CONFERENCES AND SYMPOSIA

PACS number: 01.10.Fv

Scientific session of the Physical Sciences Division of the Russian Academy of Sciences dedicated to the centenary of L D Landau's birth (22, 23 January 2008)

A scientific session of the Physical Sciences Division of the Russian Academy of Sciences dedicated to the centenary of L D Landau's birth was held in the Conference Hall of the Lebedev Physics Institute, Russian Academy of Sciences, on 22 and 23 January 2008. An Opening Address by A F Andreev and the following reports were presented at the session:

(1) Andreev A F (Kapitza Institute of Physical Problems, Russian Academy of Sciences) "Supersolidity of quantum glasses";

(2) **Kagan Yu M** (Russian Research Center Kurchatov Institute, Moscow) "Formation kinetics of the Bose condensate and long-range order";

(3) **Pitaevskii L P** (Kapitza Institute of Physical Problems, Russian Academy of Sciences; Dipartimento di Fisica, Universita di Trento and BDC Center, Trento, Italy) "Superfluid Fermi liquid in a unitary regime";

(4) **Lebedev V V** (Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow Region) "Kolmogorov, Landau, and the modern theory of turbulence";

(5) Khalatnikov I M (Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow), Kamenshchik A Yu (Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow; Dipartimento di Fisica and Istituto Nazionale di Fisica Nucleare, Bologna, Italy) "Lev Landau and the problem of singularities in cosmology";

(6) **Ioffe B L** (Russian State Scientific Center Alikhanov Institute for Theoretical and Experimental Physics, Moscow) "Axial anomaly in quantum electro- and chromodynamics and the structure of the vacuum in quantum chromodynamics";

(7) **Okun L B** (Russian State Scientific Center Alikhanov Institute for Theoretical and Experimental Physics, Moscow) "The theory of relativity and the Pythagorean theorem";

(8) **Lipatov L N** (St. Petersburg Nuclear Physics Institute, Gatchina, St. Petersburg) "Bjorken and Regge asymptotics of scattering amplitudes in QCD and in supersymmetric gauge models."

A brief presentation of the Opening Address by A F Andreev and reports 2, 3, and 5-8 is given below.

Uspekhi Fizicheskikh Nauk **178** (6) 631–668 (2008) DOI: 10.3367/UFNr.0178.200806f.0631

Translated by V I Kisin, E N Ragozin, A M Semikhatov, E Yankovsky; edited by A M Semikhatov and M S Aksent'eva



Lev Davidovich Landau 22.01.1908-01.04.1968

PACS number: **01.65** + g DOI: 10.1070/PU2008v051n06ABEH006547 DOI: 10.3367/UFNr.0178.200806g.0631

L D Landau: 100th anniversary (Introductory talk)

A F Andreev

Dear colleagues! Today is a memorable day for scientists: Lev Davidovich Landau was born 100 years ago. The audience present here today — an enviable audience in all respects, of course — does not really need to be told who Lev Davidovich Landau was or why we are all gathered here on this day. I would only like to emphasize the following aspect. From my point of view, Lev Davidovich was the man who was able to show and to formulate what theoretical physics is all about indeed, there is no need to remind you how broad his scope was, what his approach was to completely different fields of physics and chemistry, to everything the world contains. He embraced the entirety of theoretical physics and with his absolutely first-class work demonstrated the essence of the general approach of theoretical physics to all natural phenomena.

Together with Evgenii Mikhailovich Lifshitz (Lev Petrovich Pitaevsky also took part in it), Landau wrote a truly fundamental course of theoretical physics, in which he not only presented the essential core of problems attacked by theoretical physics but also demonstrated his approach what he meant by working in theoretical physics and what sort of argument is allowed in true theoretical physics stripped of fuzzy philosophizing about the nature of things; he gave a straightforward demonstration: here is the way it must be done.

Landau left behind a very large school, which in fact was not all that large while he was alive, just several dozen people. Nevertheless, the first generations of Landau's students not only sustained and preserved Landau's method - the theoretician's minimum, the approach to fostering and shaping theoretical physicists-but also developed and extended it; they have every reason to be proud of this achievement. Landau's students, and most of all Isaak Markovich Khalatnikov, created the Institute for Theoretical Physics, which started to 'mass-produce' theoretical physicists of an absolutely world-class stature (eventually, even Petr Leonidovich Kapitza had to agree with this), and not on a 'one-off' basis, as it was in Landau's lifetime, but in an 'industrial' fashion. As a result, the group known as Landau's school became a high-profile community of physicists. Once the Soviet Union crashed out of existence, this absolutely unique community of people spread all over the world and we could say that in this way Landau succeeded in defining what theoretical physics is on a world scale, not just in the Soviet Union.

No doubt, some physicists contributed more to physics in general and to theoretical physics in particular than did Landau. But I do not think that we can say about anyone else that they showed what the gist of theoretical physics was, how one should do it, how to help new generations to mature, and how to write books on theoretical physics. It was after Landau's death that Evgeny Mikhailovich Lifshitz showed me a letter from an outstanding theoretician—I do not remember exactly who it was but it was one the big names-and it said: "The entire wisdom of the West came from the books of Landau and Lifshitz." This was very high praise and Evgeny Mikhailovich was of course very proud to receive it. Now the last thing I wish to say is that the times that began in Landau's lifetime and lasted through the 1970s to the beginning of the 1980s, these times are gone forever, never to return; it is sad but I am sure I am right. Consequently, all this structure that you admired and took off your hat to, is now impossible: time brought us grants, funds, foundations, and so forth. In these conditions, it will definitely never be possible to recreate the atmosphere that reigned in Landau's time, and then, after he was gone, in a few places where Landau's school flourished, e.g., in the Institute for Theoretical Physics and in some other places. Still, there is some ground for optimism since there are a good many people around who absorbed these ideas, this high respect to science and to theoretical physics; I am convinced that we will be able to continue working successfully and for a long time and that Landau's name will continue to occupy pride of place in our hearts.

I wish to mention in conclusion that the magazine *Priroda* published an outstanding special issue (No. 1, 2008) devoted to the 100th anniversary of the birth of academician L D Landau and presenting new materials on the life and work of the great scientist. *Priroda* has published such special issues devoted to Nobel Prize winners in the past — to academicians P L Kapitza, I E Tamm, and N N Semenov.

I think that the Division of Physical Sciences of the Russian Academy of Sciences ought to express its gratitude to the magazine *Priroda* for its efforts in celebrating the memory of outstanding Russian scientists.



Front page and table of contents of the special issue of the magazine *Priroda* (No. 1, 2008) devoted to the 100th anniversary of Lev Davidovich Landau's birth.

PACS numbers: 03.75. – b, 05.70.Fh, 64.60.Ht DOI: 10.1070/PU2008v051n06ABEH006544 DOI: 10.3367/UFNr.0178.200806h.0633

Formation kinetics of the Bose condensate and long-range order

Yu M Kagan

The rapidly developing area of research related to ultracold gases has opened up the unique possibility of studying the formation kinetics of a Bose condensate and long-range order. The isolation of a gas from the walls in magnetic and electric traps and the possibility of observing the intrinsic real-time evolution of the system are the decisive factors in this case. Although the first theoretical papers in this field appeared in the early 1990s, it was not until 2007 that the first experimental research on the time evolution of long-range order was reported in the literature [1-2]. A vigorous study of this phenomenon was pursued between these dates, and this report is concerned with the analysis of the data and existing notions in this area.

The capability of rapid cooling by cutting off the Maxwellian tails enables studying the evolution starting from the points in time when all correlation properties of a gas are purely classical and there is not the slightest trace of a condensate. In this case, the kinetics proceed with conservation of the total energy and the number of particles in the system. As it turns out, the evolution comprises four stages.

During the first stage, which is described by the Boltzmann equation, a particle flux forms in the energy space, directed towards lower energies. When the particles that constitute the condensate in equilibrium occur in the energy range where the kinetic energy is lower than the interparticle interaction energy, the formation of collective correlations sets in and the kinetic equation is no loner valid (the number of particles that fall into this energy range, which is commonly termed the coherence interval, is comparable with the total number of particles). But even before this, the evolution goes through a stage during which all occupation numbers of individual modes become much greater than unity. As shown in Refs [3, 4], the system is then adequately described by the classical Bose field, which obeys the nonlinear Schrödinger equation in the form of the Gross-Pitaevskii equation. The solution of this equation leads to an important result: in the coherence interval, the fluctuations of density are suppressed and the single-particle density matrix depends only on phase fluctuations. At this stage, a special quasicondensate state emerges, which is equivalent to the genuine condensate in local properties, but has no long-range order. An instantaneous picture of the gas actually demonstrates the division of the system into finite-size quasicondensate domains. Each domain has a specific phase in the absence of phase correlation between different domains.

This picture underlay the prediction that the evolution during the third stage should be accompanied by the emergence of a vorticity structure. This prediction was borne out by the direct numerical solution of the nonlinear Schrödinger equation [5], which demonstrated the emergence of a vorticity ball and its temporal evolution.

The final stage is characterized by the damping of nonequilibrium regular-phase fluctuations and the relaxation of the vorticity structure. This occurs with an increase of quasicondensate domains in size, which is effectively equivalent to an increase in the density-matrix decay distance, thereby determining the evolution of the long-range order scale [3, 4] (see also Ref. [6]). The long-range order settling time τ_L increases with the domain size L: $\tau_L \sim L^n$, where n = 1-2, depending on parameter ratios.

The report presents a comparison with the theory and a comprehensive analysis of the experimental results found in Refs [1, 2], especially of the temporal evolution of long-range order formation [1]. The analysis relies on the theory elaborated for the analog of the Hanbury–Brown–Twiss effect for particles in the 'two sources, one detector' setup in the evolution of a nonequilibrium system involving a classical-to-quantum transformation of correlations [7]. Experimental data are qualitatively and quantitatively compared with theoretical predictions.

References

- 1. Ritter S et al. Phys. Rev. Lett. 98 090402 (2007)
- 2. Hugbart M et al. *Phys. Rev. A* **75** 011602(R) (2007)
- 3. Kagan Yu, Svistunov B V Phys. Rev. Lett. 79 3331 (1997)
- 4. Kagan Yu, Svistunov B V Pis'ma Zh. Eksp. Teor. Fiz. 67 495 (1998)
- [JETP Lett. 67 521 (1998)]
- 5. Berloff N G, Svistunov B V *Phys. Rev. A* **66** 013603 (2002)
- Kagan Yu, Svistunov B V Zh. Eksp. Teor. Fiz. 105 353 (1994) [JETP 78 187 (1994)]
- 7. Kagan Yu (2008) (in press)

PACS numbers: 05.30.Fk, **71.10.**–w, 74.20.Fg DOI: 10.1070/PU2008v051n06ABEH006548 DOI: 10.3367/UFNr.0178.200806i.0633

Superfluid Fermi liquid in a unitary regime

L P Pitaevskii

1. Introduction

When choosing the subject of my presentation at this session dedicated to the 100th anniversary of the birth of Landau, I wanted to speak about something that would have surprised Landau. I believe that the recently prepared physical object—a universal superfluid Fermi liquid—meets this requirement in the best way possible.

As is well known, Landau did not regard the microscopic theory of fluids as a problem worth being occupied with. I quote a well-known passage from *Statistical Physics* [1]: "In contrast to gases and solids, liquids do not permit calculating the thermodynamic quantities or at least their temperature dependences in the general form. The reason lies with the strong interaction between the molecules of a liquid and, at the same time, the absence of the smallness of oscillations, which imparts simplicity to the thermal motion in solids. Because of the high intensity of molecular interaction, the knowledge of a specific interaction law, which is different for different liquids, becomes significant for calculating thermodynamic quantities."

This statement is perfectly correct for all liquids existing in nature. However, progress in experimental techniques has recently enabled preparing liquids with properties independent of any quantities that characterize the interaction. This situation emerges because the interatomic interaction in these bodies is, in a sense, infinitely strong. The case in point is ultracold gases near the so-called Feshbach resonances. The liquids of interest to us are made from gases whose atoms obey the Fermi statistics. The gas is dilute in the sense that the average interatomic distance $n^{-1/3}$, where *n* is the atomic number density, is much greater than the characteristic range r_0 of interatomic forces:

$$r_0 \ll n^{-1/3}$$
. (1)

Condition (1) is always satisfied for the objects under consideration. However, the fulfillment of this condition does not yet signify that we are dealing with a gas in the sense that the interaction is weak. Let the temperature be sufficiently low, such that the gas is degenerate, $T \leq E_{\rm F}$.¹ It is then valid to say that all body properties depend on one parameter *f*, the amplitude of the scattering of atoms with the orbital momentum l = 0 by each other. The interaction is weak, i.e., the body is indeed a gas, if the amplitude is small in comparison with the interatomic distances:

$$|f| \ll n^{-1/3} \,. \tag{2}$$

The quantities r_0 and |f| are typically of the same order of magnitude and conditions (1) and (2) are practically equivalent. However, this is not the case when a system of two atoms has an energy level close to zero. According to the general scattering theory, the scattering amplitude is then expressed in the form (see, e.g., Landau and Lifshitz [2]) f(k) = $-(a^{-1}+ik)^{-1}$, where k is the wave vector and a = -f(0) is the scattering length, a constant that characterizes the scattering completely. When a > 0, the system of two atoms has a bound state with the negative energy $\epsilon = -\hbar^2/ma^2$. When a < 0, the system is said to have a virtual level. If |a| is high enough, $|a| \ge k^{-1} \sim n^{-1/3}$, the interaction weakness condition (2) is certainly violated and we are by definition dealing with a liquid, although a dilute liquid in the sense of condition (1). In this case, its properties are characterized by the sole parameter a. When $|a| \ge k^{-1}$, the scattering amplitude reaches its 'unitary limit' $f \approx i/k$. The length a then drops out of the theory and we are dealing with a universal liquid, whose properties do not depend on the interaction at all. Of course, the picture under discussion implies the possibility of changing the scattering amplitude. This opportunity arises in the presence of Feshbach resonances, in the vicinity of which the position of the energy level of the system of two atoms depends on the magnetic field [3]. The scattering length as a function of the magnetic field can be represented as

$$a = a_{\rm bg} \left(1 - \frac{\Delta_B}{B - B_0} \right). \tag{3}$$

Near the resonance $B \approx B_0$, the scattering length is large and the system is a universal liquid.

We qualitatively consider the properties of the system at T = 0 in different ranges of the scattering length *a*. When this length is positive and relatively small, $r_0 \ll a \ll n^{-1/3}$, the system of two atoms has a bound state and the atoms combine to form molecules with a binding energy ϵ . The system is a

¹ In all formulas, we set $k_{\rm B} = 1$.

Bose gas consisting of weakly bound diatomic molecules, or dimers. It is significant that the dimer-dimer scattering length a_{dd} is positive, i.e., these molecules experience mutual repulsion. Calculating a_{dd} is an intricate problem, which was solved in [4]. It turned out that $a_{dd} = 0.6a$. Therefore, in this regime, the system is a weakly nonideal superfluid Bose gas described by the Bogolyubov theory [6], with the obvious change $m \rightarrow 2m$, $a \rightarrow 0.6a$.

The question of the lifetime of this system is of paramount importance for the entire area of physics involved. This lifetime is limited by transitions from the weakly bound level to deep molecular levels in molecular collisions accompanied by the release of a large amount of energy. The molecule number loss in these inelastic processes is described by the equation $\dot{n}_{\rm d} = -\alpha_{\rm dd} n_{\rm d}^2$. The dependence of the recombination coefficient α_{dd} on *a* was also studied in Ref. [4]. It turned out that $\alpha_{dd} \propto a^{-2.25}$. Therefore, the system becomes more stable with an increase in the scattering length, i.e., as the resonance is approached. This paradoxical result stems from the Fermi nature of atoms or, to be more precise, from the fact that fermions with parallel spins cannot reside at the same point. In a Bose gas, which was also studied in experiments, the lifetime decreases sharply as the resonance is approached. This is the reason why only the Fermi liquid can actually be investigated in the unitary mode. The experimentally measured dependence of α_{dd} on *a* is depicted in Fig. 1. It is in satisfactory agreement with the theory.

We now consider the opposite limit case, where the scattering length is negative and small in modulus, a < 0, $r_0 \ll |a| \ll n^{-1/3}$, as is the case on the opposite side of the resonance. The system is then a weakly nonideal Fermi gas with attraction between the atoms. According to the theoretical concepts of Bardeen–Cooper–Schrieffer and Bogolyubov, the occurrence of a Fermi surface gives rise to Cooper pairs in this case. As a result, a gap appears in the fermion energy spectrum and the system becomes superfluid.

In the immediate vicinity of the resonance, the system is a universal unitary Fermi liquid. Because the system is superfluid in both limit cases considered, it is reasonable to assume that it is superfluid in all of the interval of *a* values. (Different arguments are presented below.) Of course, the system is then assumed to be stable in the unitary mode. This assumption is supported by the wealth of experimental data and theoretical calculations.



Figure 1. The dimer recombination coefficient α_{dd} as a function of the scattering length *a* (borrowed from Ref. [5]). The slope of the dashed line corresponds to the theoretical dependence $\alpha_{dd} \propto a^{-2.5}$.





Figure 2. (a) Schematic of the facility employed at Duke University to investigate the properties of a Fermi gas in an optical trap near a Feshbach resonance (borrowed from [7]). (b) Schematic of the facility used at MIT for investigating the rotation of a superfluid Fermi gas (borrowed from [8]). Two laser beams aligned with the axis set the gas in rotation. Separately shown is the scheme for observing the vortices from the shadowgraph of the expanding fermionic cloud.

Prior to discussing these results, I briefly describe the typical experimental arrangement using the example of a facility at Duke University [7] (Fig. 2a). Two types of Fermi atoms were actually used in the experiments, ⁶Li and ⁴⁰K isotopes. The isotope choice was dictated by the presence of a Feshbach resonance in a convenient range of the magnetic field and the occurrence of spectral lines in a convenient wavelength range. The atoms are confined in an optical trap formed by a focused laser beam. The chosen light frequency is somewhat lower than the absorption line frequency, and therefore the atoms are 'attracted' to the intensity peak. Because the intensity near the focus decreases rapidly in the radial direction and slowly in the axial direction, the sample

was elongated and cigar-shaped. Solenoids induce the magnetic field required to attain the resonance. Since the main objective of the experiments was to investigate superfluidity, two types of fermions were needed. In superconductivity theory, electrons with opposite values of spin projection are usually considered. In our case, atoms in different hyperfine structure states were used.

Experiments with fermions are arduous and the number of groups working with them is smaller than the number of groups investigating the Bose–Einstein condensation. The work is undertaken at the JILA (Joint Research Institute of the National Institute of Standards and Technology and the University of Colorado) (Boulder), Massachusetts Institute of Technology (MIT) (Boston), Duke University (Durham), and Rice University (Houston) in the USA, the École Normale Supérieure (Paris) in France, and the University of Innsbruck in Austria. It is a pleasure for me to mention that A Turlapov, one of the leading experimenters at Duke University, has returned to Nizhnii Novgorod and is making a facility there.

I give the typical parameters of recent experiments. The number of atoms in the trap is $N \sim 3 \times 10^6 - 10^7$ and the atom density at its center is $n \sim 2 \times 10^{12}$. Accordingly, the Fermi energy is $E_{\rm F} \sim 200-500$ nK and the magnitude of the Fermi wave vector is $k_{\rm F} \sim 0.3 \ \mu {\rm m}^{-1}$. The parameters of the trap are conveniently characterized by the frequencies of atomic oscillations in it. The radial frequency v_{\perp} normally lies in the 60-300 Hz range and the longitudinal frequency $v_z \sim 20$ Hz. The lowest attainable temperature turns out to be under $0.06E_{\rm F}$, i.e., of the order of 10 nK. As is evident from the subsequent discussion, it has been possible not only to conduct experiments at these prodigiously low temperatures but also to set up a thermodynamic temperature scale in this domain. I cannot enlarge on the techniques of gas cooling, and only mention that during the final stage, the gas is cooled due to the evaporation of the faster atoms from the trap, much like tea is cooled in a cup left on a table.

One of the most important experimental tasks was to ascertain that the system was superfluid. An immanent property of superfluidity is the existence of quantized vortices. The velocity circulation around a vortex in a Fermi liquid is $\Gamma = \pi \hbar/m$, two times smaller than in a Bose liquid. Accordingly, in the rotation with a sufficiently high angular velocity Ω , the number of vortices per unit area must be equal to $2\Omega m/(\pi\hbar)$. How can the liquid be set in rotation? MIT experimenters positioned a pair of thin laser beams along the trap axis, which were shifted from the axis (Fig. 2b) [8]. This 'mixer' rotated about the axis and entrained the liquid. At some instant, the trap was disengaged, the liquid expanded, and observations of the density distribution were made. The result is shown in Fig. 3. The vortex cores are observed as dark reduced-density domains. A simple calculation of the number of vortices confirms the theoretical value of the circulation given above.

We now consider the liquid precisely at the resonance point, when $a \to \pm \infty$. (It is pertinent to note that this is not a phase transition point.) We begin from the properties of a uniform liquid at T = 0. Apart from the density, there are no parameters at our disposal on which the thermodynamic functions may depend. Dimensionality considerations suggest, e.g., that the chemical potential of the liquid must be of the form

$$\mu(n) = \xi \mu^{\rm id}(n) \,, \tag{4}$$



Figure 3. Quantized vortices in a rotating superfluid Fermi gas (borrowed from [8]): (a) corresponds to a dilute gas of dimers, (b) to a Fermi liquid in the vicinity of the unitarity point, and (c) to a dilute Fermi gas with a weak attraction between the atoms.

where $\mu^{id}(n) = (3\pi^2 n)^{2/3} (\hbar^2/m)$ is the chemical potential of an ideal Fermi gas with the density n for T = 0 and ξ is a dimensionless coefficient independent of the kind of liquid. The theoretical task consists in the calculation of ξ and the experimental task involves its measurement. The first estimates of ξ were made proceeding from the Bardeen-Cooper-Schrieffer-Bogolyubov (BCSB) theory. This theory is a mean-field theory and, needless to say, is inapplicable near the unitarity point. But its ingenious generalization to the strong-coupling case has allowed obtaining formulas sound in both limit cases (see, e.g., [9]). At exactly the unitarity point, this theory yields $\xi = 0.59$. The most reliable result is provided by calculations involving the quantum Monte Carlo (QMC) technique: $\xi = 0.42$ [10]. It is noteworthy that the absence of a small parameter in the theory is substantially favorable to numerical calculations. It is not infrequent that the existence of such a parameter impairs convergence. An attempt has been made to apply the εexpansion technique, which relies on the fact that $\xi = 0$ in a four-dimensional space [11]. The theory is constructed in the space of $D = 4 - \varepsilon$ dimensions under the assumption that ε is small, and the results are then extrapolated to $\varepsilon = 1$. This technique, which is highly beneficial in the theory of phase transitions, supposedly yields poor accuracy in this case. It is significant that the parameter $\xi < 1$. This signifies that the interaction at the unitarity point lowers the fluid pressure, i.e., is an effective attraction. It is therefore reasonable that it leads to fermion pairing and to superfluidity. A quantitative characteristic of the pairing is the gap \varDelta in the Fermi branch of the spectrum. Once again, the dimensionality considerations suggest that

$$\Delta(n) = \theta \mu^{id}(n) \,. \tag{5}$$

QMC calculations yield $\theta = 0.5$ [10].

We now turn our attention to the experimental verification of the theory. The most direct method of determining ξ consists in the precise measurement of fluid density in the trap. In the semiclassical approximation, this distribution is given, in view of expression (4), by the equation $\xi \mu^{id}[n(\mathbf{x})] + V(\mathbf{x}) = \text{const.}$ Fitting to the observed distribution allows determining ξ . At Rice University, the value $\xi = 0.46$ was thus found for ⁶Li [12]. Another method was applied by experimenters at JILA, who worked with ⁴⁰K. They measured the density distribution and calculated the potential energy $U_{\text{pot}} = \int n(\mathbf{x}) V(\mathbf{x}) d\mathbf{x}$ of the liquid, which is proportional to $\sqrt{\xi}$ [13]. By this means, they obtained the value $\xi = 0.46$. The proximity of the values for ⁶Li and ⁴⁰K to the theoretical one confirms the universal nature of ξ . Reliable measurements of the gap Δ , in my opinion, have not been made to the present day.

Important information about the properties of the liquid may be obtained by investigating its oscillations in the trap. These oscillations are described by the Landau superfluid hydrodynamics [14]. (I emphasize that Landau believed from the outset that his equations applied both to Bose and Fermi superfluid liquids.) An especially simple result for the oscillation frequencies in a harmonic trap is obtained for a liquid with the polytropic equation of state $\mu(n) \propto n^{\gamma}$. We consider an important type of oscillation: axially symmetric radial oscillations whose frequency is $\omega = \sqrt{2(\gamma + 1)} \omega_{\perp}$ [15]. According to this formula, in the molecular limit (a > 0, $na^3 \ll 1$), when $\mu \propto a_{dd}n$, i.e., $\gamma = 1$, the frequency $\omega = 2\omega_{\perp}$. In the unitary limit and BCSB limit, $\gamma = 2/3$ as in an ideal Fermi gas and $\omega = \sqrt{10/3}\omega_{\perp} = 1.83 \omega_{\perp}$. For intermediate values of *a*, the frequency cannot be calculated analytically, but it appears reasonable that the frequency for a > 0 is monotonically decreasing with increasing a. These were precisely the indications of the first experiments. Theories that have this property and rely on the mean-field approximation have also been proposed.

However, the situation is not that simple. For $na^3 \ll 1$, the theory permits rigorous calculations of not only the first term in μ but also a correction, which was first determined in [16]. This gives a correction to the frequency equal to [17]

$$\frac{\delta\omega}{\omega} = +0.72\sqrt{n(\mathbf{x}=0)\,a_{\rm dd}^3}\,.\tag{6}$$

The positive correction sign signifies that the frequency must initially increase with increasing a and only then decrease to attain the limit value $1.83\omega_{\perp}$. This reasoning was disputed on the grounds that molecular dimers are nevertheless not entirely bosons. However, correction (6) bears a clear physical meaning. It stems from the contribution to the energy made by zero-point phonon oscillations, whose occurrence in the superfluid liquid is beyond question. This is why it is anomalously large, of the order of the square root of the gas parameter na^3 , while the 'normal' expansion is performed in this parameter. All this leads us to the statement that the author has been vigorously promoting, namely, that the monotonic behavior of the frequency would imply a catastrophe for the theory. Fortunately, the situation has recently been clarified. New experiments do yield above- $2\omega_{\perp}$ values of the frequency on the molecular side of the resonance. They are in good agreement with the calculations throughout the interval of a values performed by the QMC technique [19] (Fig. 4). We note that the disagreement with the data of previous experiments is attributable to the fact that the temperature in those experiments was not low enough. Meanwhile, correction (6) is temperature sensitive, because it is related to the excitation of relatively low energies $\hbar \omega \sim \mu$.

We now discuss the fluid properties at the unitarity point at finite temperatures. In this case, the temperature is assumed to be not too high, and therefore the wavelength of atoms in their thermal motion is long in comparison with the atomic size: $r_0 \ll \hbar/\sqrt{mT}$. We are actually dealing with temperatures of the order of $E_{\rm F}$.

The question of the temperature of transition to the superfluid state is all-important here. The most reliable data were obtained in [20] using the Monte Carlo technique:



Figure 4. (a) Frequency of radial oscillations as a function of the scattering length. The upper solid curve represents the data calculated by the quantum Monte Carlo technique and the lower one is the result of calculations by the mean-field theory. The points stand for experimental values. The upper and lower dashed straight lines show the limit frequency values in a tenuous dimer gas and the unitarity point. (b) Measured values of oscillation damping. (Borrowed from [18].)

 $T_{\rm c} = 0.16 \mu^{\rm id}$. This result is in good agreement with experiment. It is noteworthy that the transition temperature is relatively low, and hence at temperatures somewhat higher than $T_{\rm c}$, we face an interesting research object — a degenerate normal Fermi liquid in the unitary regime.

At finite temperatures, the equation of state cannot be written proceeding from only the dimensionality considerations. But these considerations lead to important similarity relations. For instance, the chemical potential is of the form $\mu(n, T) = \mu^{id}(n) f_{\mu}[T/\mu^{id}(n)]$; the entropy per atom can be written as $s(n, T) = f_S[T/\mu^{id}(n)]$. The last relation implies that under an adiabatic density variation, the temperature varies as $\propto n^{2/3}$, as in an ideal monoatomic gas.

For a fluid in the trap, these formulas lead to an important integral relation. Following the standard derivation of the virial theorem, it can be shown that

$$2U_{\rm pot} = E\,,\tag{7}$$

where U_{pot} is the potential energy and *E* is the total energy, i.e., the sum of potential, internal, and hydrodynamic kinetic energies [21]. As indicated above, U_{pot} can be calculated directly from the measured density distribution. The total energy may be changed in a controllable way. For this, the trap potential was switched off for some 'heating time' t_{heat} . During this period, the liquid was free to expand. The sum of the kinetic and internal energies was conserved in the process. Then, the trap was turned on again and the system came to equilibrium, and its potential energy, which was measured anew, turned out to be higher. This ingenious method enabled the authors of [21] to verify relation (7) with high precision and thus confirm the similarity laws formulated above.

The total entropy $S = \int n(\mathbf{x}) s(\mathbf{x}) d\mathbf{x}$ of the system as a function of its energy *E* was measured in a similar experiment in [22]. In the experiment, the energy was varied and measured as described above; to measure the entropy, the magnetic field was adiabatically increased, taking the system away from resonance, where the interaction was insignificant. Measure-

ments of the cloud dimension enabled calculating the entropy from the formulas for an ideal Fermi gas, which, due to the adiabaticity of the process, was equal to the entropy of the liquid before the increase in the magnetic field. It is noteworthy that the derivative T = dE/dS directly yields the absolute temperature of the system. I believe that the capability of measuring the absolute temperature in the nanokelvin domain is a wonderful achievement by itself. Another way of measuring the absolute temperature is described below.

The aforesaid leaves no room for doubt that the theoretical notions about the properties of a 'unitary' superfluid liquid are amply borne out by experiments. I believe, however, that the significance of the issue calls for highprecision verification. Such a possibility does exist. For this, the fluid should be placed in a trap that is harmonic and isotropic with a high degree of accuracy. Then, we can state with certainty that the spherically symmetric cloud pulsations are precisely equal to $2\omega_h$ in frequency, where ω_h is the eigenfrequency of the trap, and do not attenuate [23]. This theorem is valid both below and above the superfluid transition point and applies to oscillations of arbitrary amplitude. It is a corollary of the hidden symmetry of the system at the unitarity point. (A similar situation occurs for oscillations of a dilute Bose gas in a cylindrical trap [24, 25].) The absence of damping signifies that the second viscosity ζ of the fluid is equal to zero above the transition point. Of the three second viscosity coefficients introduced by Khalatnikov [26], ζ_1 and ζ_2 turn out to be zero [27] in the superfluid phase.

So far, we have dealt with experiments in which the numbers of atoms in two spin states were equal. Recently, active work commenced to study polarized systems in which the number of atoms in one spin state (we conventionally speak of 'spin-up' atoms) is greater than in the other state. This question had already been discussed for superconductors. In [28] and [29], the existence of spatially inhomogeneous phases (LOFF phases) was predicted, in which the superconducting gap is a periodic function of coordinates [28, 29]. In superconductors, the population difference of the spin states may exist in ferromagnetic bodies or may be induced by an external magnetic field. In both cases, the magnetic field affects the orbital motion and destroys superconductivity.

In our neutral dilute systems, the spin relaxation time is quite long and the numbers of atoms in different states are practically arbitrary parameters, determined by the initial conditions. Theoretical calculations in [30] and the experiment in [31] show that the liquid in a trap at T = 0 near the unitarity domain breaks up into three phases. At the center is the superfluid phase with equal numbers of 'spin-up' and 'spin-down' atoms. It is surrounded by the partially polarized normal phase with unequal densities of the atoms of different polarization. At the periphery is the completely polarized phase, which consists of only the atoms of excess polarization. In this case, the existence of LOFF-type phases in some parameter value ranges is not ruled out.

The measurement data are depicted in Fig. 5. The system with $N_{\uparrow} = 5.9 \times 10^6$, $T/E_{\rm F} = 0.03$ and the spin state population ratio $N_{\downarrow}/N_{\uparrow} = 0.39$ was investigated. Figure 5a shows the shadowgraph of the two-dimensional polarization distribution (the column density) $\delta n_{\rm a}(x, z) \equiv \int dy [n_{\uparrow}(\mathbf{r}) - n_{\downarrow}(\mathbf{r})]$ and Fig. 5b shows the 'weighted' distribution $\delta n_{\rm b}(x, z) \equiv \int dy [0.76n_{\uparrow}(\mathbf{r}) - 1.43n_{\downarrow}(\mathbf{r})]$, which gives a higher-contrast picture. Figure 5c shows the curves $\delta n_{\rm a}(0, z)$ and $\delta n_{\rm b}(0, z)$, and Figs 5d and 5e the plots of integrated linear densities



Figure 5. Polarization distribution in a system with unequal spin state populations (borrowed from [31]). (a) Shadowgraph of the two-dimensional polarization distribution $\delta n_a(x,z)$. (b) Weighted polarization distribution pattern $\delta n_b(x,z)$. (c) Plots of the functions $\delta n_a(0,z)$ (the upper curve) and $\delta n_b(0,z)$ (the lower curve). (d) Linear polarization density $\delta n_a(z)$ along the z axis. (e) Linear polarization density $\delta n_a(x)$ along the x axis.

 $\delta n_{\rm a}(z) \equiv \int dx n_{\rm a}(x,z)$ and $\delta n_{\rm a}(x) \equiv \int dz n_{\rm a}(x,z)$. I emphasize that the measurements were made in the trap itself, without prior expansion of the fluid. Processing the measured twodimensional distribution by the Abel transform enabled reconstructing the three-dimensional polarization distribution and confirmed the three-phase fluid structure. Measurements at different temperatures were also made. Worthy of note in this connection is the special role played by the completely polarized phase. Because slow fermions with parallel spins hardly interact with each other, this phase is an ideal Fermi gas. By measuring the density distribution of this phase and fitting it to formulas for the ideal gas, it is possible to determine the thermodynamic temperature of the system. The polarized phase plays the role of an ideal-gas thermometer contacting with other phases. It is significant in this case that the fermions of the polarized phase interact with the fermions of other phases, which ensures thermodynamic equilibrium. These temperature measurements permitted verifying the transition temperature calculated in Ref. [20]. The results under discussion are at some variance with the findings in [12], where a smaller number of atoms was considered. Conceivably, the surface tension at the phase boundaries plays a role under these conditions.

I mention several interesting possibilities for future investigations. One of them involves employing two types of fermions of different masses for which the Feshbach resonance exists [32]. Theory predicts unconventional properties for a superfluid liquid formed as a result of Cooper pairing of the fermions of different masses.

Another possibility is related to the vortex-free rotation of a Fermi liquid [33]. The vortex lattice shown in Fig. 3 is formed due to a strong fluid perturbation by the rotating mixers. If a trap asymmetric about the axis is simply set in rotation, there are grounds to believe that vortices would be formed only for a high rotation rate, when the fluid shape becomes unstable. At lower rotation rates, the fluid would

break up into two phases. The center of the weakly deformed trap would be occupied by the superfluid liquid at rest, while the normal phase of the liquid would rotate in the usual way at the periphery. The existence of the normal phase at absolute zero kept by rotation from transiting into the superfluid state raises difficult theoretical issues.

A very rich area of research opens up when the fluid is placed in a periodic lattice produced by counterpropagating laser beams (see the author's review Ref. [34]). This research in the unitary domain is still in its infancy.

We see that the investigations of a near-resonance Fermi gas in a trap have opened up entirely new theoretical and experimental opportunities in condensed matter physics, reflecting the modern trend. Work to an increasing extent is shifting to the investigation of specially fabricated objects that do not exist in nature and have surprising new properties. In view of this, I believe, no exhaustion of our realm of physics is to be expected in the foreseeable future.

Acknowledgements. I express my appreciation to S Stringari for discussions, to J Thomas for providing the original of Fig. 2a, and to R Grimm for providing the original of Fig. 4.

References

- Landau L D, Lifshitz E M Statisticheskaya Fizika (Statistical 1. Physics) (Moscow: Gostekhteorizdat, 1951) p. 223 [Translated into English: (Oxford: Pergamon Press, 1969)]
- Landau L D, Lifshitz E M Kvantovaya Mekhanika: Nerelyativists-2. kaya Teoriya (Quantum Mechanics: Non-Relativistic Theory) (Moscow: Fizmatlit, 2001) [Translated into English (Oxford: Pergamon Press, 1977)]
- 3. Feshbach H Ann. Phys. (New York) 5 357 (1958); 19 287 (1962)
- Petrov D S, Salomon C, Shlyapnikov G V Phys. Rev. Lett. 93 4. 090404 (2004)
- 5 Petrov D S, Salomon C, Shlyapnikov G V J. Phys. B: At. Mol. Opt. Phys. 38 \$645 (2005)
- 6. Bogolyubov N N Izv. Akad. Nauk SSSR, Ser. Fiz. 11 77 (1947)
- Thomas J E, Gehm M E Am. Scientist 92 238 (2004) 7.
- 8. Zwierlein M W et al. Nature 435 1047 (2005)
- 9. Nozières P, Schmitt-Rink S J. Low Temp. Phys. 59 195 (1985)
- Carlson J, Reddy S Phys. Rev. Lett. 95 060401 (2005) 10
- Nishida Y, Son D T Phys. Rev. Lett. 97 050403 (2006) 11.
- 12. Patridge G B et al. Science 311 503 (2006)
- 13. Stewart J T et al. Phys. Rev. Lett. 97 220406 (2006)
- Landau L D Zh. Eksp. Teor. Fiz. 11 592 (1941) 14.
- 15. Cozzini M, Stringari S Phys. Rev. Lett. 91 070401 (2003)
- Lee T D, Huang K, Yang C N Phys. Rev. 106 1135 (1957) 16
- 17.
- Pitaevskii L, Stringari S Phys. Rev. Lett. 81 4541 (1998) 18 Altmeyer A et al. Phys. Rev. Lett. 98 040401 (2007)
- 19.
- Astrakharchik G E et al. Phys. Rev. Lett. 95 030404 (2005)
- Burovski E et al. New J. Phys. 8 153 (2006) 20.
- 21. Thomas J E, Kinast J, Turlapov A Phys. Rev. Lett. 95 120402 (2005)
- Luo L et al. Phys. Rev. Lett. 98 080402 (2007) 22
- 23. Castin Y C.R. Physique 5 407 (2004)
- 24. Pitaevskii L P Phys. Lett. A 221 14 (1996)
- Kagan Yu, Surkov E L, Shlyapnikov G V Phys. Rev. A 54 R1753 25. (1996)
- 26. Khalatnikov I M Teoriya Sverkhtekuchesti (Theory of Superfluidity) (Moscow: Nauka, 1971)
- 27. Son D T Phys. Rev. Lett. 98 020604 (2007)
- 28 Larkin A I, Ovchinnikov Yu N Zh. Eksp. Teor. Fiz. 47 1136 (1964) [Sov. Phys. JETP 20 762 (1965)]
- Fulde P, Ferrell R A Phys. Rev. 135 A550 (1964) 29
- 30. Lobo C et al. Phys. Rev. Lett. 97 200403 (2006)
- 31. Shin Y et al., arXiv:0709.3027
- 32. Petrov D S Phys. Rev. A 67 010703 (2003)
- 33. Bausmerth I, Recati A, Stringari S Phys. Rev. Lett. 100 070401 (2008); arXiv:0711.0653
- 34. Pitaevskii L P Usp. Fiz. Nauk 176 345 (2006) [Phys. Usp. 49 333 (2006)]

PACS numbers: 02.20.Sv, **04.20.-q**, **98.80.-k** DOI: 10.1070/PU2008v051n06ABEH006550 DOI: 10.3367/UFNr.0178.200806j.0639

Lev Landau and the problem of singularities in cosmology

I M Khalatnikov, A Yu Kamenshchik

1. Introduction

We consider different aspects of the problem of cosmological singularity such as the BKL oscillatory approach to singularity, the new features of cosmological dynamics in the neighborhood of the singularity in multidimensional and superstring cosmological models, and their connections with such a modern branch of mathematics as infinite-dimensional Lie algebras. In addition, we consider some new types of cosmological singularities that have been widely discussed during the last decade, after the discovery of the phenomenon of cosmic acceleration.

Many years ago, in conversations with his students, Lev Davidovich Landau used to say that three problems were most important for theoretical physics: the problem of cosmological singularity, the problem of phase transitions, and the problem of superconductivity [1]. We now know that the great breakthrough was achieved in the explanation of phenomena of superconductivity [2] and phase transitions [3]. The problem of cosmological singularity has been widely studied during the last 50 years and many important results have been obtained, but it still preserves some intriguing aspects. Moreover, some quite unexpected facets of the problem of cosmological singularity have been discovered.

In our review published 10 years ago [4] in the issue of this journal dedicated to the 90th anniversary of Landau's birth, we discussed some questions connected with the problem of singularity in cosmology. In the present paper, we dwell on relations between well-known old results of these studies and new developments in this area.

To begin, we recall that Penrose and Hawking [5] proved the impossibility of indefinite continuation of geodesics under certain conditions. This was interpreted as pointing to the existence of a singularity in the general solution of the Einstein equations. These theorems, however, did not allow finding the particular analytic structure of the singularity. The analytic behavior of the general solutions of the Einstein equations in the neighborhood of a singularity was investigated in [6-11]. These papers revealed the enigmatic phenomenon of an oscillatory approach to the singularity, which has become known as the Mixmaster Universe [12]. The model of a closed homogeneous but anisotropic universe with three degrees of freedom (Bianchi IX cosmological model) was used to demonstrate that the universe approaches the singularity such that its contraction along two axes is accompanied by an expansion along the third axis, and the axes change their roles according to a rather complicated law revealing chaotic behavior [10, 11, 13, 14].

The study of the dynamics of the universe in the vicinity of the cosmological singularity has become an explodingly developing field of modern theoretical and mathematical physics. We first note the generalization of the study of the oscillatory approach to the cosmological singularity in multidimensional cosmological models. It has been noticed [15] that the approach to the cosmological singularity in the multidimensional (Kaluza-Klein type) cosmological models has a chaotic character in space-times whose dimension is not higher than ten, while in space-times of higher dimensions, the universe enters a monotonic Kasner-type contracting regime after undergoing a finite number of oscillations.

The development of cosmological studies based on superstring models has revealed some new aspects of the dynamics in the vicinity of the singularity [17]. First, it was shown that these models involve mechanisms for changing Kasner epochs provoked not by the gravitational interactions but by the influence of other fields present in these theories. Second, it was proved that the cosmological models based on the six main superstring models plus the D = 11 supergravity model exhibit a chaotic oscillatory approach towards the singularity. Third, the connection between cosmological models manifesting the oscillatory approach towards singularity and a special subclass of infinite-dimensional Lie algebras [18], the so called hyperbolic Kac-Moody algebras, was discovered.

Another confirmation of the importance of the problem of singularity in general relativity has come from observational cosmology. At the end of the 1990s, the study of the relation between the luminosity and redshift of type-Ia supernovae revealed that the modern Universe is expanding with an acceleration [19]. To provide such an acceleration, it is necessary to have a particular substance, which was named 'dark energy' [20]. The main feature of this kind of matter is that it should have a negative pressure p such that $\rho + 3p < 0$, where ρ is the energy density. The simplest kind of this matter is the cosmological constant, for which $p = -\rho$. The so-called standard or Λ CDM cosmological model is based on the cosmological constant, whose energy density is responsible for roughly 70 percents of the general energy density of the Universe, while the rest is occupied by dust-like matter, both the baryonic one (approximately 4 percent) and a dark one. This model is in good agreement with observations, but other candidates for the role of dark energy are being intensively studied and new observations can give some surprises already in the nearest future. First of all, we note that some observations [21] suggest the possible existence of the socalled superacceleration, which is connected with the presence of phantom dark energy [22], characterized by the inequality $p < -\rho$. Under certain conditions, a universe filled with this type of dark energy can encounter a very particular cosmological singularity, the Big Rip [23]. When an expanding universe encounters this singularity, it has an infinite cosmological radius and an infinite value of the Hubble variable. Earlier, the possibility of this type of singularity was discussed in [24].

Study of different possible candidates for the role of dark energy has stimulated the elaboration of the general theory of possible cosmological singularities [25-29]. It is remarkable that while the 'traditional' Big Bang or Big Crunch singularities are associated with the vanishing size of the universe, i.e., with a universe squeezed to a point, these new singularities occur at finite or infinite value of the cosmological radius. The physical processes occurring in the vicinity of such singularities can have rather exotic features and their study is of great interest. Thus, we see that the development of both the theoretical and the observational branches of cosmology has confirmed the importance of the problem of singularity in general relativity mentioned by Landau many years ago.

The structure of this contribution is as follows: in Section 2, we briefly discuss the Landau theorem about the singularity, which was not published in a separate paper and was reported in book [30] and in review [6]; in Section 3, we recall the main features of the oscillatory approach to the singularity in relativistic cosmology; Section 4 is devoted to the modern development of the BKL ideas and methods, including the dynamics in the presence of a massless scalar field, multidimensional cosmology, superstring cosmology, and the correspondence between chaotic cosmological dynamics and hyperbolic Kac–Moody algebras; in Section 5, we describe some new types of cosmological singularities, and in Section 6 we present some concluding remarks.

2. Landau theorem about singularity

We consider the synchronous reference frame with the metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \gamma_{\alpha\beta} \,\mathrm{d}x^{\,\alpha} \,\mathrm{d}x^{\,\beta}\,,\tag{1}$$

where $\gamma_{\alpha\beta}$ is the spatial metric. Landau pointed out that the determinant *g* of the metric tensor in a synchronous reference system must tend to zero at some finite time if the equation of state satisfies some simple conditions. To prove this statement, it is convenient to write the 0-0 component of the Ricci tensor as

$$R_0^0 = -\frac{1}{2} \frac{\partial K_\alpha^\alpha}{\partial t} - \frac{1}{4} K_\alpha^\beta K_\beta^\alpha , \qquad (2)$$

where $K_{\alpha\beta}$ is the extrinsic curvature tensor defined as

$$K_{\alpha\beta} = \frac{\mathrm{d}\gamma_{\alpha\beta}}{\mathrm{d}t}\,,\tag{3}$$

and the spatial indices are raised and lowered by the spatial metric $\gamma_{\alpha\beta}$. The Einstein equation for R_0^0 is

$$R_0^0 = T_0^0 - \frac{1}{2} T, (4)$$

where the energy-momentum tensor is

$$T_i^j = (\rho + p) u_i u^j - \delta_i^j p, \qquad (5)$$

where ρ , *p*, and *u_i* are the energy density, the pressure, and the four-velocity, respectively. The quantity $T_0^0 - 1/2T$ in the right-hand side of Eqn (4) is

$$T_0^0 - \frac{1}{2} T = \frac{1}{2} (\rho + 3p) + (\rho + p) u_{\alpha} u^{\alpha}, \qquad (6)$$

which is positive if

$$\rho + 3p > 0. \tag{7}$$

Thus, Eqn (4) implies that

$$\frac{1}{2} \frac{\partial K_{\alpha}^{\alpha}}{\partial t} + \frac{1}{4} K_{\alpha}^{\beta} K_{\beta}^{\alpha} \leqslant 0.$$
(8)

Because of the algebraic inequality

$$K^{\beta}_{\alpha}K^{\alpha}_{\beta} \ge \frac{1}{3} \left(K^{\alpha}_{\alpha}\right)^2,\tag{9}$$

we have

$$\frac{\partial K_{\alpha}^{\alpha}}{\partial t} + \frac{1}{6} \left(K_{\alpha}^{\alpha} \right)^2 \leqslant 0 \,, \tag{10}$$

or

$$\frac{\partial}{\partial t} \frac{1}{K_{\alpha}^{\alpha}} \ge \frac{1}{6} \,. \tag{11}$$

If $K_{\alpha}^{\alpha} > 0$ at some instant of time, then if *t* decreases, the quantity $1/K_{\alpha}^{\alpha}$ decreases to zero within a finite time. Hence, K_{α}^{α} tends to $+\infty$; because of the identity

$$K_{\alpha}^{\alpha} = \gamma^{\alpha\beta} \, \frac{\partial \gamma_{\alpha\beta}}{\partial t} = \frac{\partial}{\partial t} \ln \gamma \,, \tag{12}$$

this means that the determinant g tends to zero [no faster than t^6 according to inequality (11)]. If $K_{\alpha}^{\alpha} < 0$ at the initial instant, then the same result is obtained for increasing time. A similar result was obtained in [31] in the case of dust-like matter and in [32].

This result does not prove that a true physical singularity inevitably exists that belongs to space – time itself and is not connected with the character of the chosen reference system. However, this result played an important role in stimulating the discussion about the existence and generality of singularities in cosmology. We note that energy dominance condition (7) used for the proof of the Landau theorem also appears in the proof of the Penrose and Hawking singularity theorem [5]. Moreover, the breakdown of this condition is necessary for an explanation of the phenomenon of cosmic acceleration.

The Landau theorem is deeply connected with the appearance of caustics studied in [33] and was discussed between those authors and Landau in 1961. In trying to geometrically construct the synchronous reference frame, one starts from the three-dimensional Cauchy surface and designs the family of geodesics orthogonal to this surface. The length along these geodesics serves as the time measure. It is known that these geodesics intersect on some two-dimensional caustic surface. This geometry constructed for the empty space is also valid in the presence of dust-like matter (p = 0). Such matter, moving along the geodesics, concentrates on the caustics, but the increase in density cannot be unbounded because the arising pressure destroys the caustics.¹ This question was studied in [34]. Later, caustics were used in [35] to explain the initial clustering of the dust, which, although not creating physical singularities, is nevertheless responsible for the creation of so-called pancakes. These pancakes represent the initial stage of the development of the largescale structure of the universe.

3. Oscillatory approach to the singularity in relativistic cosmology

One of the first exact solutions found in the framework of general relativity was the Kasner solution [16] for the Bianchi-I cosmological model representing the gravitational field in an empty space with a Euclidean metric depending on time according to the formula

$$ds^{2} = dt^{2} - t^{2p_{1}} dx^{2} - t^{2p_{2}} dy^{2} - t^{2p_{3}} dz^{2}, \qquad (13)$$

where the exponents p_1 , p_2 , and p_3 satisfy the relations

$$p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1.$$
 (14)

¹ In an empty space, the caustic is a mathematical but not a physical singularity. This follows simply from the fact that its location can always be shifted by changing the initial Cauchy surface.

Choosing the ordering of exponents as

$$p_1 < p_2 < p_3$$
, (15)

we can parameterize them as [6]

$$p_1 = \frac{-u}{1+u+u^2}, \ p_2 = \frac{1+u}{1+u+u^2}, \ p_3 = \frac{u(1+u)}{1+u+u^2}.$$
 (16)

As the parameter *u* varies in the range $u \ge 1$, p_1 , p_2 , and p_3 take all their allowed values:

$$-\frac{1}{3} \le p_1 \le 0, \quad 0 \le p_2 \le \frac{2}{3}, \quad \frac{2}{3} \le p_3 \le 1.$$
 (17)

The values u < 1 lead to the same range of values of p_1 , p_2 , and p_3 because

$$p_1\left(\frac{1}{u}\right) = p_1(u), \ p_2\left(\frac{1}{u}\right) = p_3(u), \ p_3\left(\frac{1}{u}\right) = p_2(u).$$
 (18)

The parameter u introduced in the early 1960s turned out to be very useful and its properties are attracting the attention of researchers in different contexts. For example, in recent paper [36], a connection was established between the Lifshitz-Khalatnikov parameter u and invariants arising in the context of Petrov's classification of the Einstein spaces [37].

In the case of Bianchi-VIII or Bianchi-IX cosmological models, the Kasner regime described by (13) and (14) is not an exact solution of the Einstein equations; however, a generalized Kasner solutions can be constructed [7-11]. It is possible to construct some kind of perturbation theory where the exact Kasner solution in (13), (14) plays the role of the zeroth-order approximation, with the role of perturbations played by the terms in the Einstein equations that depend on spatial curvature tensors (apparently, such terms are absent in Bianchi-I cosmology). This perturbation theory is effective in the vicinity of a singularity or, in other words, at $t \rightarrow 0$. The remarkable feature of these perturbations is that they imply a transition from the Kasner regime with one set of parameters to the Kasner regime with another one.

The metric of the generalized Kasner solution in a synchronous reference system can be written as

$$ds^{2} = dt^{2} - (a^{2}l_{\alpha}l_{\beta} + b^{2}m_{\alpha}m_{\beta} + c^{2}n_{\alpha}n_{\beta}) dx^{\alpha} dx^{\beta}, \quad (19)$$

where

$$a = t^{p_l}, \ b = t^{p_m}, \ c = t^{p_n}.$$
 (20)

The three-dimensional vectors **l**, **m**, and **n** define the directions along which the spatial distances vary with time according to power laws (20). Let $p_1 = p_1$, $p_m = p_2$, and $p_n = p_3$, with

$$a \sim t^{p_1}, \ b \sim t^{p_2}, \ c \sim t^{p_3},$$
 (21)

which means that the Universe is contracting in directions given by the vectors \mathbf{m} and \mathbf{n} and is expanding along \mathbf{l} . It was shown in [10] that the perturbations caused by spatial curvature terms make the variables a, b, and c undergo transition to another Kasner regime characterized by the formulas

$$a \sim t^{p'_l}, \ b \sim t^{p'_m}, \ c \sim t^{p'_n},$$
 (22)

where

$$p'_{l} = \frac{|p_{1}|}{1-2|p_{1}|}, \ p'_{m} = -\frac{2|p_{1}|-p_{2}}{1-2|p_{1}|}, \ p'_{n} = -\frac{p_{3}-2|p_{1}|}{1-2|p_{1}|}.$$
(23)

Thus, the effect of the perturbation is to replace one 'Kasner epoch' by another such that the negative power of t is transformed from the l to the m direction. During the transition, the function a(t) reaches a maximum and b(t) a minimum. Hence, the previously decreasing quantity b now increases, the quantity a decreases, and c(t) remains a decreasing function. The previously increasing perturbation caused the transition from regime (21) to regime (22), and is therefore damped and eventually vanishes. Then another perturbation begins to grow, which leads to a new replacement of one Kasner epoch by another, etc.

We emphasize that just the fact that perturbation implies a change of dynamics extinguishing it allows using the perturbation theory so successfully. We note that the effect of changing the Kasner regime already exists in cosmological models that are simpler than those of Bianchi IX and Bianchi VIII. As a matter of fact, in the Bianchi II universe, there exists only one type of perturbation connected with spatial curvature and this perturbation makes one change of Kasner regime (one bounce). This fact was known to Lifshitz and Khalatnikov in the early 1960s, and they discussed this topic with Landau (just before the tragic accident), who highly appreciated it. The results describing the dynamics of the Bianchi IX model were reported by Khalatnikov in his talk given in January 1968 at the Henri Poincaré Seminar in Paris. Wheeler, who was present there, pointed out that the dynamics of the Bianchi IX universe represent a nontrivial example of the chaotic dynamic system. Later, Thorn distributed a preprint with the text of this talk.

Returning to the rules governing the bouncing of the negative power of time from one direction to another, it can be shown that they can be conveniently expressed in terms of parameterization (16),

$$p_l = p_1(u), \ p_m = p_2(u), \ p_n = p_3(u),$$
 (24)

and then

$$p'_{l} = p_{2}(u-1), \ p'_{m} = p_{1}(u-1), \ p'_{n} = p_{3}(u-1).$$
 (25)

The greater of the two positive powers remains positive.

The successive changes as in (25), accompanied by a bouncing of the negative power between the directions I and **m**, continue until the integral part of u is exhausted, i.e., until u becomes less than one. Then, according to Eqn (18), the value u < 1 transforms into u > 1; at this instant, either the exponent p_l or p_m is negative and p_n becomes the smaller of the two positive numbers ($p_n = p_2$). The next sequence of changes bounces the negative power between the directions **n** and **I** or **n** and **m**. We emphasize that the Landau–Khalatnikov parameter u is useful because it allows encoding rather complicated laws of transitions between different Kasner regimes (23) by simple rules such as $u \to u - 1$ and $u \to 1/u$.

Consequently, the evolution of our model towards a singular point consists of successive periods (called eras) in which distances oscillate along two axes and decrease monotonically along the third axis, and the volume decreases according to a law that is near $\sim t$. In the transition from one era to another, the axes along which the distances decrease monotonically are interchanged. The order in which the pairs of axes are interchanged and the order in which eras of different lengths follow each other acquire a stochastic character.

To every (*sth*) era, there corresponds a decreasing sequence of values of the parameter *u*. This sequence has the form $u_{\max}^{(s)}, u_{\max}^{(s)} - 1, \ldots, u_{\min}^{(s)}$, where $u_{\min}^{(s)} < 1$. We introduce the notation

$$u_{\min}^{(s)} = x^{(s)}, \quad u_{\max}^{(s)} = k^{(s)} + x^{(s)},$$
 (26)

i.e., $k^{(s)} \equiv [u_{\max}^{(s)}]$ (the square brackets denote the greatest integer $\leq u_{\max}^{(s)}$). The number $k^{(s)}$ defines the era length. For the next era, we obtain

$$u_{\max}^{(s+1)} = \frac{1}{x^{(s)}}, \quad k^{(s+1)} = \left[\frac{1}{x^{(s)}}\right].$$
(27)

The ordering with respect to the length of $k^{(s)}$ of the successive eras (measured by the number of Kasner epochs contained in them) acquires an asymptotically stochastic character. The random nature of this process arises because of rules (26) and (27) that define the transitions from one era to another in the infinite sequence of values of u. If all this infinite sequence begins from some initial value $u_{\text{max}}^{(0)} = k^{(0)} + x^{(0)}$, then the lengths of series $k^{(0)}, k^{(1)}, \ldots$ are the numbers occurring in the expansion into a continued fraction:

$$k^{(0)} + x^{(0)} = k^{(0)} + \frac{1}{k^{(1)} + \frac{1}{k^{(2)} + \dots}}.$$
(28)

We can statistically describe this sequence of eras if, instead of a given initial value $u_{\text{max}}^{(0)} = k^{(0)} + x^{(0)}$, we consider a distribution of $x^{(0)}$ over the interval (0, 1) governed by some probability law. Then we also obtain some distributions of the values of $x^{(s)}$ that terminate every sth series of numbers. It can be shown that with increasing s, these distributions tend to a stationary (independent of s) probability distribution w(x) in which the initial value $x^{(s)}$ is completely 'forgotten':

$$w(x) = \frac{1}{(1+x)\ln 2} \,. \tag{29}$$

It follows from Eqn (29) that the probability distribution of the lengths of series k is given by

$$W(k) = \frac{1}{\ln 2} \ln \frac{(k+1)^2}{k(k+2)}.$$
(30)

Moreover, probability distributions for other parameters describing successive eras, such as the parameter δ , can be calculated exactly, giving a relation between the amplitudes of logarithms of the functions *a*, *b*, and *c* and the logarithmic time [14].

Thus, we have seen from the results of statistical analysis of evolution in the neighborhood of a singularity [13] that the stochasticity and probability distributions of parameters already arise in classical general relativity.

At the end of this section, a historical remark is in order. Continued fraction (28) was shown in 1968 to I M Lifshitz (Landau had already passed away) and he immediately noticed that the formula for a stationary distribution of the value of x in (29) can be derived. Later, it became known that this formula was derived in the nineteenth century by Gauss, who had not published it but had described it in a letter to a colleague.

4. Oscillatory approach to the singularity: modern development

The oscillatory approach to the cosmological singularity described in the preceding section was developed for an empty space-time. It is not difficult to understand that if the universe is filled with a perfect fluid with the equation of state $p = w\rho$, where p is the pressure, ρ is the energy density, and w < 1, then the presence of this matter cannot change the dynamics in the vicinity of the singularity. Indeed, using the energy conservation equation, we can show that

$$\rho = \frac{\rho_0}{(abc)^{w+1}} = \frac{\rho_0}{t^{w+1}} \,, \tag{31}$$

where ρ_0 is a positive constant. Thus, the term representing matter in the Einstein equations behaves as $\sim 1/t^{1+w}$ and is weaker as $t \rightarrow 0$ than the terms of geometric origin coming from the time derivatives of the metric, which behave as $1/t^2$, let alone the perturbations due to the presence of spatial curvature, responsible for changes in the Kasner regime, which behave as $1/t^{2+4|p_1|}$. But the situation changes drastically if the parameter w is equal to unity, i.e., the pressure is equal to the energy density. This matter is called 'stiff matter' and can be represented by a massless scalar field. In this case, $\rho \sim 1/t^2$ and the contribution of matter is of the same order as the leading terms of geometrical origin. Hence, it is necessary to find a Kasnertype solution, with the presence of terms connected with the presence of stiff matter (a massless scalar field) taken into account. Such a study was carried out in [38]. It was shown there that the scale factors a, b, and c can again be represented as t^{2p_1}, t^{2p_2} , and t^{2p_3} , where the Kasner indices satisfy the relations

$$p_1 + p_2 + p_3 = 1, \quad p_1^2 + p_2^2 + p_3^2 = 1 - q^2,$$
 (32)

where the number q^2 reflects the presence of stiff matter and is bounded by

$$q^2 \leqslant \frac{2}{3} \,. \tag{33}$$

It can be seen that if $q^2 > 0$, then there exist combinations of the positive Kasner indices satisfying relations (32). Moreover, if $q^2 \ge 1/2$, only triples of positive Kasner indices can satisfy relations (32). If the universe finds itself in a Kasner regime with three positive indices, the perturbative terms existing due to the spatial curvatures are too weak to change this Kasner regime, and it therefore becomes stable. This means that in the presence of stiff matter, after a finite number of changes of Kasner regimes, the universe finds itself in a stable regime and the oscillations stop. Thus, the massless scalar field plays an 'antichaotizing' role in the process of cosmological evolution [38]. The Lifshitz-Khalatnikov parameter can also be used in this case. The Kasner indices satisfying relations (32) are conveniently represented as [38]

$$p_{1} = \frac{-u}{1+u+u^{2}},$$

$$p_{2} = \frac{1+u}{1+u+u^{2}} \left[u - \frac{u-1}{2} \left(1 - (1-\beta^{2})^{1/2} \right) \right],$$

$$p_{3} = \frac{1+u}{1+u+u^{2}} \left[1 + \frac{u-1}{2} \left(1 - (1-\beta^{2})^{1/2} \right) \right],$$

$$\beta^{2} = \frac{2(1+u+u^{2})^{2}}{(u^{2}-1)^{2}}.$$
(34)

The range of the parameter *u* is now $-1 \le u \le 1$, while the admissible values of *q* at a fixed *u* are

$$q^{2} \leqslant \frac{(u^{2} - 1)^{2}}{2(1 + u + u^{2})^{2}} \,. \tag{35}$$

It is easy to show that after one bounce, the value of q^2 changes according to the rule

$$q^2 \to q'^2 = q^2 \frac{1}{(1+2p_1)^2} > q^2$$
. (36)

Thus, the parameter q^2 increases and, hence, the probability of finding all three Kasner indices to be positive increases. This again confirms the statement that after a finite number of bounces, the universe in the presence of the massless scalar field finds itself in the Kasner regime with three positive indices and the oscillations stop.

In the second half of the 1980s, a series of papers was published [15] where solutions of the Einstein equations in the vicinity of a singularity for (d + 1)-dimensional space – times were studied. The multidimensional analog of a Bianchi-I universe was considered with the generalized K asner metric

$$ds^{2} = dt^{2} - \sum_{i=1}^{d} t^{2p_{i}} dx^{i2}, \qquad (37)$$

where the Kasner indices p_i satisfy the conditions

$$\sum_{i=1}^{d} p_i = \sum_{i=1}^{d} p_i^2 = 1.$$
(38)

In the presence of spatial curvature terms, the transition from one Kasner epoch to another occurs and this transition is described by the following rule: the new Kasner exponents are equal to

$$p'_1, p'_2, \dots, p'_d = (q_1, q_2, \dots, q_d),$$
 (39)

where

$$q_{1} = \frac{-p_{1} - P}{1 + 2p_{1} + P}, \quad q_{2} = \frac{p_{2}}{1 + 2p_{1} + P}, \dots,$$

$$q_{d-2} = \frac{p_{d-2}}{1 + 2p_{1} + P}, \quad q_{d-1} = \frac{2p_{1} + P + p_{d-1}}{1 + 2p_{1} + P},$$

$$q_{d} = \frac{2p_{1} + P + p_{d}}{1 + 2p_{1} + P}, \quad (40)$$

with

1

$$P = \sum_{i=2}^{d-2} p_i \,. \tag{41}$$

However, such a transition from one Kasner epoch to another occurs if at least one of the numbers α_{ijk} is negative. These numbers are defined as

$$\alpha_{ijk} \equiv 2p_i + \sum_{l \neq j, k, i} p_l, \quad (i \neq j, \ i \neq k, \ j \neq k).$$

$$(42)$$

For space – times with d < 10, one of the factors α is always negative and, hence, one change of Kasner regime is followed by another, implying the oscillatory behavior of the universe in the neighborhood of the cosmological singularity. But for space – times with $d \ge 10$, combinations of Kasner indices satisfying Eqn (38) with all the numbers α_{ijk} positive exist. If the universe enters the Kasner regime with such indices (socalled 'Kasner stability region'), its chaotic behavior disappears and this Kasner regime is preserved. Thus, the hypothesis was put forward that in space – times with $d \ge 10$, after a finite number of oscillations, the universe under consideration finds itself in the Kasner stability region and the oscillating regime is replaced by the monotonic Kasner behavior.

The discovery of the fact that the chaotic character of the approach to the cosmological singularity disappears in space – times with $d \ge 10$ was unexpected and looked like an accidental result of an interplay of real numbers satisfying generalized Kasner relations (40). Later, it became clear that underlying this fact is a deep mathematical structure, the hyperbolic Kac-Moody algebras. Indeed, in a series of works by Damour, Henneaux, Nicolai, and some other authors (see, e.g., Refs [17]) on the cosmological dynamics in models based on superstring theories living in 10-dimensional space-time and in the d + 1 = 11 supergravity model, it was shown that in the vicinity of the singularity, these models reveal oscillating behavior of the BKL type. The important new feature of the dynamics in these models is the role played by nongravitational bosonic fields (*p*-forms), which are also responsible for transitions from one Kasner regime to another. For a description of these transitions, the Hamiltonian formalism [12] becomes very convenient. In the framework of this formalism, the configuration space of the Kasner parameters describing the dynamics of the universe can be treated as a billiard system, while the curvature terms in the Einstein theory and *p*-form potentials in superstring theories play the role of the walls of these billiards. The transition from one Kasner epoch to another is a reflection from one of the walls. Thus, there is a correspondence between the rather complicated dynamics of a universe in the vicinity of the cosmological singularity and the motion of a ball on a billiard table.

However, there exists a more striking and unexpected correspondence between the chaotic behavior of the universe in the vicinity of the singularity and such an abstract mathematical object as the hyperbolic Kac-Moody algebras [17]. We briefly explain what it means. Every Lie algebra is defined by its generators h_i , e_i , f_i , i = 1, ..., r, where r is the rank of the Lie algebra, i.e., the maximal number of its generators h_i that commute with each other (these generators constitute the Cartan subalgebra). The commutation relations between the generators are

$$[e_{i}, f_{j}] = \delta_{ij} h_{i},$$

$$[h_{i}, e_{j}] = A_{ij} e_{j},$$

$$[h_{i}, f_{j}] = -A_{ij} f_{j},$$

$$[h_{i}, h_{j}] = 0.$$
(43)

The coefficients $A_{i,j}$ constitute the generalized Cartan $r \times r$ matrix such that $A_{ii} = 2$, its off-diagonal elements are nonpositive integers, and $A_{ij} = 0$ for $i \neq j$ implies $A_{ji} = 0$. We can say that the e_i are rising operators, similar to the wellknown operator $L_+ = L_x + iL_y$ in the theory of angular momentum, while the f_i are lowering operators like $L_- = L_x - iL_y$. The generators h_i of the Cartan subalgebra can be compared with the operator L_z . The generators must also satisfy the Serre relations

$$(ad e_i)^{1-A_{ij}}e_j = 0,$$

 $(ad f_i)^{1-A_{ij}}f_j = 0,$ (44)

where $(ad A)B \equiv [A, B]$.

The Lie algebras $\mathcal{G}(A)$ built on a symmetrizable Cartan matrix A have been classified according to the properties of their eigenvalues:

if A is positive definite, $\mathcal{G}(A)$ is a finite-dimensional Lie algebra;

if A allows one null eigenvalue and all the others are strictly positive, $\mathcal{G}(A)$ is an affine Kac–Moody algebra;

if A allows one negative eigenvalue and all the others are strictly positive, $\mathcal{G}(A)$ is a Lorentz Kac–Moody algebra.

A correspondence exists between the structure of a Lie algebra and a certain system of vectors in the *r*-dimensional Euclidean space, which essentially simplifies the task of classification of the Lie algebra. These vectors, called roots, represent the rising and lowering operators of the Lie algebra. The vectors corresponding to the generators e_i and f_i are called simple roots. The system of positive simple roots (i.e., roots corresponding to the rising generators e_i) can be represented by nodes of their Dynkin diagrams, while the edges connecting (or not connecting) the nodes give information about the angles between simple positive root vectors.

An important subclass of Lorentz Kac–Moody algebras can be defined as follows. A Kac–Moody algebra such that deleting one node from its Dynkin diagram gives a sum of finite or affine algebras is called a hyperbolic Kac–Moody algebra. These algebras are all known. In particular, there exists no hyperbolic algebras with the rank higher than 10.

We recall some more definitions from the theory of Lie algebras. Reflections with respect to hyperplanes orthogonal to simple roots leave the root system invariant. The corresponding finite-dimensional group is called the Weyl group. Finally, the hyperplanes mentioned above divide the *r*-dimensional Euclidean space into regions called Weyl chambers. The Weyl group transforms one Weyl chamber into another.

We can now briefly formulate the results of the approach in [17] following paper [39]: the links between the billiards describing the evolution of the universe in the neighborhood of singularity and its corresponding Kac–Moody algebra can be described as follows:

the Kasner indices describing the 'free' motion of the universe between the reflections from the walls correspond to the elements of the Cartan subalgebra of the Kac–Moody algebra;

the dominant walls, i.e., the terms in the equations of motion responsible for the transition from one Kasner epoch to another, correspond to the simple roots of the Kac–Moody algebra;

the group of reflections in the cosmological billiard system is the Weyl group of the Kac–Moody algebra; the billiard table can be identified with the Weyl chamber of the Kac-Moody algebra.

We can imagine two types of billiard tables: infinite, where the linear motion without collisions with the walls is possible (nonchaotic regime), and those where reflections from the walls are inevitable and the regime can be only chaotic. Remarkably, the Weyl chambers of the hyperbolic Kac– Moody algebras are such that infinitely repeating collisions with the walls occur. It was shown that all the theories with the oscillating approach to the singularity such as the Einstein theory in dimensions d < 10 and superstring cosmological models correspond to hyperbolic Kac–Moody algebras.

The existence of links between the BKL approach to the singularities and the structure of some infinite-dimensional Lie algebras has inspired some authors to declare a new program of development of quantum gravity and cosmology [40]. They propose "to take seriously the idea that near the singularity (i.e., when the curvature gets larger than the Planck scale) the description of a spatial continuum and space – time based (quantum) field theory breaks down, and should be replaced by a much more abstract Lie algebraic description."

5. New types of cosmological singularities

As mentioned in the Introduction, the development of the theoretical and observational cosmology and, in particular, the discovery of the cosmic acceleration have stimulated the elaboration of cosmological models where new types of singularities are described. In contrast to the 'traditional' Big Bang and Big Crunch singularities, these singularities occur not at zero but at finite or even infinite values of the cosmological radius. The most famous of these singularities is, perhaps, the Big Rip singularity [23, 24] arising if the absolute value of the negative pressure p of dark energy is larger than the energy density ρ . Indeed, we consider a flat Friedmann universe with the metric

$$ds^2 - a^2(t) dl^2, (45)$$

filled with a perfect fluid with the equation of state

$$p = w\rho, \quad w = \text{const} < -1. \tag{46}$$

The dependence of the energy density ρ on the cosmological radius *a* is, as usual,

$$\rho = \frac{C}{a^{3(1+w)}} \,, \tag{47}$$

and the Friedmann equation in this case has the form

$$\frac{\dot{a}^2}{a} = \frac{C}{a^{3(1+w)}} , \qquad (48)$$

where C is a positive constant. Integrating Eqn (48), we obtain

$$a(t) = \left(a_0^{3(1+w)/2} + \frac{2\sqrt{C}(t-t_0)}{3(1+w)}\right)^{2/3(1+w)}.$$
(49)

It is easy to see that at the finite time $t_R > t_0$ equal to

$$t_R = t_0 - \frac{3(1+w)}{2\sqrt{C}} a_0^{3(1+w)/2}, \qquad (50)$$

the cosmological radius becomes infinite and the same also occurs with the Hubble variable \dot{a}/a and, hence, with the scalar curvature. Thus, we encounter a new type of cosmological singularity, characterized by infinite values of the cosmological radius, its time derivative, the Hubble variable, and the scalar curvature. It is usually called the 'Big Rip' singularity. Its properties have attracted a considerable attention of researchers because some observational data indicate that the actual value of the equation of state parameter w is indeed smaller than -1.

There are also other types of cosmological singularities that can be encountered at finite values of the cosmological radius (see, e.g., [25-29]). For illustration, we here consider one type of singularity, the Big Brake singularity [25]. This singularity can be achieved in a finite lapse of cosmic time and is characterized by a finite value of the cosmological radius, by the vanishing first time derivative of the radius, and by the second time derivative of the cosmological radius tending to minus infinity (an infinite deceleration). We consider a perfect fluid with the equation of state

$$p = \frac{A}{\rho} \,, \tag{51}$$

where A is a positive constant. This fluid could be called an 'anti-Chaplygin' gas because the widely used Chaplygin gas cosmological model [41] is based on the equation of state $p = -A/\rho$. The dependence of the energy density on the cosmological radius for equation of state (51) is

$$\rho = \sqrt{\frac{B}{a^6} - A}\,,\tag{52}$$

where *B* is a positive constant. When *a* is small, $\rho \sim 1/a^3$ and behaves like dust. Then, as $a \to a_B$,

$$a_B = \left(\frac{B}{A}\right)^{1/6},\tag{53}$$

and the energy density tends to zero. The solution of the Friedmann equation in this limit gives

$$a(t) = a_B - C_0(t_B - t)^{4/3}, \quad C_0 = 2^{-7/3} 3^{5/3} (AB)^{1/6}.$$
 (54)

Now, it can be easily verified that as $t \to t_B$, $\dot{a} \to 0$ and $\ddot{a} \to -\infty$. Thus, we indeed encounter the Big Brake cosmological singularity.

6. Conclusions

We have shown in this short review that the opinion expressed by Landau many years ago concerning the importance of the problem of singularity in cosmology has proved to be prophetic. The study of the cosmological singularity has revealed the existence of an oscillatory behavior of the universe as the curvature of space – time increases, which in turn has a deep connection with quite new branches of modern mathematics. On the other hand, the latest successes of observational cosmology have stimulated the development of various cosmological models, which reveal new types of cosmological singularities whose investigation from both the physical and mathematical standpoints can be very promising.

This work was partially supported by the RFBR grant 05-02-17450 and by the grant LSS-1157.2006.2.

References

- 1. Landau L D, Private communication
- 2. Bardeen J, Cooper L N, Schrieffer J R Phys. Rev. 108 1175 (1957)
- Wilson K G, Kogut J "The renormalization group and the ε-expansion" *Phys. Rep.* 12 75 (1974) [Translated into Russian (Moscow: Mir, 1975)]
- Khalatnikov I M, Kamenshchik A Yu Usp. Fiz. Nauk 168 593 (1998) [Phys. Usp. 41 525 (1998)]
- Penrose R Structure of Space-Time (New York: W.A. Benjamin, 1968) [Translated into Russian (Moscow: Mir, 1972)]; Hawking SW, Ellis G F R The Large Scale Structure of Space-Time (Cambridge: Cambridge Univ. Press, 1973) [Translated into Russian (Moscow: Mir, (1977)]; Hawking SW, Penrose R Proc. R. Soc. London Ser. A 314 529 (1970)
- Lifshitz E M, Khalatnikov I M Usp. Fiz. Nauk 80 391 (1963) [Sov. Phys. Usp. 6 495 (1964)]; Adv. Phys. 12 185 (1963)
- 7. Khalatnikov I M, Lifshitz E M Phys. Rev. Lett. 24 76 (1970)
- Lifshitz E M, Khalatnikov I M Pis'ma Zh. Eksp. Teor. Fiz. 11 200 (1970) [JETP Lett. 11 123 (1970)]
- Belinskii V A, Khalatnikov I M Zh. Eksp. Teor. Fiz. 56 1700 (1969) [Sov. Phys. JETP 29 911 (1969)]
- Belinskii V A, Lifshitz E M, Khalatnikov I M Usp. Fiz. Nauk 102 463 (1970) [Sov. Phys. Usp. 13 745 (1971)]
- 11. Belinsky V A, Lifshitz E M, Khalatnikov I M Adv. Phys. 31 639 (1982)
- 12. Misner C W Phys. Rev. Lett. 22 1071 (1969)
- Lifshitz E M, Lifshits I M, Khalatnikov I M Zh. Eksp. Teor. Fiz 59 322 (1970) [Sov. Phys. JETP 32 173 (1971)]
- 14. Khalatnikov I M et al. J. Stat. Phys. 38 97 (1985)
- Demaret J, Henneaux M, Spindel P Phys. Lett. B 164 27 (1985); Demaret J, Hanquin J-L, Henneaux M, Spindel P, Taormina A Phys. Lett. B 175 129 (1986); Demaret J, de Rop Y, Henneaux M Int. J. Theor. Phys. 28 1067 (1989)
- 16. Kasner E Am. J. Math. 43 217 (1921)
- Damour T, Henneaux M Phys. Rev. Lett. 85 920 (2000); Damour T et al. Phys. Lett. B 509 323 (2001); Damour T, Henneaux M, Nicolai H Class. Quantum Grav. 20 R145 (2003)
- Kac V G Infinite Dimensional Lie Algebras 3rd ed. (Cambridge: Cambridge Univ. Press, 1990)
- Riess A et al. Astron. J. 116 1009 (1998); Perlmutter S et al. Astrophys. J. 517 565 (1999)
- Sahni V, Starobinsky A Int. J. Mod. Phys. D 9 373 (2000); 15 2105 (2006)
- Alam U, Sahni V, Starobinsky A A JCAP (02) 011 (2007); Nesseris S, Perivolaropoulos L JCAP (02) 025 (2007)
- Caldwell R R *Phys. Lett. B* 545 23 (2002)
 Caldwell R R, Kamionkowski M, Weinberg
- 23. Caldwell R R, Kamionkowski M, Weinberg N N *Phys. Rev. Lett.* **91** 071301 (2003)
- 24. Starobinsky A A Grav. Cosmol. 6 157 (2000)
- Gorini V et al. *Phys. Rev. D* 69 123512 (2004); Kamenshchik A Y, Kiefer C, Sandhöfer B *Phys. Rev. D* 76 064032 (2007)
- Barrow J D Class. Quantum Grav. 21 L79 (2004); Cattoën C, Visser M Class. Quantum Grav. 22 4913 (2005)
- Cannata F, Kamenshchik A Yu, Regoli D "Scalar field cosmological models with finite scale factor singularities", arXiv:0801.2348
- Barvinsky A O, Deffayet C, Kamenshchik A Yu "Anomaly driven cosmology: big boost scenario and AdS/CFT correspondence" *JCAP* (5) 020 (2008); arXiv:0801.2063
- Shtanov Y, Sahni V Class. Quantum Grav. 19 L101 (2002); Tretyakov P, Toporensky A, Shtanov Y, Sahni V Class. Quantum Grav. 23 3259 (2006)
- Landau L D, Lifshitz E M *Teoriya Polya* (The Classical Theory of Fields) (Moscow: Nauka, 1988) [Translated into English (Oxford: Pergamon Press, 1975)]
- 31. Raychaudhuri A Phys. Rev. 98 1123 (1955); Phys. Rev. 106 172 (1957)
- 32. Komar A Phys. Rev. 104 544 (1956)
- Lifshitz E M, Sudakov V V, Khalatnikov I M Zh. Eksp. Teor. Fiz. 40 1847 (1961) [Sov. Phys. JETP 13 1298 (1961)]; Khalatnikov I M, Lifshitz E M, Sudakov V V Phys. Rev. Lett. 6 311 (1961)
- Grishchuk L P Zh. Eksp. Teor. Fiz. 51 475 (1966) [Sov. Phys. JETP 24 320 (1967)]

- Arnold V I, Shandarin S F, Zeldovich Ya B Geophys. Astrophys. Fluid Dynamics 20 111 (1982)
- Bini D, Cherubini C, Jantzen R T Class. Quantum Grav. 24 5627 (2007)
- Petrov A Z Prostranstva Einshteina (Einstein Spaces) (Moscow: GIFML, 1961) [Translated into English (Oxford: Pergamon Press, 1969)]
- Belinskii V A, Khalatnikov I M Zh. Eksp. Teor. Fiz. 63 1121 (1972) [Sov. Phys. JETP 36 591 (1973)]
- Damour T "Cosmological singularities, billiards and Lorentzian Kac-Moody algebras", gr-qc/0412105
- Damour T, Nicolai H "Symmetries, singularities and the deemergence of space", arXiv:0705.2643
- 41. Kamenshchik A, Moschella U, Pasquier V Phys. Lett. B 511 265 (2001)

PACS numbers: **11.15.** – **q**, 11.30.Qc, 12.38.Aw DOI: 10.1070/PU2008v051n06ABEH006551 DOI: 10.3367/UFNr.0178.200806k.0647

Axial anomaly in quantum electro- and chromodynamics and the structure of the vacuum in quantum chromodynamics

B L Ioffe

1. Introduction

In this report, I discuss the current state of the problem of the axial anomaly in quantum electrodynamics (QED) and quantum chromodynamics (QCD) and the relation of the axial anomaly to the structure of the vacuum in QCD. In QCD, the vacuum average of the axial anomaly is proportional to a new quantum number n, the winding number. There are an infinite number of vacuum states $|n\rangle$. The transition amplitudes between these states are amplitudes of tunnel transitions along certain paths in the space of gauge fields. I show that the axial anomaly condition implies that there are zero modes of the Dirac equation for a massless quark and that spontaneous chiral symmetry breaking occurs in QCD, which leads to the formation of a quark condensate. The axial anomaly can be represented in the form of a sum rule for the structure function in the dispersion representation of the axial-vector vector (AVV) vertex. On the basis of this sum rule, we calculate the width of the $\pi^0 \rightarrow 2\gamma$ decay with an accuracy of 1.5%.

2. The definition of an anomaly

We suppose that the classical field-theory Lagrangian has a certain symmetry, i.e., is invariant under transformations of the fields corresponding to this symmetry. According to the Noether theorem, the symmetry corresponds to a conservation law. An anomaly is a phenomenon in which the given symmetry and the conservation law are violated as we pass to quantum theory. The reason for this violation lies in the singularity of quantum field operators at small distances, such that finding the physical quantities requires fixing not only the Lagrangian but also the renormalization procedure. (See reviews dealing with various anomalies in Refs [1-4].)

There are two types of anomalies, internal and external. In the first case, the gauge invariance of the classical Lagrangian is broken at the quantum level, the theory becomes unrenormalizable, and is not self-consistent. This problem can be resolved by a special choice of fields in the Lagrangian, for which all the internal anomalies cancel. (Such an approach is used in the standard model of electroweak interaction and is known as the Glashow– Illiopoulos–Maiani mechanism.) External anomalies emerge as a result of the interaction between the fields in the Lagrangian and external sources. It is these anomalies that appear in quantum electrodynamics and quantum chromodynamics; they are discussed in what follows. We show that anomalies play an important role in QED and especially in QCD. Hence, the term 'anomaly' should not mislead us — it is a normal and important ingredient of most quantum field theories.

3. Axial anomaly in QED

The Dirac equation for the electron in an external electromagnetic field $A_{\mu}(x)$ is

$$i\gamma_{\mu} \frac{\partial \psi(x)}{\partial x_{\mu}} = m\psi(x) - e\gamma_{\mu}A_{\mu}(x)\psi(x).$$
(1)

The axial current is defined as

$$j_{\mu 5}(x) = \psi(x) \gamma_{\mu} \gamma_5 \psi(x) .$$
⁽²⁾

Its divergence calculated in classical theory, i.e., with the use of Eqn (1), is

$$\partial_{\mu} j_{\mu 5}(x) = 2im\bar{\psi}(x)\gamma_5\psi(x) \tag{3}$$

and tends to zero as $m \to 0$. In quantum theory, the axial current must be redefined because $j_{\mu 5}(x)$ is the product of two local fermionic fields, with the result that it is singular when both fields act at the same point. (Naturally, a similar statement is true for a vector current.) To achieve a meaningful approach, we split the points where the two fermionic fields act by a distance ε , such that

$$j_{\mu 5}(x,\varepsilon) = \bar{\psi}\left(x + \frac{\varepsilon}{2}\right) \gamma_{\mu} \gamma_{5}$$
$$\times \exp\left[\mathrm{i}e \int_{x-\varepsilon/2}^{x+\varepsilon/2} \mathrm{d}y_{\alpha} \mathcal{A}_{\alpha}(y)\right] \psi\left(x - \frac{\varepsilon}{2}\right), \tag{4}$$

and take $\varepsilon \to 0$ in the final result. The exponential factor in (4) is introduced to ensure the local gauge invariance of $j_{\mu 5}(x, \varepsilon)$. The divergence of axial current (4) has the following form (we use Eqn (1) and keep only the terms that are linear in ε):

$$\partial_{\mu} j_{\mu 5}(x,\varepsilon) = 2\mathrm{i}m\bar{\psi}\left(x+\frac{\varepsilon}{2}\right)\gamma_{5}\psi\left(x-\frac{\varepsilon}{2}\right) -\mathrm{i}e\varepsilon_{\alpha}\bar{\psi}\left(x+\frac{\varepsilon}{2}\right)\gamma_{\mu}\gamma_{5}\psi\left(x-\frac{\varepsilon}{2}\right)F_{\alpha\mu}\,,\tag{5}$$

where $F_{\alpha\mu}$ is the electromagnetic field strength. For simplicity, we assume that $F_{\mu\nu} = \text{const}$ and use the fixed-point gauge (the Fock – Schwinger gauge) $x_{\mu}A_{\mu}(x) = 0$. Then $A_{\mu}(x) =$ $(1/2) x_{\nu}F_{\nu\mu}$. We calculate the vacuum average of (5). To calculate the right-hand side of (5), we use the expression for the electron propagator in an external electromagnetic field

$$(\not x = x_{\mu}\gamma_{\mu}):$$

$$S(x) = \frac{i}{2\pi^{2}} \left[\frac{\not x}{x^{4}} + i \frac{m}{2x^{2}} + \frac{1}{16x^{2}} eF_{\mu\nu}(\not x \sigma_{\mu\nu} + \sigma_{\mu\nu} \not x) \right], \quad (6)$$

$$\sigma_{\mu\nu} = \frac{1}{2} \left(\gamma_{\mu} \gamma_{\nu} - \gamma_{\nu} \gamma_{\mu} \right). \tag{7}$$

Vacuum averaging involves first-order corrections in e^2 . Substituting Eqn (6) in Eqn (5) and ignoring the electron mass, we obtain

$$\langle 0|\,\partial_{\mu}j_{\mu5}|0\rangle = \frac{e^2}{4\pi^2} \,F_{\alpha\mu}F_{\lambda\sigma}\varepsilon_{\beta\mu\lambda\sigma}\,\frac{\varepsilon_{\alpha}\varepsilon_{\beta}}{\varepsilon^2}\,. \tag{8}$$

Because there can be no preferred direction in space-time, the limit $\varepsilon \to 0$ can be achieved in a symmetric manner, and we have

$$\partial_{\mu}j_{\mu5} = \frac{e^2}{8\pi^2} F_{\alpha\beta}\tilde{F}_{\alpha\beta} , \qquad (9)$$

where

$$\tilde{F}_{\alpha\beta} = \frac{1}{2} \, \varepsilon_{\alpha\beta\lambda\sigma} F_{\lambda\sigma} \tag{10}$$

is the dual electromagnetic field tensor. In Eqn (9), the symbol of vacuum averaging is dropped because in the e^2 -order, this equation can be considered an operator equation. Equation (9) is known as the Adler-Bell-Jackiw anomaly [5-8].

To better understand the origin of an anomaly, we consider the same problem in the momentum space. In QED, the matrix element of the transition of an axial current with a momentum q into two real or virtual photons with momenta p and p' is described by the diagrams in Fig. 1. The matrix element is

$$T_{\mu\alpha\beta}(p,p') = \Gamma_{\mu\alpha\beta}(p,p') + \Gamma_{\mu\beta\alpha}(p',p), \qquad (11)$$

$$\Gamma_{\mu\alpha\beta}(p, p') = -e^2 \int \frac{d^4k}{(2\pi)^4} \operatorname{Tr} [\gamma_{\mu}\gamma_5(\not\!\!k + \not\!\!p - m)^{-1} \\ \times \gamma_{\alpha}(\not\!\!k - m)^{-1}\gamma_{\beta}(\not\!\!k - \not\!\!p' - m)^{-1}].$$
(12)

Integral (12) linearly diverges. In a linearly divergent integral, the important terms are the surface terms, which emerge as a result of integrating over an infinitely remote



Figure 1. Feynman diagrams describing the transition of an axial current with a momentum q into two real or virtual photons with momenta p and p', q = p + p': (a) the direct diagram, and (b) the crossing diagram.

surface in the momentum space. (This becomes especially clear when the vectors q, p, and p' are space-like and the integration contour over k_0 can be rotated to the imaginary axis, $k_0 \rightarrow ik_4$, such that integration over k is carried out in Euclidean space.) The result of calculations depends on the way k is chosen: we can displace k by an arbitrary constant vector a_{λ} , i.e., $k_{\lambda} \rightarrow k_{\lambda} + a_{\lambda}$. Amplitude (11) must satisfy the conditions needed for the vector-current conservation: $p_{\alpha}T_{\mu\alpha\beta}(p, p') = 0$ and $p'_{\beta}T_{\mu\alpha\beta}(p, p') = 0$.

We try to choose the vector a_{λ} such that the conditions for both axial- and vector-current conservation are satisfied. We parameterize a_{λ} as $a_{\lambda} = (a+b)p_{\lambda} + bp'_{\lambda}$. The result of calculations shows that both conditions cannot be satisfied simultaneously: the vector-current conservation can be achieved at a = -2, while the axial-current conservation requires that a = 0 [8, 9]. The vector-current conservation is a necessary condition for the existence of QED. Hence, we select a = -2. The divergence of the axial current is

$$q_{\mu}T_{\mu\alpha\beta}(p,p') = \left[2mG(p,p') - \frac{e^2}{2\pi^2}\right] \varepsilon_{\alpha\beta\lambda\sigma} p_{\lambda} p_{\sigma}'.$$
(13)

Here, we restore the term proportional to the electron mass and define G(p, p') as

$$\left\langle p, \varepsilon_{\alpha}; p', \varepsilon_{\beta}' | \bar{\psi} \gamma_{5} \psi | 0 \right\rangle = G(p, p') \varepsilon_{\alpha\beta\lambda\sigma} p_{\lambda} p_{\sigma}', \qquad (14)$$

with ε_{α} and ε'_{β} being the polarizations of the two photons. The fact that the axial current is not conserved, stated in Eqn (13), is equivalent to Eqn (9). Our discussion of the axial anomaly in QED was limited to terms of the order e^2 . Adler and Bardeen have proved (see Refs [5, 6, 10]) that the radiative corrections caused by the photon lines connecting different points inside the triangle diagrams in Fig. 1 do not alter the anomaly equation. However, the diagram in Fig. 2 yields a nonvanishing, albeit small, correction of the order e^6 to this condition [11].

4. The axial anomaly and its relation to the structure of the vacuum in quantum chromodynamics

In QCD with massless quarks, the axial anomaly is described by a formula similar to (9):

$$\partial_{\mu} j^{a}_{\mu 5} = \frac{e^{2}}{8\pi^{2}} e^{2}_{q} N_{c} F_{\mu\nu} \tilde{F}_{\mu\nu} . \qquad (15)$$



Figure 2. The e^6 -order correction to the Adler–Bell–Jackiw anomaly in QED.

Here, $N_c = 3$ is the number of colors and e_q is the quark charge. (We wrote Eqn (15) for one massless quark.) There is also another anomaly in QCD, where the external fields are not electromagnetic but gluonic:

$$\partial_{\mu}j_{\mu5} = \frac{\alpha_{\rm s}N_{\rm c}}{4\pi} G^n_{\mu\nu}\tilde{G}^n_{\mu\nu}, \qquad (16)$$

where $G_{\mu\nu}^n$ is the gluonic field strength and $\tilde{G}_{\mu\nu}^n$ is its dual. Equation (16) can be considered an operator equation, and the fields $G_{\mu\nu}^n$ and $\tilde{G}_{\mu\nu}^n$ can be considered the fields of virtual gluons. In the same way as in QED, perturbative corrections to (16) begin at α_s^3 and are described by a diagram similar to the one shown in Fig. 2. In QCD, however, the coupling constant is large, with the result that the contribution provided by this diagram is not small; the contribution of diagrams obtained from the one in Fig. 2 by attaching additional quark and gluon loops are not small either. Obviously, in QCD, the octet axial current

$$j_{\mu 5}^{i} = \sum_{q} \bar{\psi}_{q} \gamma_{\mu} \gamma_{5} \frac{\lambda^{i}}{2} \psi_{q}, \quad i = 1, ..., 8$$
(17)

is conserved in the absence of an electromagnetic field. (Here, the λ^i are the Gell-Mann matrices, and summation is over the flavors of the light quarks, q = u, d, s.) The singlet axial current

$$j_{\mu 5}^{(0)} = \sum_{\mathbf{q}} \bar{\psi}_{\mathbf{q}} \gamma_{\mu} \gamma_{5} \psi_{\mathbf{q}} \tag{18}$$

contains the anomaly

$$\partial_{\mu} j_{\mu 5}^{(0)} = 3 \, \frac{\alpha_{\rm s} N_{\rm c}}{4\pi} \, G_{\mu\nu}^n \tilde{G}_{\mu\nu}^n \,. \tag{19}$$

In view of the spontaneous breaking of chiral symmetry, the pseudoscalar mesons belonging to the octet (π, K, η) are massless (in the $m_q \rightarrow 0$ approximation), while the SU(3) singlet η' is massive. In this way, the presence of an anomaly solves what is known as the U(1) problem [12].

I now discuss the important assertion that exists in QCD and relates the structure of the anomaly to the structure of the vacuum in this theory. Because we deal with the existence of degenerate vacua and tunnel (underbarrier) transitions between them, it is convenient (just as in quantum mechanics) to introduce imaginary time by setting $t = x_0 = -ix_4$; we thus operate in the Euclidean space, where $x^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2$. In the Euclidean space, the action integral

$$S = \frac{1}{4} \int d^4 x G_{\mu\nu}^2$$
 (20)

is positive. (We temporarily ignore the quark contribution.) The transition amplitudes are determined by the matrix elements of $\exp(-S)$. A theorem first proved by Belavin, Polyakov, Schwartz, and Tyupkin [13] states that

$$\frac{\alpha_{\rm s}}{8\pi} \int \mathrm{d}^4 x G^n_{\mu\nu} \tilde{G}^n_{\mu\nu} = n \,, \tag{21}$$

where n is an integer known as the winding number. Here, we do not prove this theorem in detail; instead, we mention its main points. The integrand in (21) can be written as the total

derivative

$$G^n_{\mu\nu}\tilde{G}^n_{\mu\nu} = \partial_\mu K_\mu \,, \tag{22}$$

$$K_{\mu} = \varepsilon_{\mu\nu\gamma\delta} \left(A_{\nu}^{n} G_{\gamma\delta}^{n} - \frac{1}{3} f^{nmp} A_{\nu}^{n} A_{\gamma}^{m} A_{\delta}^{p} \right).$$
(23)

When x^2 is large, $G_{\mu\nu}(x)$ decreases faster than $1/x^2$ (i.e., there is no physical field), and A^n_{μ} is a pure-gauge field. Then, we can drop the first term in the right-hand side of (23) and keep only the second term in the general expression for the gauge transformation for A^n_{μ} ,

$$A'_{\mu} = U^{-1}A_{\mu}U + iU^{-1}\partial_{\mu}U$$
 (24)

(here, U is a unitary, unimodular matrix, $U^+ = U$, |U|= 1). We suppose that the field A^n_{μ} (n = 1, 2, 3) belongs to the SU(2) subgroup of the color group SU(3). This subgroup plays a special role in the SU(3) group because it is isomorphic to the spatial rotation group O(3). At this point, it is convenient to introduce matrix notation for the fields A_{μ} :

$$A_i = \frac{1}{2} g \tau^k A_i^k, \quad k = 1, 2, 3; \quad i = 1, 2, 3.$$
(25)

Then, according to Eqns (22) and (23), we have

$$\int d^4 x G_{\mu\nu}(x) \, \tilde{G}_{\mu\nu}(x) = -\mathrm{i} \, \frac{4}{3} \, \frac{1}{g^2} \int dV \varepsilon_{ikl} \, \mathrm{Tr} \left(A_i A_k A_l \right). \tag{26}$$

Substituting the second term in the right-hand side of Eqn (24) in (26), we see that the integrand in (26) is a total derivative with respect to the spatial coordinates, and therefore reduces to an integral over an infinitely remote surface. Because |U|=1 on this surface, the matrix U has the form

$$U = \exp\left(2\pi n\hat{r}_a \,\frac{\tau^a}{2\mathrm{i}}\right),\tag{27}$$

where *n* in an integer and \hat{r}_a is a unit radius vector, $\hat{r}_a = r_a/|\mathbf{r}|$. The invariance of *U* under spatial rotations stems from the fact that each such rotation is accompanied by a gauge transformation, a rotation in the SU(2) group. When the right-hand side of Eqn (24) is substituted in (26), we see that Eqn (21) follows from (27). Theorem (21) also follows from general mathematical considerations, because the SU(2) group is mapped onto O(3); such a map is multivalued and is determined by the number of times the O(3) group is covered. We note that the fields corresponding to different *n* cannot be transformed into each other by a continuous transformation. In the perturbation theory, we always deal with fields corresponding to n = 0. The action integral in (2) can be written as

$$S = \frac{1}{4} \int d^4 x G^n_{\mu\nu} G^n_{\mu\nu} = \frac{1}{4} \int d^4 x \left[G^n_{\mu\nu} \tilde{G}^n_{\mu\nu} + \frac{1}{2} \left(G^n_{\mu\nu} - \tilde{G}^n_{\mu\nu} \right)^2 \right].$$
(28)

Because the last term in (28) is positive, the minimum of the action is achieved with fields satisfying the self-duality condition

$$G^n_{\mu\nu} = \tilde{G}^n_{\mu\nu}, \qquad (29)$$

$$S_{\min} = \frac{1}{4} \int d^4 x G^n_{\mu\nu} \tilde{G}^n_{\mu\nu} = \frac{8\pi^2}{g^2} |n| = \frac{2\pi}{\alpha_{\rm s}} |n| \,. \tag{30}$$

(Negative *n* correspond to anti-self-dual fields $G_{\mu\nu}^n = -\tilde{G}_{\mu\nu}^n$) The solutions of the self-duality equation (for n = 1), which became known as instantons, were found in [13]. It follows from Eqn (30) that in QCD in the Euclidean space, there exists an infinite number of action minima. In Minkowski space, instantons are paths of tunnel transitions (in the field space) between vacua characterized by different winding numbers but having the same energies [14–16]. By examining n(t), which transforms into the winding number as $t \to \pm \infty$, it can be shown that the instanton solutions correspond to $n(t \to -\infty) = 0$ and $n(t \to \infty) = 1$ and that the transition amplitude between vacuum states is [17]

$$\left\langle \frac{\Omega_{n=1}(t \to \infty)}{\Omega_{n=0}(t \to -\infty)} \right\rangle = \exp\left(-\frac{2\pi}{\alpha_{\rm s}}\right).$$
 (31)

5. Structure of the vacuum in quantum chromodynamics

Above, we showed that in QCD, there is an infinite number of vacua with the same energies, vacua that are characterized by the values of the winding number *n*. We let $\Omega(n)$ denote the wave function of such a vacuum and suppose that the wave functions are normalized, $\Omega^+(n) \Omega(n) = 1$, and form a complete system. The ambiguity in the wave function resides in the phase factor, $\Omega(n) = \exp(i\theta_n) \Omega'(n)$. We separate the Euclidean space into two big parts and assume that the field strength in the space between these parts is zero and the potentials are pure gauge. Then, obviously,

$$\exp(\mathrm{i}\theta_{n_1+n_2})\,\Omega(n_1+n_2) = \exp(\mathrm{i}\theta_{n_1})\,\Omega(n_1)\exp(\mathrm{i}\theta_{n_2})\,\Omega(n_2)\,.$$
(32)

[Here, we drop the prime on $\Omega'(n)$.] Because

$$\Omega(n_1 + n_2) = \Omega(n_1) \,\Omega(n_2) \,, \tag{33}$$

we have the equation

$$\theta_{n_1+n_1} = \theta_{n_1} + \theta_{n_2} \,, \tag{34}$$

which is solved by

$$\theta_n = n\theta \,. \tag{35}$$

Thus, the vacuum wave function in QCD is a linear combination of wave functions with different winding numbers:

$$\Omega(\theta) = \sum_{n} \exp(in\theta) \,\Omega(n) \,. \tag{36}$$

The state $\Omega(\theta)$ is known as the θ -vacuum. The vacuum state $\Omega(\theta)$ is similar to the Bloch state of an electron in a crystal, with θ acting as momentum. But in contrast to a Bloch state, all transitions between states with different θ are forbidden for the θ -vacuum. The vacuum state $\Omega(\theta)$ can be reproduced if the term

$$L_{\theta} = \frac{g^2 \theta}{32\pi^2} G_{\mu\nu} \tilde{G}_{\mu\nu} \,. \tag{37}$$

is added to the QCD Lagrangian (in Minkowski space). The presence of this term in the Lagrangian demonstrates that θ is an observable. Term (37) violates the P- and CP-invariance. However, so far all attempts to discover the violation of CP-invariance in strong interactions have failed. The strongest

bound on the value of θ has been found in searches of the neutron dipole moment, $\theta < 10^{-9}$ [18].

6. Zero eigenvalues of the Dirac equation for massless quarks as a consequence of the anomaly. Spontaneous breaking of chiral symmetry in quantum chromodynamics

We consider the Dirac equation for massless quarks in QCD in Euclidean space:

$$-\mathrm{i}\gamma_{\mu}\nabla_{\mu}\psi_{k} = \lambda_{k}\psi_{k}, \quad \nabla_{\mu} = \partial_{\mu} + \mathrm{i}g\,\frac{\lambda_{n}}{2}A_{\mu}^{n}. \tag{38}$$

From anomaly condition (16) with n = 1, we have

$$\int d^4x \operatorname{Tr} \langle 0 | \hat{o}_{\mu} j_{\mu 5}(x) | 0 \rangle$$
$$= \frac{g^2}{16\pi^2} \int d^4x \langle 0 | G^a_{\mu\nu} \tilde{G}^a_{\mu\nu} | 0 \rangle = 2N_{\rm c} \,. \tag{39}$$

The left-hand side of Eqn (39) can be written as an operator as follows:

$$d^{4}x \operatorname{Tr} \left\langle 0 \left| \partial_{\mu} j_{\mu 5}(x) \right| 0 \right\rangle$$

$$= -\int d^{4}x \partial_{\mu} \operatorname{Tr} \left\langle 0 \left| i \nabla^{-1}(x,x) \gamma_{\mu} \gamma_{5} \right| 0 \right\rangle$$

$$= -\int d^{4}x \nabla_{\mu} \operatorname{Tr} \left[\sum_{k} \frac{\psi_{k}(x) \psi_{k}^{+}(x)}{\lambda_{k}} \gamma_{\mu} \gamma_{5} \right]$$

$$= -\int d^{4}x \operatorname{Tr} \left[\sum_{k} \frac{\psi_{k}(x) \psi_{k}^{+}(x)}{\lambda_{k}} 2\lambda_{k} \gamma_{5} \right]. \quad (40)$$

States with nonzero λ_k contribute nothing to (40) because each such state $\psi_k(x)$ corresponds to the state $\gamma_5 \psi_k(x)$ with the eigenvalue $-\lambda_k$, and the two states are orthogonal. Thus, only the zero modes contribute, and hence we have

$$2\int d^4x \operatorname{Tr}\left[\gamma_5 \psi_0(x) \,\psi_0^+(x)\right] = -2N_{\rm c} \,. \tag{41}$$

This implies that in the case where n = 1, i.e., in the instanton field, the zero mode is right-handed: the quark spin is directed along the quark momentum, $\gamma_5\psi_0 = -\psi_0$. (Actually, for a quark in the instanton field, only one right-handed zero mode exists, because spin is correlated with color and the factor N_c in the right-hand side of Eqn (41) disappears.) At n = -1, the resulting equation differs from (41) only in sign, i.e., a lefthanded zero mode exists in an anti-instanton field. In the general case, we have the Atiyah–Singer theorem [19], according to which

$$n = n_{\rm L} - n_{\rm R} \,, \tag{42}$$

where $n_{\rm L}$ and $n_{\rm R}$ are the respective numbers of left- and righthand zero modes. It follows from (41) that in an instanton field, the zero mode violates the chiral symmetry of the Lagrangian, i.e., the invariance under the transformations $\psi \rightarrow \gamma_5 \psi$. (We note that in passing from the Euclidean metric to Minkowski space, the function ψ^+ is replaced by $\bar{\psi}$.) Thus, the presence of instantons is an indication that a quark condensate exists in the QCD vacuum:

$$\left\langle 0 \left| \psi \psi \left| 0 \right\rangle \neq 0 \right. \right\rangle \tag{43}$$

which breaks the chiral symmetry of the Lagrangian. (Unfortunately, it is impossible to calculate the quark condensate on the basis of the instanton approach because this approach is meaningful only when the distances are small, while the condensate forms over large distances.)

The winding number n corresponds to the topological current operator

$$Q_5(x) = \frac{\alpha_s}{8\pi} G^n_{\mu\nu}(x) \tilde{G}^n_{\mu\nu}(x) .$$
(44)

It was found in [20] that the vacuum correlator of topological currents

$$\zeta(q^2) = \mathbf{i} \int \mathbf{d}^4 x \, \exp\left(\mathbf{i} q x\right) \left\langle 0 \,|\, T\left\{ Q_5(x) \,, \, Q_5(0) \right\} \,|\, \right\rangle \tag{45}$$

vanishes at $q^2 = 0$ if the theory contains at least one massless quark. Later, it was proved in [21] that in the limit as $N_c \rightarrow \infty$, the relation

$$\zeta(0) = \left\langle 0 \,|\, \bar{q}q \,|\, 0 \right\rangle \left(\sum_{i}^{N_f} \frac{1}{m_i} \right)^{-1} \tag{46}$$

holds. In the cases of two and three massless quarks, the validity of Eqn (46) was proved in Ref. [22], where the limit $N_c \rightarrow \infty$ was not used. The concept of topological current turned out to be highly effective in QCD: it has been used to establish the spin composition of the proton [23], to establish a relation between the spin structure functions for large and small Q^2 [24, 25], and to determine the axial coupling constants for the nucleon [26].

7. The sum rule for the axial anomaly in quantum chromodynamics

We consider the general representation of the transition amplitude of the axial current into two photons with momenta p and p' in terms of the structure functions (form factors) without kinematic singularities, $T_{\mu\alpha\beta}(p, p')$ [27]. We limit ourselves to the case where $p^2 = p'^2$. Then [28, 29]

$$T_{\mu\alpha\beta}(p, p') = F_1(q^2, p^2) q_{\mu} \varepsilon_{\alpha\beta\rho\sigma} p_{\rho} p'_{\sigma} - \frac{1}{2} F_2(q^2, p^2)$$
$$\times \left[\varepsilon_{\mu\alpha\beta\sigma}(p-p')_{\sigma} - \frac{p_{\alpha}}{p^2} \varepsilon_{\mu\beta\rho\sigma} p_{\rho} p'_{\sigma} + \frac{p'_{\beta}}{p^2} \varepsilon_{\mu\alpha\rho\sigma} p_{\rho} p'_{\sigma} \right].$$
(47)

The anomaly condition in QCD reduces to

$$F_{2}(q^{2}, p^{2}) + q^{2}F_{1}(q^{2}, p^{2})$$

= $2\sum_{q} m_{q}G(q^{2}, p^{2}) - \frac{e^{2}}{2\pi^{2}}\sum_{q} e_{q}^{2}N_{c}.$ (48)

Because $T_{\mu\alpha\beta}(p, p')$ is nonsingular at $p^2 = 0$, we have $F_2(q^2, 0) = 0$. The functions $F_1(q^2, p^2)$, $F_2(q^2, p^2)$, and $G(q^2, p^2)$ can be described by dispersion relations in q^2 with no subtractions. Using these relations, we can prove the sum rule

$$\int_{4m^2}^{\infty} \operatorname{Im} F_1(t, p^2) \, \mathrm{d}t = \frac{e^2}{2\pi^2} \sum e_q^2 N_c \,, \tag{49}$$

where m^2 is the smallest of quark masses. The sum rule in (49) was proved in [30] for $p^2 < 0$, m = 0, in [28] for $p^2 = p'^2$, and

in [31] in the general cases where $p^2 \neq p'^2$. We note that (49) also holds for massive quarks. We consider the most interesting case where the axial current is the third component of the isovector current:

$$\dot{I}^{(3)}_{\mu5} = \bar{u}\gamma_{\mu}\gamma_{5}u - \bar{d}\gamma_{\mu}\gamma_{5}d.$$
(50)

We ignore the masses of the u- and d-quarks and assume that $p^2 = p'^2 = 0$. Combining (47) and (48), we obtain

$$T_{\mu\alpha\beta}(p,p') = -\frac{2\alpha}{\pi} N_{\rm c} \, \frac{q_{\mu}}{q^2} \left(e_{\rm u}^2 - e_{\rm d}^2 \right) \varepsilon_{\alpha\beta\lambda\sigma} p_{\lambda} p_{\sigma}' \,. \tag{51}$$

It follows from (51) that the transition of the isovector axial current into two photons occurs through an intermediate massless state. Such a state (in the limit $m_{\rm u}$, $m_{\rm d} \rightarrow 0$) is the π^0 -meson (Fig. 3). Combining the fact that $\langle 0 | j_{\mu 5}^{(3)} | \pi^0 \rangle = \sqrt{2} i f_{\pi} q_{\mu}$ and the anomaly condition, we can find the matrix element of the $\pi^0 \rightarrow 2\gamma$ decay,

$$M(\pi^0 \to 2\gamma) = A \varepsilon_{\alpha\beta\lambda\sigma} \varepsilon_{1\alpha} \varepsilon_{2\beta} p_{\lambda} p_{\sigma}', \qquad (52)$$

find the constant A, and calculate the width of $\pi^0 \rightarrow 2\gamma$ as

$$\Gamma(\pi^0 \to 2\gamma) = \frac{\alpha^2}{32\pi^3} \frac{m_{\pi}^3}{f_{\pi}^2} \,. \tag{53}$$

This result was first obtained in [32]. Under the assumption that $f_{\pi_0} = f_{\pi^+} = 130.7$ MeV, we obtain $\Gamma(\pi^0 \to 2\gamma)_{\text{theory}} =$ 7.73 eV from (53). It is difficult to estimate the accuracy of the prediction, but apparently it varies between 5 and 10%. The experimental value of this quantity averaged over all existing measurements (data for the year 2006) is $\Gamma(\pi^0 \rightarrow 2\gamma) = 7.8 \pm 0.6 \text{ eV}$ [33]. To achieve better accuracy for the theoretical prediction, we must (a) insert f_{π^0} instead of f_{π^+} in (53), and (b) allow the contribution of excited states (in addition to π^0) to the sum rule (49) for the isovector current at $p^2 = 0$. This program was implemented in Ref. [34], where it was shown that the difference $\Delta f_{\pi} = f_{\pi^0} - f_{\pi^+}$ is small: $\Delta f_{\pi}/f_{\pi} \approx -1.0 \times 10^{-3}$. Among the excited states, only the η-meson contributes significantly. Its contribution is determined by the value of the $\pi^0 - \eta$ mixing angle [35, 36] and the width $\Gamma(\eta \rightarrow 2\gamma) = 510 \text{ eV} [33]$. It was found in Ref. [34] that $\Gamma(\pi^0 \to 2\gamma)_{\text{theory}} = 7.93 \pm 1.5\%$. The most recent measurements in [37] yield $\Gamma(\pi^0 \to 2\gamma)_{\text{exp}} = 7.93 \pm 2\%, \pm 2.1\%$, i.e., the experimental data are in extremely good agreement with the theoretical predictions.



Figure 3. The diagram describing the transition of the isovector axial current (denoted by X) into two photons).

It would seem that Eqn (51) suggests that the existence of a massless (in the limit of massless u- and d-quarks) Goldstone π^0 -meson is a consequence of the axial anomaly described by the triangle diagrams in Fig. 1. This is not the case, however. A direct calculation of Im $F_1(q^2, p^2)$ (it is to this function that the intermediate π^0 -meson contributes) for $p^2 \neq 0$ shows [9, 28] that in this case, Im $F_1(q^2, p^2)$ is a regular function of q that tends to a constant as $q^2 \rightarrow 0$ and has no singularities of the $\delta(q^2)$ type, in contrast to the case $p^2 = 0$ described above. Thus, the amplitude $T_{\mu\alpha\beta}(p, p')$ corresponding to the transition of the axial current to two virtual photons and calculated according to the diagrams in Fig. 1 has no pole in q^2 at $q^2 = 0$. On the other hand, based on a chiral effective theory (e.g., see Ref. [38]), we can state that the transition amplitude of the axial current to two virtual photons must contain the contribution provided by the intermediate massless π^0 -meson (see Fig. 3). As shown in [6], the introduction of gluon lines into the diagrams in Fig. 1 does not change the expression for the anomaly. (Actually, this was shown in [6] to be true for QED, but there is no difference between QCD and QED in this aspect.) Thus, from examining the case where $p^2 \neq 0$, we conclude that the appearance of a massless π^0 -meson in the dispersion representation of the AVV form factor is not caused by an anomaly. The presence of massless Goldstone mesons (π, K, η) stems from the spontaneous breaking of chiral symmetry in the QCD vacuum. That there is a singularity at $q^2 = 0$ in the amplitude $T_{\mu\alpha\beta}(p, p')$ when $p^2 = 0$ is sometimes interpreted as the double nature of the anomaly, the ultraviolet and the infrared (e.g., see Ref. [3]). I believe that in view of the absence of such a singularity when $p^2 \neq 0$, this interpretation is faulty: the nature of an anomaly in QED and QCD stems from ultraviolet divergences, the singularity in the amplitudes at small distances. (In this respect, QED and QCD differ dramatically from the twodimensional Schwinger model, in which the origin of an anomaly is truly double (see Ref.[3]).)

For the eighth component of the octet current, the transition amplitude of the axial current to two real photons, $F_1(q^2, 0)$, has a pole at $q^2 = 0$ if $m_u = m_d = m_s = 0$. It is only natural to associate this pole with the η -meson. However, a relation for $\Gamma(\eta \rightarrow 2\gamma)$ similar to (53) differs dramatically from the experimental result. A possible explanation of such a discrepancy is the strong nonperturbative interaction of the type of instantons in a pseudoscalar channel mixing η - and η' -mesons [39]. In the case of a singlet axial current, the amplitude $j_{\mu 5}^{(0)} \rightarrow 2\gamma$ contains diagrams of the type shown in Fig. 2 (with virtual gluons instead of photons), their extensions, and nonperturbative contributions. Hence, we cannot expect reliable predictions concerning the width of $\eta' \rightarrow 2\gamma$ based on anomalies.

't Hooft hypothesized [40] that the singularities of the amplitudes calculated in QCD on the quark-gluon basis should reproduce themselves in calculations on the hadron basis. Obviously, this is true if both perturbative and nonperturbative interactions are taken into account. However, as a rule, we know nothing about the nonperturbative interactions. In the cases discussed above (except for the decay of π^0 into two real photons), 't Hooft's hypothesis does not hold [9].

8. Conclusion

1. An anomaly is an important and necessary element of quantum field theory.

- 2. An anomaly emerges because the amplitudes of quantum field theory contain ultraviolet singularities, in view of which it is necessary to augment the Lagrangian by renormalization conditions.
- An anomaly in QCD is related to the appearance of a new quantum number, the winding number.
- 4. The vacuum in QCD is a linear combination of an infinite number of vacua with different winding numbers.
- 5. Transitions between vacua with different winding numbers are tunnel transitions occurring along classical paths in the field space, self-dual solutions of QCD equations, or instantons.
- 6. The axial anomaly in QCD results in the appearance of zero modes in the Dirac equations for light quarks and points to the existence of spontaneous breaking of chiral symmetry in the QCD vacuum, the existence of a quark condensate.
- The axial anomaly predicts the width of the $\pi^0 \rightarrow 2\gamma$ decay 7. with a high accuracy ($\sim 2\%$), a result corroborated by experiments.

References

- Treiman S B et al. Current Algebra and Anomalies (Princeton Series 1. in Physics) (Princeton, NJ: Princeton Univ. Press, 1985)
- 2. Collins J C Renormalization (Cambridge: Cambridge Univ. Press, 1984)
- 3. Shifman M A Phys. Rep. 209 341 (1991)
- 4. Peskin M E, Schroeder D V An Introduction to Quantum Field Theories (Reading, Mass.: Addison-Wesley, 1995)
- Adler S L Phys. Rev. 177 2426 (1969) 5.
- 6. Adler S L, Bardeen W A Phys. Rev. 182 1517 (1969)
- Bell J, Jackiw R Nuovo Cimento 51 47 (1969) 7.
- Jackiw R "Field theoretical investigations in current algebra", in 8. Treiman S B et al. Current Algebra and Anomalies (Princeton Series in Physics) (Princeton, NJ: Princeton Univ. Press, 1985) pp. 81-210 9
- Ioffe B L Int. J. Mod. Phys. A 21 6249 (2006)
- 10. Adler S L "Anomalies to all orders", in 50 Years of Yang-Mills Theory (Ed. G 't Hooft) (Singapore: World Scientific, 2005) pp. 187-228; hep-th/0405040
- 11. Anselm A A, Iogansen A A Pis'ma Zh. Eksp. Teor. Fiz. 49 185 (1989) [JETP Lett. 49 214 (1989)]
- 12 Weinberg S Phys. Rev. D 11 3583 (1975)
- Belavin A A et al. Phys. Lett. B 59 85 (1975) 13.
- Gribov V N unpublished 14
- 15. Jackiw R, Rebbi C Phys. Rev. Lett. 37 172 (1976)
- Callan C G (Jr), Dashen R F, Gross D J Phys. Lett. B 63 334 (1976) 16.
- 17. Bitar K M, Chang S-J Phys. Rev. D 17 486 (1978)
- Altarev I S et al. Phys. Lett. B 276 242 (1992) 18.
- Atiyah M F, Singer I M Ann. Math. 87 484 (1968); 93 119 (1971) 19
- 20. Crewther R J Phys. Lett. B 70 349 (1977)
- Di Vecchia P, Veneziano G Nucl. Phys. B 171 253 (1980) 21.
- 22. Ioffe B L Yad. Fiz. 62 2226 (1999) [Phys. At. Nucl. 62 2052 (1999)]
- 23
- Ioffe B L, Oganesian A G Phys. Rev. D 57 R6590 (1998)
- 24. Burkert V D, Ioffe B L Phys. Lett. B 296 223 (1992)
- 25. Burkert V D, Ioffe B L Zh. Eksp. Teor. Fiz. 105 1153 (1994) [JETP 78 619 (1994)]
- 26. Ioffe B L Survey High Energy Phys. 14 89 (1999); hep-ph/9804238
- 27. Eletsky V L, Ioffe B L, Kogan Ya I Phys. Lett. B 122 423 (1983)
- 28. Hořejši J Phys. Rev. D 32 1029 (1985)
- 29. Bass S D et al. Zh. Russ. Fiz. Obshch (J. Moscow Phys. Soc.) 1 317 (1991)
- 30. Frishman Y et al. Nucl. Phys. B 177 157 (1981)
- Veretin O L, Teryaev O V Yad. Fiz. 58 2266 (1995) [Phys. At. Nucl. 31. 58 2150 (1995)]
- Dolgov A D, Zakharov V I Nucl. Phys. B 27 525 (1971) 32.
- Yao W-M et al. (Particle Data Group) "Review of Particle Physics" 33. J. Phys. G: Nucl. Part. Phys. 33 1 (2006)
- 34. Ioffe B L, Oganesian A G Phys. Lett. B 647 389 (2007)
- 35. Ioffe B L Yad. Fiz. 29 1611 (1979) [Sov. J. Nucl. Phys. 20 827 (1979)]
- Gross D J, Treiman S B, Wilczek F Phys. Rev. D 19 2188 (1979) 36.
- 37. de Jager K Prog. Part. Nucl. Phys. (Available online 28 December 2007); arXiv:0801.4520

- 38. Ioffe B L Usp. Fiz. Nauk 171 1273 (2001) [Phys. Usp. 44 1211 (2001)]
- 39. Geshkenbein B V, Ioffe B L Nucl. Phys. B 166 340 (1980)
- 't Hooft G, in Recent Developments in Gauge Theories (NATO Advanced Study Inst. Series, Ser. B, Vol. 59, Eds G 't Hooft et al.) (New York: Plenum Press, 1980) p. 241

PACS number: **03.30.** + **p** DOI: 10.1070/PU2008v051n06ABEH006552 DOI: 10.3367/UFNr.0178.2008061.0663

The theory of relativity and the Pythagorean theorem

L B Okun

1. Introduction

The report "Energy and mass in the works of Einstein, Landau and Feynman" that I was preparing for the Session of the Division of Physical Sciences of the Russian Academy of Sciences (DPS RAS) on the occasion of the 100th anniversary of Lev Davidovich Landau's birth was to consist of two parts, one on history and the other on physics. The history part was absorbed into the article "Einstein's formula: $E_0 = mc^2$. 'Isn't the Lord laughing?'' that appeared in the May issue of Uspekhi Fizicheskikh Nauk [Physics-Uspekhi] journal [1]. The physics part is published in the present article. It is devoted to various, so to speak, technical aspects of the theory, such as the dimensional analysis and fundamental constants c and \hbar ; the kinematics of a single particle in the entire velocity range from 0 to c; systems of two or more free particles; and the interactions between particles: electromagnetic, gravitational, etc. The text uses the slides of the talk at the session of the Section of Nuclear Physics of the DPS RAS in November 2007 at the Institute for Theoretical and Experimental Physics (ITEP). My goal was to present the main formulas of the theory of relativity in the simplest possible way, using mostly the Pythagorean theorem.

2. Relativity

The advanced standpoint. The history of the concept of mass in physics runs to many centuries and is very interesting, but I leave it aside here. Instead, this will be an attempt to look at mass from an advanced standpoint. I borrowed the words from the famous title of Felix Klein's *Elementary Mathematics from an Advanced Standpoint* (traditionally translated into Russian incorrectly as *Elementary Mathematics from the Standpoint of Higher Mathematics*. See V G Boltyanskii's foreword to the 4th Russian edition). The advanced modern standpoint based on principles of symmetry in general and on the theory of relativity in particular makes it possible to avoid inevitable terminological confusion and paradoxes.

The principle of relativity. Ever since the time of Galileo and Newton, the concept of relativity has been connected with the impossibility of detecting, by means of any experiments, a translational (uniform and rectilinear) motion of closed space (for instance, inside a ship) while remaining within this space. At the turn of XIX and XX centuries Poincaré gave to this idea the name 'the principle of relativity'.¹ In 1905 Einstein generalized this principle to the case of the existence of the limiting velocity of propagation of signals. (The finite velocity of propagation of light has been discovered by Römer already in 1676). Planck called the theory constructed in this way 'Einstein's theory of relativity'.

Mechanics and optics. Newton tried to construct a unified theory uniting the theory of motion of massive objects (mechanics) and the theory of propagation of light (optics). In fact, it became possible to create the unified theory of particles of massive matter and of light only in the XXth century. It was established on the road to this vantage ground of truth that light is also a sort of matter, just like the massive stuff, but that its particles are massless. This interpretation of particles of light — photons — continues to face resistance from many students of physics, and even more from physics teachers.

3. Dimensions

Units in which c = 1. The maximum possible velocity is known as the speed of light and is denoted by c. When dealing with formulas of the theory of relativity it is convenient to use a system of units in which c is chosen as a unit of velocity. Since c/c = 1, using this system means that we set c = 1 in all formulas, thus simplifying them greatly. If time is measured in seconds, then distance in this system of units should be measured in light seconds: one light second equals 3×10^{10} cm.

Poincaré and c. One of the creators of the theory of relativity, Henri Poincaré, when discussing in 1904 the fact that c is found in every equation of electrodynamics, compared the situation with the geocentric theory of Ptolemy's epicycles in which every relation between motions of celestial bodies included the terrestrial year. Poincaré expressed his hope that the future Copernicus would rid electrodynamics of c[3]. However, Einstein showed already in 1905 that c was to play the key role as the limit for the velocity of signal propagation.

Two system of units: SI and c = 1**.** The unit of velocity in the International System of Units SI, 1 m s⁻¹, is forced on us by convenience arguments and by standardization of manufacturing and commerce but not by the laws of Nature. In contrast to this, c as a unit of velocity is imposed by nature itself when we wish to consider fundamental processes of Nature.

Dimensional factors. Consider some physical quantity a. Let us denote by [a] the dimension of the quantity a. The dimension of a definitely changes if it is multiplied by any power of the universal constant c but its physical meaning remains unaffected. In what follows I explain why this is so.

Velocity, momentum, energy, mass. The dimensions of momentum, mass, and velocity of a particle are usually related by the formula $[\mathbf{p}] = [m][\mathbf{v}]$ while the dimensions of energy, mass, and velocity are related by the formula $[E] = [m][\mathbf{v}^2]$.

Let us introduce dimensionless velocity \mathbf{v}/c and from now on denote this ratio as \mathbf{v} . Likewise, referring to momentum \mathbf{p} we actually mean the ratio \mathbf{p}/c . When speaking of energy, we

¹ This sentence was added by the Author in the English proof.

actually mean the ratio $e = E/c^2$. Obviously, the dimensions of **p**, *e*, and *m* become identical and therefore, these quantities can be measured in the same units, for example, in grams or electron-volts, as is customary in elementary particle physics.

On the letter *e* denoting energy. Choosing *e* as the notation for energy may invite the reader's ire since this symbol traditionally stands for electron and electric charge. However, this choice cannot cause confusion and, importantly, it will lead to a compact form of formulas for a single particle, always reminding us that these formulas were written using the system of units in which c = 1. On the other hand, it will be clear a little later that the letter *E* is a convenient notation for the energy of two or more particles.

I happened to see Einstein's formula with a lower-case *e* on a billboard on Rublevskoye highway in Moscow. I wonder, why should this *e* irritate physicists?

On the difference between energy and frequency. Two paragraphs ago I insisted that $e = E/c^2$ is energy even though its dimension is that of mass. In that case it is logical to ask why $\omega = E/\hbar$ is not energy but frequency? Indeed, the quantum of action \hbar , like the speed of light c, is a universal constant. The answer to this question can be found by considering how e and ω are measured. E and e are measured by the same procedure, say, using a calorimeter, while frequency is measured in a drastically different manner, say, using clocks. Therefore, the equality $\omega = E/\hbar$ informs us of the link between two different types of measurement, while the equality $e = E/c^2$ carries no such information. Arguments similar to those concerning frequency hold equally well for wavelength. I have to emphasize that these metrological distinctions are mostly of a historical nature since in our day atomic clocks operate on the difference between atomic energy levels.

4. Single particle

Relative and absolute quantities. The kinetic energy of any body is a relative quantity: it depends on the reference frame in which it is measured. The same is true for the momentum of a body and its velocity. In contrast to them, the mass of a body is an absolute quantity: it characterizes the body as such, irrespective of the observer. The rest energy of a body (see below) is also an absolute quantity since the frame of reference is fixed in it once and for all — 'nailed to it'.

Invariant mass. The mass of a body is defined in the theory of relativity by the formula

$$m^2 = e^2 - p^2 \,. \tag{1}$$

Here and in what follows $p = |\mathbf{p}|$. Likewise, $v = |\mathbf{v}|$.

Note that energy and momentum of a given body are not bounded from above while the mass of the body is fixed. Formula (1) is the simplest relation between energy, momentum, and mass that one could write 'off the top of one's head'. (The relation between e, \mathbf{p} , and m cannot be linear since \mathbf{p} is a vector while e and m are scalars in three-dimensional space.) We shall see now that formula (1) has another, much more profound theoretical foundation.

The 4-momentum. Minkowski was the first to point out that the theory of relativity gains the simplest form if considered in

four-dimensional space – time [4]. Energy and momentum in the theory of relativity form a four-dimensional energy-momentum vector p_i (i = 0, a), where $p_0 = e$, $p_a = \mathbf{p}$, and a = 1, 2, 3.

Mass is the Lorentz scalar that characterizes the length of the 4-vector $p_i: m^2 = p_i^2 = e^2 - \mathbf{p}^2$; four-dimensional space is pseudo-Euclidean, which explains the minus sign in the formula for length squared. (The reader will recall that $\mathbf{p}^2 = p^2$.) Another way to clarify why the sign is negative is by introducing the imaginary momentum ip. Then $m^2 = e^2 + (i\mathbf{p})^2$ and we are dealing with the Pythagorean theorem for such a pseudo-Euclidean right triangle in which the hypotenuse *m* is shorter than the larger cathetus *e*.

Relation between momentum and velocity. The momentum of a body is related to its velocity **v** by the formula

$$\mathbf{p} = e\mathbf{v}\,.\tag{2}$$

This formula satisfies in the simplest manner the requirement that the momentum 3-vector be proportional to the velocity 3-vector and that the dimensional proportionality coefficient not vanish for the massless photon.

Conservation of the thus defined momentum in the theory of relativity is implied by the uniformity of 3-space while conservation of energy is implied by the uniformity of time (Noether's theorem).

The Pythagorean theorem. Formula (1) is shown in Fig. 1 by an ordinary right triangle in which *m* and *p* are catheti and *e* is the hypotenuse.

Transition from m \neq 0 to m = 0. Formula (1) is obviously valid at m = 0 while formula (2) holds for v = 1. This implies that there is a smooth transition from massless particles to massive, when the energy of the latter particles greatly exceeds their mass.

Physics from p = 0 to p = e. Let us consider formulas (1) and (2) first at zero momentum, then in the limit of very low momenta (when $p \ll m$), and then in the limit of very high momenta when $p \sim e \gg m$, and finally in the case of massless photons.

We will call the case of very small momenta and velocities the Newtonian case, and that of very high momenta and velocities close to the speed of light, the ultrarelativistic case. We will start with zero momentum.





5. Rest energy

Zero momentum. If momentum is zero, then in the case of a massive particle the velocity is also zero and energy e is by definition equal to the rest energy e_0 . (The subscript 0 reminds us that here we are dealing not with the energy of a given body in general but with its energy precisely in the case when its momentum is zero!) Hence equation (1) implies

$$e_0 = m \,. \tag{3}$$

If, however, the particle is massless, then equation (1) at p = 0 implies that $e = e_0 = 0$.

Horizontal 'biangle'. If p = 0, then the triangle shown in Fig. 1 'collapses' to a horizontal 'biangle' (Fig. 2).



Figure 2.

Einstein's great discovery. In units in which $c \neq 1$, equation (3) has the form

$$E_0 = mc^2 \,. \tag{4}$$

The realization that ordinary matter at rest stores an enormous amount of energy in its mass was Einstein's great discovery.

The 'famous formula'. Equation (4) is very often written (especially in popular physics literature) in the form of 'Einstein's famous equation' that drops the subscript 0:

$$E = mc^2 \,. \tag{5}$$

This simplification, to which Einstein himself sometimes resorted, might seem innocuous at first glance, but it results in unacceptable confusion in understanding the foundations of physics. In particular, it generates a totally false idea that 'according to the theory of relativity' the mass of a body is equivalent to its total energy and, as an inevitable result, depends on its velocity. ('Wished to make it simpler, got it as always'.²)

No experiment can disprove the 'famous formula'. Very clever people thought up this formula in such a way that it never contradicts experiments. However, it contradicts the essence of the theory of relativity. In this respect, the situation with the 'famous formula' is unique — I do not know another case that could be compared with this one.

This is not a matter of taste but of understanding. You hear time and again that the introduction of momentum-dependent mass is 'a matter of taste'. Of course, one can write the letter *m* instead of E/c^2 and even call it 'mass', although it is no more sensible than writing *p* instead of E/c and calling it

² A paraphrase of former Russian Prime Minister Chernomyrdin's 'statement of the day': "Wished to make it better got as always." (*Note added by the Author in translation.*)

'momentum'. Alas, this 'dress changing' introduces unnecessary and bizarre notions — relativistic mass and rest mass m_0 — and creates an obstacle to understanding the theory of relativity. A well-known Russian proverb comes to mind: "Call me a pot if you wish but don't push me into the oven." Unfortunately, people who call E/c^2 'mass' do place this 'pot' into the 'oven' of physics teaching.

Longitudinal and transverse masses. In addition to relativistic mass, concepts of intense use at the beginning of the XXth century were the transverse and longitudinal masses: m_t and m_l . This longitudinal masses increased as $(e^3/m^3)m$ and 'explained' — in terms of Newton's formula F = ma — why a massive body cannot be accelerated to the speed of light. Then it was forgotten and such popularizers of the theory of relativity as Stephen Hawking started to persuade their readers that even much gentler growth of mass with velocity ((e/m)m) could explain why the velocity of a massive body cannot reach *c*. I single out Hawking only because, printed on the dust jacket of the Russian edition of his latest popular science book [5], which advertises the formula $E = mc^2$, we see this text: "Translated into 40 languages. More than 10 million copies sold worldwide."

False intuition. After my talk at the ITEP A N Skrinskii told me that the notion of relativistic mass hampered a wellknown physicist's understanding that a relativistic electron colliding with an electron at rest can transfer all its energy to the latter. Indeed, how could a heavy baseball bat transfer all its energy to the lightest ping-pong ball? In physics, as in daily life, people very often rely on intuition. This is why it is so important, when studying the theory of relativity, to work out the relativistic intuition and mistrust nonrelativistic intuition. (In order to 'feel' how an electron at rest can receive the entire energy of a moving electron it is sufficient to use their centerof-inertia frame to consider scattering by 180 degrees, and then return back to the laboratory frame.)

6. Newtonian mechanics

Momentum in Newtonian mechanics. Newtonian mechanics describes with high accuracy the motion of macroscopic bodies in a terrestrial environment and of massive celestial bodies because their velocities are much smaller than the speed of light. For instance, the velocity of a bullet is of the order of 1 km s⁻¹, which corresponds to v = 1/300000 and $v^2 = 10^{-11}$. In this situation equation (2) reduces to

$$\mathbf{p} = m\mathbf{v} \,. \tag{6}$$

Equation (1) is schematically shown in the Newtonian limit in Fig. 3.

The side of the triangle representing p in Fig. 3 is far too long. Scaled correctly, it should be a few microns.





Kinetic energy e_k . It is reasonable to rewrite formula (1) for low velocities so as to isolate the contribution of the short cathetus:

$$e^2 - m^2 = p^2, (7)$$

and then to present it in the form

$$(e-m)(e+m) = p^2$$
. (8)

This allows us to obtain a nonrelativistic expression for kinetic energy without resorting to the conventional series expansion of the square root. We take into account that the total energy e is the sum of rest energy e_0 and kinetic energy e_k and therefore $e = m + e_k$.

Energy in Newtonian mechanics. In the Newtonian limit we have $e_k \ll m$ (e.g. for a bullet $e_k/m = 10^{-11}$). Energy can therefore be replaced with high accuracy by mass *m* in formula (2) for momentum and in the factor (e + m) in equation (8). This last equation immediately implies an expression for kinetic energy e_k in Newtonian mechanics:

$$e_{\rm k} = \frac{p^2}{2m} = \frac{mv^2}{2} \,. \tag{9}$$

Potential energy. In addition to velocity-dependent kinetic energy, an important role in nonrelativistic mechanics is played by potential energy, which depends only on the position (coordinate) of the body. The sum of kinetic and potential energy is conserved at any instance of time. The potential energy of a body placed in an external field of force is defined to within an arbitrary additive constant because the force acting on the body equals the gradient of potential energy. In a similar manner, the potential energy of interaction of several bodies depends only on their positions at the moment of interaction. However, in the theory of relativity any interaction propagates at a finite velocity. Hence, potential energy is an essentially nonrelativistic concept.

Newton and modern physics. Newton's flash of genius marked the birth of modern science. The post-Newtonian progress of science is fantastic. Today's understanding of the structure of matter is radically different from Newton's. Nevertheless, even in the XXIst century many physics textbooks continue to use Newton's equations at energies $e_k \ge e_0$, which exceed the limits of applicability of Newton's mechanics ($e_k \le e_0$) by many orders of magnitude.

If some professors prefer to insist on keeping up with this tradition of velocity-dependent mass, they ought to at least familiarize their students with the fundamental concepts of mass and rest energy, and with the true Einstein equation $E_0 = mc^2$.

7. Ultrarelativism

High energy physics. Let us now consider in some detail the limiting case in which $e/m \ge 1$. The ratio of energy and mass characteristic for high energy physics is precisely this. For example, this ratio for electrons in the LEP (Large Electron – Positron) Collider at CERN was $e/m = 10^5$, since m = 0.5 MeV and e = 50 GeV. For protons in the LHC (Large Hadron Collider), which is located in the same tunnel where the LEP was in previous years, we find $e/m \sim 10^4$. (Here, $m \sim 938$ MeV, $e \sim 7$ TeV.)



Figure 4.

A vertical triangle. The triangle for protons in the LHC is drawn highly schematically in Fig. 4. Its base is in fact shorter than its hypotenuse by four orders of magnitude.

The neutrino. Neutrinos are even more ultrarelativistic particles: their masses are a fraction of one electron-volt and their energies reach several MeV for neutrinos emerging from the Sun and nuclear reactors, and several GeV for neutrinos generated in particle decays in cosmic rays and in accelerators. The base of the triangle shown schematically in Fig. 4 is much shorter at these energies than its vertical cathetus and its hypotenuse.

Neutrino oscillations and $m^2/2e$. Equation $(e-p)(e+p) = m^2$ immediately implies that $e - p \simeq m^2/2e$. The differences between the masses of three neutrinos v_1 , v_2 , v_3 possessing definite masses in a vacuum result in oscillations between neutrinos having no well-defined masses but possessing certain flavors: v_e , v_{μ} , v_{τ} . (This phenomenon is similar to well-known beats that occur when several frequencies interfere.) The neutrino oscillation data ³ give

$$\Delta m_{21}^2 = (0.8 \pm 0.04) \times 10^{-4} \text{ eV}^2$$
,
 $\Delta m_{32}^2 = (25 \pm 6) \times 10^{-4} \text{ eV}^2$.

The photon. The photon mass is so small that no experiment has been able to detect it. Hence, it is usually assumed that the photon mass equals zero. This means that for a photon e = p and the triangle shown in Fig. 4 collapses to a vertical biangle (Fig. 5).

The photon and rest energy? It is logical to conclude the discussion of single-particle mechanics by returning to the question: is the concept of rest energy e_0 applicable to massless photon?

It may seem at first glance that it is not, since a photon propagates at the speed c, however small its energy is, so that 'a rest for it is but a dream'⁴. This being so, how can we use the equality $e_0 = 0$ if the photon is never at rest? We can because our e_0 is defined as the energy corresponding to zero

³ These values are updated in this English proof by the Author.

⁴ This is a paraphrase of the famous line from Alexandr Block. (*Note added in translation.*)



momentum, not velocity. Obviously this energy is zero for the photon with p = 0: this is implied by equation (1). If a particle has m = 0, p = 0, e = 0 and biangle of Fig. 5 collapses to a point, we can say that it 'passed away to the state of eternal rest'. Looking at the limiting transition to zero mass, we can show that the reference frame in which a photon is 'eternally at rest' has to be rigidly connected to another 'eternally resting' photon. Consequently, the value $e_0 = 0$ at m = 0 is in perfect agreement with the limiting transition.

8. Two free particles

Collision of two particles. Colliders. If two particles collide at relativistic energies, a comparison of the reference frame in which one of them is at rest with a reference frame in which their common center of inertia is at rest demonstrates the advantages of the latter. We already saw this in the case commented on by A N Skrinskii. If the momenta of the colliding particles are equal and oppositely directed, as for example in the LHC or LEP collider, then practically the entire energy of the colliding particles may be spent on the creation of new particles.

Mass of a system of particles. The total energy E and the total momentum **P** of an isolated system of particles are conserved. Energy and momentum being additive, for two free particles we have

 $E = e_1 + e_2 \,, \tag{10}$

$$\mathbf{P} = \mathbf{p}_1 + \mathbf{p}_2 \,. \tag{11}$$

We now define the quantity M by the formula

$$M^2 = E^2 - \mathbf{P}^2 \,. \tag{12}$$

Masses are additive at v = 0. Equation (12) is invariant under Lorentz transformations, as is equation (1). Therefore, it is logical to refer to M as the mass of a system of two particles. In the static limit, when p_1 and p_2 equal zero, equation (12) implies that

$$M = e_{01} + e_{02} = m_1 + m_2 \,. \tag{13}$$

In the Newtonian limit, M equals the sum of the masses of the two particles with an accuracy of $(v/c)^2$, i.e. the masses are practically additive.

Masses are not additive at $v \neq 0$. However, *M* and the masses m_1 and m_2 are practically unrelated at high velocities. For

instance, M exceeds the electron mass in the LEP collider or the proton mass in the LHC by four orders of magnitude (see Section 7). The value of M is crucially dependent on the relative directions of the momenta of two particles, since the sum of two vectors is a function of the angle between them. Thus, we have for two photons moving in the same direction

$$P = |\mathbf{P}| = |\mathbf{p}_1 + \mathbf{p}_2| = p_1 + p_2.$$
(14)

Collinear photons. For such photons $p_1 = e_1$, and $p_2 = e_2$. Therefore, for two photons moving in the same direction we can write

$$P = p_1 + p_2 = e_1 + e_2 = E.$$
(15)

Equation (12) then implies that in this case the mass of a pair of photons M = 0. And this means that the mass of a 'needle' light beam is zero.

What if photons fly away from each other? However, if photons fly away in opposite directions with equal energies, then $\mathbf{p}_1 = -\mathbf{p}_2$ and $\mathbf{P} = 0$. In that case, the rest energy of two photons simply equals the sum of their energies and the mass of this system is

$$M = E_0 = 2e. (16)$$

Shock. Of course, the statement that a pair of two massless particles has an enormous mass may shock the unprepared reader. Is there any sense in speaking of the rest energy of two photons if 'rest is but a dream' to either of them? What is at rest in this case?

The answer is obvious. The entity at rest is the geometric point — the center of inertia of the two photons. While the rest energy for one particle is the energy hidden in its mass, for two photons it is simply the sum of their energies (kinetic energies!) in the reference frame in which their momenta are equal in magnitude and opposite in direction. There is no hidden energy in this case!

What does it mean 'to be conserved'? When saying that energy is conserved, we mean that the sum of the energies of particles entering a reaction equals the sum of the energies of particles created as a result of this reaction. The statement on the conservation of momentum has a similar meaning. However, since momentum is a vector quantity, now we are dealing with a vector sum of momenta. (In the case of momenta we speak about three independent conservation laws: conserved are the sums of projections of momenta in three mutually orthogonal directions.)

The conserved quantities are thus $E = \sum e_i$ and $\mathbf{P} = \sum \mathbf{p}_i$. As for the energies of individual particles e_i , their momenta \mathbf{p}_i in the laboratory reference frame, they are conserved only in elastic forward scattering.

Here, it is important to stress the difference between the concepts of additivity and conserved. The former concept refers to the state of a system of free particles, the latter refers to the process of interaction of the particles.

Is mass conserved? With *E* and **P** conserved, the mass *M* of a system (a set) of particles, defined by the formula $M^2 = E^2 - \mathbf{P}^2$, must be conserved as well. In contrast to energy and momentum, however, mass is not additive:

 $M \neq \sum m_i$. Some authors talk about the non-additivity of mass as if it were identical to its non-conservation (e.g. we find this statement in § 9 of *Field Theory* by Landau and Lifshitz [6].) In fact, as I emphasized above, in general neither masses nor energies or momenta are conserved for individual particles participating in a reaction; not even the particles themselves are. Hence, it is incorrect to speak of mass nonconservation as something in contrast to conservation of energy and momentum.

Einstein's thought experiment. Of course, the concept of the mass of two photons flying away from each other looks rather strange. However, it was by using this very idea that Einstein came to discover the rest energy of a massive body in 1905. He noticed that having emitted 'two amounts of light' in opposite directions, the body at rest continues to stay at rest but that its mass in this thought experiment diminishes. In the laboratory reference frame both the body and the center of inertia of the two photons are at rest. Consequently, the mass of the initial body equals the sum of two masses: that of the resulting body and that of the system of two photons.

Positronium annihilation. *Nihil* in Latin means *nothing*. A positronium is an 'atom' consisting of a positron and an electron. The reaction in which a positronium converts to two photons $e^+e^- \rightarrow \gamma\gamma$ was given the name annihilation, perhaps because at that time photons were not considered particles of matter. Annihilation conserves *M* because *E* and **P** are conserved. In the initial state *M* equals the sum of masses of the electron and the positron [minus the binding energy, which is small and in this context irrelevant (see below)]. In the final state *M* equals the sum of energies of two photons in the positron thus transforms completely into the energy (kinetic) of the photons, but the masses of the initial and final states are identical in this process, exactly as follows from the conservation of total energy and total momentum.

Meson decays. Likewise, when a K meson decays into two or three π mesons, the kaon's rest energy transforms into the sum of total energies of the pions, each of which has the form $e = e_k + m$. However, the mass of a system of two or three pions produced in the decay of a kaon equals the kaon mass.

What do we call 'matter'? In any decay the rest energy transforms into the energy of motion, while the total energy of an isolated system remains conserved. The mass of the system is also conserved but the masses of its individual particles are not. Massive particles decay into less massive particles, or sometimes into massless ones. In elementary particle physics we call 'particles of matter' not only massive particles such as protons and electrons, but also very light neutrinos and massless photons, and even gravitons (see below). Today's quantum field theory treats all of them on an equal basis.

Energy without particles? Matter does not disappear in decay and annihilation reactions leaving behind only energy like the Cheshire cat would leave behind only its smile. In all these processes the carriers of energy are particles of matter. Energy without matter ('pure energy') has never been observed in any process studied so far.

True, this is not so for so-called dark energy, which was discovered in the last years of the XXth century. Dark energy

manifests itself in the accelerating expansion of the Universe. (The evidence for this accelerating expansion is found in recession velocities of remote supernovas.) Three-fourths of the entire energy in the Universe is dark energy and its carrier appears to be the vacuum. The remaining quarter is carried by ordinary matter (5%) and dark matter (20%). Dark energy does not affect processes with ordinary matter observed in laboratories. In a laboratory experiment energy is always carried by particles.

9. Non-free particles

Bodies and particles. All physical bodies consist of elementary particles. Such elementary particles as the proton and the neutron are themselves made up of 'more elementary particles' — quarks and gluons. Such particles as the electron and the neutrino appear at our current level of understanding as truly elementary particles. The feature common for the proton and the electron is that the masses of all protons in the world are strictly identical, as are the masses of all electrons. In contrast to this, the masses of all macroscopic bodies of the same type, say, of all 10-cent coins, are only approximately equal. Practically the difference between two coins arises because the process of minting coins is far from being ideal. In principle, the mass of a coin is not well defined because different energy levels of a coin are practically degenerate, while the mass of the nearest excited state of a proton exceeds the proton mass by several hundred MeV. Therefore Nature mints ideally identical protons.⁵

Mass of gas. In all the cases discussed above, particles moved away freely when the mass of the system of particles was greater than the sum of their masses. Let us turn now to a situation in which they are not free to move away. This situation is found, for example, in the frequently discussed thought experiment with a gas of molecules or photons in a closed vessel at rest. The total momentum of this gas is zero because the gas is isotropic: $\mathbf{P} = \sum \mathbf{p}_i = 0$. Hence, the total mass M of this gas equals its total energy E (and in this case it is identical to E_0) and hence to the sum of energies of individual particles: $M = E = \sum e_i$.

Mass of a heated gas. When gas in a nonmoving vessel is heated, its total momentum remains unchanged and equal to zero while the total energy increases because the kinetic energy of every particle increases. As a result, the mass of the gas as a whole increases, while the mass of each individual particle remains unchanged. (Sometimes a wrong statement may be encountered in the literature that the masses of particles (or photons) increase as their kinetic energies are increased.)

Mass of a hot iron. In the same manner, the mass of an iron must increase as it heats up, even though the masses of the vibrating atoms remain the same. However, the set of formulas (10)-(12) written for a system of free particles cannot be applied to the iron since the particles (atoms in this case) are not free but are tied into the crystal lattice of the metal. Obviously, an increase in the iron mass is too small to be measurable.

⁵ These two sentences were modified in the English proof by the Author.

10. Atoms and atomic nuclei

On formulas (10)-(12). Why are formulas (10)-(12) unsuitable for dealing with such non-free particles as electrons in atoms and nucleons in atomic nuclei? First and foremost, on account of the uncertainty relation these particles do not possess precisely defined momenta. The smaller the volume to which they are confined, the greater is the uncertainty of their momenta.

Uncertainty relation. The laws of quantum mechanics, and the uncertainty relation as one among them, are very important both for atoms and for nuclei. As we know, the product of the momentum uncertainty Δp and the coordinate uncertainty Δx must be not smaller than the quantum of action \hbar . Hence, particles within atoms have no definite momenta and only possess a certain total momentum.

Energy of the field. Another reason why formulas (10)-(12) are not valid inside atoms is the fact that the space between individual particles in an atom is essentially not empty but filled with a material medium, i.e. physical fields. The space inside the atom is filled with an electromagnetic field and the space inside a nucleus, by a much denser and stronger field, often described as the meson field.

Real and virtual particles. In classical theory particles and fields are concepts that cannot be reduced to one another. In quantum field theory we use the language of Feynman diagrams, which reduce the concept of a field to that of a virtual particle for which $e^2 - \mathbf{p}^2 \neq m^2$. We say about such particles that they are off mass shell. (Particles that are called on mass shell are real particles and for them $e^2 - \mathbf{p}^2 = m^2$.) Also, the 4-momentum $p_i = (e, \mathbf{p})$ is conserved at each vertex of the diagram.

Binding energy. As a result of the presence of the field, we need to take into account in formula (10), $E = e_1 + e_2$, the field energy of two closely interacting particles, say, in the deuteron, the nucleus of heavy hydrogen. Consequently, $M < m_1 + m_2$. The quantity $\varepsilon = m_1 + m_2 - M$ is known as the binding energy.

The mass of the deuteron is less than the mass of the proton plus that of the neutron of which deuteron consists. The binding energy of nucleons in deuteron is 2.2 MeV. To break deuteron into nucleons we need to spend an amount of energy equal to or greater than the binding energy. The atomic nuclei of all other elements of the periodic Mendeleyev table also owe their existence to the binding energy of their nucleons in the nucleus.

Fusion and fission of nuclei. We know that the binding energy per nucleon rises to a maximum at the beginning of the periodic Mendeleyev Table for the helium nucleus and in the middle of the Table for the iron nucleus. This is why huge amounts of kinetic energy are released when helium is formed from hydrogen in fusion reactions in the Sun and in hydrogen bombs. In nuclear reactors and atomic bombs, kinetic energy is released by fission reactions when heavy nuclei of uranium and plutonium break into lighter nuclei from the middle of the periodic Mendeleyev Table.

Chemical reactions. Substantially lower energy, on the order of electron-volts, is released in chemical reactions. It is caused

by differences in binding energies in various chemical compounds. However, the source of kinetic energy in both chemical and nuclear reactions is the difference between the masses of initial and final particles (molecules or nuclei) that take part in these reactions.

Since molecules and even atomic nuclei are nonrelativistic bound systems and the concept of potential energy is applicable to their components, the corresponding mass differences can be calculated using this concept. Thus, one can explain the released energy in terms of potential energy transforming into kinetic energy.

Coulomb's law. The binding energy of electrons in atoms is much lower than the electron mass. Hence, the concept of binding energy in atoms can be explained in terms of the nonrelativistic concept of potential energy. The binding energy ε equals (with a minus sign) the sum of positive kinetic energy of the bound particle and its negative potential energy. The potential energy of, say, an electron in a hydrogen atom is given by Coulomb's law (in units, where \hbar , c = 1):

$$U = -\frac{\alpha}{r} \,, \tag{17}$$

where $\alpha = e^2/\hbar c = 1/137$ and *e* is the electron charge.

More about potential energy. The concept of potential energy is defined only in the Newtonian limit (see Landau and Lifshitz, *Mechanics*, § 5 "The Lagrange function of a system of material points" and § 6 "Energy") [7]. The sum of kinetic and potential energies is conserved. If one of the two interacting particles is essentially relativistic, or both are, the concept of potential energy is inapplicable.

Electromagnetic field. ⁶ The Coulomb field in the theory of relativity is the 0th component of the 4-potential of the electromagnetic field A_i (i = 0, 1, 2, 3). The source of the field of a particle with charge e is the 4-dimensional electromagnetic current given in the next paragraph. The interaction between two moving particles works through propagation of the field from one charge to the other. It is described by the so-called Green's function or the propagator of an electromagnetic field. (In quantum electrodynamics, we speak of propagation of virtual photons. The potential A_i is a 4-vector because the spin of the photon equals unity.)

Important clarification. If a virtual photon carries away a 4-momentum q, then 4-momenta of the charged particle prior to the emission of a photon p_{in} and after its emission p_{fi} satisfy the condition $p_{in} - p_{fi} = q$. The 4-vector p in the expression ep_i/E for the conserved current is $p = (p_{in} + p_{fi})/2$, and $E = \sqrt{E_{in}E_{fi}}$. As $p_{in}^2 = p_{fi}^2 = m^2$, so qp = 0. (I denoted energy here by the letter E because e in the expression for current stands for charge. We are clearly short of letters.)

Gluons and quarks. A gluon's spin also equals unity. At first glance, the interaction between gluons and quarks is completely analogous to the interaction between photons and electrons. Not at second glance, though. The point is that all electrons carry the same electric charge while quarks have three different color charges. A quark emitting or

⁶ This and next paragraph were substantially improved in the English proof by the Author.

June 2008

absorbing a gluon may change its color. Clearly, this means that gluons must themselves be colored. It can be shown that there must be eight different color species of gluons. While photons are electrically neutral, gluons carry color charges.

Quantum chromodynamics. It might seem that color charged gluons must be intense emitters of gluons, being a sort of 'luminous light'. In fact, quantum chromodynamics — the theory of interaction between quarks and gluons — has a spectacular property known as confinement. In contrast to electrons and photons, colored quarks and gluons do not exist in a free state. These colored particles are locked 'for life' inside colorless (white) hadrons. They can only change their incarceration locality. There are no Feynman diagrams with lines of free gluons or free quarks.

11. Gravitation

Gravitational orbits. Various emblems often show the orbits of electrons in atoms resembling the orbits of planets. It should be clear from the above that according to quantum mechanics, there are no such orbits in atoms. On the other hand, quantum effects are absolutely infinitesimal for macroscopic bodies, all the more so for such heavy ones as planets. Consequently, their orbits are excellently described by classical mechanics.

Newton's constant. The potential energy of the Earth in the gravitational field of the Sun is given by Newton's law

$$U = -\frac{GMm}{r}, \qquad (18)$$

where M is the solar mass, m is the mass of the Earth, r is distance between their centers, and G is Newton's constant:

$$G = 6.71 \times 10^{-39} \hbar c \left[\text{GeV}/c^2 \right]^{-2}.$$
 (19)

(Here we used units in which $c \neq 1$.)

The quantity $p_i p_k/e$. The source of gravitation in Newton's physics is mass. In the theory of relativity the source of gravitation is the quantity $p_i p_k/e$, which plays the role of a kind of 'gravitational current'. (The reader will recall that p_i is the energy-momentum 4-vector, and i = 0, 1, 2, 3. Consequently, the 'gravitational current' has four independent components instead of the ten that a most general symmetrical four-dimensional tensor would have.)

The propagation of the field from the source to the 'sink' is described by Green's function of the gravitational field or the propagator of the graviton — a massless spin-2 particle. This propagator is proportional to $g^{il}g^{km} + g^{im}g^{kl} - g^{ik}g^{lm}$, where g^{ik} is a metric tensor. (As in the case of the photon discussed above, the 4-momentum of the graviton is $q = p_i - p_f$ and the 4-momentum in the expression for current is $p = 1/2(p_i + p_f)$, while $e = \sqrt{e_i e_f}$. We are again short of letters! This time, letters for indices.)

The graviton. Like the photon, the graviton is a massless particle. This is the reason why Newton's and Coulomb's potentials have the form 1/r. However, in contrast to the photon, which cannot emit photons, the graviton can and must emit gravitons. In this respect the graviton resembles gluons, which emit gluons.

The Planck mass. Elementary particle physics often uses the concept of the Planck mass:

$$m_{\rm P} = \sqrt{\frac{\hbar c}{G}}.$$
 (20)

In units in which c = 1 and $\hbar = 1$ we have $m_{\rm P} = 1/\sqrt{G} = 1.22 \times 10^{19}$ GeV.

The gravitational interaction between two ultrarelativistic particles increases as the square of their energy E in the center-of-inertia reference frame. It reaches maximum strength at $E \sim m_{\rm P}$ as the distance between the particles approaches $r \sim 1/m_{\rm P}$. However, let us return from these fantastically large energies and short distances to apples and photons in gravitational fields of the Earth and the Sun.

An apple and a photon. Consider a particle in a static gravitational field, for instance, that of the Sun. The source of the field is the quantity $P_l P_m / E$ where P_l is the 4-momentum of the Sun and E is its energy. In the rest frame of the Sun *l*, m = 0 and $P_l P_m / E = M$, where *M* is the solar mass. In this case the numerator of the propagator of the gravitational field $g^{il}g^{km} + g^{im}g^{kl} - g^{ik}g^{lm}$ is $2g^{i0}g^{k0} - g^{ik}g^{00}$, and the tensor quantity $p_i p_k$ times the numerator of the propagator reduces to a simple expression $2e^2 - m^2$. Hence, for a nonrelativistic apple of mass m the 'gravitational charge' equals m while for a photon with energy *e* it equals 2*e*. Note the coefficient 2. Kinetic energy is attracted twice as strongly as the hidden energy locked in mass. This simple derivation of the coefficient 2 makes unnecessary the complicated derivation of paper [8] using isotropic coordinates.

A photon in the field of the Sun. The interaction of photons with the gravitational field must cause a deflection of a ray of light propagating from a remote star and passing close to the solar disk. In 1915 Einstein calculated the deflection angle and showed that it must be $4GM/c^2R \simeq 1.75''$. (Here, *M* and *R* denote the solar mass and solar radius, respectively.) This prediction was confirmed during the solar eclipse of 1919, which stimulated a huge surge of interest in the theory of relativity.

An atom in the field of the Earth. As a nonrelativistic body on the Earth moves upwards, its potential energy increases in proportion to its mass. Correspondingly, the difference between energies of two levels of an atomic nucleus must be higher, the higher the floor of the building in which this nucleus is located.

A photon's energy is conserved. On the other hand, the frequency ω of a photon propagating through a static gravitational field, and correspondingly its total energy $e = h\omega$, should remain unchanged.

As a result, a photon emitted on the ground floor of a building from a transition between two energy levels of a nucleus will be unable to produce a reverse transition in the same nucleus on the upper floor. This theoretical prediction was confirmed in the 1960s by Pound and Rebka [9] who used the just discovered Mössbauer effect, which makes it possible to measure the tiniest shifts in nuclear energy levels.

However, the wavelength changes. A photon propagating through a static gravitational field like a stone has its total energy e and frequency ω conserved. However, its momentum

and therefore wavelength change as the distance to the gravitating body changes.

Refractive index. As a photon moves away from the source of a gravitational field, its velocity increases and tends to c, and when it approaches the source, it decreases. Hence, the gravitational field, like a transparent medium, has a refractive index. This is a visually clear explanation of the deflection of light in the field of the Sun and in the gravitational lenses of galaxies. Shapiro experimentally discovered the decrease in the velocity of photons near the Sun when measuring the delay of the radar echo returned by planets.

Clocks and gravitation. Ordinary clocks, like atomic clocks, tick faster, the higher they are lifted. Let two synchronized clocks A and B be placed on the first floor. If we move clock A to the second floor and then, say, a day later, move clock B to the second floor as well, clock A will be ahead of B as A has been ticking faster than B for 24 hours. Nevertheless, both A and B will continue to serve as identically reliable stopwatches.

When every point in space is assigned an individual clock, one in fact assumes that all clocks tick at a rate that is independent of the distance to gravitating bodies (in our case, on which floor of the building they are). However, this is not true for ordinary clocks. In order to distinguish extraordinary clocks from ordinary clocks, we will refer to extraordinary ones as 'cloned'.

As we saw above, the frequency of light measured using clocks placed on various floors is independent of the floor number. If, however, it is measured with 'cloned' local clocks, we discover that it is lower, the higher the floor. One interpretation of the Pound – Rebka experiment, stating that the energy of a vertically moving photon decreases with height, like the kinetic energy of a stone thrown upwards, is based on precisely this argument. However, a drop in kinetic energy of the stone is accompanied with an increase in its potential energy, so that the total energy is conserved. Now, a photon has no potential energy, so that its energy in a static gravitational field remains constant.

12. Epistemology and linguistics

Physics and epistemology. *Episteme* in Greek means knowledge. Epistemics is the science of knowledge, a relatively young branch of epistemology, the theory of knowledge and cognition. Obviously, the problems I discuss in this talk concern not only physics but epistemology, too.

Physics and semantics. The Greek attribute 'semanticos' (signifying) was used in linguistics already by Aristotle. However, what are the links tying the science of languages — linguistics — and semantics — the science of words and symbols, an element of linguistics — to physics?

This is the right moment to recall the words allegedly said by V A Fock: "Physics is an essentially simple science. The most important problem in it is to understand what each letter denotes."

XXth century physics drastically changed our understanding of what a vacuum and matter are, and connected in a new way such properties of matter as energy, momentum, and mass. The elaboration of the fundamental concepts of physics has not been completed and is unlikely to end in the foreseeable future. This is one of the reasons why it is so important to choose the adequate words and letters when discussing physical phenomena and theories.

'Concepts glued together'. Newton's *Principia* 'glued together' the concepts of mass and matter (substance): "mass is proportional to density and volume." In Einstein's papers mass is 'glued together' with inertia and gravitation (the inertial and gravitational masses). And energy is glued to matter.

The archetype. According to dictionaries, an archetype is the historically original form (the protoform), the original concept or word, or the original type (prototype). The concept of the archetype keenly interested Pauli, who in 1952 published a paper on the effect of archetypical notions on the creation of natural-science theories by Kepler. It is possible that the concept of mass is just the archetypical notion that glued together the concepts of matter, inertia, and weight.

Atom and archetype. Atom and Archetype — that was the title chosen for the English translation from German of the book [10] presenting the correspondence between Wolfgang Pauli and the leading German proponent of psychoanalysis Carl Jung, covering the period from 1932 to 1958. W Pauli and C Jung discussed, among other things, the material nature of time and the possibility of communicating with people who lived several centuries or millennia before us. It is widely known that Pauli treated rather seriously the effect named after him: when he walked into an experimental laboratory, measuring equipment broke down.

Poets on terminology. David Samoilov on words: "We wipe them clean as we clean glass. This is our trade." Vladimir Mayakovskii: "The street is writhing for want of tongue. It has no nothing for yelling or talking." (*Translated by Nina Iskandaryan.*)

Many an author responds to the dearth of precise terms and inability to use them by resorting to meaningless words like 'rest mass' which impart smoothness and 'energetics' to texts, just as 'blin'⁷ does to ordinary speech.

How to teach physics. Terms need 'wiping clean' and 'unglueing'.

The 'umbilical cord' connecting the modern physical theory with the preceding 'mother theory' needs careful cutting in teaching. (In the case of the theory of relativity the mother was the 'centaur' composed of Maxwell's field theory and Newton's mechanics, with relativistic mass serving as the umbilical cord.)

Let us recall the title of F Klein's famous book *Elementary Mathematics from an Advanced Standpoint*. The landscape of modern physics must be contemplated from an advanced standpoint: not from a historical gully but from the pinnacle of symmetry principles. I firmly believe that it is unacceptable to claim that the dependence of mass on velocity is an experimental fact and thus hide from the student that it is a mere interpretational 'factoid'. (Dictionaries explain that a factoid looks very much like a fact but is trusted only because we find it in printed texts.)

⁷ 'Blin' is a slang euphemism for a 'four-letter word' in vulgar Russian.

13. Concluding remarks

The ' $E = mc^2$ problem': could it be avoided? One is tempted to think that the ' $E = mc^2$ problem' would not arise from the first place if the quantity E/c^2 — the proportionality coefficient between velocity and momentum — were identified with a new physical quantity christened as, say, 'inertia' or 'iner'; it would be identical to mass as momentum tended to zero. As a result, mass would become 'rest inertia'. Likewise, another new quantity could be introduced — 'heaviness' or 'grav' — $p_i p_k / E$ reducing to mass at zero momentum. But physicists preferred 'to refrain from multiplying entities' and from introducing new physical quantities. They formulated instead new, more general relations between old quantities, for example $E^2 - \mathbf{p}^2 c^2 = m^2 c^4$ and $\mathbf{p} = \mathbf{v} E/c^2$.

Unfortunately, many authors attempt to retain even in relativistic physics such nonrelativistic equations as $\mathbf{p} = m\mathbf{v}$, and such nonrelativistic glued-up concepts as 'mass is a measure of inertia' and 'mass is a measure of gravitation'; as a result, they prefer to use the notion of velocity-dependent mass.

It is amazing how again and again a physicist would choose the first of these paths (new equations) in his research papers and the second one (old glued-up concepts) in sciencepopularizing and pedagogical activities. This could of course only produce unbelievable confusion in the minds of those who read popular texts and blindly follow the authority.

On the reliability of science. An opinion that has become widely publicized recently is that science in general and physics in particular are untrustworthy. Many popularizers of science create the impression that the theory of relativity proved Newton's mechanics wrong just as chemistry proved alchemy wrong and astronomy proved astrology wrong. Such declarations are a crude distortion of the essence of scientific revolutions. Newton's mechanics remains a correct science today, in the XXIst century, and will continue to be correct forever. The discovery of the theory of relativity only put bounds on the domain of applicability of Newton's mechanics to velocities much smaller than the speed of light c. It also demonstrated its approximate nature in this domain (to within corrections of the order of v^2/c^2).

Similarly, the discovery of quantum mechanics put bounds on the domain of applicability of classical mechanics to phenomena for which the quantity of action is large in comparison with the quantum of action \hbar . Quite to the contrary, the domain where astrology and alchemy exist is that of prejudice, superstition, and ignorance. It is rather funny that those who compare Newton's mechanics with astrology typically believe that mass depends on velocity.

Recent publications. Additional information on the aspects discussed above can be found in [11, 12].

On the title. My good friend and expert in the theory of relativity read the slides of this talk and advised me to drop Pythagoras's name from the title. I chose not to follow his advice as in the relativity-related literature I had never come across a discussion of right-angled triangles without the approximate extraction of square roots.

Acknowledgements. I am grateful to T Basaglia, A Bettini, S I Blinnikov, V F Chub, M A Gottlieb, E G Gulyaeva, E A Ilyina, C Jarlscog, V I Kisin, B A Klumov, B L Okun, S G Tikhodeev, M B Voloshin, V R Zoller for their advice and help. This work was supported by the grants NSh-5603.2006.2, NSh-4568.2006.2 and RFBR-07-02-00830-a.

References

- 1. Okun L B "Formula Einshteina: $E_0 = mc^2$. 'Ne smeetsya li Gospod' Bog'?" Usp. Fiz. Nauk **178** 541 (2008) ["The Einstein formula $E_0 = mc^2$ 'Isn't the Lord laughing'?" Phys. Usp. **51** (5) (2008)]
- Klein F Elementarmathematik vom höeheren Standpunkte aus. Erster Band, Arithmetik, Algebra, Analysis 3 Auflage (Berlin: Verlag von Julius Springer, 1924) [Translated into English: Elementary Mathematics from an Advanced Standpoint, Arithmetics, Algebra, Analysis (New York: Dover Publ., 2007); Translated into Russian (Moscow: Nauka, 1987)]
- Poincaré H "Sur la dynamique de l'électron" *Rendiconti del Circolo Matematico di Palermo* 21 129 (1906) [Translated into Russian: "O dinamike elektrona" ("On the dynamics of electron") *Izbrannye Trudy* (Selected Works) Vol. 3 (Moscow: Nauka, 1974) p. 433]
- Minkowski H "Raum und Zeit" Phys. Z. 10 104-111 (1909) [Translated into Russian: "Prostranstvo i Vremya" ("Space and time"), in Lorentz H A, Poincaré H, Einstein A, Minkowski H Printsip Otnositel'nosti. Sbornik Rabot Klassikov Relyativizma (The Principle of Relativity. Collected Papers of Classics of Relativism) (Eds V K Frederiks, D D Ivanenko) (Moscow-Leningrad: ONTI, 1935) pp. 181-203]
- Hawking S *The Universe in a Nutshell* (New York: Bantam Books, 2001) [Translated into Russian (Translated from English by A Sergeev) St.-Petersburg: Amfora, 2007)]
- Landau L D, Lifshitz E M *Teoriya polya* (The Classical Theory of Fields) (Moscow: Nauka, 1988) [Translated into English (Amsterdam: Reed Elsevier, 2000)]
- Landau L D, Lifshitz E M Mekhanika (Mechanics) (Moscow: Nauka, 1988) [Translated into English (Amsterdam: Elsevier Sci., 2003)]
- Okun L B "The concept of mass" *Phys. Today* 42 (6) 31-36 (1989);
 Okun L B "Putting to rest mass misconceptions" *Phys. Today* 43 (5) 13, 15, 115, 117 (1990)
- Pound R V, Rebka G A (Jr.) "Apparent weight of photons" *Phys. Rev. Lett.* 4 337–341 (1960)
- Pauli W, Jung C Atom and Archetype: Pauli/Jung Letters. 1932– 1958. (Ed. C A Meyer) (Princeton, NJ: Princeton Univ. Press, 2001); Meier C A (Herausgegeben) Wolfgang Pauli, C. G. Jung: Ein Briefwechsel 1932–1958 (Berlin: Springer-Verlag, 1992)
- Okun L B "Chto takoe massa? (Iz istorii teorii otnositel'nosti)" ('What is mass? (From the history of relativity theory)'), in *Issledovaniya po Istorii Fiziki i Mekhaniki*. 2007 (Research on the History of Physics and Mechanics. 2007) (Executive Ed. G M Idlis) (Moscow: Nauka, 2007)
- Okun L B "The evolution of the concepts of energy, momentum and mass from Newton and Lomonosov to Einstein and Feynman", in *Proc. of the 13th Lomonosov Conf. August 23, 2007* (Singapore: World Scientific) (in press)

PACS number: **03.30.** + **p** DOI: 10.1070/PU2008v051n06ABEH006553 DOI: 10.3367/UFNr.0178.2008061.0663

Bjorken and Regge asymptotics of scattering amplitudes in QCD and in supersymmetric gauge models

L N Lipatov

1. Introduction

We review theoretical approaches to the investigation of deep-inelastic lepton-hadron interactions and high-energy hadron-hadron scattering in the Regge kinematics. It is demonstrated that the gluon in QCD is Reggeized and the Pomeron is a composite state of the Reggeized gluons. Remarkable properties of the BFKL equation for the Pomeron wave function in QCD and supersymmetric gauge theories are outlined. It is shown that by the AdS/CFT correspondence, the BFKL Pomeron is equivalent to the Reggeized graviton in the N = 4 extended supersymmetric model. The maximal transcendentality and integrability properties realized in this model allow calculating the anomalous dimension of twist-2 operators up to 4 loops.

2. Deep-inelastic ep scattering

The inclusive electron-proton scattering in the Bjorken kinematics (see Fig. 1),

$$2pq \sim Q^2 = -q^2 \rightarrow \infty, \quad x = \frac{Q^2}{2pq}, \quad 0 \le x \le 1,$$
 (1)

is very important because it gives direct information about the distribution $n^{q}(x)$ of quarks inside the proton as a function of their energy ratio $x = |\mathbf{k}| / |\mathbf{p}|$ ($|\mathbf{p} \to \infty|$). Indeed, in the framework of the Feynman-Bjorken quark-parton model [1, 2], we can obtain the following simple expression for the structure functions $F_{1,2}(x)$ of this process:

$$\frac{1}{x}F_2(x) = 2F_1(x) = \sum_{i=q,\bar{q}} Q_i^2 n^i(x), \qquad (2)$$

where the quark charges are $Q_{\rm u} = 2/3$, $Q_{\rm d} = -1/3$.

It turns out that the partonic picture is also valid in renormalizable field theories if the parton transverse momenta \mathbf{k}_{\perp} are restricted by an ultraviolet cut-off $k_{\perp}^2 < \Lambda^2 \sim Q^2$ [3]. In these theories, the running coupling constant $\alpha = g^2/(4\pi)$ in the leading logarithmic approximation (LLA) is

$$\alpha(Q^2) = \frac{\alpha_{\mu}}{1 + \beta \alpha_{\mu} / (4\pi) \ln \left(Q^2 / \mu^2\right)} , \qquad (3)$$

where α_{μ} is its value at the renormalization point μ . In quantum electrodynamics (QED) and quantum chromodynamics (QCD), the coefficients β have opposite signs,

$$\beta_{\text{QED}} = -n_{\text{e}} \frac{4}{3}, \quad \beta_{\text{QCD}} = \frac{11}{3} N_{\text{c}} - n_{\text{f}} \frac{2}{3}, \quad (4)$$

where N_c is the rank of the gauge group ($N_c = 3$ for QCD), and n_e and n_f are the numbers of leptons and quarks, which can be considered massless for a given Q^2 .



Landau and Pomeranchuk argued that because of the negative sign of β_{QED} , a Landau pole is generated in the photon propagator, which leads to the vanishing of the physical electric charge in the local limit. On the other hand, in QCD, the non-Abelian interaction disappears at large Q^2 and, as a result of the asymptotic freedom, we have an approximate Bjorken scaling: the structure functions depend on Q^2 only logarithmically [4]. Thus, the experiments on deep-inelasic ep scattering performed at SLAC at the end of the 1960s discovered that the Landau 'zero charge' problem is absent in strong interactions.

In the infinite-momentum frame $|\mathbf{p}| \rightarrow \infty$, it is helpful to introduce the Sudakov variables for parton momenta as

$$\mathbf{k}_i = \beta_i \, \mathbf{p} + \mathbf{k}_i^{\perp}, \quad (\mathbf{k}_i^{\perp}, \mathbf{p}) = 0, \quad \sum_i \, \mathbf{k}_i = \mathbf{p}.$$
 (5)

The parton distributions are defined in terms of the proton wave function Ψ_m as

$$n^{i}(x) = \sum_{m} \int \prod_{r=1}^{m-1} \frac{\mathrm{d}\beta_{r} \,\mathrm{d}^{2}k_{r}^{\perp}}{(2\pi)^{2}} |\Psi_{m}|^{2} \sum_{r \in i} \delta(\beta_{r} - x) \,. \tag{6}$$

They are functions of $\Lambda \sim Q$ because the factor $|\Psi_m|^2 \sim \prod_{r=1}^m Z_r$ depends on Λ through the wave-function renormalization constants $\sqrt{Z_r}$ and Λ is the upper limit in integrals over the transverse momenta k_r^{\perp} . With the cascade-type dynamics of the parton number growth and with Λ taken into account, we can obtain the evolution equations of Dokshizer, Gribov, Lipatov, Altarelli, and Parisi (DGLAP) [3, 5] in the LLA,

$$\frac{\mathrm{d}}{\mathrm{d}\xi(\mathcal{Q}^2)} n_i(x) = -w_i n_i(x) + \sum_r \int_x^1 \frac{\mathrm{d}y}{y} w_{r \to i}\left(\frac{x}{y}\right) n_r(y) \,,$$
(7)

$$w_{i} = \sum_{k} \int_{0}^{1} dx \, x \, w_{i \to k}(x) \,, \tag{8}$$

where

$$\xi(Q^2) = \frac{N_c}{2\pi} \int_{\mu^2}^{Q^2} \frac{\mathbf{d}\mathbf{k}_{\perp}^2}{\mathbf{k}_{\perp}^2} \,\alpha(\mathbf{k}_{\perp}^2) \,. \tag{9}$$

Equation (7) has a clear probabilistic interpretation: the number of partons n_i decreases because of their decay into other partons in the opening phase space $d\xi(Q^2)$ and increases because the decay products of other partons *r* can contain partons of the type *i* [3].

The momenta of parton distributions

$$n_i^{j} = \int_0^1 \mathrm{d}x \, x^{j-1} \, n_i(x) \tag{10}$$

satisfy the renormalization-group equations

$$\frac{\mathrm{d}}{\mathrm{d}\xi(Q^2)} n_i^j = \sum_r w_{r \to i}^j n_r^j, \qquad (11)$$

and are related to the matrix elements of twist-2 operators

$$n_i(j) = \left\langle p \,|\, O_i^j \,|\, p \right\rangle. \tag{12}$$

The twist *t* is defined as the difference between their canonical dimension *d* measured in units of mass and the Lorentz spin *j* of the corresponding tensor. The quantities $w_{r\to i}^{j}$ are elements of the anomalous dimension matrix for the operators O_{i}^{j} .

Hadron-hadron scattering in the Regge kinematics (see Fig. 2)

$$s = (p_A + p_B)^2 = (2E)^2 \gg \mathbf{q}^2 = -(p_{A'} - p_A)^2 \sim m^2$$
 (13)

is usually described in terms of a *t*-channel exchange of the Reggeon (see Fig. 3),

$$A_p(s,t) = \xi_p(t) g(t) s^{j_p(t)} g(t), \quad j_p(t) = j_0 + \alpha' t, \quad (14)$$

$$\xi_p = \frac{\exp\left(-i\pi j_p(t)\right) + p}{\sin(\pi j_p)}, \qquad (15)$$

where $j_p(t)$ is the Regge trajectory, assumed to be linear, and j_0 and α' are its intercept and slope. The signature factor ξ_p is a complex quantity depending on the Reggeon signature $p = \pm 1$. A special Reggeon–Pomeron is introduced to explain the approximately constant behavior of total cross sections at high energies and the fulfillment of the Pomeranchuk theorem $\sigma_{h\bar{h}}/\sigma_{hh} \rightarrow 1$. Its signature *p* is positive and its intercept is close to unity: $j_0^p = 1 + \Delta$, $\Delta \ll 1$. The field theory of Pomeron interactions was constructed by Gribov around 40 years ago.

Particle production at high energies can be investigated in the multi-Regge kinematics (see Fig. 4)

$$s \gg s_1, \quad s_2, \dots, \quad s_{n+1} \gg t_1, \quad t_2, \dots, \quad t_{n+1},$$
 (16)

where s_r are squares of the sums of neighboring particle momenta k_{r-1} and k_r , and $-t_r$ are squares of the momentum transfers \mathbf{q}_r . This amplitude can also be expressed in terms of the Reggeon exchanges in each of the t_r -channels:

$$A_{2\to 2+n} \sim \prod_{r=1}^{n+1} s_r^{j_p(t_r)}$$
. (17)









4. Gluon Reggeization in QCD

In the Born approximation in QCD, the scattering amplitude for two-colored particle scattering is factored (see Fig. 2),

$$M_{AB}^{A'B'}(s,t)\big|_{\text{Born}} = \Gamma_{A'A}^{c} \frac{2s}{t} \Gamma_{B'B}^{c}, \quad \Gamma_{A'A}^{c} = g T_{A'A}^{c} \delta_{\lambda_{A'}\lambda_{A}},$$
(18)

where T^c are the generators of the color group SU(N_c) in the corresponding representation and λ_r are helicities of the colliding and final-state particles. In the LLA, the scattering amplitude in QCD can be written as [6]

$$M_{AB}^{A'B'}(s,t) = M_{AB}^{A'B'}(s,t) \big|_{\text{Born}} s^{\omega(t)}, \quad \alpha_s \ln s \sim 1,$$
(19)

where the gluon Regge trajectory is

$$\omega(-|q|^2) = -\int \frac{\mathrm{d}^2 k}{4\pi^2} \frac{\alpha_s N_c |q|^2}{|k|^2 |q-k|^2} \approx -\frac{\alpha_s N_c}{2\pi} \ln \frac{|q^2|}{\lambda^2} \,. \tag{20}$$

The fictitious gluon mass λ is introduced here to regularize the infrared divergence. This trajectory was also calculated in the two-loop approximation in QCD [7] and in supersymmetric gauge theories [8].

Further, the gluon production amplitude in the multi-Regge kinematics can be written in the factored form [6]

$$M_{2\to 1+n} = 2s \Gamma_{A'A}^{c_1} \frac{s_1^{\omega_1}}{|q_1|^2} g T_{c_2c_1}^{d_1} C(q_2, q_1) \times \frac{s_2^{\omega_2}}{|q_2|^2} \dots C(q_n, q_{n-1}) \frac{s_n^{\omega_n}}{|q_n|^2} \Gamma_{B'B}^{c_n}.$$
(21)

The Reggeon-Reggeon-gluon vertex for the produced gluon with a definite helicity is

$$C(q_2, q_1) = \frac{q_2 q_1^*}{q_2^* - q_1^*}, \qquad (22)$$

where we use the complex notation for the transverse components of particle momenta. This allows calculating the total cross section [6]

$$\sigma_{\rm t} = \sum_{n} \int \mathrm{d}\Gamma_n |M_{2\to 1+n}|^2 \,, \tag{23}$$

where Γ_n is the phase space for the produced particle momenta in the multi-Regge kinematics.

5. The BFKL equation

Using the fact that the production amplitudes in QCD are factored, we can write a Bethe–Salpeter-type equation for the total cross section σ_t . Also using the optical theorem, we can represent this equation as the Balitsky–Fadin–Kuraev–Lipatov (BFKL) equation for the Pomeron wave function [6]:

$$E \Psi(\mathbf{\rho}_1, \mathbf{\rho}_2) = H_{12} \Psi(\mathbf{\rho}_1, \mathbf{\rho}_2), \quad \Delta = -\frac{\alpha_s N_c}{2\pi} E, \qquad (24)$$

where $\sigma_t \sim s^4$ and the BFKL Hamiltonian in the coordinate representation is

$$H_{12} = \ln |p_1 p_2|^2 + \frac{1}{p_1 p_2^*} \left(\ln |\rho_{12}|^2 \right) p_1 p_2^*$$
$$+ \frac{1}{p_1^* p_2} \left(\ln |\rho_{12}|^2 \right) p_1^* p_2 - 4\psi(1) , \quad \rho_{12} = \rho_1 - \rho_2 . \tag{25}$$

It is invariant under the Möbius transformations [9, 10]

$$\rho_k \to \frac{a\rho_k + b}{c\rho_k + d} \,. \tag{26}$$

We use the complex notation for transverse coordinates and their canonically conjugate momenta. The conformal weights for the principal series of unitary representations of the Möbius group are

$$m = \gamma + \frac{n}{2}, \quad \widetilde{m} = \gamma - \frac{n}{2}, \quad \gamma = \frac{1}{2} + \mathrm{i}\nu,$$
 (27)

where γ is the anomalous dimension of the twist-2 operators and *n* is the conformal spin.

The Bartels – Kwiecinski – Praszalowicz equation for colorless composite states of several Reggeized gluons has the form [11]

$$E \Psi(\mathbf{p}_1, \ldots) = H \Psi(\mathbf{p}_1, \ldots) , \quad H = \sum_{k < l} \frac{\mathbf{T}_k \mathbf{T}_l}{-N_c} H_{kl} , \quad (28)$$

where H_{kl} is the BFKL Hamiltonian. In addition to the Möbius invariance, its wave function in the multi-color QCD $(N_c \rightarrow \infty)$ has the holomorphic factorization property [12]

$$\Psi(\mathbf{\rho}_1,\ldots,\mathbf{\rho}_n) = \sum_{r,s} a_{r,s} \,\Psi_r(\rho_1,\ldots,\rho_n) \,\Psi_s(\,\rho_1^*,\ldots,\rho_n^*)\,,$$
(29)

where the sum is taken over the degenerate set of solutions of the corresponding holomorphic and antiholomorphic BFKL equations. These equations have the duality symmetry $p_k \rightarrow \rho_{k,k+1} \rightarrow p_{k+1}$ (k = 1, 2, ..., n) [13] and *n* integrals of motion q_r , q_r^* [14]. The corresponding Hamiltonians *h* and h^* are local Hamiltonians of an integrable Heisenberg spin model in which spins are generators of the Möbius group [15]. We can introduce the transfer (T) and monodromy (t) matrices according to the definitions [14]

$$T(u) = \operatorname{tr} t(u), \quad t(u) = L_1 L_2 \dots L_n = \sum_{r=0}^n u^{n-r} q_r, \quad (30)$$

$$L_k = \begin{pmatrix} u + \rho_k p_k & p_k \\ -\rho_k^2 p_k & u - \rho_k p_k \end{pmatrix}.$$
 (31)

Then the monodromy matrix t(u) satisfies the Yang–Baxter equation [14]

$$\begin{aligned} t_{r_{1}'}^{s_{1}}(u) t_{r_{2}'}^{s_{2}'}(v) l_{r_{1}r_{2}'}^{r_{1}'r_{2}'}(v-u) &= l_{s_{1}'s_{2}'}^{s_{1}s_{2}}(v-u) t_{r_{2}}^{s_{2}'}(v) t_{r_{1}}^{s_{1}'}(u) ,\\ \hat{l}(u) &= u\hat{1} + i\hat{P} , \end{aligned}$$
(32)

where $\hat{l}(u)$ is the monodromy matrix for the usual Heisenberg spin model and \hat{P} is the permutation operator. This equation can be solved with the use of the Bethe ansatz and the Baxter-Sklyanin approach.

6. Pomeron in the N=4 SUSY

We can also calculate the integral kernel for the BFKL equation in two loops [16]. Its eigenvalue can be written as

$$\omega = 4\hat{a}\chi(n,\gamma) + 4\hat{a}^2\Delta(n,\gamma), \quad \hat{a} = \frac{g^2N_c}{16\pi^2}, \quad (33)$$

where

$$\chi(n,\gamma) = 2\psi(1) - \psi\left(\gamma + \frac{|n|}{2}\right) - \psi\left(1 - \gamma + \frac{|n|}{2}\right)$$
(34)

and $\psi(x) = \Gamma'(x)/\Gamma(x)$. The one-loop correction $\Delta(n, \gamma)$ in QCD contains nonanalytic terms, the Kronecker symbols $\delta_{|n|,0}$ and $\delta_{|n|,2}$ [8]. But in the N = 4 SUSY, they cancel and we obtain the following result for $\Delta(n, \gamma)$ in the Hermitian separable form [8, 17]:

$$\Delta(n,\gamma) = \phi(M) + \phi(M^*) - \frac{\rho(M) + \rho(M^*)}{2\hat{a}/\omega},$$

$$M = \gamma + \frac{|n|}{2},$$
(35)

$$\rho(M) = \beta'(M) + \frac{1}{2}\zeta(2),$$

$$\beta'(z) = \frac{1}{4} \left[\psi'\left(\frac{z+1}{2}\right) - \psi'\left(\frac{z}{2}\right) \right].$$
(36)

It is interesting that all functions entering these expressions have the maximal transcendentality property [17]. Moreover, $\phi(M)$ can be written as

$$\phi(M) = 3\zeta(3) + \psi''(M) - 2\Phi(M) + 2\beta'(M)(\psi(1) - \psi(M)), \qquad (37)$$

$$\Phi(M) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k+M} \left(\psi'(k+1) - \frac{\psi(k+1) - \psi(1)}{k+M} \right), \quad (38)$$

where $\psi(M)$ has the transcedentality equal to 1, its derivatives $\psi^{(n)}$ have transcedentalities n + 1, and the additional poles in the sum over k increase the transcedentality of $\Phi(M)$ up to 3, which is also the transcendentality of $\zeta(3)$. The maximal transcendentality hypothesis is also valid for the anomalous dimensions of twist-2 operators in the N = 4 SUSY [18, 19], in contrast to the case of QCD [20]. This result is discussed in the next section.

Generally, the BFKL equation in the diffusion approximation can be written in the simple form [6]

$$j = 2 - \Delta - Dv^2, \tag{39}$$

where v is related to the anomalous dimension of the twist-2 operators as [16]

$$\gamma = 1 + \frac{j-2}{2} + i\nu \,. \tag{40}$$

The parameters Δ and D are functions of the coupling constant \hat{a} and are known up to two loops. Higher-order perturbative corrections can be obtained with the use of the effective action [21, 22]. For large coupling constants, we can expect that the leading Pomeron singularity in the N = 4SUSY is moved to the point j = 2 and the Pomeron asymptotically coincides with the graviton Regge pole. This assumption is related to the AdS/CFT correspondence, formulated in the framework of the Maldacena hypothesis that the N = 4 SUSY is equivalent to a superstring model living on the 10-dimensional anti-de Sitter space [23-25]. For the BFKL equation in the diffusion approximation, it is therefore natural to impose the physical condition that γ is zero for the conserved energy-momentum tensor $\vartheta_{\mu\nu}(x)$ having the Lorents spin j = 2. As a result, we obtain that the parameters \varDelta and D coincide [19]. In this case, we can solve the above BFKL equation for γ :

$$\gamma = (j-2) \left(\frac{1}{2} - \frac{1/\Delta}{1 + \sqrt{1 + (j-2)/\Delta}} \right).$$
(41)

Using the dictionary developed in the framework of the AdS/ CFT correspondence [24], we can rewrite the BFKL equation in the form of the graviton Regge trajectory [19]

$$j = 2 + \frac{\alpha'}{2} t, \quad t = \frac{E^2}{R^2}, \quad \alpha' = \frac{R^2}{2} \Delta.$$
 (42)

On the other hand, Gubser, Klebanov, and Polyakov predicted the following asymptotic form of the anomalous dimension at large \hat{a} and j [26]:

$$\gamma_{|\hat{a},j\to\infty} = -\sqrt{j-2} \, \varDelta_{|j\to\infty}^{-1/2} = \sqrt{2\pi j} \, \hat{a}^{1/4} \,. \tag{43}$$

As a result, we can obtain the explicit expression for the Pomeron intercept at large coupling constants [19, 27],

$$j = 2 - \Delta$$
, $\Delta = \frac{1}{2\pi} \hat{a}^{-1/2}$. (44)

7. Maximal transcedentality

According to the hypothesis discussed above, the anomalous dimension

$$\gamma(j) = \hat{a}\gamma_1(j) + \hat{a}^2\gamma_2(j) + \hat{a}^3\gamma_3(j) + \dots$$
(45)

should contain the maximally transcendental functions [17]. Indeed, we have

$$\gamma_1(j+2) = -4S_1(j), \qquad (46)$$

$$\frac{\gamma_2(j+2)}{8} = 2S_1 \left(S_2 + S_{-2} \right) - 2S_{-2,1} + S_3 + S_{-3}$$
(47)

in two loops [17, 18], and

$$\frac{\gamma_{3}(j+2)}{32} = -12(S_{-3,1,1} + S_{-2,1,2} + S_{-2,2,1}) + 6(S_{-4,1} + S_{-3,2} + S_{-2,3}) - 3S_{-5} - 2S_{3}S_{-2} - S_{5} - 2S_{1}^{2}(3S_{-3} + S_{3} - 2S_{-2,1}) - S_{2}(S_{-3} + S_{3} - 2S_{-2,1}) + 24S_{-2,1,1,1} - S_{1}(8S_{-4} + S_{-2}^{2} + 4S_{2}S_{-2} + 2S_{2}^{2}) - S_{1}(3S_{4} - 12S_{-3,1} - 10S_{-2,2} + 16S_{-2,1,1})$$
(48)

in three loops [19], where the harmonic sums are defined as

$$S_{a}(j) = \sum_{m=1}^{j} \frac{1}{m^{a}}, \quad S_{a,b,c,\dots}(j) = \sum_{m=1}^{j} \frac{1}{m^{a}} S_{b,c,\dots}(m),$$

$$S_{-a}(j) = \sum_{m=1}^{j} \frac{(-1)^{m}}{m^{a}}, \quad S_{-a,b,\dots}(j) = \sum_{m=1}^{j} \frac{(-1)^{m}}{m^{a}} S_{b,\dots}(m),$$

$$\overline{S}_{-a,b,c,\dots}(j) = (-1)^{j} S_{-a,b,\dots}(j) + S_{-a,b,\dots}(\infty) \left(1 - (-1)^{j}\right).$$
(49)

It was argued in Ref. [28] that for the N = 4 SUSY, the evolution equations for anomalous dimensions of quasipartonic operators are integrable in the LLA. Later, such an integrability was generalized to other operators [29] and to higher loops [30]. With the additional use of the maximal transcendentality hypothesis, the integral equation for the socalled casp anomalous dimension was constructed in all orders of the perturbation theory [31, 32].

To calculate the anomalous dimension of the twist-2 operators in 4 loops, we can apply the integrability approach based on the asymptotic Bethe ansatz [30]. The corresponding equations for the Bethe roots u_k are

$$\left(\frac{x_k^+}{x_k^-}\right)^2 = \prod_{r=1}^{j-2} \frac{x_k^- - x_r^+}{x_k^+ - x_r^-} \frac{1 - g^2/x_k^+ x_r^-}{1 - g^2/x_k^- x_r^+} \exp\left(2\mathrm{i}\theta(u_k, u_r)\right),$$
(50)

where we use the notation

$$x_k^{\pm} = \frac{u_k^{\pm}}{2} + \sqrt{\frac{(u_k^{\pm})^2}{4} - g^2}, \quad u^{\pm} = u \pm \frac{i}{2}$$
(51)

and the dressing phase expansion [32]

$$\theta(u_k, u_j) = 4\zeta(3) g^6 (q_2(u_k) q_3(u_j) - q_3(u_k) q_2(u_j)) + \dots$$
(52)

The solution for u_k^{\pm} allows finding the anomalous dimensions

$$\gamma(g, M) = 2g^2 \sum_{k=1}^{M} \left(\frac{i}{x_k^+} - \frac{i}{x_k^-} \right).$$
(53)

In four loops, in particular, we can obtain [33]

$$\frac{\gamma_4}{256} = 4S_{-7} + 6S_7 + 2(S_{-3,1,3} + S_{-3,2,2} + S_{-3,3,1} + S_{-2,4,1}) + \dots - 80S_{1,1,-4,1} - \zeta(3)S_1(S_3 - S_{-3} + 2S_{-2,1}),$$
(54)

where the harmonic sums depend on j - 2 and the dots denote the omitted terms (their number exceeds 200). All these terms satisfy the maximal transcendentality property. The last term appears from the dressing phase.

It turns out that after the analytic continuation of this expression in the complex *j*-plane, the first two terms give rise to the pole $1/\omega^7$ for $\omega = j - 1 \rightarrow 0$, which contradicts the singularity at this point predicted in 4 loops from the BFKL equation,

$$\lim_{j \to 1} \gamma_4(j) = -\frac{32}{\omega^4} \left(32\zeta_{(3)} + \frac{\pi^4}{9} \, \omega \right) + \dots \,. \tag{55}$$

This means that the asymptotic Bethe ansatz should be modified starting from 4 loops. Specifically, wrapping effects should be taken into account [33].

Interesting results were also obtained for the scattering amplitudes in the N = 4 SUSY for particles on the mass shell [34]. These amplitudes were used in Ref. [35] for the construction of higher-loop corrections to the BFKL kernel in this model. But it was shown in [35] that the BDS ansatz in [34] does not satisfy the correct factorization properties in the multi-Regge kinematics.

8. Discussion of the obtained results

It was demonstrated that the Pomeron in QCD is a composite state of reggeized gluons. The BFKL dynamics is integrable in the LLA. In the next-to-leading approximation in the N = 4SUSY, the equation for the Pomeron wave function has remarkable properties, including analyticity in the conformal spin *n* and maximal transcendentality. In this model, the BFKL Pomeron coincides with the Reggeized graviton. The anomalous dimension for twist-2 operators has the maximal transcendentality property, which allows calculating it analytically in 2 and 3 loops. The integrability based on the asymptotic Bethe ansatz reproduces these results, but fails to reproduce the BFKL prediction in 4 loops due to the presence of wrapping effects. The BDS ansatz for scattering amplitudes in the N = 4 SUSY does not agree with the BFKL approach in the multi-Regge kinematics.

This work was supported in part by the grants 06-02-72041-MNTI-a, 07-02-00902-a, and RSGSS-5788.2006.2

References

- Bjorken J D, in Selected Topics in Particle Physics: Proc. of Intern. School of Physics "Enrico Fermi", Course 41 (Ed. J Steinberger) (New York: Academic Press, 1968); Bjorken J D, Paschos E A Phys. Rev. 185 1975 (1969)
- 2. Feynman R P Phys. Rev. Lett. 23 1415 (1969)
- 3. Lipatov L N Yad. Fiz. 20 181 (1975) [Sov. J. Nucl. Phys. 20 94 (1975)]
- Gross D J, Wilczek F Phys. Rev. Lett. 30 1343 (1973); Politzer H D Phys. Rev. Lett. 30 1346 (1973)
- Gribov V N, Lipatov L N Yad. Fiz. 15 781 (1972) [Sov. J. Nucl. Phys. 15 438 (1972)]; Altarelli G, Parisi G Nucl. Phys. B 126 298 (1977); Dokshitzer Yu L Zh. Eksp. Teor. Fiz. 73 1216 (1977) [Sov. Phys. JETP 46 641 (1977)]
- Lipatov L N Yad. Fiz. 23 642 (1976) [Sov. J. Nucl. Phys. 23 338 (1976)]; Fadin V S, Kuraev E A, Lipatov L N Phys. Lett. B 60 50 (1975); Kuraev E A, Lipatov L N, Fadin V S Zh. Eksp. Teor. Fiz. 71 840 (1976); 72 377 (1977) [Sov. Phys. JETP 44 443 (1976); 45 199 (1977)]; Balitsky I I, Lipatov L N Yad. Fiz. 28 1597 (1978) [Sov. J. Nucl. Phys. 28 822 (1978)]
- 7. Fadin V S, Fiore R, Kotsky M I Phys. Lett. B 387 593 (1996)
- 8. Kotikov A V, Lipatov L N Nucl. Phys. B 582 19 (2000)
- 9. Lipatov L N Phys. Lett. B 309 394 (1993)

- 10. Lipatov L N Zh. Eksp. Teor. Fiz. 90 1536 (1986) [Sov. Phys. JETP 63 904 (1986)]
- Bartels J Nucl. Phys. B 175 365 (1980); Kwieciński J, Prasza Jowicz M Phys. Lett. B 94 413 (1980)
- 12. Lipatov L N Phys. Lett. B 251 284 (1990)
- 13. Lipatov L N Nucl. Phys. B 548 328 (1999)
- Lipatov L N "High energy asymptotics of multi-colour QCD and exactly solvable lattice models", Preprint DFPD/93/TH/70 (Padua: Inst. Naz. Fis. Nucleare, 1993); hep-th/9311037
- Lipatov L N Pis'ma Zh. Eksp. Teor. Fiz. 59 571 (1994) [JETP Lett. 59 596 (1994)]; Faddeev L D, Korchemsky G P Phys. Lett. B 342 311 (1995)
- Fadin V S, Lipatov L N Phys. Lett. B 429 127 (1998); Ciafaloni M, Camici G Phys. Lett. B 430 349 (1998)
- 17. Kotikov A V, Lipatov L N Nucl. Phys. B 661 19 (2003)
- Kotikov A V, Lipatov L N, Velizhanin V N Phys. Lett. B 557 114 (2003)
- Kotikov A V et al. *Phys. Lett. B* 595 521 (2004); 'Erratum' 632 754 (2006)
- 20. Moch S, Vermaseren J A M, Vogt A Nucl. Phys. B 688 101 (2004)
- 21. Lipatov L N Nucl. Phys. B 452 369 (1995); Phys. Rep. 286 131 (1997)
- 22. Antonov E N et al. Nucl. Phys. B 721 111 (2005)
- 23. Maldacena J M Adv. Theor. Math. Phys. 2 231 (1998)
- 24. Gubser S S, Klebanov I R, Polyakov A M *Phys. Lett. B* **428** 105 (1998)
- 25. Witten E Adv. Theor. Math. Phys. 2 253 (1998)
- Gubser S S, Klebanov I R, Polyakov A M Nucl. Phys. B 636 99 (2002)
- 27. Brower R C et al. JHEP (12) 005 (2007)
- Lipatov L N, in *Perspectives in Hadronic Physics: Proc. of Conf. ICTP*, *Triest*, *Italy*, *May 1997* (Eds S Boffi et al.) (Singapore: World Scientific, 1998)
- 29. Minahan J A, Zarembo K JHEP (03) 013 (2003)
- 30. Beisert N, Staudacher M Nucl. Phys. B 670 439 (2003)
- 31. Eden B, Staudacher M J. Stat. Mech. P11014 (2006)
- 32. Beisert N, Eden B, Staudacher M J. Stat. Mech. P01021 (2007)
- 33. Kotikov A V et al. J. Stat. Mech. P10003 (2007)
- 34. Bern Z, Dixon L J, Smirnov V A Phys. Rev. D 72 085001 (2005)
- 35. Bartels J, Lipatov L N, Sabio Vera A, arXiv:0802.2065

636