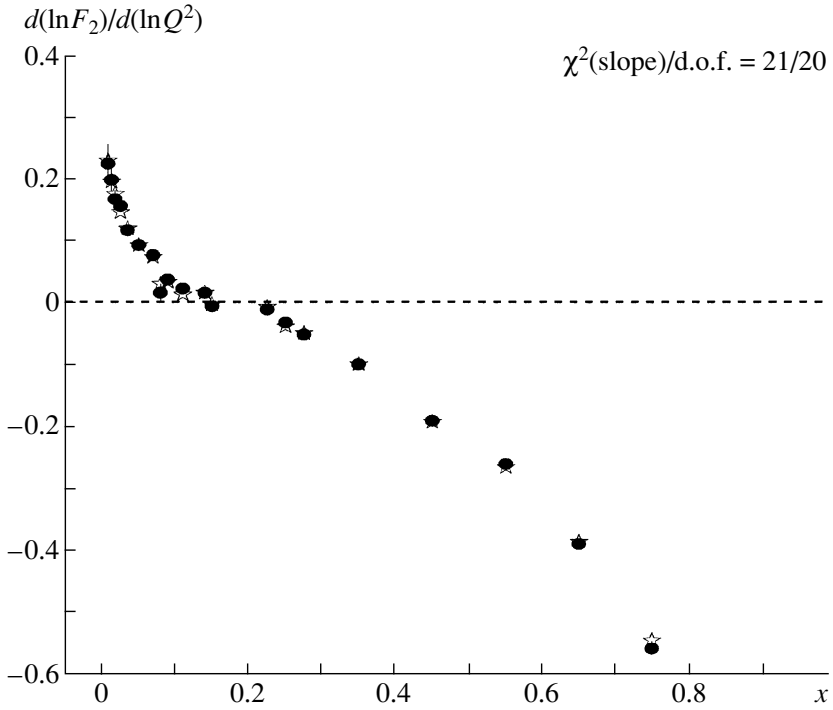

Fig. 8. The same as in Fig. 6 with one exception: TMC and twist-4 corrections are taken into account for theoretical predictions.

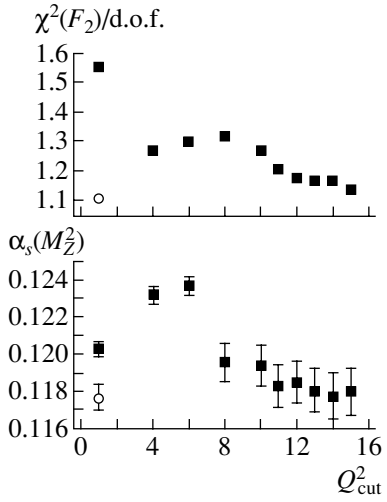

Fig. 9. The values of $\alpha_s(M_Z^2)$ and χ^2 at different Q^2 values of data cuts in the fits based on combined singlet and nonsinglet evolution. The closed squares show the analyses of data without twist-4 contributions; the circles correspond to the case where twist-4 contributions were added. Only statistical errors are shown.

when a heavy quark with mass m_f takes a possibility to be born (in the framework of the photon–gluon fusion process, for example). The second one lies close to the position of the Euclidean-reflected threshold of heavy quarks. It should play a significant role (see [55]) in the $\alpha_s(Q^2)$ evolution.

(i) Let thresholds appear at $Q_f^2 = 4m_f^2$. Then we split the range of data into three separate ranges (see Section 4.3.2). The results are shown in Table 15. The average $\alpha_s(M_Z^2)$ can be calculated and it has the following value:

$$\alpha_s(M_Z^2) = 0.1158 \pm 0.0010\,(\text{stat.}). \qquad (67)$$

(ii) Let thresholds appear at $Q_f^2 = m_f^2$. Then we split the range of data into two separate ranges (see Section 4.3.2). The results are shown in Table 16. The average $\alpha_s(M_Z^2)$ can be calculated and it has the following value:

$$\alpha_s(M_Z^2) = 0.1157 \pm 0.0020\,(\text{stat.}). \qquad (68)$$

The results are very surprising. On one hand, all variables, the coefficient functions of F_2 and anomalous dimensions, depend on the number f of active quarks. However, we do not find a strong dependence on exact value of thresholds of heavy quarks. On the other hand, the central values of the average $\alpha_s(M_Z^2)$ obtained here are substantially lower than in our other analyses [see Eqs. (66)–(68)].

Thus, the theoretical uncertainties due to threshold effects can be estimated in the case of combined singlet and nonsinglet evolution for $\alpha_s(M_Z^2)$ as

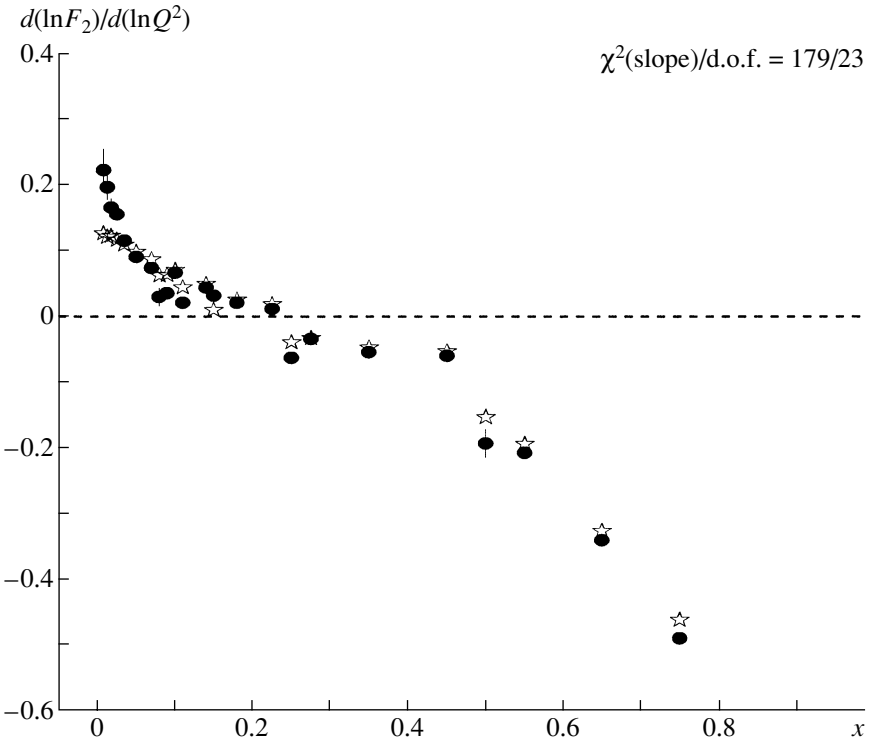

Fig. 10. The values of the slope $d(\ln F_2)/d(\ln Q^2)$ at $Q^2 = 20$ GeV^2. The stars correspond to the theoretical predictions based on perturbative QCD (with TMC taken into account) and combined singlet and nonsinglet evolution. The closed circles show SLAC, BCDMS, NM, and BFP experimental data without a Q^2 cut.

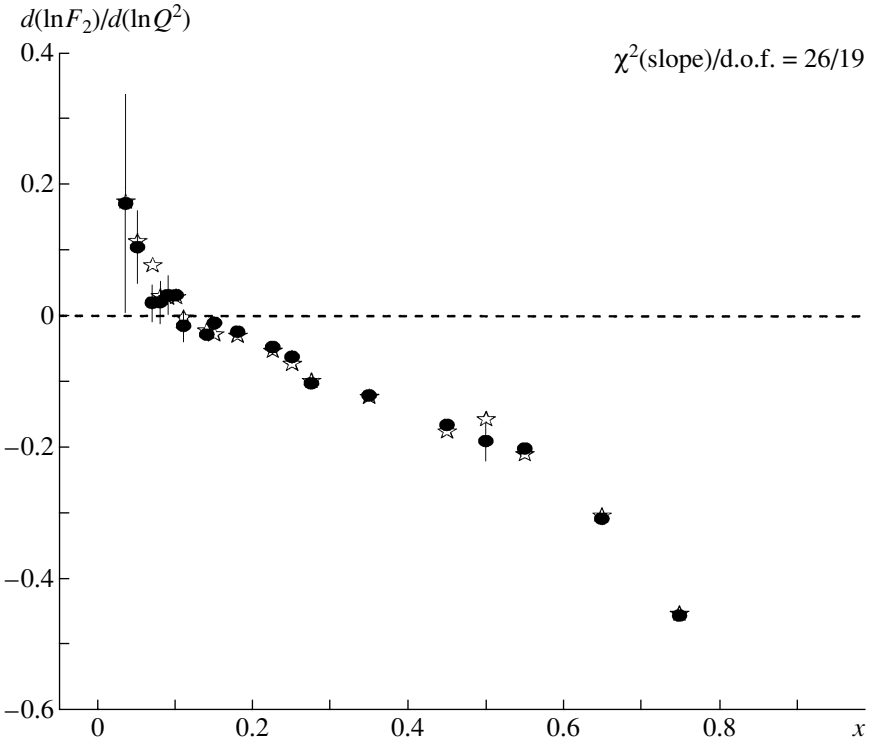

Fig. 11. The same as in Fig. 10 with one exception: the cut $Q^2 \geq 15$ GeV^2 is taken into account for experimental data.

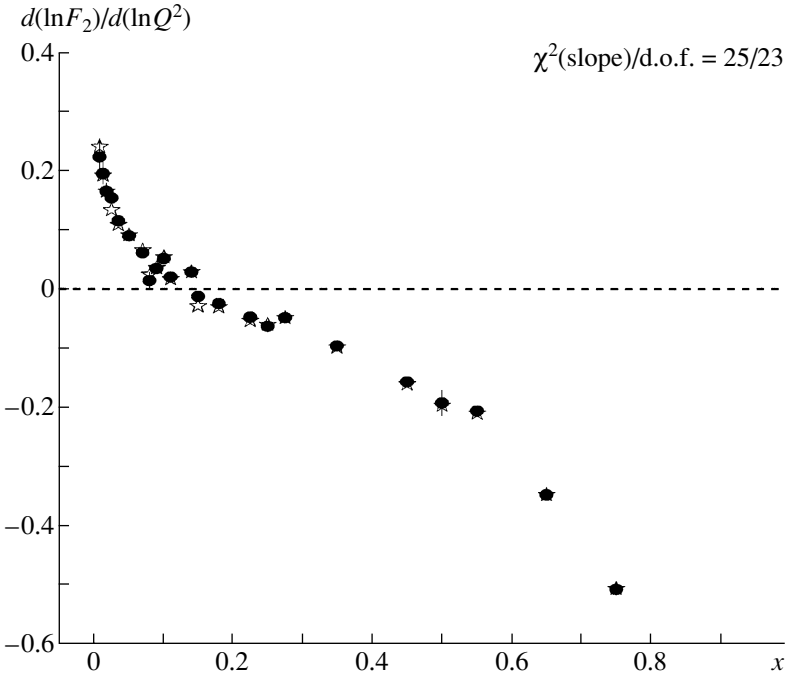

Fig. 12. The same as in Fig. 10 with one exception: the twist-4 corrections are taken into account for theoretical predictions.

0.0020, mostly due to the shift of central values of the average $\alpha_s(M_Z^2)$ in Eqs. (67) and (68).

5.3.3. The values of fitted parameters. We have obtained the following values for parameters in parametrizations (26), (27), and (39) with $a_G = a_{\rm sea} = 0$ of parton distributions (at $Q_0^2 = 20\,{\rm GeV}^2$):[20)]

$$\begin{aligned} a_u(20) &= 0.72, \quad b_u(20) = 3.72, \\ a_d(20) &= 0.69, \quad b_d(20) = 5.81, \\ C_{\rm sea}(20) &= 0.375, \quad b_{\rm sea}(20) = 13.8, \\ P_G(20) &= 0.519, \quad b_G(20) = 11.4, \\ K_1^C(20) &= 1.222, \quad K_2^C(20) = 0.554, \\ K_3^C(20) &= 0.253, \\ K_1^F(20) &= 1.10, \quad K_2^F(20) = -0.081, \\ K_3^F(20) &= -0.58. \end{aligned} \tag{69}$$

For the coefficients $a_u(20)$ and $a_d(20)$, we find good agreement between their values and the double-logarithmic estimations in [75, 76], based on [77]. We would like to note that the estimations in [75] have been given in another set of parameters that changes effectively only the value of normalization point Q_0^2. As has been shown in [78, 42], the value of $a_u(20)$ and $a_d(20)$ should be nearly Q^2 independent (if the values are not too close to 1).[21)] This Q^2 independence of values of $a_u(20)$ and $a_d(20)$ explains our good agreement with the results of [75]. The values of $a_u(20)$ and $a_d(20)$ are supported also by recent fits (see discussions in [27]).

The value $b_u(20)$ is in agreement with Eqs. (48) and (52) and with other fits [12, 13, 26, 27], which supports its slow Q^2 dependence (see [73, 42]). The value of $b_d(20)$ is higher than $b_u(20)$, which is supported by other fits (see, for example, [12, 13] and references therein) and by quark-counting rules [72]. The values of $b_G(20)$ and $b_{\rm sea}(20)$ are very high, which is in agreement with BCDMS analyses [7–9] and demonstrates the difficulties in studying the large-x asymptotics of sea-quark and gluon distributions in analyses of inclusive deep-inelastic data.[22)]

The value of $P_G(20)$ shows that, at $Q_0^2 = 20\,{\rm GeV}^2$, gluons contain about half of the nucleon momentum.

[20)] Here and in the following subsection, we give the results for the coefficient $P_G(Q_0^2)$, but not for $C_G(Q_0^2)$. They are connected because of Eq. (31): $P_G(Q_0^2) = C_G(Q_0^2)B(a_g(Q_0^2)+1, b_g(Q_0^2)+1)$, where the beta function $B(a,b)$ has been defined in Eq. (26).

[21)] This Q^2 independence is very similar to corresponding Q^2 independence of the coefficients $a_{\rm sea}(20)$ and $a_G(20)$ in the powerlike small-x asymptotics $\sim x^{a_{\rm sea}}$ and $\sim x^{a_G}$ of singlet parton distributions if $a_{\rm sea}$ and a_G are not close numerically to 0 (see studies in [47, 79–82] and references therein).

[22)] In the semi-inclusive case of DIS, the gluons give large contributions, essentially at low-x values (see, for example, the recent study of open charm production in [83] and references therein), and, thus, the gluon distribution can be perfectly extracted.

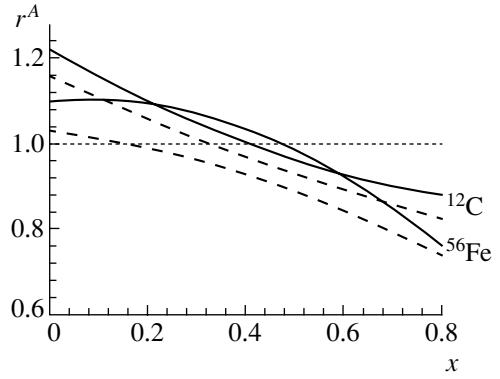

Fig. 13. The values of the nuclear-effect ratio r^A for $A = C$ and F. The solid and dashed curves correspond to the small-x asymptotics $\sim x^{-\omega}$ of sea-quark and gluon distributions with $\omega = 0$ and 0.18, respectively.

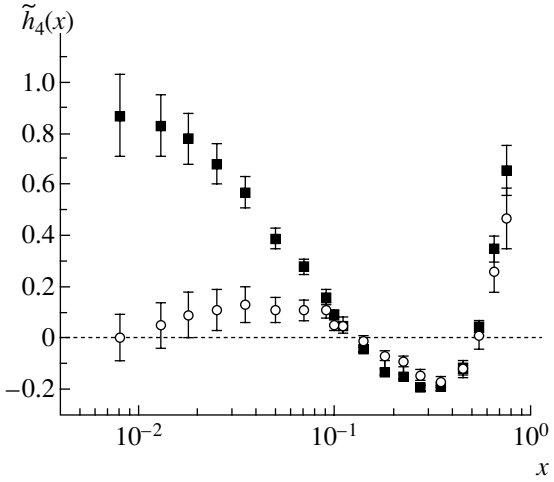

Fig. 14. The values of the twist-4 terms. The closed and open symbols correspond to the small-x asymptotics $\sim x^{-\omega}$ of sea-quark and gluon distributions with $\omega = 0$ and $\omega = 0.18$, respectively. Only statistical errors are displayed.

The coefficients $K_i^C(20)$ and $K_i^F(20)$ $(i = 1, 2, 3)$ demonstrate nonzero values of nuclear effects for bound nucleons in ^{12}C and Fe nuclei. Equations (39) with the values of coefficients $K_i^C(20)$ and $K_i^F(20)$ given in (69) demonstrate the shapes of nuclear effects, which are represented in Fig. 13, where we see a reasonable agreement of our curves with the experimental data from [57, 60].

The values of twist-4 terms are given in Table 17. To obtain the values, we used the approximate equality of twist-4 terms for H_2 and D_2 targets that has been obtained in our studies in the previous section (see Tables 5 and 7). This is also in agreement with [14]. The values of twist-4 terms are also represented in Fig. 14.

We would like to note (see Table 17 and Fig. 14) a quite strong rise of twist-4 terms in lower x bins. The necessity of a large magnitude of twist-4 corrections at low-x values can be observed also in Figs. 6, 7, and 10, where there is a quite strong difference between experimental data and theoretical predictions (based on perturbative QCD) for the slope $d(\ln F_2)/d(\ln Q^2)$. The rise is in good agreement with theoretical predictions [84] and with the recent analyses of H1 and ZEUS data at low values of x and Q^2 (see [33]).

5.3.4. BFKL-like parametrizations of gluon and sea-quark distributions. As we have already discussed in Section 2, we would like to try to study the parameters of the sea-quark and gluon distributions when the terms $\sim x^{a_{\rm sea}(Q_0^2)}$ and $\sim x^{a_G(Q_0^2)}$ are incorporated. These terms take into account a possible rise of the sea-quark and gluon distributions at low-x values. As has been already noted in Section 2, from DGLAP-like analyses [80, 47, 48], the parameters $a_{\rm sea}$ and a_G should be the same, because they are mixed together into the "plus" component of the Q^2 evolution (see [47]). Moreover, the parameter $\omega = -a_{\rm sea} = -a_G$ should be Q^2 independent (see, for example, [80, 48]), if it is not small, i.e., $x^{-\omega} \gg$ const at small x.

In the fit with free nonzero ω value, we have obtained the following values for parameters (26), (27), and (39) with $a_G \neq 0$ and $a_{\rm sea} \neq 0$ in parametrizations of parton distributions (at $Q_0^2 = 20$ GeV^2):[23)]

$$\begin{gathered} a_u(20) = 0.72, \quad b_u(20) = 3.69, \\ a_d(20) = 0.68, \quad b_d(20) = 5.44, \\ a_{\rm sea}(20) = -0.18, \quad a_G(20) = -0.18, \\ C_{\rm sea}(20) = 0.185, \quad b_{\rm sea}(20) = 10.4, \\ P_G(20) = 0.524, \quad b_G(20) = 7.31, \\ K_1^C(20) = 1.160, \quad K_2^C(20) = 0.472, \\ K_3^C(20) = 0.141, \\ K_1^F(20) = 1.03, \quad K_2^F(20) = 0.131, \\ K_3^F(20) = -0.28. \end{gathered} \tag{70}$$

We would like to note that the values of parameters of valent-quark distributions are not really changed. The values of $b_G(20)$ and $b_{\rm sea}(20)$ are still high, but they are closer to predictions of quark-counting

[23)] We would like to note that the fit contains strong correlations between the values of ω, the coupling constant, and twist-4 terms. These correlations occur because a very limited amount of experimental data used here lie in the low-x region. Indeed, only the NMC experimental data contribute there. Then, the results (70) can be considered seriously only when H1 and ZEUS data [43, 44] have been taken into account. We hope to incorporate the HERA data [43, 44] in our future investigations.

rules [72] than the corresponding values obtained in the previous subsection.

The values of the parameters of the nuclear-effect ratio are not changed within the considered errors. The similarity of the results for the nuclear-effect ratio is shown in Fig. 13.

The value of ω is equal to 0.18, which is in perfect agreement with the recent studies based on BFKL dynamics [85] when NLO corrections [86, 87] were taken into account (see, for example, studies [88] and review [89] and references therein). Moreover, this value is in good agreement also with recent phenomenological studies (see recent review in [90]) of Pomeron intercept values and also with recent H1 and L3 data [43, 91].

As one can see in Tables 17 and 18 and also in Fig. 14, the effect of the strong rise of twist-4 magnitude at small-x values observed in the previous subsection is completely absent here.[24)] So, the rise is replaced by the small-x rise of twist-2 gluon and sea-quark distributions. This replacement seems to be due to the small number of experimental points in the low-x range and the narrow range of Q^2 values there. The cancellation of twist-4 corrections at low x is in good agreement with the recent studies [32, 92]. This demonstrates the fact that the strong rise of twist-4 corrections coming from BFKL-like approaches [84] has negligible magnitude (see [92, 33]).

The value of $\alpha_s(M_Z^2)$ in the fit (with the number of points 1309 and $\chi^2/\text{d.o.f.} = 1.1$) is as follows:

$$\alpha_s(M_Z^2) = 0.1187 \pm 0.0015\,(\text{stat.}); \qquad (71)$$

i.e., it is in good agreement within statistical errors with fits performed earlier, but the middle value is slightly higher.

5.4. The Results of Analyses with Combined Singlet and Nonsinglet Evolution

Thus, using singlet analyses of the SLAC, NM, BCDMS, and BFP experimental data for SFs F_2 we obtain for $\alpha_s(M_Z^2)$ the following expression:

$$\begin{aligned}\alpha_s(20\ \text{GeV}^2) &= 0.2167 \pm 0.0024\,(\text{stat.}) \qquad (72)\\ &\pm 0.0080\,(\text{syst.}) \pm 0.0012\,(\text{norm.}),\\ \alpha_s(M_Z^2) &= 0.1177 \pm 0.0007\,(\text{stat.})\\ &\pm 0.0021\,(\text{syst.}) \pm 0.0005\,(\text{norm.}).\end{aligned}$$

[24)] As in the previous subsection, to obtain the values, we used the approximate equality of twist-4 terms for H_2 and D_2 targets that have been obtained in our studies in the previous section (see Tables 5 and 7). This is also in agreement with [14].

Looking at the results obtained in previous sections, we see a very good agreement between the value of coupling constant $\alpha_s(M_Z^2)$ obtained in the fits of combined SLAC, BCDMS, NM, and BFP data and the values of $\alpha_s(M_Z^2)$ obtained separately in the fits of BCDMS data and in ones of SLAC, BCDMS, NM, and BFP data.

6. THE DEPENDENCE ON FACTORIZATION AND RENORMALIZATION SCALES

In this section, we study the dependence of our results on the different choice of the factorization scale μ_F and the renormalization one μ_R. Following studies [14, 36], we choose three values (1/2, 1, 2) for the coefficients k_F and k_R.

6.1. Nonsinglet Evolution Case

The results are given in Table 19. We do fits here without higher twist corrections (no HTC), with the number of points 596, at $Q^2 > 10.5\ \text{GeV}^2$ and for free normalization of different sets of data. The change of the value of coupling constant $\alpha_s(M_Z^2)$ at some k_F and k_R values is denoted by the difference

$$\Delta\alpha_s(M_Z^2) = \alpha_s(M_Z^2) - \alpha_s(M_Z^2)|_{k_F=k_R=1}. \qquad (73)$$

We find similar variation of $\alpha_s(M_Z^2)$ with the variations of k_F and k_R: $\alpha_s(M_Z^2)$ increases (falls) with increasing (decreasing) values of k_F and/or k_R. So, the dependence is quite similar to the one which has been obtained in [39, 26, 27] by the variation of k scale from 1/4 to 4 ($k \equiv k_F = k_R$ in [39, 26, 27]).

Taking the maximal and minimal values (which corresponds to $k_R = k_F = 1/2$ and 2, respectively) of the coupling constant, we obtain the scale uncertainties $+0.0050$ and -0.0034 for $\alpha_s(M_Z^2)$. In the case when the replacement (22) has been used also in NLO corrections to the coefficient functions [i.e., when Eq. (22) replaces Eq. (20) there], the theoretical uncertainties for $\alpha_s(M_Z^2)$ are a little higher: $+0.0070$ and -0.0041.

Thus, using the analyses with NS evolution of the SLAC, NMC, BCDMS, and BFP experimental data for SFs F_2, we obtain for $\alpha_s(M_Z^2)$ the following expressions (when no HTC, $Q^2 > 10\ \text{GeV}^2$, and $\chi^2 = 0.98$):

$$\begin{aligned}\alpha_s(M_Z^2) &= 0.1170 \pm 0.0009\,(\text{stat.}) \qquad (74)\\ &\pm 0.0019\,(\text{syst.}) \pm 0.0010\,(\text{norm.})\\ &+ \begin{cases}+0.0050\\ -0.0034\end{cases}(\text{theor.})\end{aligned}$$

Table 17. The values of the twist-4 terms

x_i	$\tilde{h}_4(x_i) \pm$ stat.	x_i	$\tilde{h}_4(x_i) \pm$ stat.	x_i	$\tilde{h}_4(x_i) \pm$ stat.
0.008	0.87 ± 0.16	0.090	0.16 ± 0.03	0.275	-0.19 ± 0.01
0.013	0.83 ± 0.12	0.100	0.09 ± 0.02	0.350	-0.19 ± 0.01
0.018	0.78 ± 0.10	0.110	0.05 ± 0.03	0.450	-0.12 ± 0.02
0.025	0.68 ± 0.08	0.140	-0.04 ± 0.01	0.500	0.45 ± 0.23
0.035	0.57 ± 0.06	0.150	0.43 ± 0.11	0.550	0.04 ± 0.03
0.050	0.39 ± 0.04	0.180	-0.13 ± 0.01	0.650	0.35 ± 0.05
0.070	0.28 ± 0.03	0.225	-0.15 ± 0.01	0.750	0.66 ± 0.10
0.080	0.30 ± 0.15	0.250	-0.27 ± 0.13		

Table 18. The values of the twist-4 terms for BFKL-like PDF form

x_i	$\tilde{h}_4(x_i) \pm$ stat.	x_i	$\tilde{h}_4(x_i) \pm$ stat.	x_i	$\tilde{h}_4(x_i) \pm$ stat.
0.008	0.004 ± 0.090	0.090	0.11 ± 0.03	0.275	-0.14 ± 0.02
0.013	0.05 ± 0.09	0.100	0.05 ± 0.02	0.350	-0.17 ± 0.02
0.018	0.09 ± 0.09	0.110	0.05 ± 0.03	0.450	-0.12 ± 0.03
0.025	0.11 ± 0.08	0.140	-0.01 ± 0.02	0.500	0.43 ± 0.23
0.035	0.13 ± 0.07	0.150	0.62 ± 0.12	0.550	0.01 ± 0.05
0.050	0.11 ± 0.05	0.180	-0.07 ± 0.02	0.650	0.26 ± 0.08
0.070	0.11 ± 0.04	0.225	-0.09 ± 0.02	0.750	0.47 ± 0.12
0.080	0.31 ± 0.16	0.250	-0.16 ± 0.14		

or

$$\alpha_s(M_Z^2) = 0.1170 \quad (75)$$
$$\pm 0.0023 \text{ (total experimental error)}$$
$$+ \begin{cases} +0.0050 \\ -0.0034 \end{cases} \text{(theor.)},$$

where "theor." marks the theoretical uncertainties which contain the sum of the scale uncertainties, threshold error (± 0.0004), and the method error (± 0.0002) in quadratures.

6.2. Combined Singlet and Nonsinglet Evolution

The results are given in Table 20. We do fits with HTC with the number of points 1309 at $Q^2 > 1\ \text{GeV}^2$ and for free normalization of different sets of data.

We find that variations of $\alpha_s(M_Z^2)$ with the variations of k_F and k_R are very similar to those obtained in the previous subsection. However, there is quite a big difference in the cases $k_R = 2$, $k_F = 1/2$ and $k_R = 1/2$, $k_F = 2$ between the results in Table 20 in and without parentheses. The difference seems to come from the correlations between the values of higher order contributions (mimicked by scale dependences) and twist-4 corrections, i.e., so-called duality effect (see [27] and references therein).

As in the case of nonsinglet evolution, the dependence of $\alpha_s(M_Z^2)$ with the variations of k_F and k_R is quite similar to the one which has been obtained in [39] by the variation of k_R scale from 1/4 to 4.

Taking maximal and minimal values (which corresponds to $k_R = k_F = 1/2$ and 2, respectively) of the coupling constant, we obtain the scale uncertainties $+0.0047$ and -0.0057 for $\alpha_s(M_Z^2)$. In the case when the replacement (22) has been used also in NLO corrections to the coefficient functions (i.e., when Eq. (22) replaces Eq. (20) there), the theoretical uncertainties for $\alpha_s(M_Z^2)$ are changed very little, but $\chi^2(F_2)$ is higher.

Thus, using these analyses of the SLAC, NM, BCDMS, and BFP experimental data for SF F_2, we obtain

$$\alpha_s(M_Z^2) = 0.1177 \pm 0.0007\,(\text{stat.}) \quad (76)$$
$$\pm 0.0021\,(\text{syst.}) \pm 0.0005\,(\text{norm.})$$
$$+ \begin{cases} +0.0051 \\ -0.0061 \end{cases} \text{(theor.)},$$

where the theoretical uncertainties contain the scale errors (see above), the errors due to threshold effects (± 0.0020), and the method error (± 0.0002) in quadratures.

Table 19. The values of $\alpha_s(M_Z^2)$ at different values of k_F and k_R for nonsinglet evolution (the values in parentheses correspond to the case when Eq. (22) replaces Eq. (20) in the NLO corrections to coefficient functions)

k_R	k_F	$\chi^2(F_2)$	$\alpha_s(90\ \mathrm{GeV}^2) \pm$ stat.	$\alpha_s(M_Z^2)$	$\Delta\alpha_s(M_Z^2)$
1	1	556	0.1789 ± 0.0023	0.1175	0
1/2	1	558	0.1769 ± 0.0022	0.1167	−0.0008
			(0.1745)	(0.1155)	(−0.0020)
1	1/2	545	0.1730 ± 0.0021	0.1150	−0.0025
1	2	568	0.1876 ± 0.0025	0.1211	+0.0036
2	1	555	0.1826 ± 0.0025	0.1191	+0.0016
			(0.1858)	(0.1203)	(+0.0028)
1/2	2	570	0.1856 ± 0.0026	0.1203	+0.0028
			(0.1817)	(0.1186)	(+0.0011)
2	1/2	554	0.1770 ± 0.0022	0.1167	−0.0008
			(0.1784)	(0.1173)	(−0.0002)
1/2	1/2	556	0.1789 ± 0.0023	0.1175	−0.0034
			(0.1694)	(0.1134)	(−0.0041)
2	2	567	0.1912 ± 0.0028	0.1225	+0.0050
			(0.1965)	(0.1245)	(+0.0070)

Table 20. The values of $\alpha_s(M_Z^2)$ at different values of k_F and k_R for combined evolution (the values in parentheses correspond to the case when Eq. (22) replaces Eq. (20) in the NLO corrections to coefficient functions)

k_R	k_F	$\chi^2(F_2)$	$\alpha_s(20\ \mathrm{GeV}^2) \pm$ stat.	$\Lambda\frac{(4)}{MS}$, MeV	$\Lambda\frac{(5)}{MS}$, MeV	$\alpha_s(M_Z^2)$	$\Delta\alpha_s(M_Z^2)$
1	1	1410	0.2167 ± 0.0024	293	209	0.1178	0
1/2	1	1410	0.2112 ± 0.0019	270	191	0.1162	−0.0016
		(1443)	(0.2104 ± 0.0029)	(267)	(189)	(0.1160)	(−0.0018)
1	1/2	1423	0.2040 ± 0.0020	241	168	0.1140	−0.0038
1	2	1447	0.2300 ± 0.0031	351	256	0.1215	+0.0037
2	1	1413	0.2204 ± 0.0024	309	222	0.1189	+0.0011
		(1500)	(0.2263 ± 0.0030)	(334)	(242)	(0.1204)	(+0.0026)
1/2	2	1422	0.2190 ± 0.0029	303	217	0.1185	+0.0007
		(1500)	(0.2132 ± 0.0031)	(278)	(197)	(0.1167)	(−0.0011)
2	1/2	1460	0.2021 ± 0.0022	233	162	0.1134	−0.0044
		(1496)	(0.2323 ± 0.0030)	(361)	(264)	(0.1220)	(+0.0042)
1/2	1/2	1436	0.1975 ± 0.0012	216	149	0.1120	−0.0058
		(1450)	(0.1970 ± 0.0018)	(214)	(148)	(0.1120)	(−0.0058)
2	2	1447	0.2340 ± 0.0033	369	271	0.1225	+0.0047
		(1460)	(0.2343 ± 0.0032)	(370)	(271)	(0.1226)	(+0.0048)

In conclusion of this section, we would like to note that the theoretical uncertainties in both types of analyses (based on nonsinglet evolution and on combined singlet and nonsinglet evolution) are substantially larger than the corresponding total experimental errors.

Indeed, the total experimental errors are as follows: in the analyses with nonsinglet evolution,

$$\text{(total experimental error)} \tag{77}$$
$$= \begin{cases} \text{(stat.)} + \text{(syst.)} + \text{(norm.)} = 0.0038 \\ \text{(total linear experimental error)} \\ \sqrt{\text{(stat.)}^2 + \text{(syst.)}^2 + \text{(norm.)}^2} = 0.0023 \\ \text{(total quadratic experimental error);} \end{cases}$$

in the analyses with combined singlet and nonsinglet evolution,

$$\text{(total experimental error)} \tag{78}$$
$$= \begin{cases} 0.0033\ \text{(total linear experimental error)} \\ 0.0023\ \text{(total quadratic experimental error);} \end{cases}$$

i.e., they are less by a factor of 1.5–2 compared with the corresponding theoretical uncertainties.

As has been shown in [26, 27, 39, 40], the theoretical uncertainties decrease significantly (by a factor around 2.5) when NNLO corrections have been taken into account. So, the fits of combined data show the real necessity for the analysis of DIS data in the NNLO approximation.

7. SUMMARY

As a conclusion, we would like to stress again that, using the Jacobi polynomial expansion method, developed in [17–19], we have studied the Q^2 evolution of DIS structure function F_2 fitting all modern experimental data existing at the values of Bjorken variable x: $x \geq 10^{-2}$.

1. From the fits, we have obtained the value of the normalization $\alpha_s(M_Z^2)$ of the QCD coupling constant. First of all, we have reanalyzed the BCDMS data, cutting the range with large systematic errors. As one can see in Sections 4.1 and 5.1 (and also Figs. 1 and 3), the values of $\alpha_s(M_Z^2)$ rise strongly when the cuts of systematics were incorporated. On the other hand, the values of $\alpha_s(M_Z^2)$ do not depend on the concrete type of cut within modern statistical errors.

The values $\alpha_s(M_Z^2)$ obtained in various fits are in good agreement with one another. Indeed, we have very similar results for $\alpha_s(M_Z^2)$ in separate analyses of BCDMS data (with the cuts of systematics) and other data. This makes it possible for us to fit all the data together.

We have found that, at $Q^2 \geq 10$–15 GeV^2, the formulas of pure perturbative QCD (i.e., twist-2 approximation together with target mass corrections) are in good agreement with all data. The results for $\alpha_s(M_Z^2)$ are very similar for both types of analyses: based on nonsinglet evolution and on combined singlet and nonsinglet evolution. They have the following form:

$$\alpha_s(M_Z^2) = 0.1170 \pm 0.0009\,\text{(stat.)} \pm 0.0019\,\text{(syst.)} \pm 0.0010\,\text{(norm.)} \tag{79}$$

from fits based on nonsinglet evolution,

$$\alpha_s(M_Z^2) = 0.1180 \pm 0.0013\,\text{(stat.)} \pm 0.0021\,\text{(syst.)} \pm 0.0009\,\text{(norm.)} \tag{80}$$

from fits based on combined singlet and nonsinglet evolution.

When we add twist-4 corrections, we have very good agreement between QCD (i.e., the first two coefficients of the Wilson expansion) and data starting already with $Q^2 = 1$ GeV^2, where the Wilson expansion should begin to be applicable. The results for $\alpha_s(M_Z^2)$ coincide for both types of analyses: based on nonsinglet evolution and on combined singlet and nonsinglet evolution. They have the following form:

$$\alpha_s(M_Z^2) = 0.1174 \pm 0.0007\,\text{(stat.)} \pm 0.0019\,\text{(syst.)} \pm 0.0010\,\text{(norm.)} \tag{81}$$

from fits based on nonsinglet evolution,

$$\alpha_s(M_Z^2) = 0.1177 \pm 0.0007\,\text{(stat.)} \pm 0.0021\,\text{(syst.)} \pm 0.0009\,\text{(norm.)} \tag{82}$$

from fits based on combined singlet and nonsinglet evolution.

Thus, there is a very good agreement [see Eqs. (79)–(82)] between the results based on pure perturbative QCD at quite large Q^2 values (i.e., at $Q^2 \geq 10$–15 GeV^2) and the results based on the first two twist terms of the Wilson expansion (at $Q^2 \geq 1$ GeV^2, where the Wilson expansion should be applicable).

We would like to note that we have good agreement also with the analysis [69] of combined H1 and BCDMS data, which has been given by the H1 Collaboration very recently. The shapes of twist-4 corrections are very similar to the ones from [14, 93]. Our results for $\alpha_s(M_Z^2)$ are in good agreement also with the average value for the coupling constant presented in recent studies (see [26, 39, 40, 65, 74, 94, 95] and references therein) and in the famous review by Bethke [96].

2. As the second item of our summary, we would like to note the real importance of NNLO corrections in analyses of DIS experimental data. The incorporation of the NNLO corrections started several years ago in various ways (see Introduction for discussions).

The results are based on the studies of the effect of high-order corrections, which can be estimated from the dependence of our results on the factorization scale μ_F and renormalization one μ_R. As has already been pointed out in the previous section, the value of the theoretical uncertainties[25)] coming from this dependence of the results for $\alpha_s(M_Z^2)$ [given by Eqs. (74) and (76) for two types of Q^2 evolution] is equal to

$$\Delta\alpha_s(M_Z^2)|_{\text{theor}} = \begin{cases} +0.0051 \\ -0.0061. \end{cases} \qquad (83)$$

Thus, the theoretical uncertainties are substantially higher than the total experimental error (78). Similar values of the theoretical error can be found in recent analyses of the DIS process (see [39, 40, 26]) and of the e^+e^- process in [94, 95]. As has been studied recently by van Neerven and Vogt [39, 40], the value of the theoretical error decreases sharply (by a factor around 2.5) when the NNLO corrections have been taken into account. Thus, our fits of combined data performed here and also other analyses [94, 95] show the real necessity to include the NNLO corrections in the study of DIS experimental data.

As has been noted in the Introduction, using partial information about NNLO QCD corrections, several fits of experimental data have been performed (see [22–27, 39, 40, 65, 97] and references therein). In order to do the analyses of experimental data in the full range of x values, it is necessary to know all NNLO QCD corrections exactly.[26)] At present, three-loop corrections to anomalous dimensions of Wilson operators are still unknown. These calculations, which are known only for a finite number of fixed Mellin moments [101], will be performed [102] in the near future by using modern approaches (see [37, 102, 103]) to evaluate complicated Feynman diagrams.

3. At the end of our paper, we would like to discuss the contributions of higher twist corrections.

In our study here, we have reproduced the well-known x shape of the twist-4 corrections at large and intermediate values of Bjorken variable x (see, for example, Tables 5, 7, and 17 and also, for example, the results of a very popular article [14]).

We would like to note the small-x rise of the magnitude of twist-4 corrections when we use flat parton distributions at $x \to 0$. The rise is in full agreement with the theoretical predictions [84]. As we have already discussed in Section 5, there is a strong correlation between the small-x behavior of twist-4 corrections and the type of corresponding asymptotics of the leading-twist parton distributions. The possibility of having a singular type of asymptotics leads (in our fits) to the appearance of the rise of sea-quark and gluon distributions as $\sim x^{-0.18}$ at low-x values. In this case, the rise of the magnitude of twist-4 corrections is completely canceled. This cancellation is in full agreement with theoretical and phenomenological studies and low-x experimental data of the H1 and L3 Collaborations (see discussions in Section 5.3.4).

We would also like to give a few words concerning the IRR-model predictions for the twist-4 and twist-6 corrections.

In our previous study [21] based on the IRR-model predictions for higher twist corrections, we have found a strong correlation between these corrections and the value of the coupling constant. The $\alpha_s(M_Z^2)$ value tends to be very small: $\alpha_s(M_Z^2) = 0.103 \pm 0.002$ (stat.). This study has been supported by the fits of the DELPHI Collaboration (see [104]) and by some other analyses [95]. There is, however, a disagreement with the results of [105], where the twist-4 corrections in the framework of the IRR model do not lead to a decrease in the $\alpha_s(M_Z^2)$ value. In our opinion, the situation is not so clear here and it needs more investigation. We hope to return to this problem in our future studies.

ACKNOWLEDGMENTS

We are grateful to Sergei Alekhin, Sergei Mikhailov, and Alexander Nagaitsev for useful discussions. We are grateful also to the anonymous referee for his very useful notes and comments.

One of us (A.V.K.) was supported in part by an Alexander von Humboldt fellowship and INTAS, grant no. 366.

[25)] As has been already shown, the scale choices $\mu_F = \mu_R = 2Q^2$ and $\mu_F = \mu_R = Q^2/2$ give the maximal and minimal values of $\alpha_s(M_Z^2)$ (for the various choices of values $k_F = 1/2$, $k_F = 2$, $k_R = 1/2$, and $k_R = 2$ separately) and, thus, give the basic part of the theoretical error. The additional theoretical uncertainties due to our method error and choice of threshold points are negligible.

[26)] The three-loop corrections to anomalous dimensions of Wilson operators and to the longitudinal structure function $F_L(x, Q^2)$ have been calculated during the time of publication of the paper (see [98] and [99], respectively). So, at the moment, the complete NNLO analysis of the DIS SFs $F_2(x, Q^2)$ and $F_L(x, Q^2)$ can be performed. We plan to do it with the help of the Jacobi polynomial expansion method (see Section 3) and the results [98, 99], adopted in [100].

REFERENCES

1. M. Maul, E. Stein, L. Mankiewicz, *et al.*, Phys. Lett. B **401**, 100 (1997); hep-ph/9710392; V. M. Braun, Preprint NORDITA 97/53-P (1997); hep-ph/9708386; B. R. Webber, Nucl. Phys. B (Proc. Suppl.) **71**, 66 (1999).
2. M. Beneke, Phys. Rep. **317**, 1 (1999); M. Beneke and V. M. Braun, Preprints PITHA-00-25, TPR-00-19 (2000); hep-ph/0010208.
3. K. Long, R. Nisius, and W. J. Stirling, in *Proceedings of the 9th International Workshop on Deep Inelastic Scattering, Bologna, Italy, 2001*, p. 233; hep-ph/0109092.
4. SLAC Collab. (L. W. Whitlow *et al.*), Phys. Lett. B **282**, 475 (1992).
5. SLAC Collab. (L. W. Whitlow), PhD Thesis (Standford University, Standford, 1990); SLAC Report No. 357 (1990).
6. NM Collab. (M. Arneodo *et al.*), Nucl. Phys. B **483**, 3 (1997).
7. BCDMS Collab. (A. C. Benevenuti *et al.*), Phys. Lett. B **223**, 485 (1989); Preprint CERN-EP/89-06 (1989).
8. BCDMS Collab. (A. C. Benevenuti *et al.*), Phys. Lett. B **237**, 592 (1990); Preprint CERN-EP/89-170 (1989).
9. BCDMS Collab. (A. C. Benevenuti *et al.*), Phys. Lett. B **195**, 91 (1987); Preprint CERN-EP/87-100 (1987).
10. BFP Collab. (P. D. Mayers *et al.*), Phys. Rev. D **34**, 1265 (1986).
11. A. P. Bukhvostov, E. A. Kuraev, and L. N. Lipatov, Yad. Fiz. **38**, 429 (1983) [Sov. J. Nucl. Phys. **38**, 263 (1983)]; A. B. Bukhvostov, V. S. Fadin, E. A. Kuraev, *et al.*, Nucl. Phys. B **258**, 601 (1985).
12. S. I. Alekhin, Eur. Phys. J. C **10**, 395 (1999).
13. S. I. Alekhin, Phys. Rev. D **63**, 094 022 (2001).
14. M. Virchaux and A. Milsztajn, Phys. Lett. B **274**, 221 (1992).
15. A. D. Martin, R. G. Roberts, W. J. Stirling, *et al.*, Eur. Phys. J. C **14**, 155 (2000); M. Glueck, E. Reya, and A. Vogt, Eur. Phys. J. C **5**, 4611 (1998); CTEQ Collab. (H. Lai *et al.*), Eur. Phys. J. C **12**, 375 (2000).
16. V. N. Gribov and L. N. Lipatov, Yad. Fiz. **15**, 781 (1972) [Sov. J. Nucl. Phys. **15**, 438 (1972)]; L. N. Lipatov, Yad. Fiz. **20**, 181 (1975) [Sov. J. Nucl. Phys. **20**, 94 (1975)]; G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977); Yu. L. Dokshitzer, Zh. Éksp. Teor. Fiz. **73**, 1216 (1977) [Sov. Phys. JETP **46**, 641 (1977)].
17. G. Parisi and N. Sourlas, Nucl. Phys. B **151**, 421 (1979); I. S. Barker, C. B. Langensiepen, and G. Shaw, Nucl. Phys. B **186**, 61 (1981); I. S. Barker, B. R. Martin, and G. Shaw, Z. Phys. C **19**, 147 (1983); I. S. Barker and B. R. Martin, Z. Phys. C **24**, 255 (1984).
18. V. G. Krivokhizhin, S. P. Kurlovich, V. V. Sanadze, *et al.*, Z. Phys. C **36**, 51 (1987).
19. V. G. Krivokhizhin, S. P. Kurlovich, R. Lednicky, *et al.*, Z. Phys. C **48**, 347 (1990).
20. V. I. Vovk, Z. Phys. C **47**, 57 (1990); A. V. Kotikov, G. Parente, and J. Sanchez Guillen, Z. Phys. C **58**, 465 (1993).
21. A. V. Kotikov and V. G. Krivokhijine, in *Proceedings of the 6th International Workshop on Deep Inelastic Scattering and QCD, Brussels, Belgium, 1998*, p. 242; hep-ph/9805353.
22. G. Parente, A. V. Kotikov, and V. G. Krivokhizhin, Phys. Lett. B **333**, 190 (1994).
23. A. L. Kataev, A. V. Kotikov, G. Parente, *et al.*, Phys. Lett. B **388**, 179 (1996); A. V. Sidorov, Phys. Lett. B **389**, 379 (1996).
24. A. V. Sidorov and M. V. Tokarev, Phys. Lett. B **358**, 353 (1995); Nuovo Cimento A **109**, 1591 (1996); **110**, 1401 (1997); S. A. Kulagin and A. V. Sidorov, Eur. Phys. J. A **9**, 261 (2000).
25. A. L. Kataev, A. V. Kotikov, G. Parente, *et al.*, Phys. Lett. B **417**, 374 (1998); Nucl. Phys. B (Proc. Suppl.) **64**, 138 (1998).
26. A. L. Kataev, G. Parente, and A. V. Sidorov, Nucl. Phys. B **573**, 405 (2000).
27. A. L. Kataev, G. Parente, and A. V. Sidorov, Phys. Part. Nucl. **34**, 20 (2003).
28. A. Buras, Rev. Mod. Phys. **52**, 199 (1980).
29. K. G. Chetyrkin, B. A. Kniehl, and M. Steinhauser, Phys. Rev. Lett. **79**, 2184 (1997).
30. F. J. Yndurain, *Quantum Chromodynamics (An Introduction to the Theory of Quarks and Gluons)* (Springer-Verlag, Berlin, 1983).
31. A. V. Kotikov, Yad. Fiz. **56** (9), 217 (1993) [Phys. At. Nucl. **56**, 1276 (1993)].
32. A. V. Kotikov and G. Parente, Nucl. Phys. B **549**, 242 (1999); Nucl. Phys. B (Proc. Suppl.) **99A**, 196 (2001); JETP **97**, 859 (2003); in *Proceedings of the International Conference PQFT98, Dubna, 1998*; hep-ph/9810223; in *Proceedings of the 8th International Workshop on Deep Inelastic Scattering and QCD (DIS 2000), Liverpool, England, 2000*, p. 198; hep-ph/0006197.
33. A. V. Kotikov and G. Parente, in *Proceedings of the XV International Seminar on High-Energy Physics Problem, Dubna, 2000*, p. 211; hep-ph/0012299; in *Proceedings of the 9th International Workshop on Deep Inelastic Scattering (DIS 2001), Bologna, Italy, 2001*, p. 313; hep-ph/0106175; A. Yu. Illarionov, A. V. Kotikov, and G. Parente, hep-ph/0402173.
34. M. Gluck and E. Reya, Phys. Rev. D **19**, 1382 (1979).
35. E. G. Floratos, C. Kounnas, and R. Lacage, Nucl. Phys. B **192**, 417 (1981).
36. R. M. Barnett *et al.*, Phys. Rev. D **54**, 1 (1996); A. D. Martin, W. J. Stirling, and R. G. Roberts, Phys. Lett. B **266**, 173 (1991).
37. D. I. Kazakov and A. V. Kotikov, Nucl. Phys. B **307**, 791 (1988); **345**, 299(E) (1990).
38. J. Blumlein and W. L. van Neerven, Phys. Lett. B **450**, 417 (1999).

39. W. L. van Neerven and A. Vogt, Nucl. Phys. B **568**, 263 (2000); **603**, 42 (2001).
40. W. L. van Neerven and A. Vogt, Nucl. Phys. B **588**, 345 (2000).
41. A. M. Cooper-Sarkar, R. G. E. Devenish, and A. de Roeck, Int. J. Mod. Phys. A **13**, 3385 (1998).
42. V. I. Vovk, A. V. Kotikov, and S. I. Maximov, Theor. Math. Phys. **84**, 744 (1990).
43. H1 Collab. (S. Aid *et al.*), Nucl. Phys. B **470**, 3 (1996).
44. ZEUS Collab. (M. Derrick *et al.*), Z. Phys. C **72**, 399 (1996).
45. V. G. Krivokhijine and A. V. Kotikov (in press).
46. H. Abramowitz, E. M. Levin, A. Levy, *et al.*, Phys. Lett. B **269**, 465 (1991); A. Levy, DESY Preprint No. 95-003 (1995); hep-ph/9501346.
47. A. V. Kotikov, Mod. Phys. Lett. A **11**, 103 (1996); Yad. Fiz. **59**, 2219 (1996) [Phys. At. Nucl. **59**, 2137 (1996)].
48. L. L. Jenkovszky, A. V. Kotikov, and F. Paccanoni, Yad. Fiz. **55**, 2205 (1992) [Sov. J. Nucl. Phys. **55**, 1224 (1992)]; Pis'ma Zh. Éksp. Teor. Fiz. **58**, 165 (1993) [JETP Lett. **58**, 163 (1993)]; Phys. Lett. B **314**, 421 (1993).
49. H. Georgi and H. D. Politzer, Phys. Rev. D **14**, 1829 (1976); R. Barbieri, J. Ellis, M. K. Gaillard, *et al.*, Phys. Rev. Lett. **64**, 171 (1990); Nucl. Phys. B **117**, 50 (1976); O. Nachtmann, Nucl. Phys. B **63**, 237 (1973); S. Wandzura, Nucl. Phys. B **122**, 412 (1977).
50. A. V. Kotikov, Yad. Fiz. **49**, 1068 (1989) [Sov. J. Nucl. Phys. **49**, 1068 (1989)].
51. Yu. L. Dokshitzer and D. V. Shirkov, Z. Phys. C **67**, 449 (1995); D. V. Shirkov and S. V. Mikhailov, Z. Phys. C **63**, 463 (1994).
52. B. J. Edwards and T. D. Gottschalk, Nucl. Phys. B **196**, 328 (1982); D. V. Shirkov, Theor. Math. Phys. **49**, 1039 (1981); Nucl. Phys. B **371**, 267 (1992).
53. W. Bernreuther and W. Wetzel, Nucl. Phys. B **197**, 228 (1982); W. Wetzel, Nucl. Phys. B **196**, 259 (1982); W. Bernreuther, Ann. Phys. (N.Y.) **151**, 127 (1983).
54. W. Marciano, Phys. Rev. D **29**, 580 (1984).
55. D. V. Shirkov, A. V. Sidorov, and S. V. Mikhailov, Preprint No. E2-96-285, OIYaI (Joint Institute for Nuclear Research, Dubna, 1996); hep-ph/9607472; hep-ph/9707514.
56. G. Rodrigo and A. Santamaria, Phys. Lett. B **313**, 441 (1993).
57. EM Collab. (J. J. Aubert *et al.*), Phys. Lett. B **123B**, 275 (1983).
58. F. E. Close, R. G. Roberts, and G. G. Ross, Phys. Lett. B **129B**, 346 (1983); R. L. Jaffe, F. E. Close, R. G. Roberts, *et al.*, Phys. Lett. B **134B**, 449 (1984); F. E. Close, R. L. Jaffe, R. G. Roberts, *et al.*, Phys. Rev. D **31**, 1004 (1985).
59. A. V. Kotikov, Yad. Fiz. **50**, 201 (1989) [Sov. J. Nucl. Phys. **50**, 127 (1989)].
60. BCDMS Collab. (A. C. Benevenuti *et al.*), Phys. Lett. B **189**, 483 (1987).
61. M. Dasgupta and B. R. Webber, Phys. Lett. B **382**, 273 (1996).
62. E. Stein, M. Maul, L. Mankiewicz, *et al.*, Nucl. Phys. B **536**, 318 (1998).
63. F. J. Yndurain, Phys. Lett. B **74B**, 68 (1978).
64. B. Escobles, M. J. Herrero, C. Lopez, *et al.*, Nucl. Phys. B **242**, 329 (1984); D. I. Kazakov and A. V. Kotikov, Yad. Fiz. **46**, 1767 (1987) [Sov. J. Nucl. Phys. **46**, 1057 (1987)].
65. J. Santiago and F. J. Yndurain, Nucl. Phys. B **563**, 45 (1999); **611**, 447 (2001).
66. D. W. Duke and R. G. Roberts, Nucl. Phys. B **166**, 243 (1980).
67. O. Benhar, S. Fantoni, and G. I. Lykasov, Phys. Lett. B **502**, 69 (2001); C. D. Cothran, D. B. Day, and S. Liuti, Phys. Lett. B **421**, 46 (1998).
68. F. James and M. Ross, *MINUIT*, CERN Computer Center Library, D 505 (Geneve, 1987).
69. H1 Collab. (C. Adloff *et al.*), Eur. Phys. J. C **21**, 33 (2001).
70. V. Genchev *et al.*, in *Proceedings of the VII International Seminar on High-Energy Physics Problem, Dubna, Sept. 1988*, Vol. 2, p. 6.
71. A. Milsztajn, A. Staude, K. M. Teichert, *et al.*, Z. Phys. C **49**, 527 (1991).
72. V. A. Matveev, R. M. Muradian, and A. N. Tavkhelidze, Lett. Nuovo Cimento **7**, 719 (1973); S. J. Brodsky and G. R. Farrar, Phys. Rev. Lett. **31**, 1153 (1973); S. J. Brodsky, J. Ellis, E. Cardi, *et al.*, Phys. Rev. D **56**, 6980 (1997).
73. D. I. Gross, Phys. Rev. Lett. **32**, 1071 (1974); D. I. Gross and S. B. Treiman, Phys. Rev. Lett. **32**, 1145 (1974).
74. S. Schaefer, A. Schafer, and M. Stratmann, Phys. Lett. B **514**, 284 (2001).
75. B. I. Ermolaev, M. Greco, and S. I. Troyan, Nucl. Phys. B **594**, 71 (2001); **571**, 137 (2000); hep-ph/0104082; in *Proceedings of the 9th International Workshop on Deep Inelastic Scattering (DIS 2001), Bologna, Italy*, p. 303; hep-ph/0106317.
76. J. Bartels, B. I. Ermolaev, and M. G. Ryskin, Z. Phys. C **72**, 627 (1996); **70**, 273 (1996); B. I. Ermolaev, S. I. Manaenkov, and M. G. Ryskin, Z. Phys. C **69**, 259 (1996).
77. R. Kirschner and L. N. Lipatov, Nucl. Phys. B **213**, 122 (1983).
78. F. Martin, Phys. Rev. D **19**, 1382 (1979); C. Lopez and F. I. Yndurain, Nucl. Phys. B **171**, 231 (1980).
79. C. Lopez and F. I. Yndurain, Nucl. Phys. B **183**, 157 (1981).
80. A. V. Kotikov, S. I. Maximov, and I. S. Parobij, Theor. Math. Phys. **111**, 442 (1997).
81. A. V. Kotikov, Yad. Fiz. **57**, 142 (1994) [Phys. At. Nucl. **57**, 133 (1994)]; Phys. Rev. D **49**, 5746 (1994).
82. A. V. Kotikov, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 635 (1994) [JETP Lett. **59**, 667 (1994)]; Zh. Éksp. Teor. Fiz. **107**, 1761 (1995) [JETP **80**, 979 (1995)]; A. V. Kotikov and G. Parente, Phys. Lett. B **379**, 195 (1996); Mod. Phys. Lett. A **12**, 963 (1997);

Zh. Éksp. Teor. Fiz. **112**, 32 (1997) [JETP **85**, 17 (1997)]; hep-ph/9609439; in *Proceedings of International Workshop on Deep Inelastic Scattering and Related Phenomena (DIS 96), Rome, Italy, 1996*, p. 237; hep-ph/9608409.

83. A. V. Kotikov, A. V. Lipatov, G. Parente, *et al.*, Eur. Phys. J. C **25**, 77 (2002); A. V. Kotikov, A. V. Lipatov, and N. P. Zotov, Eur. Phys. J. C **27**, 219 (2003); hep-ph/0403135.
84. J. Bartels, DESY-91 074 (1991); Phys. Lett. B **298**, 204 (1993); Z. Phys. C **60**, 471 (1993); E. M. Levin, M. G. Ryskin, and A. G. Shuvaev, Nucl. Phys. B **387**, 589 (1992); J. Bartels and C. Bontus, Phys. Rev. D **61**, 034009 (2000); J. Bartels, C. Bontus, and H. Spiesberger, Preprints DESY-99-118, MZ-TH/99-33 (1999); hep-ph/9908411; A. D. Martin and M. G. Ryskin, Phys. Lett. B **431**, 395 (1998).
85. L. N. Lipatov, Yad. Fiz. **23**, 642 (1976) [Sov. J. Nucl. Phys. **23**, 338 (1976)]; E. A. Kuraev, L. N. Lipatov, and V. S. Fadin, Zh. Éksp. Teor. Fiz. **71**, 88 (1976) [Sov. Phys. JETP **44**, 443 (1976)]; Zh. Éksp. Teor. Fiz. **72**, 377 (1977) [Sov. Phys. JETP **45**, 199 (1977)]; Ya. Ya. Balitzki and L. N. Lipatov, Yad. Fiz. **28**, 822 (1978) [Sov. J. Nucl. Phys. **28**, 822 (1978)]; L. N. Lipatov, Zh. Éksp. Teor. Fiz. **90**, 1536 (1986) [Sov. Phys. JETP **63**, 904 (1986)].
86. V. N. Fadin and L. N. Lipatov, Phys. Lett. B **429**, 127 (1998); M. Ciafaloni and G. Camici, Phys. Lett. B **430**, 349 (1998).
87. A. V. Kotikov and L. N. Lipatov, Nucl. Phys. B **582**, 19 (2000); **661**, 19 (2003).
88. S. J. Brodsky, V. S. Fadin, V. T. Kim, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 161 (1999) [JETP Lett. **70**, 155 (1999)]; G. Salam, J. High Energy Phys. **9807**, 019 (1998); Acta Phys. Pol. B **30**, 3679 (1999); M. Ciafaloni, D. Colferai, and G. Salam, Phys. Rev. D **56**, 6980 (1997).
89. C. R. Schmidt, Preprint MSUHEP-10615; hep-ph/0106181.
90. A. B. Kaidalov, hep-ph/0103011.
91. L3 Collab. (M. Acciarri *et al.*), Phys. Lett. B **453**, 333 (1999); M. N. Kienzle-Focacci, in *Proceedings of the International Symposium on Evolution Equations and Large Order Estimates in QCD, Gatchina, Russia, May 2000*.
92. J. Bartels, K. Golec-Biernat, and K. Peters, Eur. Phys. J. C **17**, 121 (2000).
93. S. Liuti, Nucl. Phys. B (Proc. Suppl.) **74**, 380 (1999).
94. G. Dissertori, hep-ex/0105070; CDF Collab. (T. Affolder *et al.*), Phys. Rev. Lett. **88**, 042001 (2002).
95. P. A. Movilla Fernandez, S. Bethke, O. Biebel, *et al.*, Eur. Phys. J. C **22**, 1 (2001).
96. S. Bethke, J. Phys. G **26**, R27 (2000).
97. S. I. Alekhin, Phys. Lett. B **519**, 57 (2001).
98. S. Moch, J. A. M. Vermaseren, and A. Vogt, Nucl. Phys. B **688**, 101 (2004); A. Vogt, S. Moch, and J. A. M. Vermaseren, Nucl. Phys. B **691**, 129 (2004).
99. S. Moch, J. A. M. Vermaseren, and A. Vogt, Phys. Lett. B **606**, 123 (2005).
100. A. V. Kotikov and V. N. Velizhanin, hep-ph/0501274.
101. S. A. Larin, T. van Ritbergen, and J. A. M. Vermaseren, Nucl. Phys. B **427**, 41 (1994); S. A. Larin, P. Nogueira, T. van Ritbergen, *et al.*, Nucl. Phys. B **492**, 338 (1997); A. Retey and J. A. M. Vermaseren, Nucl. Phys. B **604**, 281 (2001).
102. J. A. M. Vermaseren, Int. J. Mod. Phys. A **14**, 2037 (1999); Acta Phys. Pol. B **29**, 2599 (1998); S. Moch and J. A. M. Vermaseren, Nucl. Phys. B **573**, 853 (2000); Nucl. Phys. B (Proc. Suppl.) **86**, 78 (2000); **89**, 131 (2000); **89**, 137 (2000); Preprint NIKHEF-2001-008, TTP-01-18; hep-ph/0108033.
103. D. I. Kazakov and A. V. Kotikov, Theor. Math. Phys. **73**, 1264 (1987); A. V. Kotikov, Theor. Math. Phys. **78**, 134 (1989); Phys. Lett. B **375**, 240 (1996); in *Proceedings of the 15th International Workshop on High-Energy Physics and Quantum Field Theory (QFTHEP 2000), Tver', Russia, 2000*, p. 211; hep-ph/0102177.
104. DELPHI Collab. (P. Abreu *et al.*), Eur. Phys. J. C **6**, 19 (1999).
105. S. I. Alekhin, Phys. Rev. D **59**, 114 016 (1999).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1904–1915. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1966–1977.

ELEMENTARY PARTICLES AND FIELDS
Theory

Gravity as the Affine Goldstone Phenomenon and Beyond*

Yu. F. Pirogov**

Institute for High Energy Physics, Protvino, Moscow oblast, 142284 Russia

Received October 20, 2004

Abstract—The two-phase structure is imposed on the world continuum, with the graviton emerging as the tensor Goldstone boson during the spontaneous transition from the affinely connected phase to the metric one. The physics principle of metarelativity, extending the respective principle of special relativity, is postulated. The theory of metagravitation as the general nonlinear model $GL(4,R)/SO(1,3)$ in an arbitrary background continuum is built. The concept of the Metauniverse as the ensemble of the regions of the metric phase inside the affinely connected phase is introduced, and the possible bearing of the emerging multiple universes to the fine tuning of our Universe is conjectured. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

General relativity (GR) is the well-stated theory attributing gravity to the Riemannian geometry of the spacetime. Nevertheless, the ultimate nature of gravity awaits, conceivably, its future explanation. In this respect, of great interest is the approach to gravity as the Goldstone phenomenon corresponding to the broken global affine symmetry [1, 2]. Originally, this approach was realized as the nonlinear model $GL(4,R)/SO(1,3)$ in the Minkowski background spacetime, as distinct from the geometrical framework of GR.

In the present paper, we adhere to the viewpoint that the above construction is more than just the mathematical one, but has a deeper physics foundation underlying it. In this respect, new insights motivating and extending the Goldstone approach to gravity are put forward. In principle, we go beyond the framework of the Riemannian geometry. Namely, we start with the world continuum considered as the affinely connected manifold without metric and end up in the spacetime with the effective Riemannian geometry.

Our main results are threefold:

(i) The physics principle of extended relativity is introduced as a substitution for that of special relativity. It states the physics invariance, at an underlying level, relative to the choice within the extended set of the local coordinates, including the inertial ones. The principle justifies the pattern of the affine symmetry breaking $GL(4,R) \to SO(1,3)$ required for the Goldstone approach to gravity.

(ii) The extended theory of gravity, with GR as the lowest approximation, is built as the proper nonlinear model in an arbitrary background continuum. The natural hierarchy of the possible GR extensions, according to the accuracy of the affine symmetry realization, is put forward.

(iii) The extended Universe, as the ensemble of the Riemannian metric universes inside the affinely connected world continuum, is considered. It is conjectured that the multiple universes may clarify the fine-tuning problem of our Universe.

The content of the paper is as follows. In Section 2, the principle of extended relativity is introduced. The spontaneous breaking of the ensuing global symmetry, the affine one, with the residual Poincare symmetry and the emerging tensor Goldstone boson is then considered. In Section 3, the nonlinear realization of the broken affine symmetry is studied. In Section 4, the respective nonlinear model in the tangent space is developed. Its prolongation to the spacetime, the extended gravitation, is presented in Section 5. Finally, the concept of the extended Universe is discussed in Section 6, with some remarks in the Conclusion.

2. METARELATIVITY

2.1. Affine Symmetry

Conventionally, GR starts by postulating that the world continuum, i.e., the set of the world events (points), is the Riemannian manifold. In other words, a metric is imposed on the world ab initio. The metric specifies all the fine properties of the continuum converting the latter into the spacetime. Nevertheless,

*The text was submitted by the author in English.

**e-mail: pirogov@mx.ihep.su

not all of the properties of the spacetime depend crucially on the metric.[1)] To appreciate the deeper meaning of the gravity and the very spacetime, one needs possibly to go beyond the Riemannian geometry.

To this end, consider the spacetime not as a priori existing but as emerging in the processes of the world structure formation. Namely, suppose that at an underlying level the continuum is endowed only with the topological structure (without metric, yet). More particularly, it is the affinely connected manifold. The affine connection supports the detailed continuity properties, such as the parallel transport of the vector fields and their covariant derivatives. In particular, the connection produces the curvature tensor as a result of the parallel transport of a vector around the infinitesimal closed contour. But there is as yet no geometrical structures which would be inherent in the metric, such as the interval, distances, and angles.

Let x^μ, $\mu = 0, \ldots, 3$, be the world coordinates, generally, in the patches. There being, in absence of the metric, no partition of the continuum onto the space and time, the index 0 has as yet no particular meaning and is just a notational one. In ignorance of the underlying "dynamics," consider all the structures related to the underlying level of the world continuum as the background ones. Let $\bar\phi^\lambda{}_{\mu\nu}(x)$ be the background affine connection and let $\bar\xi^\alpha$, $\alpha = 0, \ldots, 3$, be the background affiliated coordinates, where the connection has a particular, to be defined, form $\bar\phi^\gamma{}_{\alpha\beta}(\bar\xi)$.[2)] The connections are related as usual:

$$\bar\phi^\gamma{}_{\alpha\beta}(\bar\xi) = \frac{\partial x^\mu}{\partial\bar\xi^\alpha}\frac{\partial x^\nu}{\partial\bar\xi^\beta}\left(\frac{\partial\bar\xi^\gamma}{\partial x^\lambda}\,\bar\phi^\lambda{}_{\mu\nu}(x) - \frac{\partial^2\bar\xi^\gamma}{\partial x^\mu\partial x^\nu}\right). \tag{1}$$

It follows therefore that the parts antisymmetric and symmetric in the lower indices transform independently, respectively, homogeneously and inhomogeneously. In particular, being zero at a point in some coordinates, the antisymmetric part (the torsion) remains zero independent of the coordinates. Thus, one can adopt the background torsion to be absent identically. As for the symmetric part, one is free to choose the special coordinates to make the physics description as transparent as possible.

So, let P be a fixed but otherwise arbitrary point (the reference point) with the world coordinates X^μ. Adjust to this point the local coordinates as follows:

$$\bar\xi^\alpha = \bar\Xi^\alpha + \bar e^\alpha_\lambda(X)\Bigg((x-X)^\lambda + \frac{1}{2}\bar\phi^\lambda{}_{\mu\nu}(X)(x-X)^\mu(x-X)^\nu\Bigg) + \mathcal{O}((x-X)^3). \tag{2}$$

Here, $\bar\Xi^\alpha \equiv \bar\xi^\alpha(X)$ and $\bar e^\alpha_\lambda(X) \equiv \partial\bar\xi^\alpha/\partial x^\lambda|_{x=X}$ is the tetrad, with $\bar e^\lambda_\alpha(X)$ being the inverse one. These parameters are still arbitrary and liable to further specification. In the vicinity of P, the affine connection now looks like

$$\bar\phi^\gamma_{\alpha\beta}(\bar\xi) = \frac{1}{2}\bar\rho^\gamma{}_{\alpha\delta\beta}(\bar\Xi)(\bar\xi-\bar\Xi)^\delta + \mathcal{O}((\bar\xi-\bar\Xi)^2), \tag{3}$$

with $\bar\rho^\gamma_{\alpha\delta\beta}(\bar\Xi)$ being the background curvature tensor in the reference point. In the coordinates chosen, the affine connection vanishes at the reference point: $\bar\phi^\gamma{}_{\alpha\beta}(\bar\Xi) = 0$.

Now, consider the whole set of local coordinates nullifying the affine connection at the reference point P. Under the world coordinates x^μ being fixed, one can choose a priori any one of the local coordinates $\bar\xi^\alpha$. The group of replacements $\bar\xi^\alpha \to \bar\xi'^\alpha$ among the latter ones is given by the transformations

$$(A,a): \bar\Xi^\alpha \to \bar\Xi'^\alpha = A^\alpha{}_\beta\bar\Xi^\beta + a^\alpha, \qquad \bar e^\alpha_\mu \to \bar e'^\alpha_\mu = A^\alpha{}_\beta\bar e^\beta_\mu, \tag{4}$$

with A^α_β being an arbitrary nondegenerate matrix and a^α being an arbitrary vector. The transformations (A,a) being independent of $\bar\xi$, the group is the global one. This is the inhomogeneous general linear group $IGL(4,R) = T_4 \odot GL(4,R)$ (the affine one). Under these and only under these transformations, the affine connection still remains zero at the reference point. The respective coordinates will be called the local affine ones.[3)] In these coordinates, the world continuum is approximated by the affinely flat manifold in a neighborhood of the reference point P. Particularly, the covariant derivative in the affine coordinates at the point P coincides with the ordinary one. Being changeable under A, the nonzero background torsion at the point P would violate explicitly the affine symmetry. Just to disregard this, the torsion is adopted to be zero identically.

The affine group $IGL(4,R)$ is 20-parametric and extends the ten-parameter Poincare group $ISO(1,3) = T_4 \odot SO(1,3)$ by the transformations varying the scales, the affine dilatations,[4)] the rest being the scale preserving unimodular affine transformations. The dilatations $A = \Delta$ belong to the one-parametric multiplicative group of positive real

[1)] Compare in this respect the reflections on the spacetime structure due to Schrödinger [3].

[2)] The bar sign refers in what follows to the background. The indices α, β, etc., are those of the special coordinates, while the indices λ, μ, etc., are the arbitrary world ones.

[3)] Being understood, the term "local" will be omitted in what follows.

[4)] The world coordinates being fixed, the affine dilatations are to be distinguished from the conventional ones in the spacetime [4].

numbers, $\Delta^\alpha_\beta = e^{-\lambda}\delta^\alpha_\beta$, with any real λ. The unimodular affine transformations are the nine-parametric part of the special linear group $SL(4,R) \ni A_0$, with $\det A_0 = 1$ (times reflections).

According to the well-known principle of special relativity, the present-day local physics laws are unchanged under the choice of the inertial coordinates, the Poincare group being the physics invariance symmetry. Now, introduce the principle of extended relativity, or *metarelativity*, stating that the local physics laws at the underlying level are unchanged relative to the choice of the affine coordinates. This extends the physics invariance symmetry from the Poincare symmetry to the affine one.

2.2. Spontaneous Symmetry Breaking

Presently, there is known no exact affine symmetry. Thus, the latter should be broken in transition from the underlying level to the effective one. Postulate that this is achieved due to the spontaneous emergence of the background metric $\bar\varphi_{\mu\nu}$ in the world continuum. The metric, with the Minkowskian signature, is assumed to be correlated with the background affine connection, so that it appears in the affine coordinates as

$$\bar\varphi_{\alpha\beta}(\bar\xi) = \bar\eta_{\alpha\beta} - \frac{1}{2}\bar\rho_{\gamma\alpha\delta\beta}(\bar\Xi) \times (\bar\xi-\bar\Xi)^\gamma(\bar\xi-\bar\Xi)^\delta + \mathcal{O}((\bar\xi-\bar\Xi)^3). \quad (5)$$

Here, one sets $\bar\eta_{\alpha\beta} \equiv \bar\varphi_{\alpha\beta}(\bar\Xi)$ and $\bar\rho_{\gamma\alpha\delta\beta}(\bar\Xi) = \bar\eta_{\gamma\gamma'}\bar\rho^{\gamma'}{}_{\alpha\delta\beta}(\bar\Xi)$. The metric Eq. (5) is such that the Christoffel connection $\bar\chi^\gamma{}_{\alpha\beta}(\varphi)$, determined by the metric, matches with the affine connection $\bar\phi^\gamma{}_{\alpha\beta}$ in the sense that the connections coincide locally up to the first derivative: $\bar\chi^\gamma{}_{\alpha\beta} = \bar\phi^\gamma{}_{\alpha\beta} + \mathcal{O}((\bar\xi-\bar\Xi)^2)$. This is quite reminiscent of the well-known fact that the metric in the Riemannian manifold may be approximated locally up to the first derivative by the Euclidean metric. In the wake of the emerging background metric, there appears the (still primordial) partition of the world continuum onto the space and time.

Under the linearly realized affine symmetry, the background metric ceases to be invariant. But it still possesses an invariance subgroup. To find it note that, without any loss of generality, one can choose among the affine coordinates the particular ones with $\bar\eta_{\alpha\beta}$ being the Minkowski tensor $\eta = \mathrm{diag}(1,-1,-1,-1)$. The respective coordinates will be called the background inertial ones.[5)] Under the affine transformations, one has

$$(A,a): \eta \to \eta' = A^{-1T}\eta A^{-1} \neq \eta, \quad (6)$$

whereas the Poincare transformations still leave η invariant:

$$(\Lambda,a): \eta \to \eta' = \Lambda^{-1T}\eta\Lambda^{-1} = \eta. \quad (7)$$

It follows that the group of invariance is isomorphous to the Poincare group $ISO(1,3) \in IGL(4,R)$ for any fixed $\bar\eta_{\alpha\beta}$. Physically, the spontaneous symmetry breaking corresponds to fixing, modulo the Poincare transformations, the class of the distinguished coordinates among the affine ones. These coordinates correspond to the particular choice for $\bar\eta_{\alpha\beta}$. Of course, the fact that the distinguished coordinates are precisely those with the Minkowskian $\eta_{\alpha\beta}$ is no more than a matter of convention, corresponding to the proper inner automorphism of the affine group.

Thus under the appearance of the metric, the affine symmetry is broken spontaneously to the residual Poincare one:

$$IGL(4,R) \xrightarrow{M_A} ISO(1,3). \quad (8)$$

For the symmetry breaking scale M_A, one expects a priori $M_A \sim M_{\mathrm{Pl}}$, with M_{Pl} being the Planck mass. More particularly, the relation between the scales is discussed in Section 5. Due to the spontaneous breaking, the affine symmetry should be realized in a nonlinear manner [5], with the nonlinearity scale M_A, the residual Poincare symmetry still being realized linearly. The unitary linear representations of the latter correspond to the matter particles, as usual. The translation subgroup being intact, the broken part coincides with $GL(4,R)/SO(1,3)$. The latter should be realized in the Nambu–Goldstone mode. Accompanying the spontaneous emergence of the metric, there should appear the ten-component affine Goldstone boson which corresponds to the ten generators of the broken affine transformations. The effective field theory of the Goldstone boson is given by the relevant nonlinear model, to be studied in what follows.

2.3. Lorentz Symmetry

The group $GL(4,R)$ possesses the 16 generators $\Sigma^\alpha{}_\beta$. By means of $\eta_{\alpha\beta}$, one can redefine the generators as $\Sigma^{\alpha\beta} \equiv \Sigma^\alpha_\gamma\eta^{\gamma\beta}$ and replace the latter ones with the symmetric and antisymmetric combinations $\Sigma^{\alpha\beta}_\pm = \Sigma^{\alpha\beta} \pm \Sigma^{\beta\alpha}$. Clearly, this partition is affine noncovariant. The respective commutation relations read as follows:

$$\frac{1}{i}[\Sigma^{\alpha\beta}_\pm, \Sigma^{\gamma\delta}_\pm] = \eta^{\alpha\gamma}\Sigma^{\beta\delta}_- \pm \eta^{\alpha\delta}\Sigma^{\beta\gamma}_- \pm (\alpha\leftrightarrow\beta), \quad (9)$$

$$\frac{1}{i}[\Sigma^{\alpha\beta}_-, \Sigma^{\gamma\delta}_+] = \eta^{\alpha\gamma}\Sigma^{\beta\delta}_+ + \eta^{\alpha\delta}\Sigma^{\beta\gamma}_+ - (\alpha\leftrightarrow\beta).$$

[5)] The latter ones are to be distinguished from the effective inertial coordinates, to be introduced.

The generators $\Sigma_-^{\alpha\beta}$ correspond to the residual Lorentz symmetry, whereas $\Sigma_+^{\alpha\beta}$ correspond to the broken part of the affine symmetry. The broken generators contain, in turn, the dilatation one $i\eta_{\alpha\beta}\Sigma_+^{\alpha\beta}$. The latter commutes with all the generators and is thus proportional to unity in any irreducible representation. For the generators $\sigma^\alpha{}_\beta$ in the adjoint representation, one has $(\sigma^\alpha{}_\beta)^\gamma{}_\delta = (1/i)\delta^\alpha_\delta\delta^\gamma_\beta$, so that the respective generators $\sigma_\pm^{\alpha\beta}$ are as follows:

$$(\sigma_\pm^{\alpha\beta})^\gamma{}_\delta = \frac{1}{i}(\delta^\alpha_\delta\eta^{\beta\gamma} \pm \delta^\beta_\delta\eta^{\alpha\gamma}). \qquad (10)$$

The above partition of generators is used in what follows in constructing the nonlinear model. First, we study the three kinds of substances, i.e., the affine Goldstone boson, matter, and radiation, which are characterized by three distinct types of nonlinear realization. With these building blocks, we then construct the nonlinear model itself.

3. NONLINEAR REALIZATION

3.1. Affine Goldstone Boson

Let $\bar\xi^\alpha$ be the background inertial coordinates adjusted to the spacetime point P. Attach to this point the auxiliary linear space T, the tangent space at the point. By definition, T is isomorphous to the Minkowski spacetime. The tangent space is the structure space of the theory, whereupon the realizations of the physics spacetime symmetries, the affine and the Poincare ones, are defined. Introduce in T the coordinates ξ^α, the counterpart of the background inertial coordinates $\bar\xi^\alpha$ in the spacetime. By construction, the connection in the tangent space is zero identically. For the connection at the spacetime point P in the coordinates $\bar\xi^\alpha$ to be zero too, the coordinates in the vicinity of the reference point have to be related as $\xi^\alpha = \bar\xi^\alpha + \mathcal{O}((\bar\xi - \bar\Xi)^3)$. The coordinates ξ^α are the ones wherein all the constructions in T are originally built. The latter ones being done, one can use in T arbitrary coordinates.

According to [5], the nonlinear realization of the symmetry G spontaneously broken to the symmetry $H \subset G$ can be built on the quotient space $K = G/H$, the residual subgroup H serving as the classification group. In the present case, one is interested in the pattern $GL(4,R)/SO(1,3)$, with the quotient space consisting of all the broken affine transformations. Let $æ(\xi) \in K$ be the coset function on the tangent space. It can be represented by a group element $k(\xi) \in G$. Under the affine transformations (A, a), the representative group element is to transform in the vicinity of the reference point as

$$(A,a) : k(\xi) \to k'(\xi') = Ak(\xi)\Lambda^{-1}, \qquad (11)$$

where Λ is the appropriate element of the residual group, here the Lorentz one $SO(1,3)$. One has similarly $k^{-1} \to \Lambda k^{-1}A^{-1}$. At the same time, by its very construction, the Minkowskian η stays invariant under the nonlinearly realized affine symmetry

$$(A,a) : \eta \to \eta' = \Lambda^{-1T}\eta\Lambda^{-1} = \eta, \qquad (12)$$

in contrast to the linear realization, Eq. (6). Accounting for Eq. (12), one gets in the other terms

$$(A,a) : k(\xi)\eta \to k'(\xi')\eta = Ak(\xi)\eta\Lambda^T. \qquad (13)$$

To represent unambiguously the coset by the group element k, one should impose on the latter some auxiliary condition, e.g., requiring k to be pseudosymmetric in the sense that $k\eta = (k\eta)^T$ (and similarly for k^{-1}). This ensures that k has 10 independent components, indeed, in accord with the 10 broken generators. Under the affine transformations, this results in the restriction $Ak\eta\Lambda^T = \Lambda\eta k^T A^T$. This entails implicitly the dependence of the Lorentz transformation on the Goldstone boson: $\Lambda = \Lambda(A,k)$. Hereof, the term "nonlinear" follows. This construction implements the nonlinear realization of the whole broken group $GL(4,R)$, the residual Lorentz subgroup $SO(1,3)$ still being realized linearly, i.e., $\Lambda(A,k)|_{A=\Lambda} \equiv \Lambda$. And what is more, the dilatations Δ being Abelian, one gets $\Lambda(\Delta,k) = 1$, so that $\Lambda(A,k) = \Lambda(A_0,k)$, with $A \equiv \pm A_0\Delta$ and $A_0 \in SL(4,R)$.

By doing as above, one retains only the independent Goldstone components but loses the local Lorentz symmetry.[6] For this reason, we will not impose any auxiliary condition. Instead, we extend the affine symmetry by the hidden local symmetry[7] $\hat H \simeq H = SO(1,3)$, with the symmetry-breaking pattern $G \times \hat H \to H$. For quantities in the tangent space, one should distinguish now two types of indices: the affine ones, acted on by the global affine transformations $A \in G$, and the Lorentz ones, acted on by the local Lorentz transformations $\Lambda(\xi) \in \hat H$. To make this difference explicit, designate the affine indices in the tangent space, as before, as α, β, etc., and designate the Lorentz ones as a, b, etc. The Goldstone field is represented in this case by the arbitrary 4×4 matrix $\kappa^\alpha{}_a(\xi)$ (respectively, $\kappa^{-1a}{}_\alpha$), which transforms similar to k by Eq. (11) but with arbitrary $\Lambda(\xi)$. In what follows, it is understood that the Lorentz indices are manipulated by means of the Minkowskian η_{ab} (respectively, η^{ab}). So, in the component notation, $\kappa\eta$ looks like $\kappa^{\alpha a}$ (similarly, $\eta\kappa^{-1}$ is $\kappa^{-1}{}_{a\alpha}$). This is the

[6] Such a procedure was adopted in [1, 2].

[7] The hat sign refers in what follows to the local Lorentz symmetry.

linearization of the nonlinear model, with the extra Goldstone degrees of freedom being unphysical due to the gauge Lorentz transformations $\Lambda(\xi)$. The auxiliary gauge boson, corresponding to the generators $\hat{\Sigma}^{ab}_{-}$ of the local Lorentz symmetry, is expressed owing to the equation of motion through κ and its derivative. With this in mind, the abrupt expressions entirely in terms of κ and its derivative are used in what follows. The versions differ in the higher order corrections.

3.2. Matter

The affine symmetry contains the Abelian, though broken, subgroup of the affine dilatations. For this reason, the generic matter fields ϕ may additionally be classified by their affine scale dimension l_ϕ,[8] so that

$$(A,a):\phi(\xi)\to\phi'(\xi')=e^{l_\phi\lambda}\hat{\rho}_\phi(\Lambda)\phi(\xi), \qquad (14)$$

with $\hat{\rho}_\phi(\Lambda)$ taken in the proper Lorentz representation. According to Eq. (11), the scale dimension of κ is $l_\kappa=-1$ (respectively, $l_{\kappa^{-1}}=1$). Thus, accounting for the transformation $\det\kappa\to e^{-4\lambda}\det\kappa$ under dilatations, one can rescale the matter fields to the effective ones $\hat{\phi}=(\det\kappa)^{l_\phi/4}\phi$. The new fields are affine scale invariant, i.e., correspond to $l_{\hat{\phi}}=0$, and transform simply as the local Lorentz representations. They are to be used in constructing the nonlinear model. If the affine symmetry is not explicitly violated, only the rescaled matter fields enter the action. In any case, one can choose $\hat{\phi}$ and $\det\kappa$ as the independent field variables. Thus, instead of $\hat{\phi}$, the matter fields will be designated in what follows simply as ϕ with $l_\phi=0$.

3.3. Radiation

From the point of view of the nonlinear realization, the gauge bosons of the internal symmetries constitute one more separate kind of substance, radiation. By definition, the gauge boson fields V_α transform under the affine transformations linearly as the derivative $\partial_\alpha\equiv\partial/\partial\xi^\alpha$,

$$(A,a):V(\xi)\to V'(\xi')=A^{-1T}V(\xi), \qquad (15)$$

thus corresponding to the scale dimension $l_V=1$. For this reason, redefine the gauge fields as $\hat{V}_a=\kappa^\alpha{}_a V_\alpha$. The new fields transform as the local Lorentz vectors

$$\hat{V}(\xi)\to\hat{V}'(\xi')=\Lambda^{-1T}\hat{V}(\xi) \qquad (16)$$

and correspond to $\hat{l}_V=0$. These gauge fields are to be used in the model building. Altogether, this exhausts the description of all three kinds of substances: the affine Goldstone boson, matter, and radiation.

4. NONLINEAR MODEL

4.1. Nonlinear Connection

To explicitly account for both the affine symmetry and the local Lorentz one, it is convenient to start with the composite objects transforming only under the latter symmetry. Clearly, any nontrivial combinations of κ and κ^{-1} alone transform explicitly under A. Thus, the derivative terms are inevitable. To describe the latter ones, introduce the Maurer–Cartan one-form

$$\hat{\Omega}=\eta\kappa^{-1}d\kappa, \qquad (17)$$

with $d\kappa$ being the ordinary differential of κ. Under the affine transformations $\kappa\to\kappa'=Ak\Lambda^{-1}$, the one-form transforms as the local Lorentz representation

$$\hat{\Omega}(\xi)\to\hat{\Omega}'(\xi')=\Lambda^{-1T}\hat{\Omega}(\xi)\Lambda^{-1}+\Lambda^{-1T}\eta d\Lambda^{-1}, \qquad (18)$$

with $d\Lambda$ being the ordinary differential of $\Lambda(\xi)$. Here, use is made of the relation $\eta\Lambda\eta=\Lambda^{-1T}$ for the Lorentz transformations.

In the component notation, the so-defined one-form looks like $\hat{\Omega}_{ab}$. Decompose it as

$$\hat{\Omega}_{ab}\equiv\sum_{\pm}\hat{\Omega}^{\pm}_{ab}=\sum_{\pm}[\eta\kappa^{-1}d\kappa]^{\pm}_{ab}, \qquad (19)$$

where $[\ldots]^{\pm}$ means the symmetric and antisymmetric parts, respectively. One sees that $\hat{\Omega}^{\pm}_{ab}$ transform independently of each other:

$$\hat{\Omega}^{\pm}(\xi)\to\hat{\Omega}'^{\pm}(\xi')=\Lambda^{-1T}\hat{\Omega}^{\pm}(\xi)\Lambda^{-1}+\delta^{\pm}, \qquad (20)$$

where

$$\delta^{-}=\Lambda^{-1T}\eta d\Lambda^{-1}, \qquad (21)$$
$$\delta^{+}=0.$$

Transforming homogeneously, the symmetric part $\hat{\Omega}^{+}$ can naturally be associated with the nonlinear covariant differential of the Goldstone field. At the same time, the antisymmetric part $\hat{\Omega}^{-}$ transforms inhomogeneously and allows one to define the nonlinear covariant differential of the matter fields: $D\phi=(d+(i/2)\hat{\Omega}^{-}_{ab}\hat{\Sigma}^{ab}_{\phi})\phi$, with $\hat{\Sigma}^{ab}_{\phi}$ being the Lorentz generators in the representation $\hat{\rho}_\phi$. The so-defined $D\phi$ transform homogeneously, like ϕ themselves.

The generic nonlinear covariant derivative $D_\alpha\equiv D/d\xi^\alpha$ transforms as the affine vector. The effective

[8] The latter is to be distinguished from the conventional scale dimension.

covariant derivative, which transforms as the local Lorentz vector, can be constructed as follows:

$$\hat{D}_a \equiv \kappa_a^\alpha D_\alpha = \kappa^\alpha{}_a D/d\xi^\alpha. \quad (22)$$

Thus, one gets for the covariant derivative of the one-form

$$\hat{\Omega}^\pm_{abc} = \kappa^\gamma{}_c \hat{\Omega}^\pm_{ab}/d\xi^\gamma = [\eta\kappa^{-1}\hat{\partial}_c\ \kappa]^\pm_{ab}, \quad (23)$$

where

$$\hat{\partial}_c \equiv \kappa_c^\gamma \partial_\gamma = \kappa_c^\gamma \partial/\partial\xi^\gamma \quad (24)$$

is the effective, Goldstone boson dependent, partial derivative. It follows that $\hat{\Omega}^-_{abc}$ could be used as the connection for the nonlinear realization. Note that this expression precisely corresponds to the case of the nonlinear realization of the spontaneously broken internal symmetry, where this connection is determined uniquely. But in the present case of the space-time symmetry, the coordinates transform under the same group as the fields. This results in the possible ambiguity of the nonlinear connection.

Namely, the transformation properties of the covariant derivative do not change if one adds to the above minimal connection the properly modified terms $\hat{\Omega}^+_{abc}$, the latter ones transforming homogeneously. For reasons justified later in this section, we choose for the nonminimal connection the following special combination:

$$\hat{\omega}_{abc} = \hat{\Omega}^-_{abc} + \hat{\Omega}^+_{cab} - \hat{\Omega}^+_{cba} \quad (25)$$
$$= [\eta\kappa^{-1}\hat{\partial}_c\ \kappa]^-_{ab} + [\eta\kappa^{-1}\hat{\partial}_b\ \kappa]^+_{ca} - [\eta\kappa^{-1}\hat{\partial}_a\ \kappa]^+_{cb}.$$

The nonlinear covariant derivative of the matter fields now becomes

$$\hat{D}_c\phi = \left(\hat{\partial}_c + \frac{i}{2}\hat{\omega}_{abc}\hat{\Sigma}^{ab}_\phi\right)\phi. \quad (26)$$

$\hat{D}_c\phi$ transforms homogeneously and can be used in model building.

4.2. Gauge Interactions

Internal symmetry. Let $\hat{V}_a$ be the generator-valued gauge fields for the internal gauge symmetry. The gauge fields are supposed to be coupled universally via the nonlinear connection [Eq. (25)]. Accounting for the Lorentz generators $-(\hat{\sigma}^{ab}_-)^T$ in the contravariant adjoint representation [being given by Eq. (10) with the obvious substitution for the indices], one gets for the nonlinear derivative of the fields

$$\hat{D}_a\hat{V}_b = \left(\delta^c_b\hat{\partial}_a + \hat{\omega}^c_{ba}\right)\hat{V}_c. \quad (27)$$

It follows thereof, in particular, that $\hat{D}_c\eta_{ab} = 0$. Define the gauge strength as usual:

$$\hat{F}_{ab} = \left(\hat{D}_a + i\hat{V}_a\right)\hat{V}_b - (a \leftrightarrow b). \quad (28)$$

It follows that the so-defined gauge strength takes the form $\hat{F}_{ab} = \kappa^\alpha{}_a\kappa^\beta{}_b F_{\alpha\beta}$, with

$$F_{\alpha\beta} = \left(\partial_\alpha + iV_\alpha\right)V_\beta - (\alpha \leftrightarrow \beta). \quad (29)$$

Thus, $\hat{F}_{ab}$ does possess the correct transformation properties with respect to both the affine symmetry and the internal gauge symmetry.

Lorentz symmetry. Further, consider the local Lorentz symmetry as the gauge one with the connection $\hat{\omega}_c \equiv (1/2)\hat{\omega}_{abc}\hat{\Sigma}^{ab}_-$, where $\hat{\Sigma}^{ab}_-$ are some generic Lorentz generators. Define the corresponding gauge strength for the affine Goldstone boson as

$$\hat{G}_{cd} = (\hat{\partial}_c + i\hat{\omega}_c)\hat{\omega}_d - (c \leftrightarrow d) \equiv \frac{1}{2}\hat{R}_{abcd}\hat{\Sigma}^{ab}_-. \quad (30)$$

This gives

$$\hat{R}_{abcd} = \hat{\partial}_c\hat{\omega}_{abd} - \hat{\omega}^f{}_{ac}\hat{\omega}_{fbd} - (c \leftrightarrow d). \quad (31)$$

This quantity transforms homogeneously as the local Lorentz tensor (and similarly for its partial contraction $\hat{R}_{bd} \equiv \hat{R}^a{}_{bad}$). The total contraction

$$\hat{R} \equiv \hat{R}^{ab}{}_{ab} = 2\hat{\partial}_a\hat{\omega}^{ab}{}_b - \hat{\omega}^{fa}{}_a\hat{\omega}_f{}^b{}_b + \hat{\omega}^{fab}\hat{\omega}_{fba} \quad (32)$$

is the local Lorentz scalar and can be used in building the Lagrangian for the Goldstone boson.

4.3. Lagrangian

Lorentz-invariant form. The constructed objects can serve as the building blocks for the nonlinear model $GL(4,R)/SO(1,3)$ in the tangent space. Postulate the equivalence principle in the sense that the tangent space Lagrangian should not depend explicitly on the tangent space counterpart of the background curvature $\bar{\rho}_{\gamma\alpha\delta\beta}$ [Eq. (5)]. Thus, the Lagrangian may be written as the general Lorentz-invariant function built of $\hat{R}$, $\hat{F}_{ab}$, $\hat{D}_a\phi$, and ϕ themselves. As usual, we restrict ourselves to terms containing two derivatives at most.

The generic Lorentz-invariant (and thus affine) Lagrangian in the tangent space is

$$L = L_g(\hat{R}) + L_r(\hat{F}_{ab}) + L_m(\hat{D}_a\phi, \phi). \quad (33)$$

In the above, the basic Goldstone Lagrangian L_g is as follows:

$$L_g = c_g M_A^2\left(-\frac{1}{2}\hat{R}(\hat{\omega}_{abc}) + \Lambda\right), \quad (34)$$

with c_g being a dimensionless constant, to be chosen, and Λ proving to be in what follows the cosmological constant. The radiation Lagrangian L_r is, as usual,

$$L_r = -\frac{1}{4}\mathrm{tr}(\hat{F}^{ab}\hat{F}_{ab}). \tag{35}$$

Finally, L_m is the proper matter Lagrangian.[9] As for the radiation and matter, their Lagrangian could well be the affine invariant Lagrangian of the Standard Model or of any of its extensions. In fact, the given nonlinear model can accommodate any field theory.

Affine invariant form. The Lagrangian above gives the basic dynamical description of the affine Goldstone boson, radiation, and matter. The local Lorentz representations are necessary to construct the Lagrangian. The latter being built, one can rewrite it in terms of the affine representations. This allows one to make explicit the geometrical structure of the theory and to relate it with the gravity. This is achieved by the proper regrouping of the factors $\kappa^{\alpha}{}_{a}$ and $\kappa^{-1a}{}_{\alpha}$, in order to make the affine indices explicit. The Lagrangian now becomes

$$L = c_g M_A^2\left(-\frac{1}{2}R(\gamma_{\alpha\beta}) + \Lambda\right) + L_r(F_{\alpha\beta}) + L_m(D_\alpha\phi, \phi). \tag{36}$$

Here,

$$\gamma_{\alpha\beta} = \kappa^{-1a}{}_{\alpha}\eta_{ab}\kappa^{-1b}{}_{\beta} \tag{37}$$

transforms as the affine tensor

$$(A, a) : \gamma_{\alpha\beta} \to \gamma'_{\alpha\beta} = A^{-1\gamma}_{\alpha}\gamma_{\gamma\delta}A^{-1\delta}_{\beta}. \tag{38}$$

It proves that $R(\gamma_{\alpha\beta}) = \hat{R}(\hat{\omega}_{abc})$ can be expressed as the contraction $R = R^{\alpha\beta}{}_{\alpha\beta}$ of the tensor $R^{\gamma}{}_{\alpha\delta\beta} \equiv \kappa^{\gamma c}\kappa^{-1a}_{\alpha}\kappa^{-1d}_{\delta}\kappa^{-1b}_{\beta}\hat{R}_{cadb}$, the latter in turn being related with $\gamma_{\alpha\beta}$ as the Riemann–Christoffel curvature tensor with the metric. In this, all the contractions of the affine indices are understood with $\gamma_{\alpha\beta}$ (respectively, $\gamma^{\alpha\beta}$).

Similarly, $D_\alpha\phi$ looks like the generally covariant derivative for the matter fields

$$D_\gamma\phi = \left(\partial_\gamma + \frac{i}{2}\omega_{ab\gamma}\hat{\Sigma}^{ab}_{\phi}\right)\phi, \tag{39}$$

with the spin connection

$$\omega_{ab\gamma} \equiv \kappa^{-1c}_{\gamma}\hat{\omega}_{abc} = \kappa^{\beta}{}_{a}\nabla_\gamma\kappa^{-1}{}_{b\beta} - (a \leftrightarrow b). \tag{40}$$

In the above, $\nabla_\gamma\kappa^{-1}_{b\beta} \equiv (\delta^{\alpha}_{\beta}\partial_\gamma - \Gamma^{\alpha}_{\beta\gamma})\kappa^{-1}_{b\alpha}$ is the covariant derivative calculated with the Christoffel connection

$$\Gamma^{\alpha}_{\beta\gamma} = \kappa^{\alpha a}\kappa^{-1b}_{\beta}\kappa^{-1c}_{\gamma}\hat{\omega}_{abc} + \kappa^{\alpha}_{a}\partial_\gamma\kappa^{-1a}_{\beta} = \frac{1}{2}\gamma^{\alpha\delta}\left(\partial_\beta\gamma_{\delta\gamma} + \partial_\gamma\gamma_{\delta\beta} - \partial_\delta\gamma_{\beta\gamma}\right). \tag{41}$$

In particular, one gets $\nabla_\gamma\gamma_{\alpha\beta} = 0$ as the affine counterpart of the Lorentz relation $\hat{D}_c\eta_{ab} = 0$. For the radiation Lagrangian, one has the usual expression

$$L_r = -\frac{1}{4}\mathrm{tr}(F^{\alpha\beta}F_{\alpha\beta}), \tag{42}$$

with $F_{\alpha\beta}$ given by Eq. (29). Finally, the matter Lagrangian is obtained straightforwardly from L_m [Eq. (33)] with account of Eq. (37) and the relation $\hat{D}_a = \kappa^{\alpha}{}_{a}D_\alpha$ [Eq. (22)].

Clearly, L_g looks like the GR Lagrangian in the tangent space considered as the effective[10] Riemannian manifold with the metric $\gamma_{\alpha\beta}$, the Christoffel connection $\Gamma^{\gamma}{}_{\alpha\beta}$, the Riemann–Christoffel curvature tensor $R^{\gamma}{}_{\alpha\delta\beta}$, the Ricci tensor $R_{\alpha\beta}$, the Ricci scalar R, and the tetrad $\kappa^{-1a}{}_{\alpha}$ (the inverse one $\kappa^{\alpha}{}_{a}$). This is in no way accidental. Namely, as is shown in [2], under the special choice of the nonlinear connection Eq. (25), the Lagrangian becomes conformally invariant too. In this, the dilaton of the conformal symmetry coincides with the affine dilaton, while the vector Goldstone boson of the conformal symmetry, proving to be the derivative of the dilaton, is auxiliary. Further, according to the theorem due to Ogievetsky [6], it follows that the theory which is invariant under both the conformal symmetry and the global affine one is generally covariant, as well. After the proper choice of the metric, this imposes the effective Riemannian structure onto the tangent space. In the world coordinates, this will result in the generally covariant theory (GR, in particular). Precisely this property justifies the special choice of Eq. (25) for the nonlinear connection, with the Goldstone boson being the graviton in disguise.

5. METAGRAVITATION

5.1. General Covariance

Take the tangent space Lagrangian as that for the spacetime, being valid in the background inertial coordinates in the infinitesimal neighborhood of the reference point P. After the subsequent multiplication of the Lagrangian by the covariant volume element $(-\gamma)^{1/2}d^4\Xi$, with $\gamma \equiv \det\gamma_{\alpha\beta}$, one gets the contribution to the action of the infinitesimal neighborhood of the point P. Now one has to convert this contribution

[9] The matter Lagrangian is normalized so that $L_m|_{\phi=0} = 0$.

[10] For brevity, the term "effective" will be omitted, while the term "background" will, in contrast, be retained.

to the arbitrary world coordinates and to sum over the whole spacetime.

The relation between the background inertial and world coordinates is achieved by means of the background tetrad $\bar{e}^\alpha_\mu(X)$ [Eq. (2)], with the world metric being as follows:

$$g_{\mu\nu}(X) = \bar{e}^\alpha_\mu(X)\gamma_{\alpha\beta}(\bar{\Xi})\bar{e}^\beta_\nu(X). \quad (43)$$

With account of Eqs. (4), (38), this metric is invariant under the affine transformations

$$(A, a) : g_{\mu\nu} \to g_{\mu\nu}. \quad (44)$$

By their very construction, the world coordinates are unchanged as well:

$$(A, a) : X^\mu \to X^\mu. \quad (45)$$

As a result, the effective interval $ds^2 = g_{\mu\nu}dX^\mu dX^\nu$ remains invariant too.

Now, introduce the effective tetrad related with the background one as

$$e^a_\mu(X) = \kappa^{-1a}{}_\alpha(\bar{\Xi})\bar{e}^\alpha_\mu(X). \quad (46)$$

The effective tetrad transforms as the local Lorentz vector:

$$e_\mu(X) \to e'_\mu(X) = \Lambda(X)e_\mu(X). \quad (47)$$

Due to the local Lorentz transformations $\Lambda(X)$, one can eliminate six components from e^a_μ, the latter thus having ten independent components. In these terms, the world metric is

$$g_{\mu\nu}(X) = e^a_\mu(X)\eta_{ab}e^b_\nu(X). \quad (48)$$

In other words, the tetrad e^a_μ defines the effective inertial coordinates. Physically, Eq. (46) describes the disorientation of the effective inertial and background inertial frames depending on the distribution of the affine Goldstone boson (and thus the gravity).

Accounting for the relation $d\bar{\Xi}^\alpha = \bar{e}^\alpha_\mu dX^\mu$ between the displacements of the point P in the background inertial and world coordinates, and thus $\partial\bar{\Xi}^\alpha/\partial X^\mu = \bar{e}^\alpha_\mu$, one has

$$\Gamma^\lambda{}_{\mu\nu} = \bar{e}^\lambda_\alpha\bar{e}^\beta_\mu\bar{e}^\gamma_\nu\Gamma^\alpha_{\beta\gamma} + \bar{e}^\lambda_\alpha\partial_\mu\bar{e}^\alpha_\nu, \quad (49)$$

where $\partial_\mu = \partial/\partial X^\mu$. This can be rewritten, as usual, as

$$\Gamma^\lambda{}_{\mu\nu} = \frac{1}{2}g^{\lambda\rho}\Big(\partial_\mu g_{\rho\nu} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}\Big). \quad (50)$$

By construction, the world indices are manipulated via $g_{\mu\nu}$ and $g^{\mu\nu}$. The spin connection looks in the world coordinates like

$$\omega_{ab\mu} = \omega_{ab\gamma}\bar{e}^\gamma_\mu = e^\nu_a\nabla_\mu e_{b\nu} - (a \leftrightarrow b), \quad (51)$$

with the generally covariant derivative ∇_μ defined via the Christoffel connection $\Gamma^\lambda{}_{\mu\nu}$, as usual. Correspondingly, the covariant derivative of the matter fields looks like

$$D_\mu\phi = \Big(\partial_\mu + \frac{i}{2}\omega_{ab\mu}\hat{\Sigma}^{ab}_\phi\Big)\phi. \quad (52)$$

In a similar way, one finds the usual expressions for the Riemann–Christoffel tensor $R^\lambda{}_{\mu\rho\nu}(g)$, the Ricci tensor $R_{\mu\nu} = R^\lambda{}_{\mu\lambda\nu}$, and the Ricci scalar $R = g^{\mu\nu}R_{\mu\nu}$. The same is true for the gauge strength:

$$F_{\mu\nu} = (\partial_\mu + iV_\mu)V_\nu - (\mu \leftrightarrow \nu). \quad (53)$$

Plugging the above-modified objects into the Lagrangians for the affine Goldstone boson, radiation, and matter and integrating with the invariant volume element, one gets the total action, the Einstein–Hilbert one including

$$S = \int\Bigg[M^2_{\rm Pl}\left(-\frac{1}{2}R(g_{\mu\nu}) + \Lambda\right) + L_r(F_{\mu\nu}) + L_m(D_\mu\phi, \phi)\Bigg](-g)^{1/2}d^4X, \quad (54)$$

with $g \equiv \det g_{\mu\nu}$. In the above, the constant c_g in Eq. (36) is chosen so that $c_g M^2_A = 1/(8\pi G_{\rm N}) \equiv M^2_{\rm Pl}$, with G_N being Newton's constant and $M_{\rm Pl}$ being the Planck mass. Varying the action with respect to the metric $g^{\mu\nu}$, one arrives at the well-known equation of motion for gravity:

$$G_{\mu\nu} = M^{-2}_{\rm Pl}T_{\mu\nu}. \quad (55)$$

In the above, $G_{\mu\nu}$ is the gravity tensor,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}, \quad (56)$$

and $T_{\mu\nu} = T^r_{\mu\nu} + T^m_{\mu\nu}$ is the conventional energy–momentum tensor of the radiation and matter, produced by L_r and L_m.

5.2. General Covariance Violation

By choosing the generally covariant Lagrangian in the tangent space, one arrives at the generally covariant theory in the spacetime. Modulo the choice of the Lagrangian, such a theory is unique, independent of the choice of the coordinates. In particular, one manages to express everything exclusively in the internal dynamical terms (but for the numerical parameters). Under extension of the tangent space Lagrangian beyond the generally covariant one, the theory in the spacetime ceases to be generally covariant and thus unique. It depends not only on the Lagrangian but also on the choice of the coordinates. Relative to

the general coordinate transformations, the variety of theories divides into observationally inequivalent classes, each of which is characterized by a particular set of background parameter functions. Precisely the latter ones make the coordinates distinguishable. A priori, no one of the sets of parameter functions is preferable. Which one is suitable (if any) should be determined by observations. Each class of theories consists of the equivalent theories related by the residual covariance group. The latter consists of the coordinate transformations leaving the background parameter functions invariant. On the contrary, one class can be obtained from another by the coordinate transformations changing these parameter functions. Weakening the requirements on the bundling of the tangent spaces, one extends the set of the admissible theories, but arrives, instead, at the dependence of the theory in the spacetime on the more elaborate properties of the background.

To clarify the corresponding parameter functions, construct the background metric

$$\bar{g}_{\mu\nu}(X) = \bar{e}^A_\mu(X)\eta_{AB}\bar{e}^B_\nu(X), \tag{57}$$

with the inverse one

$$\bar{g}^{-1\mu\nu}(X) = \bar{e}^\mu_A(X)\eta^{AB}\bar{e}^\nu_B(X). \tag{58}$$

Here, the generic index A means a or α, as appropriate (and similarly for B, b, β, etc.). This metric transforms intricately under arbitrary affine transformations,

$$(A,a) : \bar{g}_{\mu\nu} \to \bar{g}'_{\mu\nu} = \bar{e}^T_\mu A^T \eta A \bar{e}_\nu \neq \bar{g}_{\mu\nu}, \tag{59}$$

though being invariant under the Poincare transformations (Λ, a). The metric $\bar{g}_{\mu\nu}$ is the next of kin to the primordial one $\bar{\varphi}_{\mu\nu}$ [Eq. (5)]. The former approximates the latter as closely as possible in the lack of knowledge of the primordial background curvature $\bar{\rho}^\gamma{}_{\alpha\delta\beta}$ [Eq. (5)]. According to the equivalence principle, this curvature does not enter the tangent space Lagrangian and thus is inessential. Correspondingly, the Christoffel connection $\bar{\Gamma}^\lambda{}_{\mu\nu}(\bar{g})$ approximates with the same accuracy the Christoffel connection $\bar{\chi}^\lambda{}_{\mu\nu}(\bar{\varphi})$ and thus the primordial affine connection $\bar{\phi}^\lambda{}_{\mu\nu}$, i.e., $\bar{\Gamma}^\lambda{}_{\mu\nu} \simeq \bar{\chi}^\lambda{}_{\mu\nu} \simeq \bar{\phi}^\lambda{}_{\mu\nu}$. So, under reasonable assumptions, it suffices to know only the background metric $\bar{g}_{\mu\nu}$.

Affine symmetry preservation. To be more specific, consider the extension of the tangent space Lagrangian for the Goldstone boson by means of the terms depending explicitly on $\hat{\Omega}^+{}_{abc}$ [Eq. (23)]. For example, one can add to the basic Goldstone Lagrangian Eq. (34) the quadratic piece

$$\Delta L^{(0)}_g = \frac{1}{2}\varepsilon_0 M^2_{\rm Pl}\hat{\Omega}^{+b}{}_{ba}\hat{\Omega}^+{}_c{}^{ca}, \tag{60}$$

where ε_0 is a dimensionless constant. Accounting for Eqs. (23) and (41), one gets for $\sigma_\alpha \equiv -\kappa^{-1a}_\alpha \hat{\Omega}^{+b}_{ba}$ the relation $\sigma_\alpha = \Gamma^\beta_{\beta\alpha}(\gamma) = \partial_\alpha\sigma$, where $\sigma \equiv (1/2)\ln(-\gamma)$ and $\gamma \equiv \det\gamma_{\alpha\beta} = -(\det\kappa^a{}_\alpha)^{-2}$. In the affine terms, one has

$$\Delta L^{(0)}_g = \frac{1}{2}\varepsilon_0 M^2_{\rm Pl}\gamma^{\alpha\beta}\partial_\alpha\sigma\partial_\beta\sigma. \tag{61}$$

This Lagrangian violates the conformal symmetry in the tangent space (more particularly, the local dilatation), as well as the general covariance, though not violating the global affine symmetry.

It follows from Eq. (43) that $\gamma = -g/\bar{g}$, where $\bar{g} = \det\bar{g}_{\mu\nu} = -(\det\bar{e}^\alpha_\mu)^2$. In the world coordinates, the Lagrangian $\Delta L^{(0)}_g$ becomes

$$\Delta L^{(0)}_g = \frac{1}{2}\varepsilon_0 M^2_{\rm Pl} g^{\mu\nu}\partial_\mu\sigma\partial_\nu\sigma, \tag{62}$$

with

$$\sigma \equiv \frac{1}{2}\ln(g/\bar{g}) \tag{63}$$

and

$$\partial_\mu\sigma = \Gamma^\lambda{}_{\lambda\mu}(g) - \bar{\Gamma}^\lambda{}_{\lambda\mu}(\bar{g}). \tag{64}$$

Thus, all the background dependence in the given case is determined only by the scalar density $\bar{g}$. Note that $\partial_\mu\sigma$ transforms homogeneously and thus cannot be eliminated by the coordinate transformations, though each one of the contributions could separately be nullified by the choice of coordinates.[11)]

Varying the total action (the Lagrangian $\Delta L^{(0)}_g$ included), one arrives at the modification of the equation of motion for gravity [Eq. (55)], with an extra piece in the gravity tensor $\Delta G^{(0)}_{\mu\nu}$. Introducing the derivative couplings of σ with matter, not violating explicitly the affine symmetry, one would get the extra piece $\Delta T^m_{\mu\nu}$ in the energy–momentum tensor for matter. Clearly, the modified theory, though not generally being covariant, is consistent with the unimodular covariance, i.e., that leaving $\bar{g}$ (as well as g) invariant. The unimodular covariance is next-of-kin to the general one. Due to this residual covariance, the given GR extension describes only three physical degrees of freedom corresponding to the "scalar" and massless tensor gravitons. In the case $\varepsilon_0 = 0$, the general covariance is restored, thus eliminating one more degree of freedom. This leaves just two of them with helicities $\lambda = \pm 2$, as should be the case for the massless spin-2 particle.

[11)] Under $\bar{g} = -1$, the given GR extension reduces to that of [7].

The extra terms in the Goldstone boson Lagrangian would make physical the other latent degrees of freedom of the gravity field, but at the cost of further violating the general covariance. For example, one could supplement the Goldstone Lagrangian by the other independent quadratic pieces:

$$\Delta L_g^{(1)} = \varepsilon_1 M_{\rm Pl}^2 \hat{\Omega}^+{}_{ab}{}^b \hat{\Omega}^{+ac}{}_c, \qquad (65)$$
$$\Delta L_g^{(2)} = \varepsilon_2' M_{\rm Pl}^2 \hat{\Omega}^+_{abc} \hat{\Omega}^{+abc} + \varepsilon_2'' M_{\rm Pl}^2 \hat{\Omega}^+_{abc} \hat{\Omega}^{+cab}.$$

This would, in particular, violate causality for the "vector" graviton, as well as modify interactions for the tensor graviton. Phenomenologically, these and similar modifications could be done as small as necessary by the choice of the numerical parameters ε. This is insured by the fact that, in the limit where these parameters vanish, the symmetry of the theory increases up to the general covariance.

Affine symmetry violation. The derivative couplings above preserve the affine symmetry, though violating the general covariance. One can conceive another way to violate the general covariance by introducing into the tangent space Lagrangian the potential $U_g(\kappa)$, which contains only the derivativeless couplings of the Goldstone boson. Of necessity, this would explicitly violate the affine symmetry too. To preserve the local Lorentz symmetry, the potential should depend only on $\gamma_{\alpha\beta}$ (and/or $\gamma^{\alpha\beta}$). In order not to violate the global Lorentz symmetry too, the potential is to be chosen as a Lorentz-invariant function as follows:

$$U_g = U_g\Big(\det\gamma, \mathrm{tr}(\gamma\eta)^n\Big), \qquad (66)$$

with any degree n. In the above, one sets $(\gamma\eta)_A{}^B \equiv \gamma_{AA'}\eta^{A'B}$, where as before $A = a$ or α, etc., as appropriate. At $n < 0$, one uses the relation $(\gamma\eta)^n = (\eta\gamma^{-1})^{|n|}$, with $\gamma^{-1\alpha\beta} \equiv \gamma^{\alpha\beta}$. It follows therefore that, in the world terms, the potential should depend on $g\bar{g}^{-1}$:

$$U_g = U_g\Big(\det(g\bar{g}^{-1}), \mathrm{tr}(g\bar{g}^{-1})^n\Big), \qquad (67)$$

with the background metric given by Eqs. (57), (58). Generally, one has $\bar{g}^{-1\mu\nu} \neq \bar{g}^{\mu\nu} \equiv g^{\mu\mu'}g^{\nu\nu'}\bar{g}_{\mu'\nu'}$ (and similarly, $\bar{g}^{-1}_{\mu\nu} \equiv g_{\mu\mu'}g_{\nu\nu'}\bar{g}^{-1\mu'\nu'} \neq \bar{g}_{\mu\nu}$). At negative n, one sets $(g\bar{g}^{-1})^n = (\bar{g}g^{-1})^{|n|}$, with $g^{-1\mu\nu} \equiv g^{\mu\nu}$. Here, the terms depending only on $\det(g\bar{g}^{-1}) = e^{2\sigma}$ are unimodular covariant. The potential above corresponds to the case of the most general graviton mass with Lorentz symmetry preservation.[12)]

With advent of the potential, the only modification of the gravity equation of motion is the appearance of the extra piece $\Delta G_{\mu\nu}^{(U)}$ on the left-hand side of Eq. (55). The Bianchi identity states the covariant divergenceless of the gravity tensor $G_{\mu\nu}$ [Eq. (56)]. Owing to this identity, there appear four constraints on the metric field and its first derivative. These constraints substitute the Lorentz–Hilbert gauge condition. Thus, at the level of the equation of motion, the theory describes six physical degrees of freedom, the massive scalar and tensor gravitons. Choosing different contributions to U_g, one can vary the relation between the respective masses. In the limit of vanishing potential, the general covariance is restored and one recovers smoothly the GR with the massless two-component tensor graviton.

One more similar source of the general covariance violation could be due to the derivativeless couplings of the affine Goldstone boson with matter. Violating the affine symmetry, all the derivativeless couplings are expected naturally to be suppressed (if any). This is in contrast to the extra terms depending on the derivatives of the Goldstone boson. The latter terms also result in the general covariance violation. Nevertheless, being affine invariant, they are not expected a priori to be small.

This exhausts the foundations of the effective field theory of gravity, radiation, and matter. The above theory, embodying GR and its extensions in the framework of the affine symmetry and general relativity, may be called *metagravitation*.

6. METAUNIVERSE

6.1. World Continuum

The ultimate goal of the Goldstone approach to gravity is to go beyond the effective metric theory and to build the underlying premetric one. In what follows, we present some hints of the respective scenario. Of necessity, we will be very concise, just to indicate the idea.

The forebear of the spacetime is supposed to be the world continuum. At the very least, the latter is to be endowed with the defining structure, the continuity in the topological sense. Being covered additionally with the patches of the smooth real coordinates x^μ,

[12)] For the theory of the massive tensor field in the Minkowski background spacetime, see, e.g., [8]. For the phenomenology of the graviton mass and for further references on the subject, cf., e.g., [9].

$\mu = 0, 1, \ldots, d-1$ (index 0 having as yet no particular meaning), the continuum acquires the structure of the differentiable manifold of the dimension d (4, for definiteness). There exist in the continuum the tensor densities, in particular, the volume element. Thus, integration over the manifold is allowed. But it does not suffice to define the covariant derivative and thus to get the covariant differential equations, etc. Suppose now that the continuum can exist in two phases with the following affinity properties.

Affine connection. Being endowed with the primordial affine connection $\bar{\phi}^{\lambda}{}_{\mu\nu}$, the continuum becomes the affinely connected manifold. Generally, the connection is a 64-parametric structure. It defines the parallel transport of the world vectors, as well as their covariant derivatives. The parallel transport along the infinitesimal closed contour defines, in turn, the background curvature tensor $\bar{\rho}^{\lambda}{}_{\mu\rho\nu}$ and thus its contraction $\bar{\rho}_{\mu\nu} = \bar{\rho}^{\lambda}{}_{\mu\lambda\nu}$ (but not yet the scalar $\bar{\rho}$). To every point P, there can be attached the coordinates $\bar{\xi}^{\alpha}$, where the symmetric part of the connection locally nullifies, the manifold becoming thus locally affinely flat. This defines the global affine symmetry. For the symmetry to be exact, the antisymmetric part of the connection, the torsion, should be trivial, with the connection being just 40-parametric. In this phase, there is as yet no metric and thus no space and time directions, even no definite spacetime signature, no lengths and angles, no preferred Lorentz group and thus no finite-dimensional spinors, no preferred Poincare group and thus no conventional particles, no invariant intervals, no quadratic invariants, no causality, etc. Though there can be implemented the principle of the least action with the primitive Lagrangians, the world structure is still rather dull. Nevertheless, it should ultimately lead to the spontaneous transition from the given phase to the metric one.

Metric. Further, being endowed spontaneously with the metric $\bar{\varphi}_{\mu\nu}$ having the Minkowskian signature, the continuum becomes the metric space, i.e., the spacetime. The metric is a much more restrictive ten-parametric structure. It defines the background Riemannian geometry. Accompanying the emergence of the metric and the spontaneous breaking of the affine symmetry, there appears the affine Goldstone boson serving as the graviton in disguise. This results in the effective Riemannian geometry with the effective metric $g_{\mu\nu}$, etc. Now there appear the preferred time and space directions, the lengths and angles, the definite Lorentz group and thus the finite-dimensional spinors, the definite Poincare group and thus the particles, the invariant intervals, the quadratic invariants, the causality, etc. The world structure becomes now very flourishing. In the wake of the gravity, there appears the conventional matter. The spontaneous breaking of the affine symmetry to the Poincare one reflects the appearance of the coherent particle structure, among a lot of a priori possible ones corresponding to the various choices of the Poincare subgroup. Formally, the effective Riemannian geometry is to be valid at all the spacetime intervals. Nevertheless, its accuracy worsens when diminishing the intervals, requiring more and more terms in the decomposition over the ratio of the energy to the symmetry breaking scale M_A, as should be the case for the effective theory. Thus, the scale M_A (or, rather, the Planck mass M_{Pl}) is a kind of inverse minimal length in nature.

6.2. The Universe

Conceivably, the formation of the Universe is the result of the actual transition between the two phases of the continuum. This transition is thus the *"Grand Bang,"* the origin not only of the Universe but also of the very spacetime. At this stage, there appears the world "arrow of time" as the reflection of the spontaneous synchronization of the chaotic local times. The residual dependence of the structure of the Universe on the background parameter functions could result in the variety of the primordial effects, such as anisotropy and inhomogeneity. And what is more, there is conceivably the appearance (as well as disappearance and coalescence) of the various regions of the metric phase inside the affinely connected one (and vice versa). These regions are to be associated with the multiple universes. One of the latter ones happens to be ours. Call the ensemble of the universes the *Metauniverse*. Within the concept of the Metauniverse, there becomes sensible the notion of the wave function of the Universe. Hopefully, this may clarify the long-standing problem of the fine-tuning of our Universe.[13)]

7. CONCLUSION

To conclude, the theory proposed realizes consistently the approach to gravity as the Goldstone phenomenon. It proceeds, in essence, from the two basic symmetries: the global affine one and the general covariance. The affine symmetry is the structure symmetry which defines the theory in the small. The general covariance is the bundling symmetry which terminates the a priori admissible local theories according to their ability to be prolonged onto the spacetime. The theory embodies GR as the lowest approximation. Its distinction from GR is twofold.

[13)] Cf., e.g., [10].

At the effective level, the given theory predicts the natural hierarchy of the conceivable GR extensions, according to the accuracy of the affine symmetry realization. At the underlying level, the theory presents a new look at gravitation, the Universe, and the very spacetime.

ACKNOWLEDGMENTS

I am grateful to V.V. Kabachenko for useful discussions.

REFERENCES

1. C. Isham, A. Salam, and J. Strathdee, Ann. Phys. (N.Y.) **62**, 98 (1971).
2. A. B. Borisov and V. I. Ogievetsky, Teor. Mat. Fiz. **21**, 329 (1974).
3. E. Schrödinger, *Space–Time Structure* (Cambridge Univ. Press, Cambridge, 1950).
4. S. R. Coleman, *Dilatations*, in *Proceedings of the International Summer School of Physics Ettore Majorana*, Ed. by A. Zichichi (Erice, 1971).
5. S. R. Coleman, J. Wess, and B. Zumino, Phys. Rev. **177**, 2239 (1969); D. V. Volkov, Fiz. Élem. Chastits At. Yadra **4**, 3 (1973) [Sov. J. Part. Nucl. **4**, 1 (1973)].
6. V. I. Ogievetsky, Lett. Nuovo Cimento **8**, 988 (1973).
7. W. Buchmüller and N. Dragon, Nucl. Phys. B **321**, 207 (1989).
8. V. I. Ogievetsky and I. V. Polubarinov, Ann. Phys. (N.Y.) **35**, 167 (1965).
9. M. Visser, Gen. Relativ. Gravit. **30**, 1717 (1998); gr-qc/9705051.
10. P. C. W. Davies, *The Accidental Universe* (Cambridge Univ. Press, Cambridge, 1982; Mir, Moscow, 1985).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1916–1921. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1978–1983.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS"
ARMENIA, YEREVAN, AUGUST 13–19, 2003

An Approach to Reduction of the Real Symmetric 3 × 3 Secular Problem and Applications*

I. M. Mladenov[1] and C. D. Mladenova[2]

Received October 29, 2004

Abstract—An original analytical method for casting a real three-dimensional symmetric matrix into its diagonal form along with explicit formulas for the corresponding eigenvectors is given. This is achieved by a two-step procedure relying on a nice geometrical parametrization of the rotational group $\mathbb{SO}(3)$ and the fundamental algebraic theorem about solutions of the polynomial equations. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The main objective of the present paper is to solve the problem of diagonalization of a real three-dimensional symmetric matrix with determination of explicit formulas for the corresponding eigenvectors.

A straightforward analysis of the procedure of finding eigenvalues of a symmetric 3×3 matrix can be found in Smith [1], while Bojanczyk and Lutoborski [2] provide explicit formulas for eigenvectors. This was done using the standard result about rotational matrices in real three-dimensional space, namely, that any of them can be represented as a product of three plane rotations. An interesting algorithm for the construction of the eigenvector matrix and the spectrum of 4×4 real symmetric matrices is presented in [3]. Using the well-known fact that the Lie algebra of $\mathbb{O}(4)$ is isomorphic to the product of two copies of the Lie algebra of $\mathbb{O}(3)$, the problem is reduced eventually to the three-dimensional case.

The new procedure which we follow is based on the so-called vector parametrization that is quite different from the standard Euler parametrization of the rotational group $\mathbb{SO}(3)$. Contrary to other known coordinatizations of $\mathbb{SO}(3)$, the vector parametrization is a symmetrical and a natural one. The most important property of this parametrization is that the vector parameters constitute a Lie group with a simple composition law and clear geometrical sense. Our aim here is to find the vector parameter defining the orthogonal matrix which diagonalizes a given symmetric matrix. This process is realized in two stages. At each level, we obtain the corresponding vector parameter and the resultant vector parameter is their composition.

Why have we chosen the vector parameter to reach the above aim? The answer to this question is quite simple—this is the only parametrization which deals with algebraic parameters that are free of redundancy and are not transcendental. This is not the case with the Eulerian angles, Bryant angles, Eulerian parameters, or quaternions, or with Euler–Rodrigues parameters or Gibbs vectors which receive standard consideration in the literature. In particular, the last two are classical parametrizations which play an important role in the geometrical and kinematical descriptions of motion, especially in the dynamics of spacecraft and aircraft. But the vector parametrization is the best among others coordinatizations for all these purposes. On the basis of vector parametrization, we have developed a unified and numerically efficient approach for kinematical and dynamical modeling and control of a rigid body and mechanical systems of rigid and elastic bodies [4]. Because of the decoupling of the differential equations of motion, the problem of diagonalization of the inertia matrix and the stress matrices is of great importance for modeling of mechanical systems with rigid and elastic links.

The structure of the paper is as follows. Some basic results about special orthogonal transformations in ordinary three-dimensional space are presented in the next section. In Sections 3 and 4, the principal moments of the diagonalization procedure are discussed. As an example of the proposed method, in Section 5, we consider in some detail the stress tensor associated with Boussinesq solution in elasticity theory. We end the paper with a few concluding remarks in Section 6.

*The text was submitted by the authors in English.

[1]Institute of Biophysics, Bulgarian Academy of Sciences, Sofia, Bulgaria; e-mail: mladenov@obzor.bio21.bas.bg

[2]Institute of Mechanics, Bulgarian Academy of Sciences, Sofia, Bulgaria; e-mail: clem@imbm.bas.bg

1063-7788/05/6811-1916$26.00

2. DIAGONALIZATION AND FEDOROV'S PARAMETRIZATION

Let $A = [a_{ij}]$, $i, j = 1, 2, 3$, be a real, symmetric 3×3 matrix. Due to the spectral theorem [5], A is diagonalizable; i.e., there exists an orthogonal matrix $O = [o_1, o_2, o_3]$ such that

$$O^T A O = \Lambda = \mathrm{diag}\{\lambda_{\pi(1)}, \lambda_{\pi(2)}, \lambda_{\pi(3)}\}, \quad \pi \in \Sigma_3, \tag{1}$$

where T means the transposed matrix, $\lambda_1 \le \lambda_2 \le \lambda_3$ are the eigenvalues of the matrix A, π is an element of the group Σ_3 of permutations of the three element set 1, 2, 3, and the column vectors o_1, o_2, o_3 of O are the orthonormal eigenvectors associated with these eigenvalues.

The special orthogonal group in the real Euclidean space $\mathbb{R}^3$ is the group

$$\mathbb{SO}(3) = \{O \in \mathrm{Mat}(3, \mathbb{R}); OO^T = I, \ \det O = 1\}, \tag{2}$$

where $\mathrm{Mat}(3, \mathbb{R})$ is the set of 3×3 real matrices and I is the 3×3 identity matrix. The Lie algebra (or infinitesimal generators) of the $\mathbb{SO}(3)$ group is the vector space of real skew-symmetrical 3×3 matrices and will be denoted by $\mathfrak{so}(3)$. Now, if X belongs to $\mathfrak{so}(3)$, the matrix $I - X$ is invertible.

The Cayley transformation providing an explicit connection between the Lie algebra $\mathfrak{so}(3)$ and the Lie group $\mathbb{SO}(3)$ is given by the formula

$$O = (I + X)(I - X)^{-1} \tag{3}$$
$$= (2I - (I - X))(I - X)^{-1} = 2(I - X)^{-1} - I.$$

As an exception in the three-dimensional space, there exists a map between vectors and skew-symmetrical matrices [6]; i.e., if $c \in \mathbb{R}^3$, then $c \to X \equiv c^\times$, where $c^\times$ is the corresponding skew-symmetric matrix. Using this fact, we may write the $\mathbb{SO}(3)$ matrix in the form (see [7] and also [8–10])

$$O = O(c) = (I + c^\times)(I - c^\times)^{-1} \tag{4}$$
$$= \frac{(1 - c^2)I + 2c \otimes c + 2c^\times}{1 + c^2},$$

which can be considered as a mapping from $\mathbb{R}^3$ to $\mathbb{SO}(3)$. On some open domain $\mathcal{D}$ of the group, it can be inverted and this yields the associated vector-parameter

$$c^\times = \frac{O - O^T}{1 + \mathrm{tr}O}, \tag{5}$$

where $c \otimes c$ means diadic and $\mathrm{tr}\,O$ is the sum of the diagonal elements of the matrix O. This formula provides us with an explicit parametrization of the rotational group $\mathbb{SO}(3)$. Being orthogonal, matrix O also satisfies the following relations:

$$1 + \mathrm{tr}O = \frac{4}{1 + c^2}, \quad O - O^T = \frac{4c^\times}{1 + c^2}. \tag{6}$$

The vector c is called vector parameter. It is directed along the rotation axis and its module is equal to

$$| \, c \, | = \tan(\alpha/2). \tag{7}$$

The vector parameters form a Lie group [7] with a group composition law following from the defining relation:

$$O(c)O(c') = O(c''), \tag{8}$$

where

$$c'' = \langle c, c' \rangle = \frac{c + c' + c \times c'}{1 - cc'}. \tag{9}$$

The symbol "$\times$" here means the cross product of vectors. More about the topological structure of the Lie group $\mathbb{SO}(3)$ and the superposition laws under above-mentioned parametrizations can be found in [11].

Our objective is the following: given a symmetric 3×3 matrix A, construct a diagonalizing rotation matrix $O(c)$, where c is the vector parameter of resultant rotation, such that

$$O^T(c) A O(c) = \Lambda = \mathrm{diag}\{\lambda_{\pi(1)}, \lambda_{\pi(2)}, \lambda_{\pi(3)}\}. \tag{10}$$

3. FIRST LEVEL OF DIAGONALIZATION

Let $A = [a_{ij}]$, $i, j = 1, 2, 3$, be a real, symmetric 3×3 matrix. Consider the vector parameter $c = (x, 0, z)$. The corresponding orthogonal matrix is $O(c) = O(x, z)$ and $O^T(c) = O^T(x, z)$. We denote by $B = [b_{ij}]$, the real symmetric 3×3 matrix

$$B = O^T(x, z) A O(x, z). \tag{11}$$

We are going to eliminate the elements b_{12} and b_{13}. For this purpose, we have to solve a pair of coupled highly nonlinear equations. Several methods on how to deal with such problems exist in computational kinematics. Most promising in our case seem to be the elimination methods. Here, we shall use the so called dialytic elimination method [12]. The basic steps in this method are the following:

(1) Rewrite equations with one variable suppressed.

(2) Define the remaining power products as new linear, homogeneous unknowns.

(3) Obtain new linear equations so as to have as many linearly independent homogeneous equations as linear unknowns.

(4) Set the determinant of the coefficient matrix to zero, and obtain a polynomial in the suppressed variable. (If one is interested in numerical solutions, this step is omitted and we can go directly to the next step.)

(5) Determine the roots of the characteristic polynomial or the eigenvalues of the matrix. This yields all possible values of the suppressed variable.

(6) Substitute (one of the roots or eigenvalues) the suppressed variable and solve the linear system for the remaining unknowns. Repeat this for each value of the suppressed variable.

Guided by this strategy, we present the elements b_{12} and b_{13} as polynomials of z whose coefficients depend on x. In this form, we denote them as B_{12} and B_{13}, respectively, i.e.,

$$B_{12} = F(z) = \mathrm{Pol}[b_{12}, z], \qquad (12)$$
$$B_{13} = G(z) = \mathrm{Pol}[b_{13}, z].$$

The degree of the polynomials B_{12} and B_{13} in the variable z is four. With the letters given below, we denote the coefficients of the polynomials B_{12} and B_{13} in front of the corresponding power of z:

$$C_0 = \mathrm{coeff}[B_{12}, z^0], \quad D_0 = \mathrm{coeff}[B_{13}, z^0], \qquad (13)$$
$$C_1 = \mathrm{coeff}[B_{12}, z^1], \quad D_1 = \mathrm{coeff}[B_{13}, z^1],$$
$$C_2 = \mathrm{coeff}[B_{12}, z^2], \quad D_2 = \mathrm{coeff}[B_{13}, z^2],$$
$$C_3 = \mathrm{coeff}[B_{12}, z^3], \quad D_3 = \mathrm{coeff}[B_{13}, z^3],$$
$$C_4 = \mathrm{coeff}[B_{12}, z^4], \quad D_4 = \mathrm{coeff}[B_{13}, z^4],$$

where

$$C_0 = -a_{12}x^4 + 2a_{13}x^3 + 2a_{13}x + a_{12}, \qquad (14)$$
$$C_1 = -2a_{23}x^3 + 4a_{33}x^2 - 2a_{11}x^2 - 2a_{22}x^2 + 6a_{23}x + 2a_{22} - 2a_{11},$$
$$C_2 = -6a_{13}x - 6a_{12},$$
$$C_3 = -2a_{23}x + 2a_{11} - 2a_{22}, \quad C_4 = a_{12},$$

and

$$D_0 = -a_{13}x^4 - 2a_{12}x^3 - 2a_{12}x + a_{13}, \qquad (15)$$
$$D_1 = -2a_{33}x^3 + 2a_{11}x^3 - 6a_{23}x^2 + 2a_{11}x - 4a_{22}x + 2a_{33}x + 2a_{23},$$
$$D_2 = \quad 6a_{13}x^2 + 6a_{12}x,$$
$$D_3 = -2a_{11}x + 2a_{33}x + 2a_{23},$$
$$D_4 = -a_{13}.$$

So, we have

$$B_{12} = C_4z^4 + C_3z^3 + C_2z^2 + C_1z + C_0, \qquad (16)$$
$$B_{13} = D_4z^4 + D_3z^3 + D_2z^2 + D_1z + D_0,$$

where one should take into account that the new coefficients $C_0, D_0, C_1, D_1, \ldots$ contain the suppressed variables x. In step two, we consider each power of z as separate independent linear indeterminate. We have to note also that the number one is counted as a variable as well since it is always convenient to have homogeneous equations and it provides a rationale to discard trivial solutions. The coefficient of the "variable" 1 is the constant term. Having in mind all these arguments, we rewrite Eq. (16) as the following linear set:

$$C_4Z_1 + C_3Z_2 + C_2Z_3 + C_1Z_4 + C_0Z_5 = 0, \qquad (17)$$
$$D_4Z_1 + D_3Z_2 + D_2Z_3 + D_1Z_4 + D_0Z_5 = 0.$$

Since we have two equations with five unknowns, we need additional equations. In our case, this can be accomplished by multiplying Eqs. (17) first by z, after that by z^2, and finally by z^3. So, we obtain eight equations with eight unknowns since three new power products appear. Using the concept of step two, we invoke new independent variables $Z_6 = z^5$, $Z_7 = z^6$, and $Z_8 = z^7$. As a result, we obtain a system of eight homogeneous linear equations in eight unknowns:

$$C_4Z_1 + C_3Z_2 + C_2Z_3 + C_1Z_4 + C_0Z_5 = 0, \qquad (18)$$
$$D_4Z_1 + D_3Z_2 + D_2Z_3 + D_1Z_4 + D_0Z_5 = 0,$$
$$C_4Z_6 + C_3Z_1 + C_2Z_2 + C_1Z_3 + C_0Z_4 = 0,$$
$$D_4Z_6 + D_3Z_1 + D_2Z_2 + D_1Z_3 + D_0Z_4 = 0,$$
$$C_4Z_7 + C_3Z_6 + C_2Z_1 + C_1Z_2 + C_0Z_3 = 0,$$
$$D_4Z_7 + D_3Z_6 + D_2Z_1 + D_1Z_2 + D_0Z_3 = 0,$$
$$C_4Z_8 + C_3Z_7 + C_2Z_6 + C_1Z_1 + C_0Z_2 = 0,$$
$$D_4Z_8 + D_3Z_7 + D_2Z_6 + D_1Z_1 + D_0Z_2 = 0.$$

This is the main idea in the dialytic elimination method, namely, that, even though the new equations are dependent on the original equations, their dependence is not linear but encoded into a linear system. We go to step four, where we obtain a single polynomial equation in the suppressed variable x. We rewrite the system (18) in a matrix form,

$$\begin{bmatrix} C_4 & C_3 & C_2 & C_1 & C_0 & 0 & 0 & 0 \\ D_4 & D_3 & D_2 & D_1 & D_0 & 0 & 0 & 0 \\ C_3 & C_2 & C_1 & C_0 & 0 & C_4 & 0 & 0 \\ D_3 & D_2 & D_1 & D_0 & 0 & D_4 & 0 & 0 \\ C_2 & C_1 & C_0 & 0 & 0 & C_3 & C_4 & 0 \\ D_2 & D_1 & D_0 & 0 & 0 & D_3 & D_4 & 0 \\ C_1 & C_0 & 0 & 0 & 0 & C_2 & C_3 & C_4 \\ D_1 & D_0 & 0 & 0 & 0 & D_2 & D_3 & D_4 \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \\ Z_5 \\ Z_6 \\ Z_7 \\ Z_8 \end{bmatrix} = 0, \qquad (19)$$

and in more compact notation, we have $UZ = 0$. Since we know that $Z_5 = 1$, then the trivial solution $Z_i \equiv 0 (i = 1, 2, ..., 8)$ is not admissible and therefore the determinant of the coefficient matrix U must be equal to zero, i.e., $\det U \equiv 0$. Having in mind that the entries of the matrix U contain the suppressed variable x, expansion of its determinant produces a polynomial of sixteenth degree which factorizes into the form

$$\det U = 256(1 + x^2)^3 P^2(x) Q(x), \qquad (20)$$

where

$$P(x) = (a_{11}a_{12}a_{23} - a_{12}^2 a_{13} + a_{13}a_{23}^2 - a_{12}a_{23}a_{33})x^3 + (a_{11}a_{12}a_{22} - a_{12}^3 + 2a_{12}a_{13}^2 + 2a_{13}a_{22}a_{23} - a_{11}a_{13}a_{23} - a_{12}a_{23}^2 - a_{11}a_{12}a_{33} - a_{12}a_{22}a_{33} - a_{13}a_{23}a_{33} + a_{12}a_{33}^2)x^2 + (2a_{12}^2 a_{13} - a_{13}^3 - a_{11}a_{13}a_{22} + a_{13}a_{22}^2 - a_{11}a_{12}a_{23} - a_{12}a_{22}a_{23} - a_{13}a_{23}^2 - a_{13}a_{22}a_{33} + a_{11}a_{13}a_{33} + 2a_{12}a_{23}a_{33})x + a_{11}a_{13}a_{23} - a_{12}a_{13}^2 - a_{13}a_{22}a_{23} + a_{12}a_{23}^2$$

and

$$Q(x) = (a_{11}^2 + 4a_{13}^2 - 2a_{11}a_{33} + a_{33}^2)x^4 + (8a_{12}a_{13} - 4a_{11}a_{23} + 4a_{23}a_{33})x^3 + (2a_{11}^2 + 4a_{12}^2 + 4a_{13}^2 - 2a_{11}a_{22} + 4a_{23}^2 - 2a_{11}a_{33} + 2a_{22}a_{33})x^2 + (8a_{12}a_{13} - 4a_{11}a_{23} + 4a_{22}a_{23})x + a_{11}^2 + 4a_{12}^2 - 2a_{11}a_{22} + a_{22}^2.$$

Abel's fundamental theorem in algebra says that it is always possible to write down the solutions of polynomial equations up to fourth degree in analytical form using rational operations and radicals, and this means that our equations

$$P(x) = 0 \quad \text{and} \quad Q(x) = 0 \qquad (21)$$

can be solved explicitly in any concrete case. In both analytical and numerical cases, we are only interested in real roots (at least one coming from the first of the above equations exists always). Therefore, any complex or purely imaginary roots which meet the determinant condition (as those coming from the multiplier $(1 + x^2)^3$) reduce the number of admissible solutions to the maximal possible value of seven.

Finally, in step six, we substitute the variable x into the linear set of equations and solve them for the other original variables, which in this case is z. Substituting any of the real roots of x into (19) and setting $Z_5 = 1$, we obtain the corresponding variable z. Since the system is linear, this yields just one z for each x [when the rank of the matrix U in (19) is maximal].

It is worth noting here that the introduction of new power products and the so-obtained additional equations is optimized in our approach and the proof of this fact is just the form of both polynomials which we have obtained. In spite of the fact that U is an 8×8 matrix, we manage to derive analytically solvable equations. Our experience shows that, if we start the procedure of diagonalization with another vector parameter (e.g., $\check{c} = (x, y, 0)$), the elimination procedure does not give polynomials of such low degrees.

So, in principle, one can associate with the seven couples (x_i, z_i) the corresponding seven vector parameters c_i $(i = 1, 2, \ldots, 7)$. We substitute their values in the matrix B and we continue the procedure towards elimination of the third nonzero (in the general case) element b_{23} of the new symmetric matrix B.

Remark 3.1. If we exchange the suppressed variable x with z, the relevant polynomials $\tilde{P}(z)$ and $\tilde{Q}(z)$ are of degree four and six, respectively, and there is no guarantee that the equations $\tilde{P}(z) = 0$ and $\tilde{Q}(z) = 0$ allow any real root.

4. SECOND LEVEL OF DIAGONALIZATION

After actualization of the matrix B, the resultant matrix will be denoted by $R = [r_{ij}]$, i.e., $R = \text{Actual}[B]$, which is again a 3×3 matrix. Now we continue the process of diagonalization keeping

$$r_{12} = 0 \quad \text{and} \quad r_{13} = 0. \qquad (22)$$

Consider the vector parameter $\tilde{c} = (u, 0, 0)$. There exists an orthogonal matrix $O_{\tilde{c}} = O(\tilde{c})$ and $O^T(\tilde{c}) = O_{\tilde{c}}^T$. Now we form the matrix $S = [s_{ij}]$ as follows,

$$S = O_{\tilde{c}}^T R O_{\tilde{c}} = O_{\tilde{c}}^T O_c^T A O_c O_{\tilde{c}}, \qquad (23)$$

and set $s_{23} = 0$. In this way, we obtain a polynomial in u whose power is not greater than four. Assuming that we are in the generic case, this gives

$$u^4 - 2au^3 - 6u^2 + 2au + 1 = 0, \qquad (24)$$
$$a = (r_{33} - r_{22})/r_{23}.$$

Solving this equation, we find the following solutions:

$$u_1 = (a + M - \sqrt{2(M^2 + aM)})/2, \qquad (25)$$
$$u_2 = (a - M + \sqrt{2(M^2 - aM)})/2,$$
$$u_3 = (a - M - \sqrt{2(M^2 - aM)})/2,$$
$$u_4 = (a + M + \sqrt{2(M^2 + aM)})/2;$$

and all of them are real because $M = \sqrt{a^2 + 4} > |a|$. Choosing any of these roots, we actually fix the vector $\tilde{c}$. Composing the so-obtained vector parameters

$(x,0,z)$ and $(u,0,0)$, we get the vector parameter of the resultant rotation

$$\ddot{c} = \langle c, \tilde{c} \rangle, \tag{26}$$

and actually we have proved the following theorem:

Theorem 4.1. Let $A = [a_{ij}], i,j = 1,2,3$, be a real symmetric 3×3 matrix. Then there exists a vector $\ddot{c} \in \mathbb{R}^3$ given in (26) and orthogonal matrix O uniquely associated with it via (4) such that (1)

$$O^T AO = \Lambda = \mathrm{diag}\{\lambda_{\pi(1)}, \lambda_{\pi(2)}, \lambda_{\pi(3)}\}, \quad \pi \in \Sigma_3,$$

where $O = O(\ddot{c})$ and $O^T = O^T(\ddot{c})$.

5. THE BOUSSINESQ STRESS TENSOR

As an illustration of the procedure described in the previous two sections, we will consider the diagonalization of the Boussinesq stress tensor which appears in elasticity theory. Its components in Cartesian coordinates ξ, η, ζ are given by the simple equations

$$\sigma_{\xi\xi} = 2\mu A \frac{\partial^2 \phi}{\partial \xi^2}, \quad \sigma_{\xi\eta} = \sigma_{\eta\xi} = 2\mu A \frac{\partial^2 \phi}{\partial \xi \partial \eta}, \tag{27}$$
$$\sigma_{\eta\eta} = 2\mu A \frac{\partial^2 \phi}{\partial \eta^2}, \quad \sigma_{\xi\zeta} = \sigma_{\zeta\xi} = 2\mu A \frac{\partial^2 \phi}{\partial \xi \partial \zeta},$$
$$\sigma_{\zeta\zeta} = 2\mu A \frac{\partial^2 \phi}{\partial \zeta^2}, \quad \sigma_{\eta\zeta} = \sigma_{\zeta\eta} = 2\mu A \frac{\partial^2 \phi}{\partial \eta \partial \zeta}.$$

Assuming that the solid occupies the half-space $\zeta \geq 0$ and choosing

$$\phi = \log(\rho + \zeta), \quad \rho^2 = \xi^2 + \eta^2 + \zeta^2, \tag{28}$$

one gets the symmetric matrix (the scalar multiplier $2\mu A/\rho^3$ has been omitted)

$$\sigma(\xi,\eta,\zeta) \tag{29}$$
$$= \begin{bmatrix} \dfrac{\eta^2+\zeta^2}{\rho+\zeta} - \dfrac{\rho\xi^2}{(\rho+\zeta)^2} & -\dfrac{(2\rho+\zeta)\xi\eta}{(\rho+\zeta)^2} & -\xi \\ -\dfrac{(2\rho+\zeta)\xi\eta}{(\rho+\zeta)^2} & \dfrac{\xi^2+\zeta^2}{\rho+\zeta} - \dfrac{\rho\eta^2}{(\rho+\zeta)^2} & -\eta \\ -\xi & -\eta & -\zeta \end{bmatrix}.$$

We begin its diagonalization with determination of x and z which appear in (11). At this stage, we recognize immediately that $P(x)$ is a homogeneous polynomial and therefore $x = 0$ is one of its roots. This situation is just a manifestation of the hidden symmetries of our matrix. Thus, at the first level, we are left with the problem of finding an appropriate $z \neq 0$. This can be done either by solving the linear system (19) or by identifying the common roots of the equations in (12). The result of the second procedure is

$$z = \frac{\eta + \sqrt{\xi^2+\eta^2}}{\xi} = \frac{\eta + \sqrt{\rho^2-\zeta^2}}{\xi}. \tag{30}$$

Accordingly, the rotation matrix $O(c)$ generated with the vector parameter $c = (0,0,z)$ is

$$O(c) = \begin{bmatrix} -\dfrac{\eta}{\sqrt{\rho^2-\zeta^2}} & -\dfrac{\xi}{\sqrt{\rho^2-\zeta^2}} & 0 \\ \dfrac{\xi}{\sqrt{\rho^2-\zeta^2}} & -\dfrac{\eta}{\sqrt{\rho^2-\zeta^2}} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{31}$$

and the conjugation of σ with $O(c)$ produces

$$\sigma \quad \longrightarrow \quad B = O^T(c)\sigma O(c) \tag{32}$$
$$= \begin{bmatrix} \dfrac{\rho^2}{\rho+\zeta} & 0 & 0 \\ 0 & \dfrac{\zeta^2+\rho(\zeta-\rho)}{\rho+\zeta} & \sqrt{\rho^2-\zeta^2} \\ 0 & \sqrt{\rho^2-\zeta^2} & -\zeta \end{bmatrix}.$$

From here, we obtain the four real roots of Eq. (24) by plugging into (25)

$$a = \frac{\rho^2 - 2\zeta(\rho+\zeta)}{(\rho+\zeta)\sqrt{\rho^2-\zeta^2}}$$

and

$$M = \frac{\rho}{\rho+\zeta}\sqrt{\frac{\rho(5\rho+4\zeta)}{\rho^2-\zeta^2}}.$$

Especially if we take the last root u_4, the rotation matrix generated by the vector $\tilde{c} = (u_4, 0, 0)$ is

$$O(\tilde{c}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\dfrac{\sqrt{2}(\rho+\zeta)\sqrt{\rho^2-\zeta^2}}{\rho^{3/4}(5\rho+4\zeta)^{1/4}\sqrt{2\zeta^2+2\rho\zeta-\rho^2+\rho^{3/2}\sqrt{5\rho+4\zeta}}} & -\dfrac{\sqrt{2\zeta^2+2\rho\zeta-\rho^2+\rho^{3/2}\sqrt{5\rho+4\zeta}}}{\sqrt{2}\rho^{3/4}(5\rho+4\zeta)^{1/4}} \\ 0 & \dfrac{\sqrt{2\zeta^2+2\rho\zeta-\rho^2+\rho^{3/2}\sqrt{5\rho+4\zeta}}}{\sqrt{2}\rho^{3/4}(5\rho+4\zeta)^{1/4}} & -\dfrac{\sqrt{2}(\rho+\zeta)\sqrt{\rho^2-\zeta^2}}{\rho^{3/4}(5\rho+4\zeta)^{1/4}\sqrt{2\zeta^2+2\rho\zeta-\rho^2+\rho^{3/2}\sqrt{5\rho+4\zeta}}} \end{bmatrix}. \tag{33}$$

Performing the similarity transformation of B with $O(\tilde{c})$ brings it into the diagonal form

$$B \quad \longrightarrow \quad S = O^T(\tilde{c})BO(\tilde{c}) = \begin{bmatrix} \dfrac{\rho^2}{\rho+\zeta} & 0 & 0 \\ 0 & -\dfrac{\rho^{3/2}\left(\sqrt{\rho}+\sqrt{5\rho+4n}\right)}{2\left(\rho+\zeta\right)} & 0 \\ 0 & 0 & \dfrac{2\rho^{3/2}}{\sqrt{\rho}+\sqrt{5\rho+4n}} \end{bmatrix}. \qquad (34)$$

It has to be noted that the same result can be obtained using u_1, while the other two solutions, u_2 and u_3, do not diagonalize B. The rotational matrices which furnish this diagonalization can be explicitly built via (4) using

$$c_i = (u_i, zu_i, z), \qquad (35)$$

where c_i is the composition vector of the vectors $(0,0,z)$ and $(u_i,0,0)$ when $i = 1$ or 4.

In addition, a few remarks are in order here. The first one is that the actual spectrum of (29) is displayed by (34) up to a factor $2\mu A/\rho^3$. The second one concerns the fact that implicitly we have tastefully assumed in our considerations that both a_{12} and a_{13} are nonzero elements. If this is not the case and one of them vanishes, the proposed procedure is simplified considerably. If both are zero, one simply goes directly to the second stage.

Finally, it is worth noting here that our analytical algorithm is realized using version 4.2 of the Mathematica® package for symbolic calculations.

6. CONCLUDING REMARKS

The purely algebraic feature of the presented approach allows us to find eigenvalues and eigenvectors of an arbitrary real 3×3 symmetric matrix in a closed analytical form following a straightforward algorithm. It offers a means of studying in full detail various models of theoretical and experimental relevance. Contrary to Jacobi's method which is based on three consecutive plane rotations, our method is based upon general two-parameter rotation (first level of diagonalization) followed by one-parameter plane rotation (second level of diagonalization). This is the first main point in the diagonalization procedure. The second interesting moment is that the information about these rotations is encoded in a vector form, their composition is expressed by simple vector operations, and any use of transcendental functions is avoided. In addition, one will be able to examine the behavior of critical parameters as functions of input data. And the third essential element is that, in the first level of the diagonalization procedure, a pair of coupled highly nonlinear equations has to be solved, which is realized in the paper by using the so-called dialytic elimination method. Having in mind the abundance of contexts in physics, mechanics, crystallography, elasticity, hydromechanics, robotics, etc., where symmetric matrices appear, we hope that potential users will find this method useful in any concrete situation.

ACKNOWLEDGMENTS

One of us (C.M.) would like to thank the Alexander von Humboldt Foundation for donating the computer equipment and computer algebra system Mathematica®, which were indispensable in realizing the above program algorithm in symbolic form.

REFERENCES

1. O. Smith, Commun. ACM **4**, 168 (1961).
2. A. Bojanczyk and A. Lutoborski, SIAM J. Matrix Anal. Appl. **12**, 41 (1991).
3. J. Linderberg, Int. J. Quantum Chem. **19**, 237 (1981).
4. C. Mladenova, Int. J. Robotics, Comput. Integrated Manufact. **8**, 233 (1991).
5. P. Lancaster, *Theory of Matrices* (Academic, New York, 1969; Nauka, Moscow, 1982).
6. V. Arnold, *Mathematical Methods of Classical Mechanics* (Nauka, Moscow, 1978; Springer, New York, 1978).
7. F. Fedorov, *The Lorentz Group* (Nauka, Moscow, 1979) [in Russian].
8. N. Gromov, Proc. Byeloruss. Acad. Sci., Phys. Math. Ser. **2**, 108 (1984).
9. M. Schuster, J. Astron. Sci. **41**, 439 (1993).
10. H. Schaub, P. Tsiotras, and J. Junkins, Int. J. Eng. Sci. **33**, 2277 (1995).
11. C. Mladenova, Acta Mech. **112**, 159 (1995).
12. B. Roth, in *Computational Kinematics*, Ed. by J. Angeles *et al.* (Kluwer, Dordrecht, 1993), p. 3.

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1922–1925. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1984–1987.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS"
ARMENIA, YEREVAN, AUGUST 13–19, 2003

Order–Disorder Transition in One-Dimensional System with Disorder in Composition*

V. F. Morozov[1)], A. V. Badasyan[1)], E. Sh. Mamasakhlisov[1)], and A. V. Grigoryan[1)]**

Received October 29, 2004

Abstract—The influence of structural inhomogeneity onto order–disorder transitions in 1D systems is considered in the scope of the Potts-like model with many-particle interactions. The helix–coil transition in DNAs, heterogeneous by hydrogen bonding energy, is considered as an example. The microcanonical method is employed to evaluate the free energy. The secular equation for the heteropolymer is constructed. Both the melting temperature and interval of DNA melting are obtained. In the limit of small difference between inverse melting temperatures of poly(A–T) and poly(G–C), the coincidence with classical results is obtained. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The phenomenon of helix-coil transition in biopolymers is known from the middle of the last century and a lot of authors have treated it both theoretically and experimentally (see books [1–5] and reviews [6, 7]). However, this problem is interesting even now [7–19].

Most of the investigations have been conducted using approaches with phenomenological parameters [1, 5, 6]. In [5] was expressed the necessity of the consistent analytical theory of transition in biopolymers, which would straightforwardly take into account the structural inhomogeneity of biopolymers and would be based on a model with microscopic parameters. Using the transfer matrix approach for the evaluation of the partition function, one runs into the problem of calculating the trace of the product of N noncommutative matrices, which is not an easy mathematical problem at large N [1, 5]. This paper is devoted to overcoming this difficulty and represents itself as the first step to construct the statistical physics of order–disorder-type transition on the basis of the Potts-like model in a 1D system with quenched disorder in the structure.

2. BASE MODEL

In [16, 17], the microscopic theory on the basis of a 1D Potts-like model with Δ-particle interactions was constructed to describe the helix–coil transition in homopolypeptides. This model was named the generalized model of polypeptide chain (GMPC). The Hamiltonian of the GMPC looks like

$$-\beta H = J\sum_{i=1}^{N}\prod_{k=\Delta-1}^{0}\delta\left(\gamma_{i-k},1\right) = J\sum_{i=1}^{N}\delta_i^{(\Delta)}, \quad (1)$$

where $\beta = T^{-1}$ is inverse temperature, N is the number of repeated units, and $J = U/T$ is the temperature reduced energy of interchain hydrogen bonding. The notation $\delta_j^{(\Delta)} = \prod_{k=\Delta-1}^{0}\delta\left(\gamma_{j-k},1\right)$ is introduced, with $\delta\left(x,1\right)$ being the Kronecker symbol; $\gamma_l = \overline{1,Q}$ is the variable which describes the conformation of lth repeated unit; Q is the number of possible conformations of each repeated unit and thus describes the conformational ability of each repeated unit. Conformation number 1 corresponds to the helical state. The Kronecker symbol inside the Hamiltonian ensures that the energy J emerges only when all Δ neighboring repeated units are in helical number 1 conformation.

A similar approach to DNA was employed in [18] and it was proved that a two strand problem can be reduced to a one-strand problem, ignoring the large-scale loop factor. Thus, it was shown that the same GMPC may be used to describe the helix–coil transition in both homopolymeric polypeptides and polynucleotides.

*The text was submitted by the authors in English.

[1)]Department of Molecular Physics, Yerevan State University, Yerevan, 375025 Armenia.

**e-mail: morozov@ysu.am

1063-7788/05/6811-1922$26.00

The $(\Delta \times \Delta)$ transfer matrix of the model looks like

$$\widehat{G} = \begin{pmatrix} W & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & Q-1 \\ 1 & 1 & 1 & 1 & \dots & 1 & Q-1 \end{pmatrix}, \quad (2)$$

where $W = \exp J$. The secular equation for this matrix looks like

$$\lambda^{\Delta-1}(\lambda - W)(\lambda - Q) = (W-1)(Q-1).$$

As was shown in [16], the two-particle correlation function of this model in thermodynamic limit can be represented as

$$g_2(r) = \left\langle \delta_i^{(\Delta)} \delta_{i+r}^{(\Delta)} \right\rangle - \left\langle \delta_i^{(\Delta)} \right\rangle \left\langle \delta_{i+r}^{(\Delta)} \right\rangle \sim \exp\left[-\frac{r}{\xi}\right],$$

where r is the distance (in repeated units), $\xi = [\ln \lambda_1/\lambda_2]^{-1}$ is the correlation length, λ_1 is the maximal eigenvalue, and λ_2 is the second in magnitude. Near the transition point, which is estimated as $T_m = U/\ln Q$, the correlation length ξ passes through the maximum, which can be estimated as [16–18]

$$\xi_{\max} \sim Q^{\frac{1-\Delta}{2}}.$$

The parameter σ of Zimm–Bragg theory can be put into accordance with $\xi_{\max}$ in the following way [16, 17]:

$$\sigma \sim \xi_{\max}^{-2}.$$

The following set of parameters can be estimated for DNA [18]: $Q \propto 3$–5; $\Delta \propto 10$–15. One can see that, for this set, the parameter of cooperativity σ is estimated at 10^{-5}–10^{-7}. So, the high cooperativity of homogeneous DNA can be explained as conditioned by the large value of Δ, which reflects the high rigidity of one-strand DNA.

Using this model, it was possible to find the expressions for helicity degree, for the average number of junctions between the helix and coil sections, and therefore for the mean length of the helical section. In the scope of the same model, it is also possible to describe the influence of solvent [16, 17]. This can be done with the help of redefinition of model parameters. But another case exist, the problem of heterogeneous DNA melting, which cannot be reduced to purely homogeneous model in any way. Thus, a new Hamiltonian should be constructed and new methods proposed.

3. HETEROPOLYMER MODEL

In analogy with Eqs. (1), (2), the Hamiltonian and the corresponding transfer matrix of a heteropolymer can be written as

$$H = \sum_{i=1}^{N} J_i \delta_i^{(\Delta)},$$

$$\widehat{G_j} = \begin{pmatrix} W_j & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & Q-1 \\ 1 & 1 & 1 & 1 & \dots & 1 & Q-1 \end{pmatrix},$$

where $W_j = \exp[J_j] = \exp[U_j/T]$, U_j being the energy of hydrogen bond formation in the jth repeated unit.

DNA heterogeneity is conditioned by the difference in energy of hydrogen bonding of A–T- and G–C-type nitrogen base pairs, and by conformation, these pairs are indistinguishable [1, 4, 5]. In the helical state, an A–T base pair is stabilized by two and G–C by three hydrogen bonds. It means that U_j bonding energy in the jth repeated unit is equal to U_{A-T} if the jth site is occupied by an A–T pair and is equal to U_{G-C} if it is occupied by a G–C pair; $U_{A-T} < U_{G-C}$.

4. FREE ENERGY EVALUATION USING MICROCANONICAL APPROACH

Using the transfer-matrix approach, the partition function at a given realization of disorder may be written as

$$Z_{\text{seq}} = \operatorname{Tr} \prod_{i=1}^{N} \widehat{G}_i,$$

where $\widehat{G}_i$ is equal to $\widehat{G}_{A-T}$ or $\widehat{G}_{G-C}$. As these matrices do not commute with each other, each sequence from the set of sequences of length N and disorder concentration q has unique statistical properties. Generally, each particular chain may be characterized by the sequence-dependent free energy F_{seq}. However, the free energy is believed to obey the self-averaging principle [5, 19] which says that the probability distribution of free energies for independent samples is very narrow and thus the free energy virtually coincides with the mean free energy for almost all sequences. Of course, self-averaging holds in the

thermodynamic limit, and for any real computational model, there is some sequence dependence of the free energy. However, for the analytic theory, which operates with the thermodynamic limit, we can safely ignore the sequence dependence of F_{seq} as long as the overall monomer composition is fixed. Thus, we come to the idea of averaging the free energy over the sequences [19]. In the thermodynamical limit, the free energy per base pair and temperature is

$$f = \lim_{N\to\infty} \frac{\langle F_{\text{seq}} \rangle}{NT} = -\lim_{N\to\infty} \frac{1}{N} \left\langle \ln \operatorname{Tr} \prod_{i=1}^{N} \widehat{G}_i \right\rangle.$$

The problem of evaluating such a quantity becomes much easier if we replace the "quenched" average $\left\langle \ln \operatorname{Tr} \prod_{i=1}^{N} \widehat{G}_i \right\rangle$ by "annealed" $\ln \left\langle \operatorname{Tr} \prod_{i=1}^{N} \widehat{G}_i \right\rangle$ at fixed concentration of disorder. It is the main idea of the microcanonical method. The validity and mathematical background of this method is considered in detail in the book by Crisanti *et al.* [20]. Therefrom, one can know that the "quenched" and "annealed" averages are equal up to the fluctuations in disorder concentration q.

With the help of this method, one can obtain (see [20]) the free energy of the form

$$f(y) = -q\ln q - (1-q)\ln(1-q) - \ln\left[\frac{(1+y)\,\lambda_1(y)}{y^{1-q}}\right], \tag{3}$$

$$\frac{d}{d\ln y}\ln\left[(1+y)\,\lambda_1(y)\right] = 1-q.$$

Thus, the problem of heteropolymer is reduced to the fictive homopolymeric problem with $W(y) = W_{G-C} + yW_{A-T}/(1+y)$ redefinition. Here, $\lambda_1(y)$ is the maximal eigenvalue of the $\left(\widehat{G}_{G-C} + y\widehat{G}_{A-T}\right)$ matrix. In this system of equations, y is the additional variable and should be eliminated.

5. DISCUSSION

Using the designations $c = W_{G-C}/W_{A-T}$ and $b = (1-q)/q$, the second equation of the system (3) can be written in the form

$$\frac{\theta(y)\,(1-c)\,(1+b)}{y+c} - \frac{b}{y} + 1 = 0, \tag{4}$$

where $\theta = d\ln\lambda_1/d\ln W$ is the helicity degree of the auxiliary fictive homopolymeric problem. Let us assume that the transition is sharp, i.e., $\theta = 0$ at $T < T_m$, and $\theta = 1$ at $T \geq T_m$ (T is the temperature, and T_m is the temperature at the transition point), and solve this equation with respect to y:

(i) at $\theta = 0$, from Eq. (4),

$$y^{(0)} = b = \frac{1-q}{q};$$

(ii) at $\theta = 1$, taking into account that $y > 0$,

$$y^{(1)} = bc = \frac{1-q}{q}\frac{W_{G-C}}{W_{A-T}}.$$

Inserting these solutions into the equation $W(y) = (W_{G-C} + yW_{A-T})/(1+y)$, one can obtain the explicit dependence of temperature parameter W of the heteropolymer model on homopolymer parameters. In the high-temperature limit, it looks like

$$W_{\theta=0} = qW_{G-C} + (1-q)\,W_{A-T} \tag{5}$$

and, in the low-temperature limit,

$$W_{\theta=1} = \frac{W_{A-T}W_{G-C}}{qW_{A-T} + (1-q)\,W_{G-C}}. \tag{6}$$

Thus, the dependence of the temperature parameter of the heteropolymer model on homopolymeric parameters changes throughout the transition. At high temperatures, expression (5) for W is just the arithmetic average of homopolymeric parameters, and at low temperatures, expression (6) is the harmonic average.

The transition point of the model with Hamiltonian (1) and transfer matrix (2) can be evaluated from the equation [16, 17]

$$W = Q. \tag{7}$$

Let us insert Eqs. (5), (6) and W for the $A{-}T$ and $G{-}C$ homopolymer models into Eq. (7):

$$W_{\theta=0} = q\exp\left[\frac{U_{G-C}}{T_0}\right] + (1-q)\exp\left[\frac{U_{A-T}}{T_0}\right] = Q, \tag{8}$$

$$W_{\theta=1}^{-1} = q\exp\left[-\frac{U_{G-C}}{T_1}\right] + (1-q)\exp\left[-\frac{U_{A-T}}{T_1}\right] = Q^{-1},$$

$$W_{A-T} = \exp\left[\frac{U_{A-T}}{T_{A-T}}\right] = Q,$$

$$W_{G-C} = \exp\left[\frac{U_{G-C}}{T_{G-C}}\right] = Q.$$

Here, T_0 and T_1 are transition points corresponding to the $\theta = 0$ and $\theta = 1$ cases, and T_{A-T} and T_{G-C} are transition points of pure homopolymeric cases. As the curve $W_{\theta=0}$ is always above the curve $W_{\theta=1}$, and they both, as averages of W_{A-T} and W_{G-C}, are between

them, then transition temperatures satisfy the condition $T_{A-T} < T_1 < T_0 < T_{G-C}$. Taking into account that the quantity $(1/T_{A-T} - 1/T_{G-C})$ is small as these temperatures are of the same order, one can expand Eq. (8) by small parameters $(1/T_0 - 1/T_{G-C})$ and $(1/T_{A-T} - 1/T_1)$, and then obtain the values of temperatures T_0 and T_1. The evaluations show that, in a linear approximation, the heteropolymer melting temperature looks like

$$T_0 = T_1 = T_{\text{hetero}} = qT_{G-C} + (1-q)\,T_{A-T}. \quad (9)$$

As T_1 is the temperature corresponding to the completely helical state and T_{hetero} and T_0 correspond to the completely coil state, then saying "heteropolymer melting interval" we will understand the quantity $(T_0 - T_1)$. And as T_1 and T_0 are indistinguishable in a linear approximation, then to obtain the interval we need the expansion up to the second order. Therefrom, one can obtain

$$\Delta T = T_0 - T_1 = 2q\,(1-q) \quad (10)$$
$$\times \left(\frac{T_{G-C} - T_{A-T}}{T_{\text{hetero}}}\right)^2 T_{\text{hetero}} \ln Q.$$

Equations (9), (10) coincide with classical results [1–7].

Thus the following results are obtained:

(1) During the helix–coil transition, the averaging regime of temperature parameter W is changed.

(2) The classical expressions for melting temperature and interval are derived without using any phenomenological parameter or unproven trick.

Thus, in the scope of the given approach, well-known experimentally observable results on heteropolymer melting have been obtained. The proposed theory, in principle, allows one to describe the helix–coil transition in heteropolymer DNA on the basis of the DNA structure and using new mathematical results [20]. Moreover, in first- and second-order approximations, well-known expressions are derived [1, 4–7]. The theory allows one to control the degree of approximation and, hence, the error range as well. It was also noticed that, during the helix–coil transition, the averaging regime of temperature parameter W is changed. This result needs to be further investigated.

ACKNOWLEDGMENTS

This work was supported in part by ISTC under grant nos. A-301.2 and A-092.2.

REFERENCES

1. D. C. Poland and H. A. Scheraga, *The Theory of Helix–Coil Transition* (Academic, New York, 1970).
2. P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969; Mir, Moscow, 1971).
3. M. V. Volkenstein, *Molecular Biophysics* (Academic, New York, 1977).
4. C. R. Cantor and T. R. Shimmel, *Biophysical Chemistry* (Freeman, San Francisco, 1980; Mir, Moscow, 1984).
5. A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP, New York, 1994).
6. A. A. Vedenov, A. M. Dyhne, and M. D. Frank-Kamenetskii, Usp. Fiz. Nauk **105**, 479 (1971) [Sov. Phys. Usp. **14**, 715 (1971)].
7. R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
8. T. Garel, C. Monthus, and H. Orland, Europhys. Lett. **55**, 132 (2001).
9. M. Baiesi, E. Carlon, E. Orlandini, and A. L. Stella, cond-mat/0207122(1).
10. D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
11. M. Takano, K. Nagayama, and A. Suyama, J. Chem. Phys. **116**, 2219 (2002).
12. V. Munoz and L. Serrano, Biopolymers **41**, 495 (1997).
13. M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
14. T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
15. T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
16. N. S. Ananikyan, Sh. A. Hayryan, E. Sh. Mamasakhlisov, and V. F. Morozov, Biopolymers **30**, 357 (1990).
17. Sh. A. Hayryan, E. Sh. Mamasakhlisov, and V. F. Morozov, Biopolymers **35**, 75 (1995).
18. V. F. Morozov, E. Sh. Mamasakhlisov, Sh. A. Hayryan, and Chin-Kun Hu, Physica A **281**, 51 (2000).
19. V. S. Pande, A. Yu. Grosberg, and T. Tanaka, Rev. Mod. Phys. **72**, 259 (2000).
20. A. Crisanti, G. Paladin, and A. Vulpiani, in *Products of Random Matrices in Statistical Physics* (Springer-Verlag, Berlin, 1993).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1926–1928. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1988–1990.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS"
ARMENIA, YEREVAN, AUGUST 13–19, 2003

One-Electron States of a 2D Spatially Limited System*

R. M. Movsesyan[1)] and A. S. Sahakyan[1)]**

Received October 29, 2004

Abstract—The spectrum of a particle in a 2D potential as a polar-angle-dependent well is investigated. One-electron states of 2D systems are considered within the framework of the simplest quantum-mechanical models. Such states can be realized by dusting the atoms of various substances which are posited on a substrate. © *2005 Pleiades Publishing, Inc.*

1. It is well known that the character of spatial bounding has strong impact on the one-electron spectrum of a system and, therefore, on all of its physical properties [1]. Modern technologies make it possible to obtain films of various substances with thickness less than 100 Å, up to one monolayer. It is remarkable that one can control not only the film thickness but also its geometrical shape [2]. In this connection, the 2D system of atoms dusted in the form of a sector while the complementary sector is made of atoms of another sort is of interest. The dusted atoms of both sectors can be fully different such as metal–dielectric, metal1–metal2, and metal–semiconductor. It is obvious that the one-electron spectrum and consequently all the physical properties will depend on choice of these substances. However, most of these systems have one interesting property: an electron of one of these two sectors can penetrate into the adjacent sector, and as a result, there must exist a tunnel transition of an electron from the sector to the same one. This is possible if one of the dusted substances is not a dielectric. Due to the symmetry of the system, a situation arises which is analogous to the motion of an electron in a periodic field; however, in our case, it concerns only the azimuthal degree of freedom. As examples of systems with circle symmetry, one can note the well-investigated quantum rings [3, 4], which, in contrast to the ones considered here, are one-dimensional systems.

In the present paper, the simplest models that can describe such systems are considered, and the one-electron spectrum is investigated in the framework of these models.

Let us choose the sector-shaped infinite well with central angle φ_0 and radius R as a simplest model of 2D potential,

$$V(r,\varphi)=\begin{cases}0, & 0\le\varphi\le\varphi_0\\ \infty, & \varphi_0\le\varphi\le 2\pi.\end{cases} \qquad (1)$$

The one-electron wave function must obey the following obvious boundary conditions

$$\Psi(r,0)=\Psi(r,\varphi_0)=0, \qquad \Psi(r,\varphi)=0, \qquad (2)$$

where the second one is due to bounding of the system in the radial direction. The choice of potential (1) is justified only in the case when the second sector is covered by atoms of a dielectric.

Then the wave functions take the following form:

$$\Psi(r,\varphi)=C\sin(q,\varphi)J_q(k_{nm}r), \qquad (3)$$

where $q=n\pi/\varphi_0$, n is integer, $J_\lambda(z)$ is a Bessel function of real argument, $k_{nm}=æ_{nm}/R$, $æ_{nm}$ are the zeros of Bessel functions, and the energy eigenvalues are defined by (2) and (3):

$$E_{nm}=\frac{\hbar^2k_{nm}^2}{2m^*}=\frac{\hbar^2æ_{nm}^2}{2m^*R^2}.$$

The range of $n\pi/\varphi_0\gg 1$ is physically more interesting. It corresponds to small central angles of the sector. In this region,

$$\Psi\simeq\frac{1}{\pi}\sqrt{\frac{\varphi_0}{2\pi}}\left(\frac{ek_nr}{2s}\right)^q;$$

i.e., when r tends to R, the wave function increases as r^q. The reason for such behavior is simple enough. As the particle approaches the vertex, its transversal region of localization $r\varphi_0$ decreases and its energy $E\sim\hbar^2/(m\varphi_0^2r^2)$ increases, which makes electron motion to the vertex energetically disadvantageous. On the other hand, it means that a quasiclassical force $F\sim r^3$ acts on the electron and pushes it far from the vertex. The 2D electron gas is strongly inhomogeneous

*The text was submitted by the authors in English.

[1)]State Engineering University of Armenia, Yerevan, 375009 Armenia.

**e-mail: sahakyan@web.am

and it is concentrated mainly in the region of width $a \geq R/æ_{nm}$. Taking into account that, at least for large enough n, $æ_{nm} \sim q$ holds, we get $a \geq R/q$, i.e., $a \ll R$.

2. Let us consider the same problem for a model of a potential well of finite depth:

$$V(\varphi) = \begin{cases} 0, & 0 \leq \varphi \leq \varphi_0, \quad \text{(I)} \\ V, & \varphi_0 \leq \varphi \leq 2\pi. \quad \text{(II)} \end{cases} \tag{4}$$

Potential (4) is periodic with the period 2π, and the finiteness of the well allows the tunnel connection of the well with itself. Such a choice for the potential corresponds to the case when the regions I and II are conductive. The potentials (1) and (4) are noncentral-symmetric and belong to the class of potentials considered in [5–11].

The Schrödinger equation describing the behavior of an electron in the field (4) has the form

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi}{\partial r^2} + \frac{1}{r}\frac{\partial\Psi}{\partial r} + \frac{1}{r^2}\frac{\partial^2\Psi}{\partial\varphi^2}\right) + V(\varphi)\Psi = E\Psi \tag{5}$$

with the following boundary conditions for the wave function:

$$\Psi_{\mathrm{I}}(0,r) = \Psi_{\mathrm{II}}(2\pi,r), \tag{6}$$

$$\Psi'_{\mathrm{I}}(0,r) = \Psi'_{\mathrm{II}}(2\pi,r), \quad \forall r;$$

$$\Psi_{\mathrm{I}}(\varphi_0,r) = \Psi_{\mathrm{II}}(\varphi_0,r), \quad \Psi'_{\mathrm{I}}(\varphi_0,r) = \Psi'_{\mathrm{II}}(\varphi_0,r);$$

$$\Psi_{\mathrm{I}}(\varphi,R) = 0, \quad \Psi_{\mathrm{II}}(\varphi,R) = 0, \quad \forall\varphi.$$

For the step-like potential (4), it is possible to separate the variables in (5) and the one-electron states should be represented in the form of a linear combination of a factorized wave function, but in this case it is a rather difficult problem to satisfy the boundary conditions (6). For this reason, we are going to solve Eq. (5) only approximately. We suppose that the wave function $\Psi(r,\varphi)$ can be represented in the factorized form

$$\Psi(r,\varphi) = \Psi_1(r,\varphi)\Psi_2(r), \tag{7}$$

where $\Psi_1(r,\varphi)$ is the solution to the equation

$$-\frac{\hbar^2}{2mr^2}\frac{\partial^2\Psi_1}{\partial\varphi^2} + V(\varphi)\Psi_1 = \varepsilon\Psi_1, \tag{8}$$

which describes the states of a rotator with the moment of inertia $I = mr^2$ in the field $V(\varphi)$. The physical meaning of this approach we shall discuss below. It is easy to rewrite the boundary conditions (6) for the wave function $\Psi_1(r,\varphi)$. We seek the solutions to (8) in the form

$$\Psi_1(r,\varphi) = \begin{cases} A_1 e^{i\lambda\varphi} + A_2 e^{-i\lambda\varphi}, & 0 \leq \varphi \leq \varphi_0 \quad \text{(I)} \\ B_1 e^{\mu\varphi} + B_2 e^{-\mu\varphi}, & \varphi_0 \leq \varphi \leq 2\pi - \varphi_0. \quad \text{(II)} \end{cases} \tag{9}$$

Then, due to boundary conditions (6), we come to the following dispersion equation for eigenvalues ε:

$$\cosh(\mu b)\cos(\lambda\varphi_0) + \frac{\mu^2 - \lambda^2}{2\mu\lambda}\sinh(\mu b)\sin(\lambda\varphi_0) = 1; \tag{10}$$

or, following the method of [12] developed for the case of the Kronig–Penney model, this equation can be reduced to

$$\tan\frac{\mu\varphi_0}{2} = \frac{\mu}{\lambda}\tanh\frac{\mu b}{2}, \qquad \tan\frac{\mu\varphi_0}{2} = -\frac{\mu}{\lambda}\tanh\frac{\mu b}{2}, \tag{11}$$

where

$$\lambda = \frac{\sqrt{2m\varepsilon}}{\hbar}, \quad \mu = \frac{\sqrt{2m(V-\varepsilon)}}{\hbar}r, \quad b = 2\pi - \varphi_0.$$

Equation (9) coincides with the dispersion equation of the Kronig–Penney model for the case of zero Bloch wave vector; i.e., it corresponds to the periodic azimuthal states.

Then, substituting (7) into (5) and making some transformations, we get the equation for $\Psi_2(r)$:

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi_2}{\partial r^2} + \frac{1}{r}\frac{\partial\Psi_2}{\partial r}\right) + \varepsilon(r)\Psi_2 = E\Psi_2 + \hat{C}\Psi_2, \tag{12}$$

where the operator $\hat{C}$ is defined by

$$\hat{C}\Psi_2 = \frac{\hbar^2}{2m} \times \left[\int \Psi_1^*\left(\Delta_r\Psi_2\right)\Psi_2 d\varphi + \int \Psi_1^*\frac{\partial\Psi_1}{\partial r}\frac{\partial\Psi_2}{\partial r}d\varphi\right],$$

and Δ_r is the radial part of the 2D Laplacian. We omit the term $\hat{C}\Psi_2$ in (12) and obtain

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2\Psi_2}{\partial r^2} + \frac{1}{r}\frac{\partial\Psi_2}{\partial r}\right) + \varepsilon(r)\Psi_2 = E\Psi_2, \tag{13}$$

where the eigenvalues of (8) enter as an effective potential, which corresponds to the adiabatic perturbation theory technique [13], i.e., to the assumption that the motion in the radial direction is slower than that in the azimuthal direction. In other words, the energy level spacing for spectrum of Eq. (8) is much larger than that for Eq. (13).

Equation (10) can be solved for some limiting cases. Thus, in the case of a sufficiently deep well and a narrow region of its localization ($\varphi_0 \ll 1$),

$$\varepsilon(r) = \frac{\pi^2\hbar^2 n^2}{2m\varphi^2 r^2}, \quad n = 1, 2, \ldots. \tag{14}$$

This solution corresponds to the levels of an infinite potential well. Solving Eq. (13) with effective potential (12), we obtain an expression for Ψ_2 which coincides with that of the radial part of (3). It should be noted that, in this case, parameter λ does not depend on r [see (11)], and consequently, all derivatives of $\Psi_1(r, \varphi)$ with respect to r vanish within the range (1) with a high degree of accuracy. Then, $\Psi_1(r, \varphi)$ strongly dies out in the region (II). Therefore, $\hat{C}\Psi_2$ can be set equal to zero.

3. Now let us consider the case of a barrier which is low enough that there exists a shallow level $\varepsilon = V - \delta$, $\delta \ll V$. It is obvious that such a level is absent in the limiting case (14) considered in **2**. Solving (11), we get

$$\varepsilon_1 = V + \frac{\varphi_0}{6b}V - \frac{6\hbar^2}{mb^2r^2}, \quad \varepsilon_2 = V - \frac{\varphi_0}{6b}V.$$

The second energy level ε_2 does not depend on r owing to the accepted approximation. It follows from the condition $\delta \ll V$ that the obtained result is valid in the range $r \gg 0.06\lambda\!\!\!^{-}$, where $\lambda\!\!\!^{-} = \hbar/\sqrt{2m\varepsilon}$ is the de Broglie wavelength of the electron on the shallow level. The solution to Eq. (13) can be represented in the form

$$\Psi_1(r) = AJ_{i\beta}(ikr) + BJ_{-i\beta}(-ikr),$$

where $k^2 = 2m(V - E)/\hbar^2 > 0$.

It should be mentioned that actually the vertex of the angle has a smooth rather than sharp shape. This means that the wave function must obey one additional condition besides (6), namely, $\Psi(a) = 0$, where a characterizes smoothness of the vertex. Its value is of the order of atomic size, being $a \gg 0.06\lambda\!\!\!^{-}$. Then the dispersion equation for the energy spectrum holds,

$$J_{i\beta}(ikR)J_{-i\beta}(-ika) - J_{-i\beta}(-ikR)J_{i\beta}(ika) = 0. \tag{15}$$

Consider the regions $kR \gg 1$ and $ka \ll 1$, that is, $a \ll R$. Then Eq. (15) transforms into

$$\tan\left(\beta \ln\frac{ka}{2} - \alpha\right) = 1 - e^{-2kR}. \tag{16}$$

Ignoring the exponentially small term on the right-hand side, we get

$$E = V - \frac{2\hbar^2}{ma^2} \times \exp\left[\frac{2\alpha - \pi(2n - 1/2)}{\beta}\right], \quad n \text{ is integer.}$$

It can be seen from (16) that the condition $ka \ll 1$ implies that $n \gg 1$.

In the region adjacent to the vertex, the wave function has the form

$$\Psi \cong \sin\left(\beta \ln\frac{r}{a}\right).$$

Thus, in the framework of the considered approximation, in the sectorial well there are an infinite number of levels which get denser near the top of the well.

Thus, the tunnel penetration into the adjacent sector demonstrates nonexponential behavior. It is easy to show that, in this utmost case too, one can neglect the nonadiabaticity operator $\hat{C}$. Indeed (in fact), as $V - \varepsilon \sim r^{-2}$, then μ does not depend on r and the r derivatives of Ψ_1 vanish, and the contribution of the region (I) is small due to the smallness of φ_0. Thus, in principle, an electron can penetrate into the vertex region in the case of not too deep a well.

REFERENCES

1. T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982).
2. Y. Imry, *Introduction to Mesoscopic Physics* (Oxford Univ. Press, Oxford, 2002).
3. M. Buttiker, Y. Imry, and R. Landauer, Phys. Lett. A **96A**, 365 (1983).
4. Ho-Fai Cheung, Y. Gefen, E. K. Riedel, and W.-H. Shih, Phys. Rev. B **37**, 6050 (1988).
5. M. Znojil, J. Phys. A **36**, 7825 (2003).
6. M. Sotona and J. Zofka, Phys. Rev. C **10**, 2646 (1974).
7. E. G. Kalnins, *Separation of Variables for Riemannian Spaces of Constant Curvature* (Longman, Harlow, 1986).
8. F. Calogero, J. Math. Phys. (N.Y.) **10**, 2191 (1969).
9. A. Turbiner, Mod. Phys. Lett. A **13**, 1473 (1998).
10. J. Fris, V. Mandrosov, Ya. A. Smorodinsky, *et al.*, Phys. Lett. A **16A**, 354 (1965).
11. N. W. Evans, J. Math. Phys. (N.Y.) **32**, 3369 (1991).
12. D. M. Sedrakian and A. Zh. Khachatrian, J. Contemp. Phys. **38**, 211 (2003).
13. M. Born and K. Huang, *Dynamical Theory of Crystal Lattices* (Oxford Univ. Press, Oxford, 1998; Inostrannaya Literatura, Moscow, 1958).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1929–1940. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 1991–2001.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS"
ARMENIA, YEREVAN, AUGUST 13–19, 2003

Rational Conformal Correlation Functions of Gauge-Invariant Local Fields in Four Dimensions*

N. M. Nikolov[1], Ya. S. Stanev[2], and I. T. Todorov[3]

Received October 29, 2004

Abstract—Global conformal invariance in Minkowski space and the Wightman axioms imply strong locality (Huygens principle) and rationality of correlation functions, thus providing an extension of the concept of a vertex algebra to higher (even) dimensions D. We (p)review current work on a model of a Hermitian scalar field $\mathcal{L}$ of scale dimension 4 ($D = 4$) which can be interpreted as the Lagrangian of a gauge field theory that generates the algebra of gauge-invariant local observables in a conformally invariant renormalization group fixed point. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The present paper provides a concise review and a continuation of our work [1–4] aimed at constructing a nontrivial globally conformal invariant (GCI) four-dimensional quantum field theory (QFT) model. Our attempt to build such a model is based on the following results of [4].

Invariance under finite conformal transformations in Minkowski spacetime M and local commutativity imply the Huygens principle: the commutator of two local Bose fields vanishes for nonisotropic separations. The Huygens principle and energy positivity yield rationality of correlation functions (Theorem 3.1 of [4]). These results allow one to extend any GCI QFT to compactified Minkowski space $\bar{M}$, which admits the following convenient complex variable realization [1–3, 5, 6]:

$$\bar{M} = \{z_\mu = e^{2\pi i\zeta}u_\mu, \mu = 1, \dots, 4; \zeta \in \mathbb{R}, \quad (1)$$
$$u \in \mathbb{S}^3 = \{u \in \mathbb{R}^4; u^2 = \mathbf{u}^2 + u_4^2 = 1\}\} = \frac{\mathbb{S}^1 \times \mathbb{S}^3}{\mathbb{Z}_2}.$$

Fields $\phi(z)$ are expressed as formal power series of the form

$$\phi(z) = \sum_{\nu\in\mathbb{Z}}\sum_{m\geq 0}(z^2)^\nu \phi_{\{\nu,m\}}(z), \quad (2)$$
$$\Delta\phi_{\{\nu,m\}}(z) = 0,$$
$$\phi_{\{\nu,m\}}(\lambda z) = \lambda^m \phi_{\{\nu,m\}}(z),$$
$$z^2 := \sum_{\mu=1}^4 z_\mu^2, \Delta = \sum_{\mu=1}^4 \frac{\partial^2}{\partial z_\mu^2},$$

$\phi_{\{\nu,m\}}(z)$ being an (in general, multicomponent) operator-valued (homogeneous, harmonic) polynomial in z. The expansion (2) singles out the eigenvalues of the conformal energy operator H defined up to an additive constant (the vacuum energy) by the commutation rules

$$[H, \phi(z)] = \left(d + z\frac{\partial}{\partial z}\right)\phi(z), \quad (3)$$
$$\text{implying } [H, \phi_{\{\nu,m\}}(z)] = (d + 2\nu + m)\phi_{\{\nu,m\}}(z),$$

where d is the *scale* (or conformal) dimension of the field ϕ. The Minkowski space spectral conditions (including energy positivity) imply analyticity of the vector-valued function $\phi(z)|0\rangle$ (where $|0\rangle$ is the conformally invariant vacuum vector) in the z-picture image

$$T_+ = \left\{z \in \mathbb{C}^4 : |z^2| < 1, \quad (4)\right.$$
$$\left.|z|^2 \left(:= \sum_{\mu=1}^4 |z^\mu|^2\right) < \frac{1 + |z^2|^2}{2}\right\}$$

of the forward tube

$$\mathcal{T}_+ = \left\{\xi = x + iy : x, y \in M,\right.$$
$$\left. y^0 > |\mathbf{y}| := \sqrt{y_1^2 + y_2^2 + y_3^2}\right\}$$

*The text was submitted by the authors in English.

[1] Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria; e-mail: mitov@inrne.bas.bg

[2] Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria, and INFN, Sez. di Roma 2, Italy; e-mail: stanev@roma2.infn.it

[3] Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria, and Institut des Hautes Etudes Scientifiques, Bures-sur-Yvette, France; e-mail: todorov@inrne.bas.bg

under the complex conformal transformation (with singularities)

$$h : M_{\mathbb{C}}(\ni \xi) \to E_{\mathbb{C}}(\ni z), \quad \mathbf{z} = \frac{\boldsymbol{\xi}}{\omega_\xi}, \quad z_4 = \frac{1-\xi^2}{\omega_\xi},$$

$$\xi^2 = \xi^2 - (\xi^0)^2, \quad \omega_\xi = \frac{1}{2}(1+\xi^2) - i\xi^0, \qquad (5)$$

$$z^2 = \frac{1+\xi^2+2i\xi^0}{1+\xi^2-2i\xi^0}.$$

[Note that h maps the real Minkowski space M onto the open dense subset of $\bar{M}$ (1) such that $\mathbf{z}^2 + (z_4 + 1)^2 = 2e^{2\pi i\zeta}(\cos 2\pi\zeta + u_4) \neq 0$.] It follows, in particular, that no negative powers of z^2 appear in $\phi(z)|0\rangle$:

$$\phi_{\{\nu,m\}}(z)|0\rangle = 0 \quad \text{for } \nu < 0. \qquad (6)$$

This is unambiguous [1], because, as is well known, every homogeneous polynomial $p(z)$ of degree m has a unique decomposition $p(z) = h(z) + z^2 q(z)$, where h is harmonic (of degree m) and q is homogeneous of degree $m-2$.

The resulting (analytic) z picture provides a higher dimensional generalization [1] of chiral vertex algebras (which have been an outgrowth of physicists' work on conformal field theory and dual resonance models, formalized by Borcherds [7] and since the subject of numerous studies, including several books—see, e.g., [8, 9] and references therein).

The four-dimensional vertex algebra of GCI fields with rational correlation functions corresponds to the algebra of local observables in Haag's approach [10] to QFT. Its isotypical (or factorial) representations (i.e., multiples with a finite multiplicity of an irreducible representation) give rise to the superselection sectors of the theory. The intertwiners between the vacuum and other superselection sectors are higher dimensional counterparts of primary fields (which typically have fractional dimensions and multivalued n-point distributions).

We shall add to the traditional assumption that the conserved (symmetric, traceless) stress-energy tensor $T_{\mu\nu}(z)$ is a local observable the requirement that so is the scalar, gauge-invariant Lagrangian density $\mathcal{L}(z)$ (of dimension $d = 4$). Then the construction of a GCI QFT model becomes a rather concrete program of writing down rational (conformally invariant and "crossing symmetric") correlation functions and studying the associated operator product expansions (OPE).

After a brief review (in Section 2.1) of the results of [4] reformulated in the above z picture, we outline (in Section 2.2) the general truncated GCI four-point function w_4^t of a neutral scalar field of (integer) dimension d. This is a homogeneous rational function of degree $-2d$ in the (complex) Euclidean invariant variables

$$\rho_{ij} = z_{ij}^2, \quad z_{ij} = z_i - z_j, \quad z^2 = \mathbf{z}^2 + z_4^2 \qquad (7)$$

of denominator $(\rho_{12}\rho_{13}\rho_{14}\rho_{23}\rho_{24}\rho_{34})^{d-1}$ and numerator, a homogeneous polynomial of degree $4d-6$ (for $d \geq 2$) depending (linearly) on $\left[\frac{d^2}{3}\right]$ (i.e., no more than $d^2/3$) real parameters. It is just $c(\rho_{13}\rho_{24} + \rho_{12}\rho_{34} + \rho_{14}\rho_{23})$ for the simplest candidate for a nontrivial, $d = 2$, model and involves five parameters for the physically interesting case of a $d = 4$ Lagrangian density. As the model of a $d = 2$ scalar field was proven in [2] to correspond to normal products of free (massless) scalar fields, we concentrate in the rest of the paper on the $d = 4$ case. We study in Section 3 OPE organized in bilocal fields of fixed twist which provide what could be called a conformal partial wave expansion of the four-point function (a concept introduced in [11], see also [12], and recently revisited in [13]). The bilocal field $V_1(z_1, z_2)$ of (lowest) dimension $(1, 1)$, which admits a Taylor expansion in z_{12} involving only twist-2 symmetric traceless tensors, is harmonic in each argument, allowing one to compute (in Section 3.2) its (rational) four-point function. The corresponding (crossing) symmetrized contribution to w_4^t gives rise to a three-parameter subfamily of the original five-parameter family of GCI four-point functions. This provides what we call (in Section 3.3) a minimal model corresponding to a given V_1.

We argue, in Section 4.1 (summarizing results of [3]), that the Lagrangian $\mathcal{L}(z)$ of a gauge field theory should have vanishing odd point functions and should not involve a $d = 2$ scalar field in its OPE. This reduces to three the five parameters in w_4^t. One of the remaining parameters corresponds to the Lagrangian (i.e., the contracted normal square) of a free Maxwell field. Another corresponds to a twist-4 contribution that is not recovered in the minimal version of the theory. We are thus led to consider a one-parameter family of four-point functions that appears as the simplest candidate for a nontrivial GCI QFT consistent with Wightman axioms [14], which may correspond to a (conformally invariant) renormalization group fixed point of a gauge field theory. As a first step in studying this distinguished simple model, we point out, in Section 4.2, that the resulting special bilocal field $V_1 = v(z_1, z_2)$ has the properties of a bilinear combination of a free Weyl spinor and its conjugate. This permits the computation of higher point functions as displayed in Section 4.3.

The results and the challenging open problems are discussed in Section 4.4.

2. IMPLICATIONS OF GCI: GENERAL FOUR-POINT FUNCTION

2.1. Strong Locality and Energy Positivity Imply Rationality

It follows from local commutativity and GCI that, for any pair of conjugate (Bose or Fermi) fields $\psi(z_1)$ and $\psi^*(z_2)$, there is a positive integer N_ψ such that

$$\rho_{12}^N\{\psi(z_1)\psi^*(z_2) - \varepsilon_\psi\psi^*(z_2)\psi(z_1)\} = 0 \quad (8)$$
$$\text{for} \quad N \geq N_\psi, \quad \rho_{12} := z_{12}^2,$$

where $\varepsilon_\psi(=\pm 1)$ is the fermion parity of ψ. If ψ transforms under an elementary local field representation of the spinor conformal group $SU(2,2)$ (see [15, 16]), i.e., one induced by a $(2j_1+1)(2j_2+1)$-dimensional representation $(d; j_1, j_2)$ of the maximal compact subgroup $S(U(2)\times U(2))$ of $SU(2,2)$ (d being the $U(1)$ character coinciding with the scale dimension), then

$$N_\psi = d + j_1 + j_2, \quad \varepsilon_\psi = (-1)^{2j_1+2j_2} = (-1)^{2d}. \quad (9)$$

It follows that, for any n-point function of GCI local fields and for large enough $N \in \mathbb{N}$, the product

$$F_{1\ldots n}(z_1,\ldots,z_n) := \left(\prod_{1\leq i<j\leq n}\rho_{ij}\right)^N \times \langle 0|\phi_1(z_1)\ldots\phi_n(z_n)|0\rangle \quad (10)$$

($\rho_{ij} = z_{ij}^2 \equiv (z_i - z_j)^2$) is $\mathbb{Z}_2$ symmetric under any permutation of the factors within the vacuum expectation value. Energy positivity, on the other hand, implies that $\langle 0|\phi_1(z_1)\ldots\phi_n(z_n)|0\rangle$, and hence $F_{1\ldots n}(z_1,\ldots,z_n)$ do not contain negative powers of z_n^2. It then follows from the symmetry and the homogeneity of $F_{1\ldots n}$ that it is a polynomial in all z_i^μ. Thus, the (Wightman) correlation functions are rational functions of the coordinate differences. {For more detail, see [5]; an equivalent Minkowski space argument based on the support properties of the Fourier transform of (the x-space counterpart of) (10) is given in [4].}

Rationality of correlation functions implies that all dimensions of GCI fields should be integer or half-integer depending on their spin, more precisely, that sums like N_ψ (9) should be integer. This condition is, however, not sufficient for rationality even of three-point functions.

Observation 2.1. *The necessary and sufficient condition for the existence of a GCI three-point function* $\langle 0|\phi_1(z_1)\phi_2(z_2)\phi_3(z_3)|0\rangle$ *of elementary conformal fields* $\phi_i(z)$ *of* $S(U(2)\times U(2))$ *weight* $(d_i; j_{i1}, j_{i2})$ *is*

$$N_i := d_i + j_{i1} + j_{i2} \in \mathbb{N}, \quad (11)$$

$$\frac{1}{2}\sum_{i=1}^3 N_i \in \mathbb{N}, \sum_{i=1}^3 d_i \in \mathbb{N}.$$

The statement follows from the explicit knowledge of three-point functions (for reviews, see [16–18]).

In particular, there is no Yukawa-type rational conformal three-point function of a pair of conjugate canonical ($d = 3/2$) spinor fields and a canonical ($d = 1$) scalar field. Similarly, one observes that two-point functions of free massless fields in odd spacetime dimensions are not rational and, hence, cannot be GCI.

It is important for the feasibility of constructing a GCI model that the singularities of n-point functions [the integer N in (10)] is majorized by that of the two-point function [N_ψ in (9)] whenever Wightman positivity is satisfied.

2.2. General Truncated Four-Point Function of a GCI Scalar Field

Infinitesimal (or Euclidean) conformal invariance is sufficient to determine two- and three-point functions (see, e.g., [17]). One can construct, however, two independent conformally invariant cross-ratios out of four points,

$$s = \frac{\rho_{12}\rho_{34}}{\rho_{13}\rho_{24}}, \quad t = \frac{\rho_{14}\rho_{23}}{\rho_{13}\rho_{24}}, \quad (12)$$

so that a simple-minded symmetry argument does not determine the four-point functions. GCI, on the other hand, combined with Wightman axioms, yields rationality and thus allows one to construct higher point correlation functions involving just a finite number of free parameters. In particular, the truncated four-point function of a Hermitian scalar field ϕ of (integer) dimension d can be written in the form (Section 1 of [2])

$$w_4^t \equiv w^t(z_1,z_2,z_3,z_4) := \langle 1234\rangle - \langle 12\rangle\langle 34\rangle - \langle 13\rangle\langle 24\rangle - \langle 14\rangle\langle 23\rangle = \frac{(\rho_{13}\rho_{24})^{d-2}}{(\rho_{12}\rho_{23}\rho_{34}\rho_{14})^{d-1}}P_d(s,t), \quad (13)$$
$$P_d(s,t) = \sum_{i\geq 0, j\geq 0}^{i+j\leq 2d-3} c_{ij}s^i t^j,$$

where $\langle 1\ldots n\rangle$ is shorthand for the n-point function of ϕ:

$$\langle 1\ldots n\rangle = \langle 0|\phi(z_1)\ldots\phi(z_n)|0\rangle, \quad (14)$$
$$\langle 12\rangle = B_\phi\rho_{12}^{-d}.$$

In writing down (13), we have used the fact that, for spacetime dimensions $D > 2$, Hilbert space positivity

implies that the degree of singularities of the truncated n-point function ($n \geq 4$) is strictly smaller than the degree of the two-point function.

Furthermore, crossing symmetry (which is a manifestation of local commutativity) implies an $\mathcal{S}_3$ symmetry of P_d:

$$s_{ii+1}P_d(s,t) = P_d(s,t), \quad i = 1,2, \tag{15}$$
$$s_{12}P_d(s,t) := t^{2d-3}P_d\left(\frac{s}{t},\frac{1}{t}\right),$$
$$s_{23}P_d(s,t) := s^{2d-3}P_d\left(\frac{1}{s},\frac{t}{s}\right),$$

s_{ij} being the substitution exchanging the arguments z_i and z_j. The number of independent crossing symmetric polynomials P_d is $\left[\frac{d^2}{3}\right]$ (the integer part of $d^2/3$: $\left[\frac{d^2}{3}\right] = n(2d-3n)$ for $3n-1 \leq d \leq 3n+1$, $n = 0,1,2,\ldots$).

The one-parameter family of crossing symmetric polynomials for $d = 2$ is $P_2(s,t) = c(1+s+t)$, i.e.,

$$w_4^t = c\{(\rho_{12}\rho_{23}\rho_{34}\rho_{14})^{-1} + (\rho_{13}\rho_{23}\rho_{24}\rho_{14})^{-1} + (\rho_{12}\rho_{13}\rho_{24}\rho_{34})^{-1}\}, \tag{16}$$

thus corresponding to the sum of three one-loop diagrams for a sum of normal products of free massless fields:

$$\phi(z) = \frac{1}{2}\sum_{i=1}^{N} :\varphi_i^2(z): \quad ([\varphi_i(z_1),\varphi_j(z_2)] = 0 \quad \text{for} \quad i \neq j, \quad \Delta\varphi_i(z) = 0). \tag{17}$$

Indeed, it was proven in [2] that $\phi(z)$ generates under commutations a central extension of the infinite symplectic algebra $sp(\infty,\mathbb{R})$ for $d = 2$ and that the unitary vacuum representations of this algebra correspond to integer central charge $c = N(\in \mathbb{N})$. Thus, (17) is the general form of a $d = 2$ GCI field satisfying Wightman axioms (including Hilbert space positivity) and involving a unique rank-2 symmetric traceless tensor of dimension 4 in its OPE algebra.

The physically most attractive example, corresponding to a $d = 4$ scalar field $\mathcal{L}(z)$ that can be interpreted as a QFT Lagrangian density, gives rise to a five-parameter truncated four-point function [3] of type (13) with

$$P_4(s,t) = \sum_{\nu=0}^{2} a_\nu J_\nu(s,t) + st[b(Q_1(s,t) - 2Q_2(s,t)) + cQ_2(s,t)]. \tag{18}$$

As we shall see in Section 3, J_ν correspond to the symmetrized twist-2 contribution to the OPE of two $\mathcal{L}$:

$$J_0(s,t) = s^2(1+s) + t^2(1+t) + s^2t^2(s+t) \tag{19}$$

is the only one among the five polynomials on the right-hand side of (18) which does not vanish for $s = 0$, $t = 1$,

$$J_1(s,t) = s(1-s)(1-s^2) + t(1-t)(1-t^2) + st[(s-t)(s^2-t^2) - 2Q_1], \tag{20}$$

$$J_2(s,t) = (1+t)^3[(1+s-t)^2 - s] - 3s(1-t) + s^3[(1+t-s)^2 - t]; \tag{21}$$

Q_i are $\mathcal{S}_3$ symmetric polynomials of degree 2:

$$Q_1 = 1 + s^2 + t^2, \quad Q_2 = s + t + st,$$
$$Q_1(s,t) - 2Q_2(s,t) = (1-s-t)^2 - 4st; \tag{22}$$

and the contributions stQ_j to P_4 correspond to twist 4 and higher in the OPE.

The above choice of basic $\mathcal{S}_3$ symmetric polynomials is not accidental: it is essentially determined by its relation to the "partial wave" expansion of w_4^t to be displayed in the next section. We only note at this point that the term proportional to $J_2(s,t)$ is reproduced by the free Maxwell Lagrangian

$$\mathcal{L}_0(x) = -\frac{1}{4} :F_{\mu\nu}(z)F^{\mu\nu}(z): \tag{23}$$

here, $F_{\mu\nu}$ is characterized by its two-point function

$$\langle 0|F_{\mu_1\nu_1}(z_1)F_{\mu_2\nu_2}(z_2)|0\rangle = R_{\mu_1\mu_2}(z_{12})R_{\nu_1\nu_2}(z_{12}) - R_{\mu_1\nu_2}(z_{12})R_{\nu_1\mu_2}(z_{12}), \tag{24}$$

where $R_{\mu\nu}$ is related to the vector representation of the conformal inversion:

$$R_{\mu\nu}(z) = \frac{r_{\mu\nu}(z)}{z^2}, \quad r_{\mu\nu}(z) = \delta_{\mu\nu} - 2\frac{z_\mu z_\nu}{z^2}, \quad (r^2)_{\mu\nu} = \delta_{\mu\nu}. \tag{25}$$

{The calculation proving that Eqs. (23)–(25) yield a w_4^t proportional to $\rho_{13}^2\,\rho_{24}^2(\rho_{12}\rho_{23}\rho_{34}\rho_{14})^{-3}J_2(s,t)$ is given in Appendix B from [2].}

3. OPE IN TERMS OF BILOCAL FIELDS

3.1. Fixed-Twist Fields: Conformal Partial Wave Expansion

The infinite series of integrals of local tensor fields appearing in the singular part of the OPE of $\mathcal{L}(z_1)\mathcal{L}(z_2)$ can be organized into a finite sum of bilocal (scalar) fields:

$$\mathcal{L}(z_1)\mathcal{L}(z_2) = \frac{B}{\rho_{12}^4} \tag{26}$$

$$+\sum_{\kappa=1}^{3}\rho_{12}^{\kappa-4}V_\kappa(z_1,z_2)+:\mathcal{L}(z_1)\mathcal{L}(z_2):,$$

$$V_\kappa(z_1,z_2)=V_\kappa(z_2,z_1).$$

By definition, V_κ is a bilocal conformal field of dimension (κ,κ) which only involves (derivatives of) twist-2κ tensor fields in its Taylor expansion in z_{12}. More precisely, it can be written in the form (cf. [2, 11, 17, 19, 20])

$$V_\kappa(z_1,z_2)=\sum_{\ell=0}^{\infty}C_{\kappa\ell} \tag{27}$$

$$\times\int_0^1 K_{\kappa\ell}(\alpha,\rho_{12}\Delta_2)O_{2\kappa,\ell}(z_2+\alpha z_{12};z_{12})d\alpha,$$

where $O_{2\kappa,\ell}(z;w)$ are (contracted) symmetric traceless tensor fields,

$$O_{2\kappa,\ell}(z;w)=O_{2\kappa}^{\mu_1\ldots\mu_\ell}(z)w_{\mu_1}\ldots w_{\mu_\ell} \tag{28}$$

$$(\text{tracelessness}\quad\Leftrightarrow\Delta_w O_{2\kappa,\ell}(z;w)=0),$$

of scale dimension $d_{\kappa\ell}=2\kappa+\ell$ (i.e., of fixed twist $d_{\kappa\ell}-\ell=2\kappa$). The (four-dimensional) Laplacian Δ_w acts on the 4-vector w, while Δ_2 acts on z_2 (for fixed z_{12}). The integro-differential operator with kernel $K_{k\ell}$ on the right-hand side of (27) is defined to transform the two-point function of $O_{2\kappa,\ell}$ into the three-point function

$$\langle 0|V_\kappa(z_1,z_2)O_{2\kappa,\ell}(z_3;w)|0\rangle \tag{29}$$

$$=A_{\kappa\ell}\frac{(X_{12}^3w)^\ell}{(\rho_{13}\rho_{23})^\kappa}\quad\text{for}\quad w^2=0,$$

$$X_{12}^3=\frac{z_{13}}{\rho_{13}}-\frac{z_{23}}{\rho_{23}}\left((X_{12}^3)^2=\frac{\rho_{12}}{\rho_{13}\rho_{23}}\right) \tag{30}$$

{see [20] and Section of [3], where more general OPE—for any scale dimension d and for complex fields—are considered, and the kernel $K_{\kappa\ell}$ which implements (29) is written down}. For real $\mathcal{L}$, due to the symmetry property (26) of V_κ, only even rank tensors (even ℓ) appear in expansion (27).

The "normal product" $:\mathcal{L}(z_1)\mathcal{L}(z_2):$ is defined by (26) and can in turn be expanded in higher twist V_κ ($\kappa\geq 4$), accompanied by positive powers of ρ_{12}.

The usefulness of expansion (26) is enhanced by the fact that fields of different twists are mutually orthogonal under vacuum expectation values:

$$\langle 0|V_\kappa(z_1,z_2)|0\rangle=0 \tag{31}$$

$$=\langle 0|V_\kappa(z_1,z_2)V_\lambda(z_3,z_4)|0\rangle\quad\text{for}\quad 0<\kappa\neq\lambda.$$

For equal dimensions, we can write

$$\langle 0|V_\kappa(z_1,z_2)V_\kappa(z_3,z_4)|0\rangle=(\rho_{13}\rho_{24})^{-\kappa}f_\kappa(s,t) \tag{32}$$

$$\times\left(s_{12}f_\kappa(s,t)=t^{-\kappa}f_\kappa\left(\frac{s}{t},\frac{1}{t}\right)=f_\kappa(s,t)\right).$$

As a consequence of the OPE (27) and of (29), we find the following (light-cone) conformal partial wave expansion:

$$f_\kappa(0,t)=\sum_{\ell=0}^{\infty}B_{\kappa\ell}(1-t)^{2\ell} \tag{33}$$

$$\times F(2\ell+\kappa,2\ell+\kappa;4\ell+2\kappa;1-t),$$

$$F(\alpha,\beta;\gamma;x)=\sum_{n=0}^{\infty}\frac{(\alpha)_n(\beta)_n}{n!(\gamma)_n}x^n=1+\frac{\alpha\beta}{\gamma}x+\ldots,$$

being the Gauss hypergeometric function. The structure constants $B_{\kappa\ell}$, unlike those appearing in (27) and (29), are invariant under rescaling of $O_{2\kappa,2\ell}$:

$$B_{\kappa\ell}:=A_{\kappa 2\ell}C_{\kappa 2,\ell}\Big(O_{2\kappa 2,\ell}\to\lambda O_{2\kappa 2\ell} \tag{34}$$

$$\Rightarrow A_\kappa\to\lambda A_{\kappa 2\ell},C_{\kappa 2\ell}\to\frac{1}{\lambda}C_{\kappa 2\ell}\Big).$$

The fixed-twist condition allows one to restore the s dependence of f_κ. This is particularly transparent in the $\kappa=1$ case, in which $O_{2\ell}(z;w)$ are conserved symmetric tensor fields (including, for $\ell=2$, the stress-energy tensor). Writing the expression on the right-hand side of (29) for $\kappa=1$, $w^2=0$ as

$$\frac{1}{\ell!}\left(\frac{w}{2}\left(\frac{\partial}{\partial z_2}-\frac{\partial}{\partial z_1}\right)\right)^\ell,\frac{1}{\rho_{13}\rho_{23}},$$

we observe that it is harmonic in both z_1 and z_2. It then follows from (26) that V_1 is harmonic in each argument:

$$\Delta_1V_1(z_1,z_2)=0=\Delta_2V_1(z_1,z_2) \tag{35}$$

$$\left(\Delta_j=\sum_{\mu=1}^{4}\frac{\partial^2}{(\partial z_j^\mu)^2}\right).$$

This implies what may be called the conformal Laplace equation[4] for f_1:

$$\Delta_{st}f_1(s,t)=0, \tag{36}$$

$$\Delta_{st}:=s\frac{\partial^2}{\partial s^2}+t\frac{\partial^2}{\partial t^2}$$

$$+(s+t-1)\frac{\partial^2}{\partial s\partial t}+2\left(\frac{\partial}{\partial s}+\frac{\partial}{\partial t}\right).$$

[4] The operator (36) has appeared in various contexts in [20, 21] and [3].

Its general solution is expressed in terms of the "chiral variables" u, v of [20] (see also [21]),

$$f_1(s,t) = \frac{g(u)-g(v)}{u-v} \quad \text{for} \quad s=uv, \quad t=(1-u)(1-v). \tag{37}$$

The function g is expressed in terms of $f_1(0,t)$ as $g(u) = uf_1(0,1-u)$ (for $s=0=v$). Furthermore, as demonstrated in [20], the following extension of (37) is valid for $\kappa > 1$:

$$f_\kappa(s,t) = \frac{1}{u-v}\{F(\kappa-1,\kappa-1;2\kappa-2;v) \times g_\kappa(u) - F(\kappa-1,\kappa-1;2\kappa-2;u)g_\kappa(v)\}. \tag{38}$$

[We shall view (38) as a generalization of (37) which can be included as a special case if we set $F(0,0;0;u) \equiv 1$.] Then the light-cone expansion (33) (appearing as an "initial condition") implies

$$g_\kappa(u) = uf_\kappa(0,1-u) = u\sum_{\ell=0}^{\infty} B_{\kappa\ell}u^{2\ell}F(2\ell+\kappa,2\ell+\kappa;4\ell+2\kappa;u) \tag{39}$$

($g_\kappa(0)=0$). The functions $f_\kappa(0,t)$, on the other hand, can be determined [for a given P_4 (18)] inductively in κ as follows:

$$f_1(0,t) = t^{-3}P_4(0,t), \tag{40}$$

$$f_\kappa(0,t) = \lim_{s\to 0}\left\{s^{1-\kappa}\left[t^{-3}P_4(s,t) - \sum_{\nu=1}^{\kappa-1} s^{\nu-1}f_\nu(s,t)\right]\right\}.$$

It is important to realize that the kernels $K_{\kappa\ell}$, the functional form of the three-point function (29), and the hypergeometric functions defining the conformal partial waves (33) are universal; only the structure constants $B_{\kappa\ell}$ depend on $\mathcal{L}(z)$ and are, in fact, determined by its four-point function.

3.2. Symmetrized Contribution of Twist-2 (Conserved) Tensors

We now proceed to writing down the general rational solution to (36) for which $t^3f_1(0,t)$ is a polynomial of degree not exceeding five and which satisfies the symmetry condition $s_{12}f_1(s,t) = f_1(s,t)$; in particular [according to (32)], it obeys

$$f_1\left(0,\frac{1}{t}\right) = tf_1(0,t). \tag{41}$$

The first requirement follows from the comparison of the homogeneous factor multiplying P_d in (13) for $d=4$ by the factor $(\rho_{12}\rho_{34})^{-3}(\rho_{13}\rho_{24})^{-1}$ [resulting from (26) and (32)] which multiplies $f_1(s,t)$ in the expression for w_4^t. By (crossing) symmetrizing the contribution to w_4^t so obtained, we shall reproduce the basic polynomials J_ν (19)–(21) in (18).

Anticipating the result of Sections 3.2 and 4.2 of [3], we choose the following complete set of initial conditions, $f_1(0,t) = j_\nu(0,t)$, $\nu = 0,1,2$, consistent with the symmetry property (41):

$$j_0(0,t) = 1+t^{-1}, \quad j_1(0,t) = j_0(0,t)\frac{(1-t)^2}{t}, \quad j_2(0,t) = j_1(0,t)\left[\frac{(1-t)^2}{t}+1\right]. \tag{42}$$

Note that the rational function $I(t) := (1-t)^2/t$ which enters the expressions for j_1 and j_2 is the minimal degree polynomial in t and t^{-1} that is nonidentically zero and satisfies $I1/t = I(t)$ and $I(1)=0$.

The solutions to (36) with these initial conditions are given by

$$j_0(s,t) = j_0(0,t) = 1+t^{-1}, \tag{43}$$

$$j_1(s,t) = \left(\frac{1-t}{t}\right)^2(1+t-s) - 2\frac{s}{t} = j_1(0,t) - s(1+t^{-2}),$$

$$j_2(s,t) = (1+t^{-3}) \times [(1+s-t)^2 - s] - 3s(1-t)t^{-3}.$$

As none of the above solutions is symmetric under the exchange s_{23} (15), locality implies that such twist-2 contributions to the partial wave expansion should necessarily be accompanied by higher twist terms present in the symmetrized expressions

$$J_\alpha(s,t) = (1+s_{23}+s_{13})t^3j_\alpha(s,t), \alpha = 0,1, \quad J_2(s,t) = \frac{1}{2}(1+s_{23}+s_{13})t^3j_2(s,t). \tag{44}$$

The presence of different factors (1 and 1/2) in front of the symmetrizer in the two formulas calls for a precise definition of what is meant by saying that "J_ν is a symmetrized expression of t^3j_ν." We note that, for any value of λ_ν, the expression $J_\nu = \lambda_\nu(1+s_{23}+s_{13})t^3j_\nu$ is $\mathcal{S}_3$ symmetric. We shall say that such a J_ν is a symmetrization of t^3j_ν if $J_\nu - t^3j_\nu$ tends to zero for $s\to 0$. It then in fact follows that the possible values of λ_ν are 1 and 1/2.

It should be stressed that not every solution to (36) admits a symmetrized version in this sense. (The reader can verify that, e.g., $j_2 - j_1$ does not.) Thus, the above requirement restricts the choice of basis. In

fact, it follows from (19)–(21) and from (43) that the differences $J_\nu - t^3 j_\nu$ have different orders for small s:

$$J_0(s,t) - t^3 j_0(s,t) = s^2(1 + t^3 + s(1 + t^2)), \quad (45)$$

$$J_1(s,t) - t^3 j_1(s,t) = s(1-t)(1-t^3) - s^2(1+t^3) - s^3(1+t)^2 + s^4(1+t),$$

$$J_2(s,t) - t^3 j_2(s,t) = s^3(1 + t + t^2) - 2s^4(1+t) + s^5.$$

Comparing (45) with (40), we deduce that the differences $J_1 - t^3 j_1$, $J_0 - t^3 j_0$, and $J_2 - t^3 j_2$ carry twist 4 and higher, twist 6 and higher, and twist 8 and higher, respectively. The above basis is clearly unique with such a separation property.

Putting everything together, we can, in principle, determine all structure constants $B_{\kappa\ell}$. It follows from (29), (34) and from the relation $A_{\kappa\ell} = N_{\kappa\ell} C_{\kappa\ell}$, where $N_{\kappa\ell} (> 0)$ stands for the normalization of the two-point function of $O_{2\kappa,\ell}$, that $B_{\kappa\ell} = N_{\kappa 2\ell} C^2_{\kappa 2\ell}$ should be positive if Hilbert space (or Wightman) positivity holds. (The full argument uses the classification [15] of unitary positive energy representations of $SU(2,2)$ according to which the state spaces spanned by $O_{2\kappa,2\ell}(z,w)|0\rangle$, for $\kappa = 1, 2, \ldots$, $\ell \geq 0$, belong to the unitary series.) Thus, such a calculation will restrict the admissible values of the parameters a_ν, b, c, and B in (18) and (26), providing a nontrivial positivity check for the four-point function of $\mathcal{L}$. We shall display the corresponding equations and their solution for $\kappa = 1, 2, 3$ (the twists for which the two-point normalization B does not contribute).

Inserting on the left-hand side of (33) for $\kappa = 1$ the expression $f(0,t) = \sum_{\nu=0}^{2} a_\nu j_\nu(0,t)$, we can solve for $B_{1\ell}$ with the result

$$B_{1\ell} = \frac{1}{\binom{4\ell}{2\ell}} \{2a_0 + 2\ell(2\ell+1) \times [2a_1 + (2\ell-1)(\ell+1)a_2]\}. \quad (46)$$

The equation for $\kappa = 2$ involves a_1, b, and c:

$$f_2(0,1-u) = a_1 u \left(\frac{1}{(1-u)^3} - 1\right) + \frac{bu^2}{(1-u)^2} + \frac{c}{1-u} = \sum_{\ell=0}^{\infty} B_{2\ell} u^{2\ell} F(2\ell+2, 2\ell+2; 4\ell+4; u). \quad (47)$$

Its solution is

$$B_{2\ell} = \frac{1}{\binom{4\ell+1}{2\ell}} \times \{\ell(2\ell+3)[(\ell+1)(2\ell+1)a_1 + 2b] + c\}. \quad (48)$$

For $\kappa = 3$ we have to use expression (38) for $f_2(s,t)$, which involves a log term as

$$F(1,1;2;v) = \sum_{n=1}^{\infty} \frac{v^{n-1}}{n} = \frac{1}{v} \log \frac{1}{1-v}.$$

The result is

$$f_3(0,1-u) = \left(a_0 + \frac{a_1}{2}\right)(1 + (1-u)^{-3}) - \frac{3}{2}\frac{b}{1-u}\left(1 + \frac{1}{1-u}\right) + \frac{c}{2}\left\{\frac{2u-1}{u(1-u)} \times \left(1 + \frac{2}{u(1-u)}\right) - 2\frac{\log(1-u)}{u^3}\right\} = \sum_{\ell=0}^{\infty} B_{3\ell} u^{2\ell} F(2\ell+3, 2\ell+3; 4\ell+6; u). \quad (49)$$

A computer-aided calculation (using Maple) gives in this case

$$B_{3\ell} = \frac{1}{2}\binom{4\ell+3}{2\ell+1}^{-1} \{(\ell+1)(2\ell+3) \times [(\ell+2)(2\ell+1)(2a_0 + a_1) - 6b + 4c] - c\}. \quad (50)$$

The positivity of $B_{j\ell}$, $j = 1, 2, 3$, implies

$$a_\nu \geq 0, \quad \nu = 0, 1, 2; \quad 3a_1 + b \geq 0, \quad c \geq 0; \quad 6(2a_0 + a_1 - 3b) + 11c \geq 0. \quad (51)$$

This leaves a nonempty domain in the space of (four-point function) parameters in which positivity holds.

3.3. The Concept of a Minimal Model

We observe that the twist-2 contribution to the $2n$-point function,

$$w_1(1,2;3,4;\ldots;2n-1,2n) := \left(\prod_{i=1}^{n} \rho_{2i-1,2i}^{-3}\right) \times \langle 0|V_1(z_1,z_2)V_1(z_3,z_4)\ldots V_1(z_{2n-1},z_{2n})|0\rangle, \quad (52)$$

combined with locality, implies the existence of higher twist terms. The question arises whether one can define a "minimal model" obtained by an appropriate symmetrization of such a bilocal field contribution. The difficulty in making this idea precise comes

from the fact that there are two distinct notions of symmetrization—as displayed in (44). We shall therefore restrict our attention to symmetrizable V_1, which we proceed to define.

We begin by extending the notion of a truncated $2n$-point function to bilocal fields, setting

$$w_1^t(1,2;\ldots;2n-1,2n) = w_1(1,2;\ldots;2n-1,2n) \quad \text{for} \quad n<4, \tag{53}$$

$$\begin{aligned} w_1^t(1,2;\ldots;7,8) &= w_1(1,2;\ldots;7,8) \\ &- w_1(1,2;3,4)w_1(5,6;7,8) \\ &- w_1(1,2;5,6)w_1(3,4;7,8) \\ &- w_1(1,2;7,8)w_1(3,4;5,6) \end{aligned}$$

(and similar expressions involving symmetric subtractions for $n>4$). We say that V_1 is symmetrizable if, for any $n=2,3,\ldots$, there is a λ_n such that the function

$$w^t(1,2,\ldots,2n) = \lambda_n \sum_{i_{2k-1}<i_{2k}\le 2n} w_1^t(1,i_2;\ldots;i_{2n-1},i_{2n}), \tag{54}$$

where the sum is spread over all $(2n-1)!!$ permutations $(1,2,\ldots,2n)\to(1,i_2,\ \ldots,i_{2n})$ whose entries satisfy the above inequalities, involves the same twist-2 contribution with respect to the pair of arguments $(2i-1,2i)$ for $i=1,\ldots,n$ as $w_1^t(1,2;\ldots;2n-1,2n)$:

$$\lim_{\rho_{2i-1,2i}\to 0}\{\rho_{2i-1,2i}^3(w^t(z_1,\ldots,z_{2n}) - w_1^t(1,2;\ldots;2n-1,2n))\}=0. \tag{55}$$

We then define a minimal model by identifying the truncated $2n$-point correlation function with $w^t(1,\ldots,2n)$ (54).

Remark 3.1. The symmetrizability of V_1 is only a sufficient condition enabling us to construct a minimal model. For instance, the truncated four-point function (13) (18) for $b=c=0$ is "minimal," although V_1 is not symmetrizable according to the above definition, but is a sum of symmetric expressions associated with different factors λ in (54). It follows from the results summed up in Section 2 that both J_0 and J_2 have free-field realization:

$$V_1^{(0)}(z_1,z_2)=\sum_{i=1}^N :\varphi_i(z_1)\varphi_i(z_2): \ (\Delta\varphi_i(z)=0) \tag{56}$$

and to expression (4.21) of [3],

$$V_1^{(2)}(z_1,z_2)=C\left\{\frac{1}{4}z_{12}^2 : F^{\sigma\tau}(z_1)F_{\sigma\tau}(z_2) : -\delta^{\sigma\tau}z_{12}^\mu z_{12}^\nu : F_{\sigma\mu}(z_1)F_{\tau\nu}(z_2) :\right\} \tag{57}$$

$$(dF(z)=0=d^*F(z)),$$

respectively. As we shall see (in Section 4.2, below), $J_1(s,t)$ also admits such a realization. This allows us to extend the construction of

$$w_4^t=\frac{\rho_{13}^2\rho_{24}^2}{(\rho_{12}\rho_{23}\rho_{34}\rho_{14})^3} \times (a_0J_0(s,t)+a_1J_1(s,t)+a_2J_2(s,t)) \tag{58}$$

to higher point correlation functions, thus obtaining a minimal model of a more general type than the above-considered case of a symmetrizable V_1.

4. IS THERE A NONTRIVIAL GAUGE FIELD THEORY MODEL?

4.1. Restrictions on the Parameters in the Four-Point Function

We now address the question how to characterize the local gauge-invariant Lagrangian, which gives rise to a 4-form

$$\mathcal{L}(z)dz_1\wedge dz_2\wedge dz_3\wedge dz_4=\mathrm{tr}({}^*F(z)\wedge F(z)), \tag{59}$$

where F is the (Maxwell, Yang–Mills) curvature 2-form and ${}^*F$ is its Hodge dual,

$$F(z)=\frac{1}{2}F_{\mu\nu}(z)dz^\mu\wedge dz^\nu, \tag{60}$$

$${}^*F(z)=\frac{1}{4}\varepsilon_{\kappa\lambda\mu\nu}F^{\kappa\lambda}dz^\mu\wedge dz^\nu,$$

without introducing gauge-dependent quantities like F (in the non-Abelian case) or the connection 1-form (of the gauge potential) A. We first note that a pure gauge Lagrangian of type (59) (i.e., a Lagrangian without matter fields) should not allow for a scalar of dimension (= twist) 2 in the OPE of $\mathcal{L}(z_1)\mathcal{L}(z_2)$. In view of (46), this implies

$$a_0=0. \tag{61}$$

Furthermore, assuming invariance of the theory under "electric–magnetic" (or Hodge) duality[5] and noting that ${}^*({}^*F)=-F$ in Minkowski space, we deduce that the theory should be invariant under a change of sign of $\mathcal{L}$. Hence, all odd-point functions of $\mathcal{L}$ should vanish. We make the stronger assumption that *no* scalar field of dimension 4 should appear in the OPE of two $\mathcal{L}$. According to (48), this implies

$$c=0. \tag{62}$$

[5] We thank Dirk Kreimer for a discussion on this point.

(The vanishing of the three-point function of the Maxwell Lagrangian is verified by a direct calculation.) We are thus left with the three parameters a_1, a_2, and b in the truncated four-point function, the positivity restrictions (51) implying

$$a_1 \geq 0, \quad a_1 + a_2 > 0, \quad -3a_1 \leq b \leq \frac{1}{3}a_1. \quad (63)$$

Clearly, for $a_1 = 0$, we shall also have $b = 0$ and the truncated four-point function will be a multiple of that of the free electromagnetic Lagrangian (23) (Section 2.2). In order to go beyond the free-field theory we shall assume $a_1 > 0$ and will study the properties of the contribution J_1 (20) to w_4^t.

4.2. A Composite Bilocal Field Reproducing J_1

We denote by $v(z_1, z_2)$ the field V_1 which reproduces the term $j_1(s,t)$ (43) in the expansion of f_1 [defined by (32) for $\kappa = 1$]; in other words, we set

$$\langle 0|v(z_1,z_2)v(z_3,z_4)|0\rangle = \frac{a}{\rho_{13}\rho_{24}} j_1(s,t) \quad (64)$$
$$= a\left(\frac{t^{-1} - st^{-1} - 1}{\rho_{14}\rho_{23}} + \frac{t - s - 1}{\rho_{13}\rho_{24}}\right).$$

The two terms on the right-hand side are obtained from one another by the substitution $s_{12}(z_1 \rightleftarrows z_2)$, suggesting that there exists a free-field realization of $v(z_1, z_2)$ which reproduces (64) (for an appropriate choice of the constant a). Indeed, one can check that the CP-even projection of the product of two conjugate Weyl fields does the job.

In order to write it down in the above z picture, we introduce the 2×2-matrix realization of the quaternion algebra setting

$$\not{z} := \sum_{\mu=1}^{4} z_\mu Q_\mu = \begin{pmatrix} z_4 - iz_3 & -z_2 - iz_1 \\ z_2 - iz_1 & z_4 + iz_3 \end{pmatrix}, \quad (65)$$
$$\not{z}^+ := \sum_{\mu=1}^{4} z_\mu Q_\mu^+ = z_4 - \mathbf{z}\cdot\mathbf{Q}$$

(where, in the last equation, we have omitted the 2×2 unit matrix Q_4). It is characterized by the basic anticommutation relations

$$\not{z}\not{w}^+ + \not{w}\not{z}^+ = 2zw \quad (= 2(\mathbf{z}\cdot\mathbf{w} + z_4 w_4)). \quad (66)$$

A free Weyl field ψ is a complex two-component field of $S(U(2) \times U(2))$ weight $(3/2; 1/2, 0)$ satisfying the Weyl equation

$$\not{\partial}\psi(z) := Q_\mu \frac{\partial}{\partial z_\mu}\psi(z) = 0 \left(= \frac{\partial}{\partial z_\mu}\psi^+(z)Q_\mu\right); \quad (67)$$

for an appropriate normalization, its two-point function is given by

$$\langle 0|\psi(z_1)\psi^+(z_2)|0\rangle = \frac{\not{z}_{12}^+}{\rho_{12}^2}. \quad (68)$$

A straightforward computation (see Appendix) shows that the bilocal field

$$v(z_1,z_2) =: \psi^+(z_1)\not{z}_{12}\psi(z_2) : \quad (69)$$
$$- : \psi^+(z_2)\not{z}_{12}\psi(z_1) := v(z_2,z_1)$$

(where :: stands for the standard free-field normal product) indeed reproduces (64) for $a = 2$.

The bilocal field (69), being symmetric under the exchange of arguments, only involves even-rank (conserved) tensors in its expansion in local fields. The product $v(z_1, z_2)v(z_3, z_4)$, however, also gives rise to odd-rank tensors, in particular, a conserved current, which will thus contribute to the eight-point function of $\mathcal{L}$. A gauge-invariant conserved current, on the other hand, can only appear in an Abelian, say $U(1)$, gauge theory. This looks puzzling: Are we not dealing with a free Weyl theory in disguise? The answer is no: in the theory of a free Weyl field, there is no room for a dimension-4 scalar. In particular, the Lagrangian of a free Weyl field vanishes, while in the model at hand $\mathcal{L}$ is coupled to v since the OPE of two $\mathcal{L}$ involves v.

The basic question in the title of this section can then be sharpened as follows: Is there a minimal model (in the sense of Section 3.3) for a scalar field $\mathcal{L}(z)$ with the reduced OPE (in which we have set $a_2 = b = 0$)

$$\mathcal{L}(z_1)\mathcal{L}(z_2) = \frac{B}{\rho_{12}^4} + \frac{v(z_1,z_2)}{\rho_{12}^3} + O(\rho_{12}^{-2}) \quad (70)$$

[where $v(z_1, z_2)$ has correlation functions corresponding to the ansatz (69)]?

We have no definitive answer to this simple question. We know that the four-point function of $\mathcal{L}$ with truncated part

$$w_4^t = \frac{a\rho_{13}^2\rho_{24}^2}{(\rho_{12}\rho_{23}\rho_{34}\rho_{14})^3} \quad (71)$$
$$\times \{s(1-s)(1-s^2) + t(1-t)(1-t^2)$$
$$+ st[(s-t)(s^2-t^2) - 2 - 2s^2 - 2t^2]\}$$

satisfies all Wightman axioms—including positivity (for a nonempty set of positive a and B). Happily, as we shall see in the next subsection, the ansatz (69) also allows one to construct higher (even-point) correlation functions of $\mathcal{L}$. Then one should be able to attack the positivity problem in its full generality (all other Wightman axioms being obviously satisfied by rational symmetric correlation functions). Albeit we

are ultimately interested in the theory with a truncated four-point function depending on all three admissible parameters, a_1, a_2, and b, it should be easier to, first, handle the simpler model defined by (69), (70).

4.3. Elementary Contributions to w_{2n}^t, Symmetric under $\mathbb{Z}_2 \times \mathbb{Z}_n$

We observe that expression (44) for J_1 is a sum of three terms each of which splits into two, as displayed on the right-hand side of (64). This observation can be understood as follows: each of the six contributions, say

$$\mathcal{W}(12;34) = \frac{2}{\rho_{12}^3\rho_{34}^3}\frac{\rho_{13}\rho_{24} - \rho_{12}\rho_{34} - \rho_{14}\rho_{23}}{\rho_{14}^2\rho_{23}^2} \quad (72)$$

(where we have set $a = 2$), has a $\mathbb{Z}_2 \times \mathbb{Z}_2$ symmetry generated by the pair of involutive transformations $(12;34) \to (21;43)$ and $(12;34) \to (43;21)$. The number of such elementary contributions is $|\mathcal{S}_4|/|\mathbb{Z}_2 \times \mathbb{Z}_2| = 4!/4 = 3!$ (where we denote by $|G|$ the number of elements of the finite group G). Such a counting readily extends to the general case of a $2n$-point function.

The $2n$-point function

$$\langle 0|v(z_1,z_2)v(z_3,z_4)\dots v(z_{2n-1},z_{2n})|0\rangle \quad (73)$$

of the composite field (69) has a $(\mathbb{Z}_2)^{\times n} \times \mathcal{S}_n$ symmetry (consisting of exchanging the arguments of each individual factor and of permuting the v). There are, hence, $|\mathcal{S}_{2n}|/(2^n|\mathcal{S}_n|) = (2n-1)!!$ different $2n$-point functions $\langle 0|v(z_1,z_{i_2})\dots v(z_{i_{2n-1}},z_{i_{2n}})|0\rangle$. The expression for each such $2n$-point function has $(2n-2)!! = 2^{n-1}(n-1)!$ elementary contributions labeled by the set of pairs (ij) for which the corresponding rational function has a pole in ρ_{ij}. Each elementary contribution to w_{2n}^t, say $\mathcal{W}(12;34;\dots;2n-12n)$, possesses a $\mathbb{Z}_2 \times \mathbb{Z}_n$ symmetry, where $\mathbb{Z}_n$ is the group of cyclic permutations of ordered pairs of arguments generated by $(12,34,\dots,2n-12n) \to (2n-12n,12,\dots,2n-32n-2)$, while the nontrivial element of $\mathbb{Z}_2$ is the involutive permutation $(12,\dots,2n-12n) \to (2n2n-1,\dots,21)$.

For the six-point function, we have $5!! = 15$ vacuum expectation values of type (73) each consisting of $2^2 \times 2! = 8$ elementary $\mathbb{Z}_2 \times \mathbb{Z}_3$ $(=\mathbb{Z}_6)$ symmetric contributions. One of the resulting $8 \times 15 = 120$ elementary contributions to w_6^t is $\mathcal{W}(12;34;56)$; for the corresponding term in (73), we find (see Appendix)

$$\rho_{12}^3\rho_{34}^3\rho_{56}^3\mathcal{W}(12;34;56) \quad (74)$$
$$= (\rho_{16}\rho_{23}\rho_{45})^{-2}\{\rho_{12}(\rho_{34}\rho_{56} - \rho_{35}\rho_{46} + \rho_{36}\rho_{45})$$
$$- \rho_{13}(\rho_{24}\rho_{56} - \rho_{25}\rho_{46} + \rho_{26}\rho_{45})$$
$$+ \rho_{14}(\rho_{23}\rho_{56} - \rho_{25}\rho_{36} + \rho_{26}\rho_{35})$$
$$- \rho_{15}(\rho_{23}\rho_{46} - \rho_{24}\rho_{36} + \rho_{26}\rho_{34})$$
$$+ \rho_{16}(\rho_{23}\rho_{45} - \rho_{24}\rho_{35} + \rho_{25}\rho_{34})\},$$

thus exhibiting the general pattern: it displays a Wick structure of a fermionic correlation function with "propagators" ρ_{ij}. In the perspective of studying the positivity property of a sum of 120 terms of type (74) (plus 15 products of two-point functions $\langle 1i_2\rangle\langle i_3i_4\rangle\langle i_5i_6\rangle$ with $i_{2k-1} < i_{2k} \le 6$), it seems more promising to try to use the OPE (70) for the minimal model at hand together with the fact that the theory of a free Weyl field [and hence the theory of the bilocal field (69)] does satisfy Wightman positivity.

4.4. Concluding Remarks

Global conformal invariance [4] opens the way to constructing four- (or higher) dimensional QFT models satisfying all Wightman axioms (except for asymptotic completeness). Experience with gauge field theory suggests that the simplest local gauge-invariant observable is the Lagrangian density $\mathcal{L}$. The present update of our effort to construct a nonperturbative GCI gauge QFT [2, 3] displays some new features and suggests new questions (or new ways of approaching old ones).

We emphasize that the main tool for attacking the difficult problem of Wightman positivity is the conformal partial wave expansions of four-point functions. They should be extended to four-point functions of composite (tensor) fields or, alternatively, to higher point functions of $\mathcal{L}(z)$. OPE provide just a means to derive such expansions with structure constants invariant under rescaling [like (34)].

The notion of a minimal model is introduced that is fully determined by a (twist-2) symmetrizable bilocal field $V_1(z_1,z_2)$ which is harmonic in each argument.

It is demonstrated that the nontrivial twist-2 contribution to the four-point function (which is not reproduced by an $\mathcal{L}(z)$ that is a normal product of free fields) requires a bilocal field $V_1 = v(z_1,z_2)$ that can be realized by such a normal product [involving free Weyl spinors ψ and ψ^+—see (69)].

The resulting rather special minimal model appears to be sufficiently simple to enable one to decide whether it is consistent with Wightman positivity and to find out whether its properties (in particular, the existence of a conserved $U(1)$ current in the OPE of four $\mathcal{L}$) agree with the conjecture that we are dealing with a nonperturbative renormalization group fixed point of a gauge field theory.

ACKNOWLEDGMENTS

I.T. acknowledges the hospitality of l'Institut des Hautes Etudes Scientifiques (Bures-sur-Yvette), where this paper was written.

The research of Ya.S. has been supported in part by EC contracts HPRN-CT-2000-00122 and -00148. We all acknowledge partial support by the Research Training Network within Framework Programme 5 of the European Commission under contract HPRN-CT-2002-00325.

Appendix

COMPUTING CORRELATION FUNCTIONS OF $v(z_1, z_2)$

The four-point function (64) for v given by (69) and the two-point function of ψ normalized according to (68) is expressed as a symmetric combination of traces:

$$\langle 0|v(z_1,z_2)v(z_3,z_4)|0\rangle = \frac{\mathrm{tr}\{\not z_{12}(\not z^+_{23}\not z_{34}\not z^+_{14}+\not z^+_{14}\not z_{34}\not z^+_{23})\}}{\rho^2_{14}\rho^2_{23}} - \frac{\mathrm{tr}\{\not z_{12}(\not z^+_{24}\not z_{34}\not z^+_{13}+\not z^+_{13}\not z_{34}\not z^+_{24})\}}{\rho^2_{13}\rho^2_{24}}. \tag{A.1}$$

It is sufficient to compute the first term since the second can be obtained from it by the substitution $z_3 \rightleftarrows z_4$. To do that, we shall use the following trace formula for the product of any four 4-vectors a, b, c, d written as quaternions:

$$\mathrm{tr}(\not a\not b^+\not c\not d^+) = 2[(ab)(cd)-(ac)(bd)+(ad)(bc)+\det(a,b,c,d)], \qquad 2(ab)=\mathrm{tr}\not a\not b^+, \tag{A.2}$$

where $\det(a,b,c,d)$ is the determinant of the 4×4 matrix of the components of the four (column) vectors, changing sign under transposition of any two arguments. It follows that

$$\mathrm{tr}(\not z_{12}\not z^+_{23}\not z_{34}\not z^+_{14}+\not z_{12}\not z^+_{14}\not z_{34}\not z^+_{23}) = 4[(z_{12}z_{23})(z_{34}z_{14})-(z_{12}z_{34})(z_{14}z_{23})+(z_{12}z_{14})(z_{23}z_{34})]. \tag{A.3}$$

To reproduce (72), one uses the relations

$$2z_{12}z_{23}=\rho_{13}-\rho_{12}-\rho_{23}, \qquad 2z_{34}z_{14}=\rho_{34}+\rho_{14}-\rho_{13}, \quad \text{etc.}, \qquad 2z_{12}z_{34}=\rho_{14}+\rho_{23}-\rho_{13}-\rho_{24}, \qquad 2z_{14}z_{23}=\rho_{13}+\rho_{24}-\rho_{12}-\rho_{34}, \quad \text{etc.} \tag{A.4}$$

Similarly, the (polynomial) elementary contribution (74) to the six-point function

$$\rho^2_{16}\rho^2_{23}\rho^2_{45}\langle 0|v(z_1,z_2)v(z_3,z_4)v(z_5,z_6)|0\rangle$$

is given by

$$\begin{aligned} P(12;34;56) &:= \mathrm{tr}\{\not z_{12}(\not z^+_{23}\not z_{34}\not z^+_{45}\not z_{56}\not z^+_{16}+\not z^+_{16}\not z_{56}\not z^+_{45}\not z_{34}\not z^+_{23})\} \\ &= 4\{(z_{12}z_{23})[(z_{34}z_{45})(z_{56}z_{16})-(z_{34}z_{56})(z_{45}z_{16})+(z_{34}z_{16})(z_{45}z_{56})]-(z_{12}z_{34})[(z_{23}z_{45})(z_{56}z_{16}) \\ &\quad -(z_{23}z_{56})(z_{45}z_{16})+(z_{23}z_{16})(z_{45}z_{56})]+\ldots \\ &\quad +(z_{12}z_{16})[(z_{23}z_{39})(z_{45}z_{56}) \\ &\quad -(z_{23}z_{45})(z_{34}z_{56})+(z_{23}z_{56})(z_{34}z_{45})]\} \end{aligned} \tag{A.5}$$

(5×3 terms). Applying to this expression the relations of type (A.4) (and using Maple to simplify the result), we recover (74).

REFERENCES

1. N. M. Nikolov, Commun. Math. Phys. **253**, 283 (2005); hep-th/0307235.
2. N. M. Nikolov, Ya. S. Stanev, and I. T. Todorov, J. Phys. A **35**, 2985 (2002); hep-th/0110230.
3. N. M. Nikolov, Ya. S. Stanev, and I. T. Todorov, Nucl. Phys. B **670**, 373 (2003); hep-th/0305200.
4. N. M. Nikolov and I. T. Todorov, Commun. Math. Phys. **218**, 417 (2001); hep-th/0009004.
5. N. M. Nikolov and I. T. Todorov, Rev. Math. Phys. **176**, 613 (2005); hep-th/0403191.
6. I. T. Todorov, in *Conformal Groups and Related Symmetries. Physical Results and Mathematical Background*, Ed. by A. O. Barut and H.-D. Doebner (Springer, Berlin, 1986), p. 387.
7. R. E. Borcherds, Proc. Natl. Acad. Sci. USA **83**, 3068 (1983).
8. V. G. Kac, *Vertex Algebras for Beginners*, 2nd ed. (American Mathematical Society, Providence, 1998).
9. E. Frenkel and D. Ben-Zvi, *Vertex Algebras and Algebraic Curves* (American Mathematical Society, Providence, 2001).
10. R. Haag, *Local Quantum Physics, Fields, Particles, Algebras* (Springer, Berlin, 1992).
11. V. K. Dobrev, G. Mack, V. B. Petkova, S. G. Petrova, and I. T. Todorov, *Harmonic Analysis of the n-Dimensional Lorentz Group and Its Application to Conformal Quantum Field Theory* (Springer, Berlin, 1977).
12. V. K. Dobrev, V. B. Petkova, S. G. Petrova, and I. T. Todorov, Phys. Rev. D **13**, 887 (1976).
13. F. A. Dolan and H. Osborn, Nucl. Phys. B **678**, 491 (2004); hep-th/0309180.
14. R. F. Streater and A. S. Wightman, *PCT, Spin and Statistics, and All That* (Benjamin, New York, 1964; Princeton Univ. Press, Princeton, 2000; Nauka, Moscow, 1966).

15. G. Mack, Commun. Math. Phys. **55**, 1 (1977).

16. I. T. Todorov, in *Mathematics and Physics. Lectures on Recent Results*, Ed. by L. Streit (World Sci., Singapore, 1985), Vol. 1, p. 195.

17. I. T. Todorov, M. C. Mintchev, and V. B. Petkova, *Conformal Invariance in Quantum Field Theory* (Scuola Normale Superiore, Pisa, 1978).

18. E. S. Fradkin and M. Ya. Palchik, Phys. Rep. **300**, 1 (1998); *Conformal Quantum Field Theory in D Dimensions* (Kluwer, Dordrecht, 1996).

19. S. Ferrara, R. Gatto, A. Grillo, and G. Parisi, Lett. Nuovo Cimento **4**, 115 (1972).

20. F. A. Dolan and H. Osborn, Nucl. Phys. B **599**, 459 (2001); hep-th/0011040.

21. B. Eden, A. C. Petkou, C. Schubert, and E. Sokatchev, Nucl. Phys. B **607**, 191 (2001); hep-th/0009106.

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1941–1946. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 2002–2007.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS"
ARMENIA, YEREVAN, AUGUST 13–19, 2003

Spectrum Generating Algebra and Coherent States of the C_λ-Extended Oscillator*

C. Quesne[1)]

Received February 10, 2005

Abstract—C_λ-extended oscillator algebras, generalizing the Calogero–Vasiliev algebra, where C_λ is the cyclic group of order λ, have recently proved very useful in the context of supersymmetric quantum mechanics and some of its variants. Here, we determine the spectrum generating algebra of the C_λ-extended oscillator. We then construct its coherent states, study their nonclassical properties, and compare the latter with those of standard λ-photon coherent states, which are obtained as a special case. Finally, we briefly review some other types of coherent states associated with the C_λ-extended oscillator.

1. INTRODUCTION

Coherent states (CS) of the harmonic oscillator [1] are known to have properties similar to those of the classical radiation field. They may be defined in various ways, for instance, as eigenstates of the oscillator annihilation operator b. With the corresponding creation operator $b^\dagger$ and the number operator $N_b \equiv b^\dagger b$, the latter satisfies the commutation relations

$$\left[N_b, b^\dagger\right] = b^\dagger, \quad [N_b, b] = -b, \quad \left[b, b^\dagger\right] = I. \quad (1)$$

In contrast, generalized CS associated with various algebras [2] may have some nonclassical properties, such as photon antibunching or sub-Poissonian photon statistics, and squeezing. As examples of such CS, we may quote the eigenstates of b^2, which were introduced as even and odd CS or cat states [3], and are a special case of generalized CS associated with the Lie algebra $su(1,1)$ [4]. We may also mention the eigenstates of b^λ ($\lambda > 2$) or kitten states [5], which may be generated in λ-photon processes.

Other examples are provided by nonlinear CS associated with a deformed oscillator (or f oscillator). The latter is defined in terms of creation, annihilation, and number operators, $a^\dagger = f(N_b)b^\dagger$, $a = bf(N_b)$, and $N = N_b$, satisfying the commutation relations [6, 7]

$$\left[N, a^\dagger\right] = a^\dagger, \quad [N, a] = -a, \quad \left[a, a^\dagger\right] = G(N), \quad (2)$$

where f is some Hermitian operator-valued function of the number operator and $G(N) = (N+1)f^2(N+1) - Nf^2(N)$. Nonlinear CS, defined as eigenstates of a [7–9], of a^2 [10], or of an arbitrary power a^λ ($\lambda > 2$) [11], have been considered in connection with nonclassical properties. It has been shown that, for a particular class of nonlinearities, the first ones are useful in the description of a trapped ion [8].

In the present paper, we shall consider some multiphoton CS, which may be associated with the recently introduced C_λ-extended oscillator [12]. The latter may be considered as a deformed oscillator with a $\mathbb{Z}_\lambda$-graded Fock space and has proved very useful in the context of supersymmetric quantum mechanics and some of its variants [12, 13]. In particular, we shall deal here in detail with CS of the C_λ-extended oscillator spectrum generating algebra [14], which are a special case of the CS of [11] and exhibit some nonclassical properties.

2. THE C_λ-EXTENDED OSCILLATOR ALGEBRA

The C_λ-extended oscillator algebra (where $C_\lambda = \mathbb{Z}_\lambda$ is the cyclic group of order λ) was introduced as a generalization of the Calogero–Vasiliev algebra, defined by

$$\left[N, a^\dagger\right] = a^\dagger, \quad \left[a, a^\dagger\right] = I + \alpha_0 K, \quad \left\{K, a^\dagger\right\} = 0, \quad (3)$$

and their Hermitian conjugates, where α_0 is some real parameter subject to the condition $\alpha_0 > -1$, and K is some Hermitian operator. The latter may be realized as $K = (-1)^N$, so that the second equation

*The text was submitted by the author in English.

[1)]Physique Nucléaire Théorique et Physique Mathématique, Université Libre de Bruxelles, Belgium; e-mail: cquesne@ulb.ac.be

in (3) becomes equivalent to $\left[a, a^\dagger\right] = I + \alpha_0 P_0 + \alpha_1 P_1$, where $\alpha_0 + \alpha_1 = 0$ and $P_0 = \frac{1}{2}\left[I + (-1)^N\right]$, $P_1 = \frac{1}{2}\left[I - (-1)^N\right]$ project onto the even and odd subspaces of the Fock space $\mathcal{F}$, respectively.

When partitioning $\mathcal{F}$ into λ subspaces $\mathcal{F}_\mu \equiv \{|k\lambda + \mu\rangle | k = 0, 1, \ldots\}$, $\mu = 0, 1, \ldots, \lambda - 1$, instead of two, the Calogero–Vasiliev algebra is replaced by the C_λ-extended oscillator algebra, defined by [12]

$$\left[N, a^\dagger\right] = a^\dagger, \quad \left[a, a^\dagger\right] = I + \sum_{\mu=0}^{\lambda-1} \alpha_\mu P_\mu, \quad (4)$$
$$a^\dagger P_\mu = P_{\mu+1} a^\dagger,$$

and their Hermitian conjugates, where $P_\mu = \lambda^{-1} \times \sum_{\nu=0}^{\lambda-1} \exp[2\pi i \nu (N - \mu)/\lambda]$ projects onto $\mathcal{F}_\mu$, $\sum_{\mu=0}^{\lambda-1} P_\mu = I$, and α_μ are some real parameters subject to the conditions $\sum_{\mu=0}^{\lambda-1} \alpha_\mu = 0$ and $\sum_{\nu=0}^{\mu-1} \alpha_\nu > -\mu$, $\mu = 1, 2, \ldots, \lambda - 1$. Taking this form of P_μ into account, it is clear that the C_λ-extended oscillator algebra (4) is a special case of deformed oscillator algebra, as defined in (2).

The operators N, $a^\dagger$, a are related to each other through the structure function $F(N) = N + \sum_{\mu=0}^{\lambda-1} \beta_\mu P_\mu$, $\beta_\mu \equiv \sum_{\nu=0}^{\mu-1} \alpha_\nu$, which is a fundamental concept of deformed oscillators: $a^\dagger a = F(N)$, $a a^\dagger = F(N+1)$ [6, 7]. Comparing this with Eq. (2), we get $f(N) = (F(N)/N)^{1/2}$.

The Fock space basis states $|n\rangle = |k\lambda + \mu\rangle = \mathcal{N}_n^{-1/2} \left(a^\dagger\right)^n |0\rangle$, where $a|0\rangle = 0$, $k = 0, 1, \ldots$, and $\mu = 0, 1, \ldots, \lambda - 1$, satisfy the relations

$$N|n\rangle = n|n\rangle, \quad a^\dagger |n\rangle = \sqrt{F(n+1)}\,|n+1\rangle, \quad (5)$$
$$a|n\rangle = \sqrt{F(n)}\,|n-1\rangle.$$

Due to the restrictions on the range of the parameters α_μ given below Eq. (4), $F(\mu) = \beta_\mu + \mu > 0$, so that all the states $|n\rangle$ are well defined.

The C_λ-extended oscillator Hamiltonian is defined by [12]

$$H_0 = \frac{1}{2}\left\{a, a^\dagger\right\}. \quad (6)$$

Its eigenstates are the states $|n\rangle = |k\lambda + \mu\rangle$ and their eigenvalues are given by $E_{k\lambda+\mu} = k\lambda + \mu + \gamma_\mu + 1/2$, where $\gamma_\mu \equiv (1/2)(\beta_\mu + \beta_{\mu+1})$. In each $\mathcal{F}_\mu$ subspace of $\mathcal{F}$, the spectrum of H_0 is harmonic, but the λ infinite sets of equally spaced energy levels, corresponding to $\mu = 0, 1, \ldots, \lambda - 1$, are shifted with respect to each other by some amount depending upon the parameters $\alpha_0, \alpha_1, \ldots, \alpha_{\lambda-1}$.

3. SPECTRUM GENERATING ALGEBRA OF THE C_λ-EXTENDED OSCILLATOR

One can generate the whole spectrum of the C_λ-extended oscillator Hamiltonian (6) from the eigenstates $|\mu\rangle$, $\mu = 0, 1, \ldots, \lambda - 1$, by using the operators [14]

$$J_+ = \frac{1}{\lambda}\left(a^\dagger\right)^\lambda, \quad J_- = \frac{1}{\lambda} a^\lambda, \quad (7)$$
$$J_0 = \frac{1}{\lambda} H_0 = \frac{1}{2\lambda}\left\{a, a^\dagger\right\}.$$

They satisfy the commutation relations

$$[J_0, J_\pm] = \pm J_\pm, \quad [J_+, J_-] = f(J_0, P_\mu), \quad (8)$$
$$[J_0, P_\mu] = [J_\pm, P_\mu] = 0,$$

where $f(J_0, P_\mu)$ [which has nothing to do with the function $f(N)$ of Eq. (2)] is a $(\lambda - 1)$th-degree polynomial in J_0 with P_μ-dependent coefficients, $f(J_0, P_\mu) = \sum_{i=0}^{\lambda-1} s_i(P_\mu) J_0^i$. The spectrum generating algebra (SGA) of the C_λ-extended oscillator is therefore a C_λ-extended polynomial deformation of $su(1,1)$: in each $\mathcal{F}_\mu$ subspace, it reduces to a standard polynomial deformation of $su(1,1)$ [15].

Its Casimir operator can be written as

$$C = J_- J_+ + h(J_0, P_\mu) = J_+ J_- \quad (9)$$
$$+ h(J_0, P_\mu) - f(J_0, P_\mu),$$

where $h(J_0, P_\mu)$ is a λth-degree polynomial in J_0 with P_μ-dependent coefficients, $h(J_0, P_\mu) = \sum_{i=0}^{\lambda} t_i(P_\mu) \times J_0^i$. Each $\mathcal{F}_\mu$ subspace is the carrier space of a unitary irreducible representation (unirrep) of the SGA, characterized by an eigenvalue c_μ of C and by the lowest eigenvalue $(\mu + \gamma_\mu + 1/2)/\lambda$ of J_0. The explicit expressions of $f(J_0, P_\mu)$, $h(J_0, P_\mu)$, and c_μ are given in [14].

For $\lambda = 2$, for which the C_λ-extended oscillator algebra reduces to the Calogero–Vasiliev algebra, the SGA (7), (8) reduces to the Lie algebra $su(1,1)$, for which $f(J_0) = -2J_0$, $h(J_0) = -J_0(J_0 + 1)$, and $c = (1 + \alpha_\mu)(3 - \alpha_\mu)/16$ [16].

Nonlinearities make their appearance for $\lambda = 3$, for which

$$f(J_0, P_\mu) = -9J_0^2 \quad (10)$$
$$- J_0 \sum_\mu (\alpha_\mu + 2\alpha_{\mu+1}) P_\mu$$
$$- \frac{1}{12} \sum_\mu (1 + \alpha_\mu)(5 - \alpha_\mu) P_\mu,$$

$$h(J_0, P_\mu) = -J_0 \Big[3J_0^2$$

$$+\frac{1}{2}J_0\sum_\mu(9+\alpha_\mu+2\alpha_{\mu+1})P_\mu$$
$$+\frac{1}{12}\sum_\mu\left(23+10\alpha_\mu+12\alpha_{\mu+1}-\alpha_\mu^2\right)P_\mu\Big],$$
$$c_\mu=\frac{1}{72}(1+\alpha_\mu)(5-\alpha_\mu)(3+\alpha_\mu+2\alpha_{\mu+1}).$$

For $\alpha_\mu=0$ corresponding to $a^\dagger=b^\dagger$, $a=b$, the operators (7) close a polynomial deformation of $su(1,1)$, with $f(J_0)$ and $h(J_0)$ expressed in terms of some binomial coefficients and Stirling numbers [14].

4. COHERENT STATES ASSOCIATED WITH THE C_λ-EXTENDED OSCILLATOR SPECTRUM GENERATING ALGEBRA

As CS associated with the C_λ-extended oscillator SGA, let us consider generalizations of the Barut–Girardelle CS of $su(1,1)$ [4], to which they will reduce in the case $\lambda=2$. These are the eigenstates $|z;\mu\rangle$ of the operator J_- defined in (7),

$$J_-|z;\mu\rangle=z|z;\mu\rangle,\quad z\in\mathbb{C},\qquad(11)$$
$$\mu=0,1,\ldots,\lambda-1.$$

Here, μ distinguishes between the λ-independent (and orthogonal) solutions to Eq. (11), belonging to the various subspaces $\mathcal{F}_\mu$. The CS $|z;\mu\rangle$ may be considered as special cases of the nonlinear CS of [11], since Eq. (11) is equivalent to $a^\lambda|z;\mu\rangle=\lambda z|z;\mu\rangle$, for $a=bf(N_b)$ and $f(N_b)$ as given in Section 2.

It can be shown [14] that the states (11) satisfy Klauder's minimal set of conditions for generalized CS [17]: they are normalizable and continuous in the label z, and they allow a resolution of unity. The other discrete label μ is analogous to the vector components of vector (or partially) CS [18].

The states $|z;\mu\rangle$ can be written in either of the alternative forms

$$|z;\mu\rangle=[N_\mu(|z|)]^{-1/2}\sum_{k=0}^{\infty}\frac{\left(z/\lambda^{(\lambda-2)/2}\right)^k}{\left[k!\left(\prod_{\nu=1}^{\mu}(\bar\beta_\nu+1)_k\right)\left(\prod_{\nu'=\mu+1}^{\lambda-1}(\bar\beta_{\nu'})_k\right)\right]^{1/2}}|k\lambda+\mu\rangle,\qquad(12)$$

$$|z;\mu\rangle=[N_\mu(|z|)]^{-1/2}\,{}_0F_{\lambda-1}\left(\bar\beta_1+1,\ldots,\bar\beta_\mu+1,\bar\beta_{\mu+1},\ldots,\bar\beta_{\lambda-1};zJ_+/\lambda^{\lambda-2}\right)|\mu\rangle,\qquad(13)$$

where $\bar\beta_\mu\equiv(\beta_\mu+\mu)/\lambda$, $(a)_k$ denotes Pochhammer's symbol, and the normalization factor $N_\mu(|z|)$ can be expressed in terms of a generalized hypergeometric function,

$$N_\mu(|z|)\qquad(14)$$
$$={}_0F_{\lambda-1}\left(\bar\beta_1+1,\ldots,\bar\beta_\mu+1,\bar\beta_{\mu+1},\ldots,\bar\beta_{\lambda-1};y\right),$$
$$y\equiv|z|^2/\lambda^{\lambda-2}.$$

Their unity resolution relation can be written as

$$\sum_\mu\int d\rho_\mu(z,z^*)\,|z;\mu\rangle\langle z;\mu|=I,\qquad(15)$$

where $d\rho_\mu(z,z^*)$ is a positive measure given in terms of a generalized hypergeometric function and a Meijer G function by

$$d\rho_\mu(z,z^*)\qquad(16)$$
$$={}_0F_{\lambda-1}\left(\bar\beta_1+1,\ldots,\bar\beta_\mu+1,\bar\beta_{\mu+1},\ldots,\bar\beta_{\lambda-1};y\right)$$
$$\times h_\mu(y)|z|d|z|d\phi,$$
$$h_\mu(y)$$
$$=\frac{G_{0\lambda}^{\lambda0}\left(y\mid0,\bar\beta_1,\ldots,\bar\beta_\mu,\bar\beta_{\mu+1}-1,\ldots,\bar\beta_{\lambda-1}-1\right)}{\pi\lambda^{\lambda-2}\left(\prod_{\nu=1}^{\mu}\Gamma(\bar\beta_\nu+1)\right)\left(\prod_{\nu'=\mu+1}^{\lambda-1}\Gamma(\bar\beta_{\nu'})\right)},$$

with y defined in Eq. (14).

In the $\lambda=2$ case, the functions ${}_0F_1$ and G_{02}^{20} of Eqs. (13), (14), and (16) being proportional to modified Bessel functions $I_{2\nu}(2|z|)$ and $K_{2\nu}(2|z|)$, $\nu=(\alpha_0-1+2\mu)/2$, respectively, the CS defined in (11) reduce to Barut–Girardello $su(1,1)$ CS [4] for the appropriate unirreps, as expected.

For $\alpha_\mu=0$, corresponding to $a^\dagger=b^\dagger$, $a=b$, the CS defined in (11) reduce to the eigenstates of b^λ or standard λ-photon CS [5],

$$|z;\mu\rangle=[N_\mu(|z|)]^{-1/2}\qquad(17)$$
$$\times\sum_{k=0}^{\infty}\left(\frac{\mu!}{(k\lambda+\mu)!}\right)^{1/2}(\lambda z)^k|k\lambda+\mu\rangle,$$

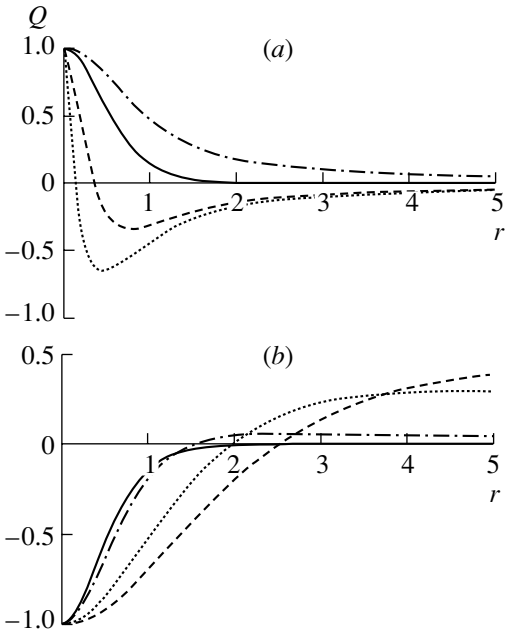

Fig. 1. Mandel's parameter Q as a function of $|z| \equiv r$ for $\lambda = 2$ and various parameters: (a) $\mu = 0$ and $\alpha_0 = 0$ (solid curve), $\alpha_0 = -4/5$ (dashed curve), $\alpha_0 = -24/25$ (dotted curve), or $\alpha_0 = 1$ (dash-dotted curve); (b) $\mu = 1$ and $\alpha_0 = 0$ (solid curve), $\alpha_0 = 1$ (dash-dotted curve), $\alpha_0 = 9$ (dotted curve), or $\alpha_0 = 19$ (dashed curve).

satisfying the resolution of unity (15) with $h_\mu(y)$ given by

$$h_\mu(y) = \lambda^{\mu-\lambda+2}\,(\pi\mu!)^{-1} \times y^{(\mu-\lambda+1)/\lambda} \exp\left(-\lambda y^{1/\lambda}\right). \quad (18)$$

The states (17) can be rewritten in the alternative form

$$|z;\mu\rangle = \left(\frac{\mu!}{E_{\lambda,\mu+1}\,(\lambda^2|z|^2)}\right)^{1/2} \times E_{\lambda,\mu+1}\left(\lambda^2 z J_+\right)|\mu\rangle, \quad (19)$$

where $E_{\alpha,\beta}(x) \equiv \sum_{k=0}^{\infty} x^k/\Gamma(\alpha k+\beta)$ is a generalized Mittag–Leffler function. Hence, they provide a simple example of the Mittag–Leffler CS considered in [19].

5. NONCLASSICAL PROPERTIES OF COHERENT STATES

The CS $|z;\mu\rangle$ may be considered as exotic states in quantum optics. Their properties may be analyzed in two different ways, by considering either "real" photons, described by the operators $b^\dagger$, b satisfying the canonical commutation relation, as given in Eq. (1), or "dressed" photons, described by the operators $a^\dagger$, a of Eq. (2), which may appear in some phenomenological models explaining some nonintuitive observable phenomena.

5.1. Photon Statistics

Since $N = N_b$, the photon-number statistics are not affected by the choice made for the type of photons. A measure of its deviation from the Poisson distribution is the Mandel parameter

$$Q = \frac{\langle(\Delta N)^2\rangle - \langle N\rangle}{\langle N\rangle}, \qquad \Delta N \equiv N - \langle N\rangle, \quad (20)$$

which vanishes for the Poisson distribution and is positive or negative according to whether the distribution is super-Poissonian (bunching effect) or sub-Poissonian (antibunching effect).

It is well known that, for $\lambda = 2$, the standard even (odd) CS, corresponding to $\alpha_0 = \alpha_1 = 0$ or $a^\dagger = b^\dagger$, $a = b$, and $\mu = 0$ ($\mu = 1$), are characterized by a super-Poissonian (sub-Poissonian) number distribution. It can be shown [14] that, for the even (odd) CS associated with the Calogero–Vasiliev algebra, i.e., for $\lambda = 2$, $\alpha_0 = -\alpha_1 \neq 0$, and $\mu = 0$ ($\mu = 1$), this trend is enhanced for positive (negative) values of α_0. However, as shown in Fig. 1, for negative (positive) values of α_0 and sufficiently high values of $|z|$, the opposite trend can be seen.

For higher values of λ, more or less similar results are obtained for $\mu = 0$, on one hand, and $\mu \neq 0$, on the other hand. However, the behavior of Q becomes more complicated for intermediate values of μ [14].

5.2. Squeezing Effect

5.2.1. "Dressed" photons. Let us define the deformed quadratures x and p as

$$x = \frac{1}{\sqrt{2}}\left(a^\dagger + a\right), \qquad p = \frac{i}{\sqrt{2}}\left(a^\dagger - a\right). \quad (21)$$

In any state belonging to $\mathcal{F}_\mu$, their dispersions $\langle(\Delta x)^2\rangle$ and $\langle(\Delta p)^2\rangle$ satisfy the uncertainty relation

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle \geq \frac{1}{4}|\langle[x,p]\rangle|^2 = \frac{\lambda^2}{4}(\bar\beta_{\mu+1} - \bar\beta_\mu)^2, \quad (22)$$

where the right-hand side becomes smaller than the conventional value $1/4$ if $\alpha_0 < 0$ for $\mu = 0$ or $-2 < \alpha_\mu < 0$ for $\mu = 1, 2, \ldots$, or $\lambda - 1$.

In $\mathcal{F}_\mu$, the role of the vacuum state is played by the number state $|\mu\rangle = |0;\mu\rangle$, which is annihilated by J_-. The corresponding dispersions are given by

$$\langle(\Delta x)^2\rangle_0 = \langle(\Delta p)^2\rangle_0 = \frac{\lambda}{2}(\bar\beta_{\mu+1} + \bar\beta_\mu). \quad (23)$$

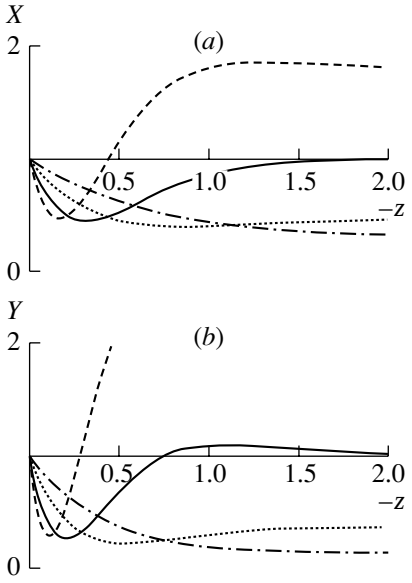

Fig. 2. The ratios (a) X and (b) Y as functions of $-z$ for real z, $\lambda = 2$, and $\mu = 0$. The parameter value is $\alpha_0 = 0$ (solid curves), $\alpha_0 = -2/5$ (dashed curves), $\alpha_0 = 1$ (dotted curves), or $\alpha_0 = 3$ (dash-dotted curves).

Comparing this with the uncertainty relation (22), we conclude that the state $|\mu\rangle$ satisfies the minimum uncertainty property in $\mathcal{F}_\mu$, i.e., gives rise to the equality in (22), only for $\mu = 0$ because $\bar{\beta}_0 = 0$ and $\bar{\beta}_\mu > 0$ for $\mu = 1, 2, \ldots, \lambda - 1$. On the other hand, the dispersions in the vacuum may be smaller than the conventional value $1/2$ for $\mu = 0, 1, \ldots, \lambda - 2$.

Let us restrict ourselves to the CS $|z; 0\rangle$, which satisfies for $z = 0$ the minimum uncertainty property. The quadrature $x(p)$ is said to be squeezed to the second order in $|z; 0\rangle$ if $X \equiv \langle(\Delta x)^2\rangle / \langle(\Delta x)^2\rangle_0$ ($P \equiv \langle(\Delta p)^2\rangle / \langle(\Delta p)^2\rangle_0$) is less than one. Similarly, it is said to be squeezed to the fourth order if $Y \equiv \langle(\Delta x)^4\rangle / \langle(\Delta x)^4\rangle_0$ ($Q \equiv \langle(\Delta p)^4\rangle / \langle(\Delta p)^4\rangle_0$) is less than one.

For $\lambda = 2$, X and P, or Y and Q, are related with each other by the transformation $\mathrm{Re}z \to -\mathrm{Re}z$. Moreover, X and Y are minimum for real, negative values of z. In Fig. 2, they are displayed for such values. We note a large squeezing effect over the whole range of real, negative values of z for positive values of α_0 (for which the conventional uncertainty relation is respected).

For $\lambda > 2$, there is no second-order squeezing, but for $\lambda = 4$, a small fourth-order squeezing is obtained in accordance with the results for standard λ-photon CS [5].

5.2.2. "Real" photons. Let us now define the quadratures x and p as

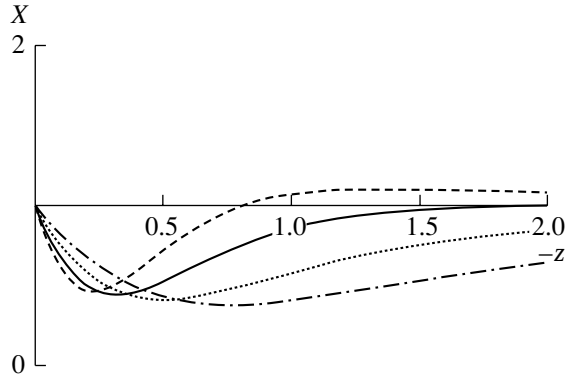

Fig. 3. The ratio X as a function of $-z$ for real z, $\lambda = 2$, and $\mu = 0$. The parameter value is $\alpha_0 = 0$ (solid curves), $\alpha_0 = -2/5$ (dashed curves), $\alpha_0 = 1$ (dotted curves), or $\alpha_0 = 3$ (dash-dotted curves).

$$x = \frac{1}{\sqrt{2}}\left(b^\dagger + b\right), \qquad p = \frac{\mathrm{i}}{\sqrt{2}}\left(b^\dagger - b\right). \quad (24)$$

Their dispersions $\langle(\Delta x)^2\rangle$ and $\langle(\Delta p)^2\rangle$ satisfy the usual uncertainty relation. Considering again the CS $|z; 0\rangle$, in Fig. 3 we observe for the ratios X and P more or less similar trends as noted in the case of "dressed" photons.

6. CONCLUDING REMARKS

In the present paper, we determined the SGA of the C_λ-extended oscillator and studied some CS associated with it, namely, the eigenstates of its lowering generator J_-.

Other types of CS may be considered and will be studied in a forthcoming publication. Let us mention here two of them:

(i) The eigenstates of the C_λ-extended oscillator annihilation operator a:

$$a|z; \mu\rangle = z|z; \mu\rangle. \quad (25)$$

These generalize the paraboson CS, which correspond to $\lambda = 2$ [20].

(ii) The solutions to the equation

$$\left[a^{\lambda-\alpha} - z\left(a^\dagger\right)^\alpha\right]|z; \mu\rangle = 0, \quad (26)$$

$$\alpha = 0, 1, \ldots, \left[\frac{\lambda}{2}\right], \quad \mu = 0, 1, \ldots, \lambda - \alpha - 1. \quad (27)$$

For $\alpha = 0$, these are the eigenstates of a^λ, which are directly related to those of J_-, considered here. Moreover, for $\lambda = 2$ and $\alpha = 1$, they reduce to the Perelomov $su(1,1)$ CS [2].

REFERENCES

1. R. J. Glauber, Phys. Rev. **131**, 2766 (1963).
2. A. P. Perelomov, *Generalized Coherent States and Their Applications* (Nauka, Moscow, 1987; Springer, Berlin, 1986).
3. V. V. Dodonov, I. A. Malkin, and V. I. Man'ko, Physica **72**, 597 (1974).
4. A. O. Barut and L. Girardello, Commun. Math. Phys. **21**, 41 (1971).
5. V. Bužek, I. Jex, and Tran Quang, J. Mod. Opt. **37**, 159 (1990).
6. C. Daskaloyannis, J. Phys. A **24**, L789 (1991).
7. A. I. Solomon, Phys. Lett. A **196**, 29 (1994).
8. R. L. de Matos Filho and W. Vogel, Phys. Rev. A **54**, 4560 (1996).
9. V. I. Man'ko, G. Marmo, F. Zaccaria, and E. C. G. Sudarshan, Phys. Scr. **55**, 528 (1997).
10. S. Mancini, Phys. Lett. A **233**, 291 (1997); S. Sivakumar, Phys. Lett. A **250**, 257 (1998).
11. X.-M. Liu, J. Phys. A **32**, 8685 (1999).
12. C. Quesne and N. Vansteenkiste, Phys. Lett. A **240**, 21 (1998).
13. C. Quesne and N. Vansteenkiste, Helv. Phys. Acta **72**, 71 (1999); math-ph/0003025; Int. J. Theor. Phys. (in press).
14. C. Quesne, Phys. Lett. A **272**, 313 (2000).
15. A. P. Polychronakos, Mod. Phys. Lett. A **5**, 2325 (1990); M. Roček, Phys. Lett. B **255**, 554 (1991).
16. T. Brzeziński, I. L. Egusquiza, and A. J. Macfarlane, Phys. Lett. B **311**, 202 (1993).
17. J. Klauder, J. Math. Phys. **4**, 1058 (1963).
18. J. Deenen and C. Quesne, J. Math. Phys. **25**, 2354 (1984); D. J. Rowe, J. Math. Phys. **25**, 2662 (1984).
19. J.-M. Sixdeniers, K. A. Penson, and A. I. Solomon, J. Phys. A **32**, 7543 (1999).
20. J. K. Sharma, C. L. Mehta, and E. C. G. Sudarshan, J. Math. Phys. (N.Y.) **19**, 2089 (1978); J. K. Sharma, C. L. Mehta, N. Mukunda, and E. C. G. Sudarshan, J. Math. Phys. **22**, 78 (1981).

Physics of Atomic Nuclei, Vol. 68, No. 11, 2005, pp. 1947–1955. From Yadernaya Fizika, Vol. 68, No. 11, 2005, pp. 2008–2016.

X INTERNATIONAL CONFERENCE "SYMMETRY METHODS IN PHYSICS" ARMENIA, YEREVAN, AUGUST 13–19, 2003

On Two Nonintegrable Cases of the Generalized Hénon–Heiles System*

S. Yu. Vernov[1)] and E. I. Timoshkova[2)]

Received October 29, 2004; in final form, February 8, 2005

Abstract—The generalized Hénon–Heiles system with an additional nonpolynomial term is considered. In two nonintegrable cases, new two-parameter solutions have been obtained in terms of elliptic functions. These solutions generalize the known one-parameter solutions. The singularity analysis shows that it is possible that three-parameter single-valued solutions exist in these two nonintegrable cases. The knowledge of the Laurent series solutions simplifies searches for the elliptic solutions and allows them to be automatized. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Beginning from papers [1–3], investigations of two-dimensional Hamiltonian systems with polynomial potentials have attracted a lot of attention due to detection of "dynamical chaos" phenomena. There is no method to find the multivalued general solution to a two-dimensional nonintegrable system in analytic form. At the same time, it is a topical problem to find single-valued special solutions in analytic form, because the investigation of the solutions with some additional properties, for example, periodic solutions, plays an important role in the study of physical phenomena. Another problem is to pick out nonintegrable cases, in which single-valued special solutions can depend on a maximal number of arbitrary parameters.

The Hénon–Heiles Hamiltonian [2]

$$H = \frac{1}{2}(x_t^2 + y_t^2 + x^2 + y^2) + x^2 y - \frac{1}{3}y^3$$

and its generalizations are one of the most actively studied two-dimensional Hamiltonians (see [4] and references therein). The generalized Hénon–Heiles system is a model widely used in astronomy [5] and physics, for example, in gravitation [6, 7].

One of the lines of investigation of this system is the search for special solutions [8–13]. The general solutions in analytic form are known only in the integrable cases [14–17]; in other cases, not only four- but even three-parameter exact solutions have yet to be found. In [12], a new type of one-parameter elliptic solutions has been obtained. Such solutions exist only in integrable cases and in two nonintegrable ones. In these nonintegrable cases, there exist three-parameter Laurent series solutions [18] which generalize the Laurent series of one-parameter elliptic solutions. In this paper, we find elliptic two-parameter solutions which generalize solutions obtained in [12].

2. BASIC EQUATIONS

The generalized Hénon–Heiles system with an additional nonpolynomial term is described by the Hamiltonian

$$H = \frac{1}{2}(x_t^2 + y_t^2 + \lambda_1 x^2 + \lambda_2 y^2) + x^2 y - \frac{C}{3}y^3 + \frac{\mu}{2x^2} \tag{1}$$

and the corresponding system of the motion equations:

$$\begin{cases} x_{tt} = -\lambda_1 x - 2xy + \dfrac{\mu}{x^3} \\ y_{tt} = -\lambda_2 y - x^2 + Cy^2, \end{cases} \tag{2}$$

where $x_{tt} \equiv d^2x/dt^2$ and $y_{tt} \equiv d^2y/dt^2$, and $\lambda_1, \lambda_2, \mu$, and C are arbitrary numerical parameters. Note that, if $\lambda_2 \neq 0$, then one can set $\lambda_2 = \mathrm{sgn}(\lambda_2)$ without loss of generality.

Due to the Painlevé analysis [19–21], the following integrable cases have been found [22]:

$$\text{(i)} \quad C = -1, \quad \lambda_1 = \lambda_2,$$

$$\text{(ii)} \quad C = -6, \quad \text{arbitrary } \lambda_1, \lambda_2,$$

$$\text{(iii)} \quad C = -16, \quad \lambda_1 = \lambda_2/16.$$

These integrable cases correspond precisely to the stationary flows of only three integrable cases of the fifth-order polynomial nonlinear evolution equations

*The text was submitted by the authors in English.

[1)]Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, 119992 Russia; e-mail: svernov@theory.sinp.msu.ru

[2)]Central Astronomical Observatory at Pulkovo, St. Petersburg, 196140 Russia; e-mail: elenatim@gao.spb.ru

of scale weight 7 (respectively, the Sawada–Kotega, the fifth-order Korteweg–de Vries, and the Kaup–Kupershmidt equations) [8, 23].

In all the above-mentioned cases, system (2) is integrable at any value of μ. The function y, solution to system (2), satisfies the following fourth-order equation [10, 12, 23]:

$$y_{tttt} = (2C-8)y_{tt}y - (4\lambda_1+\lambda_2)y_{tt} \quad (3)$$
$$+ 2(C+1)y_t^2 + \frac{20C}{3}y^3 + (4C\lambda_1 - 6\lambda_2)y^2$$
$$- 4\lambda_1\lambda_2 y - 4H,$$

where H is the energy of the system. We note that H is not an arbitrary parameter, but a function of initial data: y_0, y_{0t}, y_{0tt}, and y_{0ttt}. The form of this function depends on μ:

$$H = \frac{1}{2}(y_{0t}^2 + y_0^2) - \frac{C}{3}y_0^3$$
$$+ \left(\frac{\lambda_1}{2} + y_0\right)(Cy_0^2 - \lambda_2 y_0 - y_{0tt})$$
$$+ \frac{(\lambda_2 y_{0t} + 2Cy_0y_{0t} - y_{0ttt})^2 + \mu}{2(Cy_0^2 - \lambda_2 y_0 - y_{0tt})}.$$

This formula is correct only if $x_0 = Cy_0^2 - \lambda_2 y_0 - y_{0tt} \neq 0$. If $x_0 = 0$, which is possible only at $\mu = 0$, then we cannot express x_{0t} through y_0, y_{0t}, y_{0tt}, and y_{0ttt}, so H is not a function of the initial data. If

$$y_{0ttt} = 2Cy_0y_{0t} - \lambda_2 y_{0t},$$

then Eq. (3) with an arbitrary H corresponds to system (2) with $\mu = 0$; in the opposite case, Eq. (3) does not correspond to system (2).

To find a special solution to Eq. (3), one can assume that y satisfies some simpler equation. For example, there exist solutions in terms of the Weierstrass elliptic functions, which satisfy the following equation:

$$y_t^2 = \mathcal{A}y^3 + \mathcal{B}y^2 + \mathcal{C}y + \mathcal{D}, \quad (4)$$

where $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$ are constants yet to be determined.

The following generalization of Eq. (4)

$$y_t^2 = \tilde{\mathcal{A}}y^3 + \tilde{\mathcal{B}}y^{5/2} + \tilde{\mathcal{C}}y^2 + \tilde{\mathcal{D}}y^{3/2} + \tilde{\mathcal{E}}y + \tilde{\mathcal{G}} \quad (5)$$

gives new one-parameter solutions in two nonintegrable cases [12]: $C = -16/5$ and $C = -4/3$ (λ_1 is an arbitrary number, $\lambda_2 = 1$). It is easy to show [12] that, if $\tilde{\mathcal{B}} \neq 0$ or $\tilde{\mathcal{D}} \neq 0$, then $\tilde{\mathcal{G}} = 0$; therefore, substitution $y = \varrho^2$ transforms Eq. (4) into

$$\varrho_t^2 = \frac{1}{4}(\tilde{\mathcal{A}}\varrho^4 + \tilde{\mathcal{B}}\varrho^3 + \tilde{\mathcal{C}}\varrho^2 + \tilde{\mathcal{D}}\varrho + \tilde{\mathcal{E}}). \quad (6)$$

In [13], using the substitution $y \to y - P_0$, a new parameter P_0 has been introduced and two-parameter solutions have been constructed for the above-mentioned values of C and a few values of λ_1 ($\lambda_2 = 1$). Due to Painlevé analysis, local three-parameter solutions as the converging Laurent series have been found for an arbitrary λ_1, $\lambda_2 = 1$, and $\mu = 0$ [18]. In the present paper, we seek new elliptic solutions for arbitrary values of λ_1, λ_2, and μ.

3. NEW SOLUTIONS

Let us assume that solutions to Eq. (3) in the neighborhood of singularity point t_0 tend to infinity as $y = c_\beta(t-t_0)^\beta$, where β and c_β are some complex numbers. Of course, the real part of β has to be less than zero. From this assumption, it follows [22] that $\beta = -2$. The Laurent series of solutions to Eq. (6) begin from a term proportional to $(t-t_0)^{-1}$, so we seek solutions to Eq. (3) as a quadratic polynomial, $y = P_2\varrho^2 + P_1\varrho + P_0$, where P_2, P_1, and P_0 are arbitrary numbers, and ϱ is the general solution to Eq. (6) with arbitrary coefficients $\tilde{\mathcal{A}}$, $\tilde{\mathcal{B}}$, $\tilde{\mathcal{C}}$, $\tilde{\mathcal{D}}$, and $\tilde{\mathcal{E}}$. Because the function $\tilde{\varrho} = (\varrho - P_1/2)/\sqrt{P_2}$ is a solution to Eq. (6) as well, we can set $P_2 = 1$ and $P_1 = 0$ without loss of generality.

Substituting $y = \varrho^2 + P_0$ in Eq. (3), we obtain

$$\varrho_{tttt}\varrho = -4\varrho_{ttt}\varrho_t - 3\varrho_{tt}^2 + 2(C-4)\varrho_{tt}\varrho^3 \quad (7)$$
$$+ (2P_0(C-4) - 4\lambda_1 - \lambda_2)\varrho_{tt}\varrho + 2(3C-2)\varrho_t^2\varrho^2$$
$$+ (2CP_0 - 4\lambda_1 - 8P_0 - \lambda_2)\varrho_t^2 + \frac{10}{3}C\varrho^6$$
$$+ (2C\lambda_1 + 10CP_0 - 3\lambda_2)\varrho^4 + 2(2\lambda_1 CP_0$$
$$+ 5CP_0^2 - \lambda_1\lambda_2 - 3P_0\lambda_2)\varrho^2 + \frac{10}{3}CP_0^3$$
$$+ 2\lambda_1 CP_0^2 - 3P_0^2\lambda_2 - 2\lambda_1\lambda_2 P_0 - 2H.$$

The function ϱ is a solution to Eq. (6); hence,

Eq. (7) is equivalent to the following system:

$$\begin{cases}
(3\tilde{\mathcal{A}}+4)(-3\tilde{\mathcal{A}}+2C)=0\\
\tilde{\mathcal{B}}(9C-21\tilde{\mathcal{A}}-16)=0\\
96\tilde{\mathcal{A}}CP_0-240\tilde{\mathcal{A}}\tilde{\mathcal{C}}-192\tilde{\mathcal{A}}\lambda_1-384\tilde{\mathcal{A}}P_0\\
-48\tilde{\mathcal{A}}\lambda_2-105\tilde{\mathcal{B}}^2\\
+128\tilde{\mathcal{C}}C-192\tilde{\mathcal{C}}+128C\lambda_1\\
+640CP_0-192\lambda_2=0\\
40\tilde{\mathcal{B}}CP_0-90\tilde{\mathcal{A}}\tilde{\mathcal{D}}-65\tilde{\mathcal{B}}\tilde{\mathcal{C}}-80\tilde{\mathcal{B}}\lambda_1\\
-160\tilde{\mathcal{B}}P_0-20\tilde{\mathcal{B}}\lambda_2+56C\tilde{\mathcal{D}}-64\tilde{\mathcal{D}}=0\\
16\tilde{\mathcal{C}}CP_0-36\tilde{\mathcal{A}}\tilde{\mathcal{E}}-21\tilde{\mathcal{B}}\tilde{\mathcal{D}}-8\tilde{\mathcal{C}}^2\\
-32\tilde{\mathcal{C}}\lambda_1-64\tilde{\mathcal{C}}P_0-8\lambda_2\tilde{\mathcal{C}}+24C\tilde{\mathcal{E}}\\
+64\lambda_1CP_0+160CP_0^2-16\tilde{\mathcal{E}}\\
-32\lambda_1\lambda_2-96P_0\lambda_2=0\\
10\tilde{\mathcal{B}}\tilde{\mathcal{E}}+(5\tilde{\mathcal{C}}+8CP_0-16\lambda_1\\
-32P_0-4\lambda_2)\tilde{\mathcal{D}}=0\\
384H=-48\tilde{\mathcal{C}}\tilde{\mathcal{E}}+96C\tilde{\mathcal{E}}P_0\\
+384C\lambda_1P_0^2+640CP_0^3-9\tilde{\mathcal{D}}^2\\
-192\tilde{\mathcal{E}}\lambda_1-384\tilde{\mathcal{E}}P_0-48\tilde{\mathcal{E}}\lambda_2\\
-384\lambda_1\lambda_2P_0-576\lambda_2P_0^2.
\end{cases}\tag{8}$$

System (8) has been solved by the REDUCE computer algebra system [24].

If $\tilde{\mathcal{B}}\neq 0$, then, from two first equations of system (8), we obtain

$$C=-\frac{4}{3}\quad\text{and}\quad\tilde{\mathcal{A}}=-\frac{4}{3}$$

or

$$C=-\frac{16}{5}\quad\text{and}\quad\tilde{\mathcal{A}}=-\frac{32}{15}.$$

If $\tilde{\mathcal{B}}=0$, then solutions with $\tilde{\mathcal{D}}\neq 0$ are also possible at $C=-16$ and -1, but only in integrable cases. The obtained solutions to Eq. (3) depend on two parameters: energy H, expressed through P_0, and parameter t_0, connected to homogeneity of time.

Six solutions to system (8) correspond to each value of P_0. Two of them (with $\tilde{\mathcal{B}}=\tilde{\mathcal{D}}=0$) generate solutions to Eq. (4). Values of $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{D}}$, corresponding to other solutions, depend on λ_1 and λ_2 and are zero only at some relations between these parameters. We will consider only solutions with $\tilde{\mathcal{B}}\neq 0$ or $\tilde{\mathcal{D}}\neq 0$. They are presented in the Appendix. These solutions can be separated by pairs in such a way that solutions in one pair differ only in signs of $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{D}}$. Basic properties of the obtained solution are considered in this section. In the next section, we analyze in detail solutions to system (8) for some values of λ_1 and λ_2.

If the right-hand side of Eq. (6) is a polynomial with multiple roots, then ϱ and y can be expressed in terms of elementary functions. In the opposite case, y is an elliptic function [25, 26].

If $\varrho(t)$ is a solution to Eq. (6), then $\xi(t)\equiv-\varrho(t)$ satisfies the following equation:

$$\xi_t^2=\frac{1}{4}\left(\tilde{\mathcal{A}}\xi^4-\tilde{\mathcal{B}}\xi^3+\tilde{\mathcal{C}}\xi^2-\tilde{\mathcal{D}}\xi+\tilde{\mathcal{E}}\right).\tag{6'}$$

It is evident that $y(t)=\varrho^2(t)+P_0=\xi^2(t)+P_0$, so two solutions to system (8) correspond to one function $y(t)$. From Eq. (6), we obtain a polynomial equation for $y(t)$

$$\begin{aligned}(y_t^2-\tilde{\mathcal{A}}(y-P_0)^3-\tilde{\mathcal{C}}(y-P_0)^2\\ -\tilde{\mathcal{E}}(y-P_0))^2=(y-P_0)^3(\tilde{\mathcal{B}}(y-P_0)+\tilde{\mathcal{D}})^2.\end{aligned}\tag{9}$$

The function $\varrho(t)$ can be expressed through the Weierstrass elliptic function $\wp(t)$ ([26], Chapter 5):

$$\varrho(t-t_0)=\frac{a\wp(t-t_0)+b}{c\wp(t-t_0)+d},\quad(ad-bc=1),$$

where t_0 is an arbitrary parameter. Periods of $\wp(t)$ and the constants a, b, c, and d are determined by Eq. (6). The function

$$y(t-t_0)=\left(\frac{a\wp(t-t_0)+b}{c\wp(t-t_0)+d}\right)^2+P_0\tag{10}$$

is the fourth-order elliptic function. This function, as a solution of Eq. (3), can have only the second-order poles, therefore, in the parallelogram of periods it has two poles with opposite residues. Solutions (10) differ from solutions of Eq. (4), which are the second-order elliptic functions [26].

The function $x(t)$ satisfies the first equation of system (2) with

$$\begin{aligned}\mu=\frac{8}{3}C^2{P_0}^5+\left(2\lambda_1C^2-\frac{14}{3}\lambda_2C\right)P_0^4\\ +\left(2\lambda_2^2-\frac{10}{3}C\tilde{\mathcal{E}}-4\lambda_1\lambda_2C\right)P_0^3+(2\lambda_1\lambda_2^2\\ -2\lambda_1C\tilde{\mathcal{E}}-4CH+3\lambda_2\tilde{\mathcal{E}})P_0^2+(2\lambda_1\lambda_2\tilde{\mathcal{E}}+\tilde{\mathcal{E}}^2\\ +4\lambda_2H)P_0+2\tilde{\mathcal{E}}H+\frac{1}{2}\lambda_1\tilde{\mathcal{E}}^2+\frac{9}{128}\tilde{\mathcal{D}}^2\tilde{\mathcal{E}}.\end{aligned}\tag{11}$$

The trajectory of the motion can be derived from the second equation of system (2). Substituting y_{tt}, we obtain

$$x^2=\left(C-\frac{3}{2}\tilde{\mathcal{A}}\right)y^2+(3\tilde{\mathcal{A}}P_0-\tilde{\mathcal{C}}-1)y$$

$$-\frac{1}{4}(5\tilde{\mathcal{B}}y+3\tilde{\mathcal{D}}-5\tilde{\mathcal{B}}P_0)\sqrt{y-P_0}$$
$$-\frac{1}{2}(\tilde{\mathcal{E}}+3\tilde{\mathcal{A}}P_0^2-2\tilde{\mathcal{C}}P_0).$$

If $\tilde{\mathcal{B}}$ and $\tilde{\mathcal{D}}$ take zero values, we get simple algebraic trajectories. The full list of such trajectories is presented in [11]. The parameter P_0 is absent in these trajectory equations.

One value of the energy H can correspond to no more than three values of P_0 and, hence, no more than six different one-parameter solutions.

4. ANALYSIS OF SOLUTIONS IN A PARTICULAR CASE

4.1. The Form of Solutions

At $C=-16/5$, $\lambda_1=1/9$, and $\lambda_2=1$, one-parameter solutions ($P_0=0$) have been considered in detail in our previous papers [12, 18]. For these values of parameters, solutions to system (8) are

1. $\tilde{\mathcal{A}}=-\frac{32}{15}$, $\tilde{\mathcal{B}}=0$, $\tilde{\mathcal{C}}=-\frac{32}{5}P_0-1$, $\tilde{\mathcal{D}}=0$, $\tilde{\mathcal{E}}=-\frac{32}{5}P_0^2-2P_0$, $H=\frac{16}{15}P_0^3+\frac{1}{2}P_0^2$;

2. $\tilde{\mathcal{A}}=-\frac{4}{3}$, $\tilde{\mathcal{B}}=0$, $\tilde{\mathcal{C}}=-4P_0-\frac{17}{33}$, $\tilde{\mathcal{D}}=0$, $\tilde{\mathcal{E}}=-4P_0^2-\frac{34}{33}P_0+\frac{20}{3267}$, $H=-\frac{2}{15}P_0^3-\frac{17}{330}P_0^2+\frac{2}{3267}P_0-\frac{230}{323433}$;

3–4. $\tilde{\mathcal{A}}=-\frac{32}{15}$, $\tilde{\mathcal{B}}=\pm\frac{8i\sqrt{15}}{45}$, $\tilde{\mathcal{C}}=-\frac{32}{5}P_0-\frac{4}{9}$, $\tilde{\mathcal{D}}=\pm\frac{4i\sqrt{15}}{9}P_0$, $\tilde{\mathcal{E}}=-\frac{32}{5}P_0^2-\frac{8}{9}P_0$, $H=\frac{16}{15}P_0^3-\frac{7}{72}P_0^2$;

5–6. $\tilde{\mathcal{A}}=-\frac{32}{15}$, $\tilde{\mathcal{B}}=\pm\frac{8}{8415}\sqrt{65}\sqrt{561}$, $\tilde{\mathcal{C}}=-\frac{32}{5}P_0-\frac{1748}{1683}$, $\tilde{\mathcal{D}}=\pm\frac{\sqrt{65}\sqrt{561}}{11329956}\times(26928P_0+8125)$, $\tilde{\mathcal{E}}=-\frac{32}{5}P_0^2-\frac{3496}{1683}P_0-\frac{333125}{7553304}$, $H=\frac{16}{15}P_0^3+\frac{7291}{13464}P_0^2+\frac{6426875}{181279296}P_0+\frac{17551324375}{9762977765376}$.

If the right-hand side of Eq. (6) is a polynomial with multiple roots, then the function y can be expressed in terms of elementary functions. For example, at $P_0=0$, substitution of solution 3–4 into Eq. (5) gives

$$y=-\frac{5}{3\left(1-3\sin\left((t-t_0)/3\right)\right)^2}, \qquad (12)$$

where t_0 is an arbitrary constant.

From (11), we obtain the following values of μ:

1. $\mu=0$,

2. $\mu=\frac{160}{1089}P_0^3+\frac{680}{11979}P_0^2-\frac{800}{1185921}P_0-\frac{7000}{1056655611}$,

3–4. $\mu=\frac{4}{3}P_0^4+\frac{5}{54}P_0^3+\frac{50}{729}P_0^2$,

5–6. $\mu=-\frac{52}{561}P_0^4-\frac{81640}{944163}P_0^3-\frac{4458460825}{152546527584}P_0^2-\frac{539878421875}{128367902961936}P_0-\frac{728473377734375}{6703885364284145664}$.

4.2. Motion Trajectories

Let us consider the equations of the motion trajectories at $C=-16/5$ and $\lambda=1/9$. In the case of the solutions with $\tilde{\mathcal{B}}=\tilde{\mathcal{D}}=0$, the trajectory equation can be reduced either to $x^2=0$ (solution 1) or to

$$x^2+\frac{6}{5}\left(y+\frac{20}{99}\right)^2=\frac{50}{1089}. \qquad (13)$$

In the last case (solution 2), the motion trajectory is an ellipse. Note, however, that the real motion does not necessarily affect the whole ellipse: it depends on two arbitrary parameters. The energy H can be considered as one of them.

In the case of solution 3–4, the trajectory equation is the following:

$$\left(x^2+\frac{5}{9}y\right)^2+\frac{5}{27}(y-P_0)(2y+P_0)^2=0. \qquad (14)$$

If $P_0=0$ [see (12)], the equation for one of the trajectory branches entirely coincides with the equation obtained in [12]. The condition $y<0$ is always required for the existence of real motion along these trajectories. Formula (9) describes precisely such a solution. For solution 5–6, the trajectory equation has the same form as for solution 3–4.

Two types of dominant behavior and resonance structure of solutions to the generalized Hénon–Heiles system and Eq. (3)

Case 1	Case 2: $\beta < \mathrm{Re}(\alpha)$
$\alpha = -2$	$\alpha = \frac{1 \pm \sqrt{1 - 48/C}}{2}$
$\beta = -2$	$\beta = -2$
$a_\alpha = \pm 3\sqrt{2 + C}$	$a_\alpha = c_1$ (arbitrary)
$b_\beta = -3$	$b_\beta = \frac{6}{C}$
$r = -1, 6, \frac{5}{2} \pm \frac{\sqrt{1 - 24(1 + C)}}{2}$	$r = -1, 0, 6, \mp\sqrt{1 - \frac{48}{C}}$
$r_4 = -1, 10, \frac{5}{2} \pm \frac{\sqrt{1 - 24(1 + C)}}{2}$	$r_4 = -1, 5, 5 - \sqrt{1 - \frac{48}{C}}, 5 + \sqrt{1 - \frac{48}{C}}$

5. THREE–PARAMETER SOLUTIONS

The Ablowitz–Ramani–Segur algorithm of the Painlevé test [20] is very useful for obtaining the solutions as formal Laurent series. Let the behavior of a solution in the neighborhood of the singularity point t_0 be algebraic, i.e., x and y tend to infinity as some powers $x = a_\alpha(t - t_0)^\alpha$ and $y = b_\beta(t - t_0)^\beta$, where α, β, a_α, and b_β are some constants. If α and β are negative integer numbers, then, substituting the Laurent series expansions, one can transform nonlinear differential equations into a system of linear algebraic equations on coefficients of Laurent series. If a single-valued solution depends on more than one arbitrary parameter, then some coefficients of its Laurent series have to be arbitrary and the corresponding systems have to have zero determinants. The numbers of such systems (named *resonances* or *Kovalevskaya exponents*) can be determined due to the Painlevé test.

Two possible dominant behaviors and resonance structures of solutions to the generalized Hénon–Heiles system [22, 27] are presented in the table. The values of r denote resonances: $r = -1$ corresponds to arbitrary parameter t_0; $r = 0$ (in case 2) corresponds to arbitrary parameter c_1. Other values of r determine powers of t, to be exact, $t^{\alpha+r}$ for x and $t^{\beta+r}$ for y, at which new arbitrary parameters can appear as solutions to the linear systems with zero determinant. Note, that the dominant behavior and the resonance structure depend only on C.

It is necessary for the integrability of system (2) that all values of r be integer and that all systems with zero determinants have solutions for any values of the free parameters entering these systems. This is possible only in the integrable cases (i)–(iii) (see Section 2).

For the search for special solutions, it is interesting to consider such values of C for which r are integer numbers either only in case 1 or only in case 2. If there exists a negative integer resonance, different from $r = -1$, then such Laurent series expansion corresponds to a special rather than general solution [22]. We demand that all values of r, but one, be nonnegative integer numbers and all these values be different. From these conditions, we obtain the following values of C: $C = -1$ and $C = -4/3$ (case 1) or $C = -16/5$, $C = -6$, and $C = -16$ (case 2, $\alpha = (1 - \sqrt{1 - 48/C})/2$), and also $C = -2$, in which these two cases coincide. It is remarkable that only for these values of C do there exist solutions to system (8) with $\tilde{\mathcal{B}} \neq 0$ or $\tilde{\mathcal{D}} \neq 0$.

Let us consider the possibility of existence of single-valued three-parameter solutions in all these cases. To obtain the result for an arbitrary value of μ, we consider Eq. (3) with an arbitrary H. Note that the values of resonances obtained from Eq. (3) (in the table, they are signified as r_4) are different from r, but we obtain the same result: the condition that all values of r_4, but $r_4 = -1$, are nonnegative integer numbers gives the same values of C.

At $C = -2$, we have a contradiction: $r_4 = 0$, but b_{-2} is not arbitrary parameter: $b_{-2} = -3$. This is the consequence of the fact that, contrary to our assumption, the behavior of the general solution in the neighborhood of a singular point is not algebraic, because its dominant term includes a logarithm [22]. At $C = -6$ and any values of other parameters, the

exact four-parameter solutions are known. In the cases of $C = -1$ and $C = -16$, the substitution of an unknown function as the Laurent series leads to the conditions $\lambda_1 = \lambda_2$ or $\lambda_1 = \lambda_2/16$, accordingly. Hence, in nonintegrable cases, three-parameter local solutions have to include logarithmic terms. Single-valued three-parameter solutions can exist only in two above-mentioned nonintegrable cases: $C = -16/5$ and $C = -4/3$.

Using the method of construction of the Laurent-series solutions for nonlinear differential equations described in [18], we obtain single-valued local solutions to Eq. (3) both at $C = -16/5$ and at $C = -4/3$. The values of other parameters are arbitrary.

At $C = -4/3$, these solutions are:

$$y = -3\frac{1}{t^2} + b_{-1}\frac{1}{t} + \frac{29}{24}b_{-1}^2 + \frac{1}{2}\lambda_1 - \frac{3}{4}\lambda_2 + \left(\frac{17}{6}b_{-1}^2 + \frac{5}{3}\lambda_1 - \frac{5}{4}\lambda_2\right)b_{-1}t + b_2t^2 - \left(\frac{55}{12}\lambda_1 b_{-1}^2 + \frac{131}{90}\lambda_1^2 + \frac{33}{40}\lambda_2^2 + \frac{9359}{2592}b_{-1}^4 + b_2 - \frac{55}{16}\lambda_2 b_{-1}^2 - \frac{131}{60}\lambda_1\lambda_2\right)b_{-1}^2t^3 + \ldots. \quad (15)$$

There exist four possible values of the parameter b_{-1}:

$$b_{-1} = \pm\sqrt{\frac{\sqrt{7(1216\lambda_1^2 - 1824\lambda_1\lambda_2 + 783\lambda_2^2} - 140\lambda_1 + 105\lambda_2}{385}}$$

or

$$b_{-1} = \pm\sqrt{\frac{-\sqrt{7(1216\lambda_1^2 - 1824\lambda_1\lambda_2 + 783\lambda_2^2)} - 140\lambda_1 + 105\lambda_2}{385}}.$$

The parameters b_2 and b_8, coefficients at t^2 and t^8, respectively, are arbitrary. The energy H enters into coefficients beginning from b_4.

At $C = -16/5$, we obtain the following solutions:

$$y = -\frac{15}{8t^{-2}} + \tilde{b}_{-1} - \frac{5}{32}\lambda_2 + \frac{62}{45}\tilde{b}_{-1}^2 + \left(\frac{5}{12}\lambda_1 + \frac{632}{225}\tilde{b}_{-1}^2 - \frac{25}{192}\lambda_2\right)\tilde{b}_{-1}t + \left(\frac{29}{15}\lambda_1\tilde{b}_{-1}^2 - \frac{1}{128}\lambda_2^2 - \frac{29}{48}\lambda_2\tilde{b}_{-1}^2 + \frac{102272}{10125}\tilde{b}_{-1}^4\right)t^2 + \tilde{b}_3t^3 + \ldots, \quad (16)$$

with

$$\tilde{b}_{-1} = \frac{\pm 3\sqrt{13\,090}}{41\,888}\sqrt{525\lambda_2 - 1680\lambda_1 + 4\sqrt{35(2048\lambda_1^2 - 1280\lambda_1\lambda_2 + 387\lambda_2^2)}}$$

or

$$\tilde{b}_{-1} = \frac{\pm 3\sqrt{13\,090}}{41\,888}\sqrt{525\lambda_2 - 1680\lambda_1 - 4\sqrt{35(2048\lambda_1^2 - 1280\lambda_1\lambda_2 + 387\lambda_2^2)}}.$$

The coefficients $\tilde{b}_3$ and $\tilde{b}_8$ are arbitrary parameters. Beginning from $\tilde{b}_4$, some coefficients include the energy H. So, the obtained local solutions depend on four independent parameters: t_0, H, and two coefficients (b_2 and b_8 or $\tilde{b}_3$ and $\tilde{b}_8$).

We have found local single-valued solutions. Of course, the existence of local single-valued solutions is a necessary but not sufficient condition for the existence of global ones, because solutions which are single-valued in the neighborhood of some singularity point can be multivalued in the neighborhood of another singularity point. So, we can only assume that global three-parameter solutions are single-valued. If we assume this and moreover that these solutions are elliptic functions (or some degenerations of them), then we can seek them as solutions to some polynomial first-order equations. There are a few methods to construct such solutions [8, 10, 28, 29]. Using these

methods, one represents a solution to a nonlinear ordinary differential equation (ODE) as the sum of finite Taylor or Laurent series of elliptic functions or degenerate elliptic functions, for example, $\tanh(t)$. A similar method is applied in this paper to find two-parameter solutions. These methods use results of the Painlevé test, but do not use the obtained Laurent series solutions. In 2003, Conte and Musette [30] proposed a method which uses such solutions.

The classical theorem which was established by Briot and Bouquet [31] proves that, if the general solution to a polynomial autonomous first-order ODE is single-valued, then this solution is either an elliptic function; or a rational function of $e^{\gamma x}$, γ being some constant; or a rational function of x. Note that the third case is a degeneracy of the second one, which in turn is a degeneracy of the first one. It has been proved by Painlevé [19] that the necessary form of a polynomial autonomous first-order ODE with a single-valued general solution is

$$\sum_{k=0}^{m}\sum_{j=0}^{2m-2k} h_{jk}y^j y_t^k = 0, \quad h_{0m} = 1, \qquad (17)$$

in which m is a positive integer and h_{jk} are constants.

Rather than substituting Eq. (17) into some nonintegrable system, one can substitute the Laurent series of unknown special solutions, for example, (15) or (16), into Eq. (17) and obtain a system which is linear in h_{jk} and nonlinear in the parameters, including in the Laurent coefficients [30]. There exists a package of Maple procedures which allow one to obtain this system from the given Laurent series. Moreover, it is possible to exclude all h_{jk} from this system and to obtain a nonlinear system in parameters of a nonintegrable system and free parameters from the Laurent series. The main preference of this method is that the number of unknowns in the resulting nonlinear algebraic system does not depend on the number of coefficients of the first-order equation. For example, Eq. (17) with $m = 8$ includes 60 unknowns h_{jk}, and it is not possible to use the traditional way to find similar solutions. Using this method, we always obtain a nonlinear system in five variables: λ_1, λ_2, H, and two arbitrary coefficients of the Laurent series solutions. We hope that this method allows us to find three-parameter global solutions.

6. CONCLUSIONS

Two nonintegrable cases ($C = -16/5$ or $C = -4/3$, and λ_1, λ_2, and μ are arbitrary) of the generalized Hénon–Heiles system with a nonpolynomial term have been considered. To avoid problems with the nonpolynomial term, we have transformed the system into a fourth-order equation. Two-parameter elliptic solutions for this equation have been found in both above-mentioned cases. Two different solutions correspond to each pair of parameter values. The Painlevé test does not present any obstacle to the existence of three-parameter single-valued solutions, so the probability of finding exact, for example, elliptic, three-parameter solutions that generalize the obtained solutions is high.

ACKNOWLEDGMENTS

S.Yu.V. is grateful to F. Calogero, R. Conte, V.F. Edneral, and A.K. Pogrebkov for valuable discussions.

This work has been supported in part by the Council on Grants of the President of the Russian Federation (grant nos. NSh-1685.2003.2) and by grants from the Scientific Program "Universities of Russia" nos. UR.02.03.028 and UR.02.02.503.

Appendix

In two nonintegrable cases ($C = -16/5$ and $C = -4/3$) for arbitrary λ_1 and λ_2, we find that each value of P_0 corresponds to six solutions to system (8). Two of them (with $\tilde{\mathcal{B}} = \tilde{\mathcal{D}} = 0$) generate solutions to Eq. (4). Other solutions to system (8) can be separated by pairs such that each pair of solutions corresponds to one two-parameter function $y = \varrho^2 + P_0$, where ϱ satisfies Eq. (6) with the following values of coefficients:

$$C = -\frac{16}{5}, \qquad \tilde{\mathcal{A}} = -\frac{32}{15},$$

$$\tilde{\mathcal{B}} = -\frac{\sqrt{1122}(1120\lambda_1 + 41888P_0 + 65S_q + 6195\lambda_2)\sqrt{F_1(\lambda_1,\lambda_2,P_0)}}{29373960(3600\lambda_1^2 - 1120\lambda_1P_0 - 2425\lambda_1\lambda_2 - 20944P_0^2 - 6195\lambda_2P_0 + 225\lambda_2^2)},$$

$$\tilde{\mathcal{C}} = -\frac{240}{187}\lambda_1 - \frac{32}{5}P_0 + \frac{4}{1309}S_q - \frac{112}{187}\lambda_2,$$

$$\tilde{\mathcal{D}} = \frac{\sqrt{1122}}{5874792}\sqrt{F_1(\lambda_1, \lambda_2, P_0)},$$

$$\tilde{\mathcal{E}} = \frac{88320}{244783}\lambda_1^2 - \frac{480}{187}\lambda_1 P_0 + \frac{885}{244783}\lambda_1 S_q - \frac{153375}{244783}\lambda_1\lambda_2 - \frac{32}{5}P_0^2 + \frac{8}{1309}P_0 S_q - \frac{224}{187}\lambda_2 P_0 - \frac{685}{3916528}\lambda_2 S_q + \frac{168855}{3916528}\lambda_2^2,$$

$$H = -\frac{11516270}{45774421}\lambda_1^3 + \frac{8740}{34969}\lambda_1^2 P_0 - \frac{3296515}{2563367576}\lambda_1^2 S_q + \frac{50336425}{183097684}\lambda_1^2\lambda_2 + \frac{258}{187}\lambda_1 P_0^2 - \frac{8209}{1958264}\lambda_1 P_0 S_q + \frac{76915}{279752}\lambda_1\lambda_2 P_0 + \frac{12202395}{82027762432}\lambda_1\lambda_2 S_q - \frac{131879855}{11718251776}\lambda_1\lambda_2^2 - \frac{43}{13090}P_0^2 S_q + \frac{103}{1496}\lambda_2 P_0^2 + \frac{8881}{31332224}\lambda_2 P_0 S_q - \frac{71205}{4476032}\lambda_2^2 P_0 - \frac{12990165}{1312444198912}\lambda_2^2 S_q - \frac{168661575}{187492028416}\lambda_2^3 + \frac{16}{15}P_0^3,$$

$$C = -\frac{4}{3}, \qquad \tilde{\mathcal{A}} = -\frac{4}{3},$$

$$\tilde{\mathcal{B}} = \frac{\sqrt{330}(952\lambda_1 - 616P_0 + 13R_q - 945\lambda_2)\sqrt{F_2(\lambda_1, \lambda_2, P_0)}}{38115(432\lambda_1^2 + 952\lambda_1 P_0 - 291\lambda_1\lambda_2 - 308P_0^2 - 945P_0\lambda_2 + 27\lambda_2^2)},$$

$$\tilde{\mathcal{C}} = -\frac{4}{33}\lambda_1 - 4P_0 - \frac{1}{66}R_q - \frac{31}{22}\lambda_2,$$

$$\tilde{\mathcal{D}} = \frac{\sqrt{330}}{7623}\sqrt{F_2(\lambda_1, \lambda_2, P_0)},$$

$$\tilde{\mathcal{E}} = \frac{3394}{363}\lambda_1^2 + \frac{54}{11}\lambda_1 P_0 - \frac{1123}{10164}\lambda_1 R_q - \frac{5897}{484}\lambda_1\lambda_2 - \frac{17}{3}P_0^2 - \frac{31}{308}P_0 R_q - \frac{349}{44}\lambda_2 P_0 + \frac{1223}{27104}\lambda_2 R_q + \frac{13005}{3872}\lambda_2^2,$$

$$H = -\frac{552922}{83853}\lambda_1^3 - \frac{29801}{2541}\lambda_1^2 P_0 + \frac{173605}{2347884}\lambda_1^2 R_q + \frac{778033}{74536}\lambda_1^2\lambda_2\lambda_2 - \frac{185}{66}\lambda_1 P_0^2 + \frac{3001}{20328}\lambda_1 P_0 R_q + \frac{104959}{6776}\lambda_1\lambda_2 P_0 - \frac{695609}{12522048}\lambda_1\lambda_2 R_q - \frac{2990049}{596288}\lambda_1\lambda_2^2 + \frac{89}{1232}P_0^2 R_q + \frac{5}{2}P_0^3 + \frac{865}{176}\lambda_2 P_0^2 - \frac{3065}{54208}\lambda_2 P_0 R_q - \frac{225909}{54208}\lambda_2^2 P_0 + \frac{2733}{260876}\lambda_2^2 R_q + \frac{57699}{74536}\lambda_2^3,$$

---

where

$$F_1(\lambda_1, \lambda_2, P_0) \equiv 39474176000\lambda_1^3 + 122782105600\lambda_1^2 P_0 - 104358400\lambda_1^2 S_q - 17822336000\lambda_1^2\lambda_2 + 210552545280\lambda_1 P_0^2 - 680261120\lambda_1 P_0 S_q - 10941145600\lambda_1\lambda_2 P_0 - 41066800\lambda_1\lambda_2 S_q + 8305290000\lambda_1\lambda_2^2 - 501315584P_0^2 S_q - 65797670400\lambda_2 P_0^2 + 55920480P_0 S_q + 1611640800\lambda_2^2 P_0 + 2884725\lambda_2^2 S_q - 468507375\lambda_2^3,$$

$$S_q \equiv \pm\sqrt{35(2048\lambda_1^2 - 1280\lambda_1\lambda_2 + 387\lambda_2^2)},$$

$$F_2(\lambda_1, \lambda_2, P_0) \equiv 2099776\lambda_1^3 - 497728\lambda_1^2 P_0 - 20008\lambda_1^2 R_q - 4911144\lambda_1^2\lambda_2 + 948640\lambda_1 P_0^2 + 19096\lambda_1 P_0 R_q + 1458072\lambda_1\lambda_2 P_0 + 37173\lambda_1\lambda_2 R_q + 3943233\lambda_1\lambda_2^2 + 6776P_0^2 R_q - 711480\lambda_2 P_0^2 - 9240\lambda_2 P_0 R_q - 615384\lambda_2^2 P_0 - 13581\lambda_2^2 R_q - 1006425\lambda_2^3,$$

$$R_q \equiv \pm\sqrt{7(1216\lambda_1^2 - 1824\lambda_1\lambda_2 + 783\lambda_2^2)}.$$

REFERENCES

1. G. Contopoulos, Z. Astrophys. **49**, 273 (1960); Astron. J. **68**, 1 (1963); **68**, 673 (1963).
2. M. Hénon and C. Heiles, Astron. J. **69**, 73 (1964).
3. A. G. Gustavson, Astron. J. **71**, 670 (1966).
4. S. Yu. Vernov, math-ph/0203003.
5. C. D. Murray and S. F. Dermott, *Solar System Dynamics* (Cambridge Univ. Press, Cambridge, 1999).
6. F. Kokubun, Phys. Rev. D **57**, 2610 (1998).
7. Ji. Podolský and K. Veselý, Phys. Rev. D **58**, 081501 (1998).

8. J. Weiss, Phys. Lett. A **102A**, 329 (1984); **105A**, 387 (1984).
9. E. I. Timoshkova, Astron. Zh. **68**, 1315 (1991) [Sov. Astron. **35**, 660 (1991)].
10. R. Conte and M. Musette, J. Phys. A **25**, 5609 (1992).
11. V. A. Antonov and E. I. Timoshkova, Astron. Zh. **70**, 265 (1993) [Astron. Rep. **37**, 138 (1993)].
12. E. I. Timoshkova, Astron. Zh. **76**, 470 (1999) [Astron. Rep. **43**, 406 (1999)].
13. E. I. Timoshkova, in *Proceedings of the International Conference "Stellar Dynamics: From Classic to Modern," St. Petersburg, Russia*, Ed. by L. P. Ossipkov and I. I. Nikiforov (St. Petersburg, 2001), p. 201.
14. C. M. Cosgrove, Stud. Appl. Math. **104**, 1 (2000).
15. A. Zhivkov and I. Makaveeva, in *Proceedings of the Third International Conference on Geometry, Integrability, and Quantization, Varna, Bulgaria, 2001*, Ed. by I. M. Mladenov and G. L. Nabel (Coral, Soña, 2001), p. 454.
16. R. Conte, M. Musette, and C. Verhoeven, J. Math. Phys. (N.Y.) **43**, 1906 (2002); nlin.SI/0112030; J. Nonlin. Math. Phys. **12** (Suppl. 1), 212 (2005); nlin.SI/0412057.
17. R. Conte, M. Musette, and C. Verhoeven, Theor. Math. Phys. **134**, 128 (2003); nlin.SI/0301011.
18. S. Yu. Vernov, Theor. Math. Phys. **135**, 792 (2003).
19. P. Painlevé, in *Oeuvres de Paul Painlevé* (ed. du CNRS, Paris, 1973), Vol. 1.
20. M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. (N.Y.) **21**, 715 (1980); **21**, 1006 (1980).
21. R. Conte, *The Painlevé Property, One Century Later, Proceedings of the Cargèse School, 1996, CRM Series in Math. Phys*. (Springer-Verlag, New York, 1999), p. 810.
22. M. Tabor, *Chaos and Integrability in Nonlinear Dynamics: An Introduction* (Wiley, New York, 1989; Editorial URSS, Moscow, 2001).
23. M. Antonowicz and S. Rauch-Wojciechowski, Phys. Lett. A **163**, 167 (1992).
24. A. C. Hearn, *REDUCE. User's and Contributed Packages Manual, Vers. 3.7* (CA and Codemist Ltd, Santa Monica, 1999), p. 488; http://www.zib.de/Symbolik/reduce/more/moredocs/reduce.pdf.
25. A. Erdélyi *et al.*, *Higher Transcendental Functions* (*Bateman Manuscript Project*) (McGraw-Hill, New York, 1955), Vol. 3.
26. A. von Hurwitz and R. von Courant, *Allgemeine Funktionentheorie und Elliptische Funktionen* (Springer-Verlag, Berlin, 1964).
27. S. Melkonian, J. Nonlinear Math. Phys. **6**, 139 (1999); math.DS/9904186.
28. G. S. Santos, J. Phys. Soc. Jpn. **58**, 4301 (1989).
29. E. Fan, J. Phys. A **36**, 7009 (2003).
30. R. Conte and M. Musette, Physica D **181**, 70 (2003); nlin.PS/0302051.
31. C. Briot and T. Bouquet, *Théorie des Fonctions Doublement Périodiques* (1859).