Russian original reference
Usp. Fiz. Nauk
Vol. **183**, pages

Contents

English translation reference
Phys. Usp.
Vol. **56**, pages

◇ ◇ ◇

In the next issue

Graphene based semiconductor nanostructures
P B Sorokin, L A Chernozatonskii

From self-focusing light beams to femtosecond laser pulse filamentation
S V Chekalin, V P Kandidov

Electrohydrodynamics of charged surfaces
A I Zhakin

Radio pulsars: the search for truth
V S Beskin, Ya N Istomin, A A Filippov

On the physical interpretation of Thomson scattering in a plasma
V N Tsytovich

Conferences and symposia

Physics news on the Internet

1063-7869(2013)56:1;1-I

ISSN: 1063 – 7869

Physics–Uspekhi

Advances in Physical Sciences

January 2013

Volume 56, Number 1

Translation of the Russian journal

Успехи физических наук, Uspekhi Fizicheskikh Nauk

January 2013, Volume 183, No. 1

Turpion

PHUSEY(1) 1 – 104 (2013) ISSN 1063-7869

PHYSICS–USPEKHI

Volume 56
Number 1
2013
Pages 1 – 104

English edition of *Uspekhi Fizicheskikh Nauk*
January 2013 Volume 183 Number 1

Contents

From the Editorial Board

This is the year when the Uspekhi Fizicheskikh Nauk (*UFN*) [*Physics – Uspekhi in English Edition*] *journal celebrates the 95th anniversary of the publication of its first issue. The objectives of the journal have remained unchanged all these years: the dissemination of information, first and foremost about the latest achievements in physics and in those fields of science that are closely related to physics. Our most important consideration has always been to take good care of our readers. We adhere to the opinion that the path that our journal should steer must not only meet the demand of mature physicists who wish to learn of the latest developments in their expert and related fields. We are convinced that UFN should be accessible to young scientists — senior year and PhD students — and all those who are just starting their careers in science or are new to a field they wish to work in. This is why we request authors of even monographic reviews written for specialists to begin their paper with a popularizing introduction which would enable any physicist to understand what the review is about, and to end it with a conclusion giving a brief exposition of the results and the prospects of the field. We ask the authors to explain all notations, even those widely accepted in special literature. They are also requested not to assume that all readers are familiar with the history of their specific issue and that they are also aware of the contribution by Russian authors. In view of this, we consider a brief historical overview as very welcome. Of course, Vitalii Lazarevich Ginzburg (a UFN author between 1938 and 2010, member of UFN Editorial Board from 1964 till 2009, and UFN Editor-in-Chief from 1998 until 2009) used to repeat: "Leave well alone!" Accordingly, these wishes are fairly fuzzy. We only ask our authors to reflect on what has been said and to correctly understand the comments of reviewers and members of the editorial board and of the editorial staff, who strive to do their jobs in the best possible way.*

We also need to emphasize the cardinal difference between UFN-type journals and journals publishing original research papers. UFN cannot and will not publish papers, even very good ones, if they are too narrowly specialized or are devoted to technical issues or to problems that the journal has already outlined in sufficient detail earlier. UFN is not a place for the publication of original results or fundamental theories and ideas that have not passed the test of time and scrutiny: these should go to specialized journals.

Neither do we welcome 'self-reviews' or papers in which the bibliographic list is dominated by references to its authors' publications.

Our authors should be aware that the selection of manuscripts for publication in UFN operates on a competitive basis. The competition depends on the number of papers currently in the editorial 'portfolio'. For this reason, reproaches like "the contents of the paper have not been fully studied or fairly evaluated" can not be taken as grounds for reviewing the decision of the editorial board.

UFN cannot offer its pages for materials relevant to long-lasting discussions among several authors. Our "discussion" format is as follows: a comment on the published paper plus the author's response. After this, the editorial staff closes the discussion on the pages of the journal. UFN also exercises great caution in printing material on natural sciences contiguous to physics. Material of this sort, prepared by

well-known experts in a specific field, must be of exceptionally high interest to physicists. Such off-profile papers pass through editing by both physicists and specialists in the field involved. Interesting materials of this sort can also be uploaded to the UFN web site.

UFN publishes articles discussing the history of physics and philosophy, but only in exceptional cases. Nevertheless, such materials of great interest to a large number of our readers can be published on the UFN web site. The same is true for some other matters which, not constituting products of research effort, are nevertheless of interest to our physics community. What we mean here are Personalia, book reviews, reports on seminars and conferences, and brief information on references, news, etc. that would be useful for the readers of UFN and visitors to our web site (*www.ufn.ru*).

We are grateful to all our readers for visiting our UFN web site more than 125 million times in total, and for 20 million visits annually in 2011 and 2012. In the process, readers have copied more than two million documents. Obviously, we are very thankful to all our authors whose papers attract so much interest from our readers. These statistics lead us to believe that the editorial policy of selecting papers for UFN (*which has remained essentially unchanged all these 95 years*) *has proved fully justified and useful to all physicists. The format of presenting the material on the pages of UFN has no doubt been improving all the time — we try hard 'to move with the times'. Suffice it to remind the readers that the UFN journal was one of the first in the world to create its electronic document version, as early as 1994. In 1995, UFN started to print ''Physics News on the Internet''. On V L Ginzburg's initiative, in 2000 the UFN web site incorporated a section, ''UFN's Tribune'', where UFN authors and readers could discuss issues significant for the physics community. In 1999, the journal started to post archived papers on its web site, and by 2005 we completed uploading the entire archive of our journal in Russian* (*from 1918 onwards*). *In 2002, it became possible to include full-color illustrations in the electronic version of the journal, and in 2004, video appendices to papers also appeared. On the 90th anniversary of the UFN journal* (*in 2008*), *personal pages of UFN authors were created on the web site, to which authors* (*should they so wish*) *can upload any information they consider interesting and useful to their readers. Photographs of many authors and interviews with the authors of the most cited papers in UFN were thus added. In 2007, UFN, was the first Russian journal to become a member of the Publishers International Linking Association* (*PILA*) *doing business as CrossRef, and being assigned DOI* (*Digital Object Index*) *to Russian-language versions of the papers in UFN. We now have the option to attach to each paper published online in UFN a list of publications that cited the given paper. Obviously, only citations from journals within the CrossRef system can be automatically taken into account. In 2012, we sent UFN authors a recommendation to accompany their reviews with a video annotation* (*video presentation*) *for the electronic version of UFN. We are especially grateful to those authors who responded to our request and provided wonderful video presentations of their papers.*

Clearly, these 'technological' innovations also aim at the execution of our primary objectives — to take good care of our readers and to provide them with materials not only of high scientific quality but also presented in the most convenient and convincing format. Of course, much remains to be improved, and we ask the readers and authors to help the editorial board in our endeavor. Many thanks to all our readers and authors for their critical comments and suggestions: we will try to take them into account.

Please, do write to us — every improvement in the UFN journal
is in the common interest of all physicists!

Physics – Uspekhi **56** (1) 3 – 28 (2013)

REVIEWS OF TOPICAL PROBLEMS PACS numbers: 13.75.Cs, 13.85.Dz

Elastic scattering of hadrons

I M Dremin

DOI: 10.3367/UFNe.0183.201301a.0003

Contents

Abstract. Colliding high-energy hadrons either produce new particles or scatter elastically with their quantum numbers conserved and no other particles produced. We consider the latter case here. Although inelastic processes dominate at high energies, elastic scattering contributes considerably (18–25%) to the total cross section. Its share first decreases and then increases at higher energies. Small-angle scattering prevails at all energies. Some characteristic features can be seen that provide information on the geometrical structure of the colliding particles and the relevant dynamical mechanisms. The steep Gaussian peak at small angles is followed by the exponential (Orear) regime with some shoulders and dips, and then by a power-law decrease. Results from various theoretical approaches are compared with experimental data. Phenomenological models claiming to describe this process are reviewed. The unitarity condition predicts an exponential fall for the differential cross section with an additional substructure to occur exactly between the low momentum transfer diffraction cone and a power-law, hard parton scattering regime under high momentum transfer. Data on the interference of the Coulomb and nuclear parts of amplitudes at extremely small angles provide the value of the real part of the forward scattering amplitude. The real part of the elastic scattering amplitude and the contribution of inelastic processes to the imaginary part of this amplitude (the so-called overlap function) are also discussed. Problems related to the scaling behavior of the differential cross section are considered. The power-law regime at highest momentum transfer is briefly described.

I M Dremin Lebedev Physical Institute, Russian Academy of Sciences, Leninskii prosp. 53, 119991 Moscow, Russian Federation
Tel. +7 (499) 783 37 19. Fax +7 (499) 135 78 80
E-mail: dremin@lpi.ru

Received 18 June 2012
Uspekhi Fizicheskikh Nauk **183** (1) 3 – 32 (2013)
DOI: 10.3367/UFNr.0183.201301a.0003
Translated by the author; edited by A M Semikhatov

"If only you knew what trash gives rise
To verses that are not ashamed uprise..."
Anna Akhmatova

1. Introduction

Hadron interactions are strong and, in principle, should be described by quantum chromodynamics (QCD). However, experimental data show that their main features originate from the nonperturbative sector of QCD. Only comparatively rare processes with large transferred momenta can be treated theoretically rather successfully by perturbative methods due to the well-known property of the asymptotic freedom of QCD. Hence, in the absence of methods for a rigorous solution of QCD equations, our understanding of the dynamics of the main bulk of strong interactions is severely limited by model building or some rare rigorous relations. In fact, our approach to high-energy hadronic processes at present is at best still in its infancy.

As has been learned from experiment, strong interactions of colliding high-energy particles give rise to inelastic and elastic processes. Some new particles (mostly pions) are produced in inelastic processes, which are the most probable ones, comprising 75% to 80% of all processes at high energies. Most created particles have comparatively small transverse momenta.

At the same time, in 25% to 20% of events, the colliding particles do not change their nature and scatter elastically, declining at some angle from their initial trajectories. The only information about this process available from experiment is obtained by the measurement of the differential cross section (proportional to the probability) of elastic scattering at some angle at a given energy.

In a very tiny range of extremely small angles, the charged particles scatter due to electromagnetic forces. But the dominant process of elastic scattering due to strong interactions proceeds at somewhat larger angles in the so-called diffraction cone. The differential cross sections are heavily weighted toward small transferred momenta exhibiting a huge peak. The scattering angle is still rather small there and becomes smaller and smaller as the energy increases. The probability of scattering at a given angle in this region

decreases steeply, similarly to a Gaussian exponential. Noticeably less than one percent of particles are elastically scattered to larger angles outside this diffraction cone. The Gaussian behavior is replaced there by a simple exponential one with some shoulders and (or) dips. At ever larger angles (or transferred momenta), a power-like decrease has been observed. At angles close to $\pi/2$, some additional flattening is seen.

The elastic cross section (the integral of the differential distribution over angles or transverse momentum) depends on the energy of the colliding partners. At high energies, it shows a steady tendency to become larger with an increase in energy. We note that the inelastic cross section also increases, such that their sum (the total cross section) increases as well.

The process of elastic scattering of hadrons has been studied experimentally in a wide energy range with different initial particles. At high energies of colliding partners, the most detailed results are available for the scattering of protons (pp) and antiprotons ($p\bar{p}$) on protons. We mainly discuss these data, sometimes referring to other colliding partners of protons such as pions and kaons.

Some surprises in the behavior of differential cross sections appeared in the 1960s when the very first experimental data on elastic pp and πp scattering were obtained at energies between 6.8 and 19.2 GeV in laboratory system [1–11] (the total energy in the center-of-mass system (cms) is only $\sqrt{s} \approx 4-6$ GeV!). The diffraction cone behavior changed at larger transferred momenta $|t|$ to a slower t-dependence. Somewhat later, the energy range was extended to 50 GeV [12–14]. With the advent of new accelerators, the data for pp scattering at energies $\sqrt{s} \approx 19$, 20, 23, 28, 31, 45, 53, 62 GeV were published [15–30], and the data for $p\bar{p}$ at 31, 53, 62, 546, 630, 1800, 1980 GeV [31–44] appeared. The early results are reviewed in Refs [45, 46]. The compilation of the data can be found in [47]. Only recently, the results of the TOTEM collaboration at the LHC on elastic pp scattering processes at $\sqrt{s} = 7$ TeV were published [48, 49].

Surely, these results called for their understanding and theoretical interpretation. The most important task is to acquire some knowledge about the internal structure of colliding particles by deciphering the information supplied by experimental data about the dependence on energy and transferred momentum. The transferred momentum is directly related to the size and the structure of those regions inside the hadron that participate in the interaction.

Many phenomenological models have been proposed. Most of them aspire to be a 'phenomenology of everything' related to elastic scattering of hadrons in a wide energy range. Doing so in the absence of applicable laws and methods of the fundamental theory, they have to use a large number of adjustable parameters. The free parameters have been determined by fitting the model results to the available experimental data. Even then, model predictions often fail when a new energy domain becomes accessible. And the 'verse' does not grow anymore! (If not recultivated.) Independent of their success and failure, we are sure that "in the long run, the physical picture may be expected to be much more important than most of the detailed computations" [50]. In what follows, we mention and discuss many of them.

The scattering of charged particles at extremely small angles is completely dominated by the Coulomb amplitude. The absolute value of the Born amplitude is well known. The phase of the Coulomb amplitude varies depending on the model chosen. However, this variation is rather mild in the considered tiny region of extremely small angles. The interference of the Coulomb amplitude with the strong-interaction (nuclear) amplitude in the transition region where they are almost equal has been used for the experimental determination of the ratio of the real to imaginary parts of the nuclear amplitude. This interference also depends on the chosen form of the nuclear amplitude. Theoretically, this ratio can be estimated with the help of dispersion relations. We briefly discuss this problem and show how the obtained results influence our analysis of scattering at somewhat larger angles.

The most numerous group of models deals with phenomenological attempts to describe the main bulk of elastic scattering at small angles in the diffraction cone. In general, they are based on some geometric models of particle substructure, with peripheral regions playing the decisive role. The approach using Reggeon (Pomeron) exchanges is the most popular among them. The approximately Gaussian (in angles) shape of the experimentally measured differential cross section in this region has been fitted just in this way. In addition to it, the simplest classical expressions for diffractive processes and results on the electromagnetic form factors are also used. However, the bold extension of the obtained results to larger angles is usually not very successful, even if some new parameters are introduced.

Particles scattered at larger angles give insight into the deeper internal regions of particle structure. The multiple iteration (rescattering) of diffractive processes can explain the region of angles that are somewhat larger than the diffractive ones. Without any additional model building, it can be described as a consequence of the unitarity condition. The only necessary input is the experimentally known energy behavior of the diffraction cone slope and the total cross section. This allows predicting the observed exponential fall-off with angles and damped oscillations imposed on it, which, depending on their amplitudes, lead to shoulders or dips of the differential cross sections.

At somewhat larger angles, the elastic processes can be considered to be dominated by the innermost constituents of the colliding particles. The perturbative QCD approach to hard parton scattering convoluted with some results or the parton structure of colliding particles is then used to describe experimental data. This approach predicts a power-like angular dependence of the differential cross sections. It has been seen in experiment. The dimensional (or quark) counting of the number of participating partons has been successful. The convolution with the internal structure of particles implies some coherence in the behavior of its constituents: all of them should coherently turn at the same angle. A particle should not be destroyed during the collision, and its internal wave function must be left intact. Therefore, we can call such processes coherent large-angle scattering.

At angles close to $\pi/2$, the effects of symmetrization of the corresponding amplitudes can become important and lead to some flattening of the differential distribution.

There are no strict definitions of the lower and upper bounds of these regions. The diffraction peak shrinks with energy, such that the exponential fall-off with the squared transferred momenta t terminates at ever smaller values. Correspondingly, the dip after it shifts to smaller values of $|t|$, as does the $\sqrt{|t|}$-exponential. At low energies, this regime approximately occupies the interval between 0.8 and 2 GeV^2, while in the LHC, it has moved to 0.4–1.5 GeV^2. According to the QCD prejudice, the scale for parton scattering should be

set above 1 GeV^2. This is actually observed with a power-like decrease starting somewhere around $|t| > 1.5-2$ GeV^2 at the LHC.

Hence, we can speak, at least, about five subregions of elastic scattering. We mainly discuss three of them: the diffraction cone, the Orear regime, and coherent hard parton scattering. The diffraction cone is well known to us from semiclassical effects. The regions beyond it became noticeable only at energies of colliding particles above several GeV, where processes of scattering at sufficiently large angles or transferred momenta are observable. They persist up to the present LHC energy of 7 TeV. Who ordered them and whether they will survive at ever higher energies are also questions to be discussed in this review.

The structure of this paper is as follows. The main relations between different characteristics of elastic scattering are presented in Section 2. Then, in Section 3, their global dependences on energy and transferred momenta are discussed, together with our attempts to understand their implications within the simplest approaches. A more detailed analysis of experimental data in the framework of different theoretical ideas and approximations is the content of Section 4. Finally, the general picture is briefly discussed in Section 5.

We do not consider the scattering of polarized particles, and the spin structure of the amplitude is ignored.

2. Main relations

As discussed above, the measurement of the differential cross section is the only source of experimental information about a process. Hence, the main characteristics of hadron interactions directly related to the elastic scattering amplitude, such as the total cross section, the elastic scattering cross section, the ratio of the real to the imaginary part of the amplitude, and the slope of the diffraction cone, are obtained. The first two are functions of the total energy only, while the others depend on two variables: the total energy and the transferred momentum (or the scattering angle).

The dimensionless elastic scattering amplitude A defines the differential cross section as

$$\frac{\mathrm{d}\sigma(s)}{\mathrm{d}t} = \frac{1}{16\pi s^2}|A|^2 = \frac{1}{16\pi s^2}\left(\operatorname{Im} A(s,t)\right)^2\left(1+\rho^2(s,t)\right), \quad (1)$$

where the ratio of the real to imaginary parts of the amplitude is defined:

$$\rho(s,t) = \frac{\operatorname{Re} A(s,t)}{\operatorname{Im} A(s,t)}. \quad (2)$$

In what follows, we consider very high energy processes. Therefore, the masses of the colliding particles can be neglected, and we use the expression $s = 4E^2 \approx 4p^2$, where E and p are the energy and momentum in the center-of-mass system. The four-momentum transfer squared is

$$-t = 2p^2(1-\cos\theta) \approx p^2\theta^2 \approx p_\mathrm{t}^2, \quad \theta \ll 1, \quad (3)$$

with θ denoting the scattering angle in the center-of-mass system and p_t being the transverse momentum.

The elastic scattering cross section is given by the integral of differential cross section (1) over all transferred momenta:

$$\sigma_\mathrm{el}(s) = \int_{t_\mathrm{min}}^{0} \mathrm{d}t\, \frac{\mathrm{d}\sigma(s)}{\mathrm{d}t}. \quad (4)$$

The total cross section σ_t is related by the optical theorem to the imaginary part of the forward scattering amplitude as

$$\sigma_\mathrm{t}(s) = \frac{\operatorname{Im} A(p,\theta=0)}{s}. \quad (5)$$

Elastically scattered hadrons escape from the interaction region declining mostly at quite small angles within the so-called diffraction cone.[1] Therefore, the main focus has been on this region. As is known from experiment, the diffraction peak has a Gaussian shape in the scattering angles or decreases exponentially as a function of the transferred momentum squared:

$$\frac{\mathrm{d}\sigma/\mathrm{d}t}{(\mathrm{d}\sigma/\mathrm{d}t)_{t=0}} = \exp(Bt) \approx \exp(-Bp^2\theta^2). \quad (6)$$

In view of relations (4)–(6), any successful theoretical description of the differential distribution must also work in fitting the energy dependences of the total and elastic cross sections. The diffraction cone slope B is given by

$$B(s,t) \approx \frac{\mathrm{d}}{\mathrm{d}t}\left(\ln\frac{\mathrm{d}\sigma(s,t)}{\mathrm{d}t}\right). \quad (7)$$

Actually, the slope B depends slightly on t at a given energy s, e.g., at the LHC, its value changes by about 10% within the cone for $|\Delta t| \approx 0.3$ GeV^2. We neglect this in the first approximation. The normalization factor in Eqn (6) is

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}t}\right)_{t=0} = \frac{\sigma_\mathrm{t}^2(s)\left(1+\rho_0^2(s)\right)}{16\pi}, \quad (8)$$

where ρ_0 is defined as the ratio of the real and imaginary parts of the amplitude in the forward direction at $\theta = t = 0$. Equation (8) follows from formula (1) and optical theorem (5) at $t = 0$.

According to the dispersion relations, which connect the real and imaginary parts of the amplitude, and optical theorem (5), the value ρ_0 can be expressed as an integral of the total cross section over the whole energy range. In practice, ρ_0 is mainly sensitive to the local derivative of the total cross section. In the first approximation, the result of the dispersion relation can then be written in the form [51–54]

$$\rho_0(s) \approx \frac{1}{\sigma_\mathrm{t}}\left[\tan\left(\frac{\pi}{2}\frac{\mathrm{d}}{\mathrm{d}\ln s}\right)\right]\sigma_\mathrm{t} = \frac{1}{\sigma_\mathrm{t}}\left[\frac{\pi}{2}\frac{\mathrm{d}}{\mathrm{d}\ln s} + \frac{1}{3}\left(\frac{\pi}{2}\right)^3\frac{\mathrm{d}^3}{\mathrm{d}\ln s^3} + \ldots\right]\sigma_\mathrm{t}. \quad (9)$$

It follows that at high energies, $\rho_0(s)$ is mainly determined by the derivative of the logarithm of the total cross section with respect to the logarithm of energy.

The bold extension of the first term in this series to nonzero transferred momenta would look like

$$\rho(s,t) \approx \frac{\pi}{2}\left(\frac{\mathrm{d}\ln\operatorname{Im} A(s,t)}{\mathrm{d}\ln s} - 1\right). \quad (10)$$

[1] In practice, the tiny region of the interference of the Coulomb and nuclear amplitudes at extremely small angles does not contribute to the total cross section of elastic scattering. Its role in obtaining some estimates of $\rho(s,t)$ is described below.

If we neglect the high-$|t|$ tail of the differential cross section, which is several orders of magnitude lower than the optical point, and integrate in Eqn (4) using expression (6) with constant B, we obtain the approximate relation between the total cross section, the elastic cross section, and the slope:

$$\frac{\sigma_{\rm t}^2(1+\rho_0^2)}{16\pi B\sigma_{\rm el}} \approx 1\,. \tag{11}$$

We can compare this formula with the upper bound obtained in Ref. [55]:

$$\frac{\sigma_{\rm t}^2}{18\pi B\sigma_{\rm el}} \leqslant 1\,. \tag{12}$$

The phase ζ of the hadronic amplitude is often defined as

$$A(s,t) = {\rm i}\big|A(s,t)\big| \exp\big[-{\rm i}\zeta(s,t)\big]\,; \tag{13}$$

then

$$\rho(s,t) = \tan\zeta(s,t)\,. \tag{14}$$

These formulas are used for measuring the luminosity, which relates the cross section σ_i of a given process i to the corresponding number of events N_i by

$$L = \frac{N_i}{\sigma_i}\,. \tag{15}$$

A simultaneous measurement of the total number of events $N_{\rm t}$ and the number of elastic events $N_{\rm el}$ is used to define the luminosity as

$$L = \frac{1+\rho_0^2}{16\pi}\,\frac{N_{\rm t}^2}{{\rm d}N_{\rm el}/{\rm d}t\big|_{t=0}}\,. \tag{16}$$

The measured total cross section is independent of luminosity:

$$\sigma_{\rm t} = \frac{16\pi}{1+\rho_0^2}\,\frac{{\rm d}N_{\rm el}/{\rm d}t\big|_{t=0}}{N_{\rm t}}\,. \tag{17}$$

The elastic scattering amplitude must satisfy the general properties of analyticity, crossing symmetry, and unitarity. The unitarity of the S-matrix, $SS^+ = 1$, imposes certain requirements on it. In the s-channel, we have

$$\begin{aligned}{\rm Im}\,A(p,\theta) &= I_2(p,\theta) + F(p,\theta)\\ &= \frac{1}{32\pi^2}\iint {\rm d}\theta_1\,{\rm d}\theta_2\,\sin\theta_1\sin\theta_2 A(p,\theta_1)A^*(p,\theta_2)\\ &\times\Big\{\big[\cos\theta-\cos(\theta_1+\theta_2)\big]\big[\cos(\theta_1-\theta_2)-\cos\theta\big]\Big\}^{-1/2} + F(p,\theta)\,.\end{aligned} \tag{18}$$

The region of integration in (18) is defined by the conditions

$$|\theta_1-\theta_2| \leqslant \theta\,, \qquad \theta \leqslant \theta_1+\theta_2 \leqslant 2\pi-\theta\,. \tag{19}$$

The integral term represents the two-particle intermediate states of the incoming particles. The function $F(p,\theta)$ represents the shadowing contribution of the inelastic processes to the elastic scattering amplitude. Following [56], we call it the overlap function. It determines the shape of the diffraction peak and is completely nonperturbative. Only some phenomenological models can claim to describe it.

In the forward direction $\theta = 0$, this relation, in combination with optical theorem (5), reduces to the general statement that the total cross section is the sum of cross sections of elastic and inelastic processes:

$$\sigma_{\rm t} = \sigma_{\rm el} + \sigma_{\rm in}\,. \tag{20}$$

Unitarity relation (18) has been successfully used [57–60] for the model-independent description of the Orear region between the diffraction cone and hard parton scattering, which became the crucial test for phenomenological models.

Experimentally, all characteristics of elastic scattering are measured as functions of the energy s and transferred momentum t. However, it is desirable to have concrete information on the geometric structure of scattered particles and the role of different spatial regions in the scattering process. We should use the Fourier–Bessel transform to obtain the correspondence between the transferred momenta and these space regions. The transverse distance between the centers of colliding particles, called the impact parameter **b**, determines the effective transferred momentum t. The amplitudes in the corresponding representations are related as

$$h(s,b) = \frac{1}{16\pi s}\int_{t_{\rm min}=-s}^{0} {\rm d}t\,A(s,t)\,J_0(b\sqrt{-t}\,)\,. \tag{21}$$

More peripheral collisions with large b lead to smaller transferred momenta $|t|$.

The amplitude $A(s,t)$ can be connected to the eikonal phase $\delta(s,\mathbf{b})$ and to the opaqueness (or blackness) $\Omega(s,\mathbf{b})$ at the impact parameter **b** by the Fourier–Bessel transformation

$$\begin{aligned}A(s,t=-q^2) &= \frac{2s}{\rm i}\int {\rm d}^2b\,\exp({\rm i}\mathbf{qb})\big[\exp\big(2{\rm i}\delta(s,\mathbf{b})\big)-1\big]\\ &= 2{\rm i}s\int {\rm d}^2b\,\exp\big({\rm i}\mathbf{qb}\big)\big[1-\exp\big(-\Omega(s,\mathbf{b})\big)\big]\,.\end{aligned} \tag{22}$$

The integration is over the two-dimensional space of the impact parameter **b**.

Assuming $\Omega(s,\mathbf{b})$ to be real and using Eqn (5), we obtain

$$\sigma_{\rm t} = 4\pi\int_0^\infty \big[1-\exp\big(-\Omega(s,\mathbf{b})\big)\big]b\,{\rm d}b\,. \tag{23}$$

Also,

$$\sigma_{\rm el} = 2\pi\int_0^\infty \big[1-\exp\big(-\Omega(s,\mathbf{b})\big)\big]^2 b\,{\rm d}b\,, \tag{24}$$

$$B = \frac{\int_0^\infty \big[1-\exp\big(-\Omega(s,\mathbf{b})\big)\big]b^3\,{\rm d}b}{2\int_0^\infty \big[1-\exp\big(-\Omega(s,\mathbf{b})\big)\big]b\,{\rm d}b}\,. \tag{25}$$

To apply the inverse transformation, we must know the amplitude $A(s,t)$ at all transferred momenta. Therefore, it is necessary to continue it analytically to the nonphysical region of t [61, 62]. This can be done [63]. Correspondingly, the mathematically consistent inverse formulas generally contain the sum of contributions from the physical and nonphysical parts of the amplitude $A(s,t)$. Unitarity condition (18) involves only the amplitude in the physical region; only this part of its Fourier–Bessel transform is important in the unitarity relation for the impact parameter representation. It is written as

$${\rm Im}\,h(s,b) = \big|h(s,b)\big|^2 + F(s,b)\,, \tag{26}$$

where $h(s,b)$ and $F(s,b)$ are obtained by the direct transformation of $A(s,t)$ and $F(s,t)$ integrated only over the physical transferred momenta from $t_{\min}$ to 0. They show the dependence of the intensity of elastic and inelastic interactions on the mutual impact parameter of the colliding particles. Analogously to relation (20), the integrals over all impact parameter values in this relation respectively represent the total, elastic, and inelastic cross sections. It is especially simple to calculate the overlap function from algebraic equation (26) if the real part is small in some subregion, i.e., $|h(s,b)| \approx \operatorname{Im} h(s,b)$. Then

$$\operatorname{Im} h(s,b) \approx \frac{1}{2}\left(1 - \sqrt{1 - 4F(s,b)}\,\right). \tag{27}$$

In the region where the transformed overlap function is small, $F(s,b) \ll 1$, the imaginary part is also small: $\operatorname{Im} h(s,b) \approx F(s,b)$.

However, the accuracy of the unitarity condition in b-representation (26) is still under discussion [61–65], because some corrections due to the nonphysical region enter there, even though their role may be negligible. Moreover, the further use of the approximate formulas of the quasi-eikonal unitarization often leads to failure in describing the differential cross section outside the diffraction cone.

The average values of the impact parameters for all — elastic and inelastic — processes can be estimated from the amplitude $A(s,t)$ if we assume that $\mathrm{d}\rho/\mathrm{d}t = 0$ at $t = 0$ [64]:

$$\left\langle b^2(s)\right\rangle_{\text{tot}} = \frac{\sigma_{\text{el}}}{\sigma_{\text{t}}}\left\langle b^2(s)\right\rangle_{\text{el}} + \frac{\sigma_{\text{in}}}{\sigma_{\text{t}}}\left\langle b^2(s)\right\rangle_{\text{in}} = 2B(s,0)\,, \tag{28}$$

where, e.g.,

$$\left\langle b^2(s)\right\rangle_{\text{el}} = 4\int_{t_{\min}}^{0} \mathrm{d}t\,|t|\left|\frac{\mathrm{d}}{\mathrm{d}t}\,A(s,t)\right|^2 \left(\int_{t_{\min}}^{0} \mathrm{d}t\,\left|A(s,t)\right|^2\right)^{-1}. \tag{29}$$

Nevertheless, the problem of the relative contributions of the central (small b) and peripheral (large b) regions under elastic hadron collisions is still widely disputed. We must be especially careful when considering unitarity condition (26) with small impact parameters for certain models. Slight variations of $h(s,b)$ in this region may lead to strong variations of the amplitude $A(s,t)$ at large $|t|$.

The elastic scattering at extremely small angles allows estimating the forward ratio of the real part of the amplitude to its imaginary part ρ_0 in experiment. For completeness, we show an approximate expression for the amplitude $A(s,t)$ in the region dominated by the Coulomb amplitude and its interference with the nuclear amplitude:

$$A^{\text{CN}}(s,t) = \mp\frac{8\pi\alpha}{|t|}\,s f_1\left(|t|\right) f_2\left(|t|\right)\exp\left(\mathrm{i}\alpha\Phi\right) + \left(\mathrm{i} + \rho_0(s)\right)s\sigma_{\text{t}}\exp\frac{Bt}{2}\,, \tag{30}$$

where the upper (lower) sign corresponds to the scattering of particles with the same (opposite) electric charges, the form factors of two colliding particles $f_j(|t|)$ added 'by hand' in Eqn (30) take their internal composition into account, Φ is the Coulomb phase, and $\alpha = 1/137$ is the fine structure constant. The expressions for $f_j(|t|)$ and Φ depend on various prescriptions for them obtained with different assumptions concerning the internal structure of a hadron. The most popular shapes of the form factors are either the Gaussian fall-off with an increasing angle, like $\exp(2t/\Lambda^2)$, similar to that in (6), or the dipole (power-like) approximation, like $(1 - t/\Lambda^2)^{-2}$, with some more complicated subleading factors. The phase Φ usually contains a term with the typical logarithmic dependence on the angle θ, which becomes large at very small angles, and some subleading terms. In both cases, the subleading terms have to contain additional free parameters for a more accurate description of experimental data. As we see, the ratio $\rho(s,t)$ in (2) is approximated by $\rho(s,0) = \rho_0$ in the fit (30). This implies that both real and imaginary parts of the nuclear amplitude exhibit the same purely exponential t-dependence in the interference region (with the dominance of the imaginary part for small ρ_0). More details can be found in [66–81].

3. Where do we stand now?

We first discuss the asymptotic properties of fundamental characteristics such as the total cross section σ_{t}, the elastic cross section σ_{el}, the ratio of the real part to the imaginary part of the elastic amplitude ρ, and the width of the diffraction peak B at infinite energies. Then we compare this with some trends in present experimental data.

More than half a century ago, it was claimed in [82, 83] that according to the general principles of field theory and ideas about hadron interactions, the total cross section cannot increase with energy faster than $\ln^2 s$. The upper bound was recently improved [84], with the coefficient in front of the logarithm shown to be half that in the earlier limit,

$$\sigma_{\text{t}} \leqslant \frac{\pi}{2m_\pi^2}\ln^2\frac{s}{s_0}\,, \tag{31}$$

where m_π is the pion mass. If estimated at present energies, this bound is still much higher than the experimentally measured values of the cross sections, with $s_0 = 1$ GeV^2 chosen as a 'natural' scale. Therefore, this is only a functional constraint; it forbids extremely fast growth of the total cross section, asymptotically exceeding the above limits. Both the coefficient in front of the logarithm in (31) and the constancy of s_0 are often questioned. In particular, some possible dependence of s_0 on the energy s has been pointed out (see, e.g., [85]).

The Heisenberg uncertainty relation shows that such a regime favors an exponentially bounded spatial profile of the matter density distribution $D(r)$ in colliding particles, such as $D(r) \propto \exp(-mr)$. Because the energy density is $ED(r)$ and there should be at least one created particle with mass m in the overlap region, the condition $ED(r) = m$ leads to $r \leqslant (1/m)\ln(s/m^2)$ and, consequently, to the functional dependence in (31).

It was Heisenberg who first proposed such a behavior of total cross sections [86]. He considered the pion production processes in proton–proton collisions as a shock wave problem governed by some nonlinear field theory equations.

To study the asymptotic regime, some theoretical arguments based on the general principles of field theory and the analogy of strong interactions to massive quantum electrodynamics [87] were promoted. The property that the limits as $s \to \infty$ and $M \to 0$ (where M is the photon mass) commute has been used [88], implying that the asymptotic domain of strong interactions coincides with the massless limit of quantum electrodynamics. These studies led to the general geometric picture of two hadrons colliding with asymptotically high energies and interacting as Lorentz-contracted

Table 1. The gray and Gaussian disks models ($X = \sigma_{el}/\sigma_t$, $Z = 4\pi B/\sigma_t$).

Model	$1 - \exp(-\Omega) = \Gamma(s,b)$	σ_t	B	X	Z	X/Z	XZ
Gray	$\alpha\theta(R-b), 0 \leqslant \alpha < 1$	$2\pi\alpha R^2$	$R^2/4$	$\alpha/2$	$1/2\alpha$	α^2	$1/4$
Gaussian	$\alpha \exp(-b^2/R^2), 0 \leqslant \alpha \leq 1$	$2\pi\alpha R^2$	$R^2/2$	$\alpha/4$	$1/\alpha$	$\alpha^2/4$	$1/4$

black disks (see also review paper [89]). In what follows, we discuss some other possibilities as well. But as a starting point for further reference, we describe the predictions of this proposal.

The main conclusions are:

(1) For black ($\Omega(s,\mathbf{b}) \to \infty$) and logarithmically expanding disks with finite radii R ($R = R_0 \ln s$, $R_0 = \text{const}$), it follows from (23) that σ_t asymptotically approaches infinity as

$$\sigma_t(s) = 2\pi R^2 + O(\ln s)\,, \quad R = R_0 \ln s\,, \quad R_0 = \text{const}\,. \quad (32)$$

(2) The elastic and inelastic processes make equal contributions to the total cross section:

$$\frac{\sigma_{el}(s)}{\sigma_t(s)} = \frac{\sigma_{in}(s)}{\sigma_t(s)} = \frac{1}{2} \mp O(\ln^{-1} s)\,. \quad (33)$$

This quantum mechanical result differs from 'intuitive' classical predictions.

(3) The width of the diffraction peak $B^{-1}(s)$ must shrink because its slope increases as (see also [90])

$$B(s) = \frac{R^2}{4} + O(\ln s)\,. \quad (34)$$

(4) The forward ratio of the real part to the imaginary part of the amplitude ρ_0 must vanish asymptotically as

$$\rho_0 = \frac{\pi}{\ln s} + O(\ln^{-2} s)\,. \quad (35)$$

This result follows directly from Eqn (9) for $\sigma_t \propto \ln^2 s$.

(5) The differential cross section has a shape resembling the classical diffraction of light on a disk:

$$\frac{d\sigma}{dt} = \pi R^4 \left(\frac{J_1(qR)}{qR}\right)^2, \quad (36)$$

where $q^2 = -t$.

(6) The product of σ_t with the value γ of $|t|$ at which the first dip in the differential elastic cross section occurs is a constant independent of the energy:

$$\gamma\sigma_t = 2\pi^3\beta_1^2 + O(\ln^{-1} s) = 35.92 \text{ mb GeV}^2\,, \quad (37)$$

where $\beta_1 = 1.2197$ is the first zero of $J_1(\beta\pi)$.

These are merely a few conclusions among many others, albeit model-dependent ones.

None of these asymptotic predictions have been observed yet in experiment.

Surely, there is another possibility — more realistic at present energies — that the black disk model is too extreme and the gray fringe always exists. It opens the way to much speculation, with many new parameters concerning particle shape and opacity (see, e.g., [78, 80, 81, 91–102]).

The black disk limit might be unrealistic. Therefore, in Table 1 we show the predictions of the gray disk model with the steep rigid edge described by the Heaviside step-function and the Gaussian disk model. The total cross section, the slope B, the ratio of the elastic to total cross section $X = \sigma_{el}/\sigma_t$, the ratios $Z = 4\pi B/\sigma_t$ and X/Z, and the product XZ are displayed there; $\Gamma(s,b)$ is the diffraction profile function.

The slope B is completely determined by the size of the interaction region R. Other characteristics are sensitive to the blackness of disks α. In particular, the ratio X is proportional to α. The ratio Z plays an important role for fits at larger angles, as explained in Section 4.2. It is inversely proportional to α. The corresponding formulas are given by (23)–(25). The black-disk limit follows from the gray-disk model at $\alpha = 1$. For a Gaussian distribution of matter, the disk becomes nontransparent at its center in this limit. The parameter XZ is constant in these models and does not depend on the nucleon transparency. On the contrary, the parameter X/Z is very sensitive to it, being proportional to α^2. Therefore, it would be extremely instructive to obtain knowledge about them from experimental data.

In Table 2, we show how the above ratios evolve with energy according to experimental data. Most of the entries, except the last two, are taken from Refs [91, 103], with the simple recalculation $Z = 1/4Y$. The data at Tevatron and LHC energies are taken from Refs [48, 49, 104]. All results are for pp scattering, except those at 546 and 1800 GeV for $p\bar{p}$ processes, which should be close to pp at these energies. The accuracy of the numbers listed in Table 2 can be very approximately estimated to be better than $\pm 10\%$ from known error bars for the cross sections and the slopes.

Table 2. The energy behavior of various characteristics of elastic scattering.

$\sqrt{s}$, ГэВ	2.70	4.11	4.74	6.27	7.62	13.8	62.5	546	1800	7000
X	0.42	0.28	0.27	0.24	0.22	0.18	0.17	0.21	0.23	0.25
Z	0.64	1.02	1.09	1.26	1.34	1.45	1.50	1.20	1.08	1.00
X/Z	0.66	0.27	0.25	0.21	0.17	0.16	0.11	0.18	0.21	0.25
XZ	0.27	0.28	0.29	0.30	0.30	0.26	0.25	0.26	0.25	0.25

The most interesting feature of the experimental results is the minimum of the blackness parameter α at ISR energies. It can be clearly seen in the minima of X and X/Z and in the maximum of Z at $\sqrt{s} = 62.5$ GeV. The steady decrease in ratios X proportional to α and X/Z proportional to α^2 up to the ISR energies and their increase at Sp$\bar{p}$S, Tevatron, and LHC energies means that the nucleons become more transparent up to the ISR energies and more black toward 7 TeV. The same conclusion follows from the behavior of Z, which is inversely proportional to α. The value of Z rapidly approaches its limit for the Gaussian distribution of matter in the disk. For the Gaussian shape, the parameter X/Z cannot exceed 0.25. This model is excluded only at low energies. According to Eqn (12), $XZ \approx 0.25(1 + \rho_0^2)$, which is indeed close to 0.25 within the experimental errors, the estimate of $\rho_0^2 \leqslant 0.02$, and slight variations of B inside the cone in the framework of our crude model as predicted in Table 1. This

shows that our models are not bad for qualitative estimates in a first approximation.

Before discussing various fits, we briefly comment on some important general trends in high-energy data observed in experiment.

(1) Total cross sections increase with energy. At present energies, the power-like approximation is the most preferable one. The preasymptotic behavior of σ_t proposed in earlier papers [87, 88] was

$$\sigma_t \propto s^a \ln^{-2} s\,, \tag{38}$$

where the numerical value of a was estimated to be of the order of unity in strong interactions. It was shown in [105] to lie in the range between 0.08 and 0.2, which is close to values obtained in recent phenomenological fits. The power-law increase persists in a wide interval of energies (see Ref. [106] for the recent analysis of experimental data). Consequently, the density distribution in colliding particles is closer to a power-like dependence than to an exponential one in that energy range.

(2) The ratio $X = \sigma_{el}/\sigma_t$ decreases from low energies to those of ISR, where it becomes approximately 0.17 and then strongly increases to 0.25 at the LHC energies. However, it is still quite far from the asymptotic value 0.5, corresponding to the black-disk limit.

The only higher-energy data came from the Pierre Auger collaboration, which recently reported [107] a measurement of the inelastic p–air cross section σ_{in}^{p-air} at $\sqrt{s} = 57 \pm 6$ TeV. After some corrections and Glauber model calculations, it results in the pp inelastic cross section $\sigma_{in}^{pp} \approx 90$ mb. Some models [108–110] extrapolate their predictions for the total cross section to this energy and obtain a value of about 135 mb. Hence, the ratio of the inelastic to the total cross section could become equal to 0.67, which is smaller than 0.75 at 7 TeV. However, it is premature to reach any definite conclusions because of large errors in the cosmic ray data and the underestimated value of the total cross section predicted to be 7 TeV by the model [108–110]. The extrapolation to infinite energies done in the same model leads to this ratio estimated as 0.509, which is compatible with the black-disk predictions. Still, asymptopia is but an elusive concept!

Sometimes, the modified black-disk limit is attributed to the sum of elastic and diffractive processes [111]. It may then be that

$$\frac{\sigma_{el} + \sigma_{diff}}{\sigma_t} \to \frac{1}{2}\,, \tag{39}$$

where σ_{diff} is the sum of cross sections of single and double inelastic diffraction. The fits in Ref. [106] suggest that the relations

$$\frac{\sigma_{el}}{\sigma_t} \to \frac{1}{3}\,, \qquad \frac{\sigma_{diff}}{\sigma_t} \to \frac{1}{6} \tag{40}$$

can be valid separately.

(3) The diffraction peak shrinks about twice from energies $\sqrt{s} \approx 6$ GeV, where $B \approx 10$ GeV^{-2}, to the LHC energy, where $B \approx 20$ GeV^{-2}. At the ISR energies, the slope $B(s)$ increases logarithmically. Accounting for LHC data requires a stronger dependence than a simple logarithmic one. The terms proportional to $\ln^2 s$ are usually added in phenomenological fits. Even then, predictions [112, 113] are not completely satisfactory. At present energies, in connection with the power-like preasymptotic behavior of σ_t, we could also expect a faster-than-logarithmic shrinkage of the diffraction peak.

The tendency in the peak behavior at larger $|t|$ also changes with an energy increase. In the energy region up to ISR, it becomes less steep near its end (see Figs 4 and 5 in Ref. [45]), but its slope increases at the LHC energies. Both the minimum and maximum following the peak shift to smaller $|t|$.

As regards the behavior of the differential cross section as a function of the transverse momentum behind the maximum, the t-exponential of the diffraction peak is replaced, according to experimental data, by a $(-\sqrt{|t|} \approx -p_t)$-exponential at the intermediate angles:

$$\frac{d\sigma}{dt} \propto \exp\left(-2a\sqrt{|t|}\right), \qquad a \approx \sqrt{B}\,. \tag{41}$$

The slope $2a$ in this region also increases with the energy, and the whole Orear region shifts to the ever lower transferred momenta.

In this connection, we also note an intriguing property of the ratio $Z = 4\pi B/\sigma_t$, which is closely related to the value of the slope B. From Table 2, we see that it is about 1 at $\sqrt{s} = 4$ GeV, increases to 1.5 at ISR energies, and then again drops to around 1 at 7 TeV. This ratio, in combination with values of ρ at different angles, determines the slope in the $|t|$ region beyond the diffraction peak at any s (see Ref. [58] and the discussion in Section 4.2.2). According to Eqns (32) and (34), Z should decrease and be asymptotically equal to $1/2$ in the black-disk limit, such that the relation

$$\sigma_t = 8\pi B \tag{42}$$

be asymptotically fulfilled. At the LHC energy 7 TeV, the coefficient in the right-hand side is still half as much. However, if the preasymptotic power-like increase in the total cross section accompanied by a slower increase in the slope persists, the tendency to this limit looks quite promising.

The relation between σ_t and B is also discussed in Refs [106, 114]. In particular, the fits in [106] correspond to the value $Z \approx 0.93$ at Auger energies 57 ± 6 TeV, i.e., lower than 1 at 7 TeV.

(4) As a function of energy, the ratio ρ_0 increases from negative values at comparatively low energies, crosses zero in the region of hundreds of GeV, and becomes positive at higher energies. This is a general tendency for collisions of any initial particles. For pp scattering, the prediction of (35) with values of s scaled by 1 GeV is still somewhat higher (about 0.177) than the estimates from dispersion relations (≈ 0.14 in Refs [110, 115]), even at 7 TeV, while strongly overshooting them at ISR, where $\pi/\ln s \approx 0.37$. No logarithmic decrease is seen in these predictions, which, however, depend on the behavior of the total cross section at higher energies. Moreover, the value 0.14 can only be reached according to (35) at the energy of 75 TeV. Probably, at energies higher than 75 TeV, the first signs of approach to the asymptotic regime will become visible. No data about ρ_0 at the LHC energies exist yet. The local value of ρ_0 estimated from Eqn (9) with a power-like fit of the total cross section, proportional to s^{Δ}, is $\rho_0 \approx \pi\Delta/2$. That agrees quite well with the soft Pomeron intercept $\Delta \approx 0.08$.

(5) To describe the shape of the differential cross section in the diffraction cone, significant corrections to Eqn (36) must be added at present energies. This is discussed in Section 4.1.

(6) The product $\gamma\sigma_{\rm t}$ changes from 39.5 mb GeV^2 at $\sqrt{s} = 6.2$ GeV to 51.9 mb GeV^2 at $\sqrt{s} = 7$ TeV and strongly deviates from the predicted asymptotic value (37). The total cross section $\sigma_{\rm t}$ increases faster than γ decreases.

From the geometrical standpoint, the general picture is one of protons becoming blacker, edgier, and larger (BEL) [116]. We conclude that even though the qualitative trends may be considered rather satisfactory, we are still quite far from the asymptotic regime, even at the LHC energies. This feature may be connected [117, 118] with the strong evolution of the parton content of strong interactions at present energies, revealing itself in an increase in the number of active parton pairs inside each proton with energy increase (higher density) and a softening of the structure functions, which leads to lower energy shares x for each parton pair (larger radii).

4. Experimental data and phenomenological models

As always, our knowledge about particular physical processes is limited by the practical possibility of measuring their characteristics. As mentioned above, numerous experimental data on elastic scattering of hadrons at various angles and at different energies have been obtained. Unfortunately, in some of them, the available region of angles is strongly limited by the experimental setup. Therefore, a comparison with theoretical proposals is possible only in the corresponding range of angles and energies.

The data and their fits at various energies and in different intervals of transferred momenta for different participating particles are so numerous that it is impossible to show all of them in a single review paper. Therefore, from the very beginning, we use the latest results of the TOTEM collaboration at the highest LHC energy, 7 TeV, as a reference point [48, 49]. The discussion of theoretical models is also concentrated around these data.

The total and elastic cross sections at 7 TeV are respectively estimated as 98.3 mb and 24.8 mb.[2] The figures from published papers [48, 49] demonstrating the behavior of the differential cross section as a function of the transferred momentum are displayed below. They clearly confirm the existence of the three regions discussed above.

The cross section shape in the region of the diffraction cone [48] is shown in Fig. 1. The t-exponential behavior with $B \approx 20.1$ GeV^{-2} is clearly seen at $|t| < 0.3$ GeV^2. The peak steepens at the end of the diffraction cone, and its slope becomes approximately 23.6 GeV^{-2} in the $|t|$ interval (0.36–0.47) GeV^2. The results at somewhat larger angles [49] in the Orear region are presented in Fig. 2. The dip at $|t| \approx 0.53$ GeV^2 with a subsequent maximum at $|t| \approx 0.7$ GeV^2 and the $\sqrt{|t|}$-exponential behavior are demonstrated. Some curves corresponding to different model predictions are also drawn here. The same data as in Fig. 2 are shown in Fig. 3, but with more details, including the steepened slope, the dip position, and the region of $|t|^{-8}$-behavior. The last one is ascribed to the hard parton scattering processes.

We congratulate all members of the TOTEM collaboration with this fantastic achievement! Their efforts are truly appreciated when estimating the values of angles at which the measurements had to be done. They were even smaller than 10^{-4}! Detectors had to be installed at very long distances from the collision point to obtain results at low transferred momenta. These data revived interest in elastic scattering.

[2] Here, we do not reproduce the statistical and systematic errors. They are shown in the original papers.

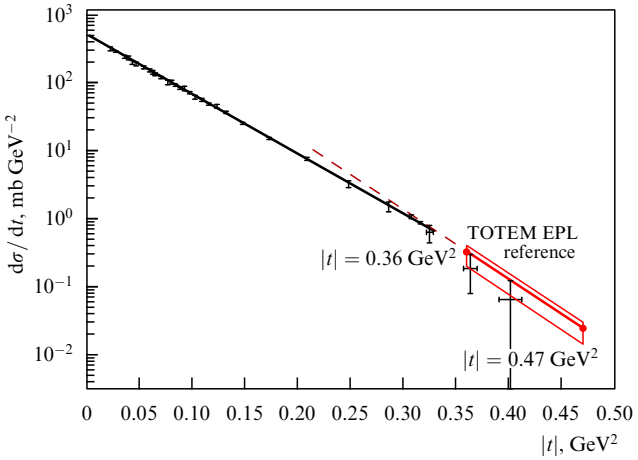


Figure 1. The differential cross section of elastic proton–proton scattering at $\sqrt{s} = 7$ TeV measured by the TOTEM collaboration (Fig. 4 in [49]). The region of the diffraction cone with the $|t|$-exponential decrease is shown.

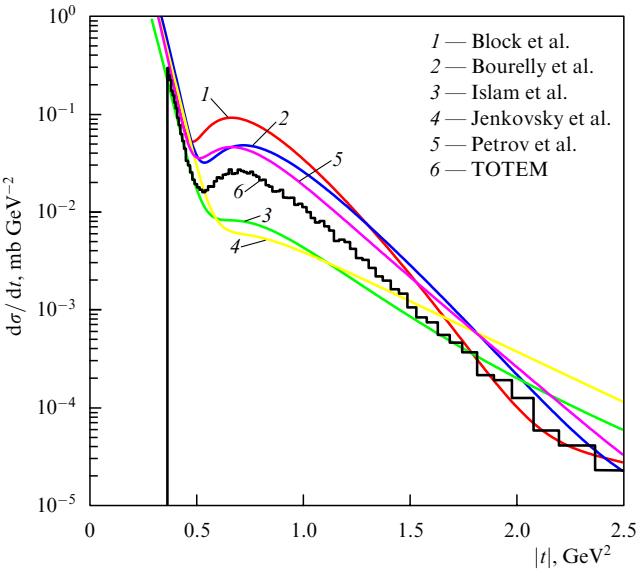


Figure 2. The differential cross section of elastic proton–proton scattering at $\sqrt{s} = 7$ TeV measured by the TOTEM collaboration (Fig. 4 in [48]). The region beyond the diffraction peak is shown. Predictions of five models are demonstrated.

Theoretical models usually describe the diffraction cone and values of the total and elastic cross sections related to it more or less precisely (therefore, their fits are almost indistinguishable in that region and are not drawn in Fig. 1). However, all of them fail to quantitatively predict the behavior of the differential cross section outside the diffraction cone, as can be seen in Fig. 2. The predictions of five models [79, 92, 110, 119, 120] are drawn here. They are very widely spread around the experimental line. We can conclude that just this region becomes the Occam razor for all models. In what follows, we consider these models, as well as some others, in more detail.

The three intervals of $|t|$ (the diffraction cone, the Orear regime, and the region of hard parton scattering) are

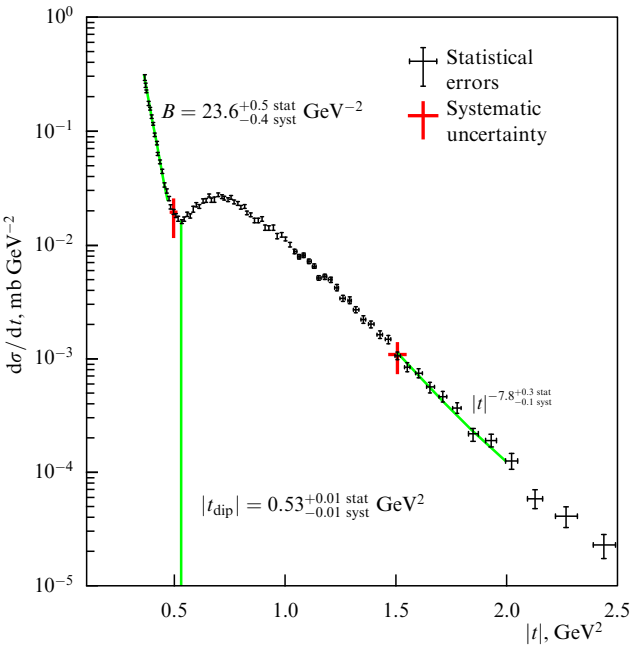


Figure 3. The differential cross section of elastic proton–proton scattering at $\sqrt{s} = 7$ TeV measured by the TOTEM collaboration (Fig. 3 in [48]). The same regions as in Fig. 2 are shown with the values of the steepened slope near the diffraction peak, the position of the dip, and the power-like behavior at the largest transferred momenta.

characterized by different dynamical content, as we understand it now. They require separate approaches to their descriptions. It seems reasonable that these regions are governed by different but interrelated physical mechanisms. In particular, different spatial regions of overlapping colliding objects are responsible for corresponding effects. Sections 4.1–4.3 are devoted to theoretical approaches to their explanation.

4.1 Diffraction cone and geometric approach

The internal structure of colliding, strongly interacting particles plays a crucial role in the outcome of their collisions. In high-energy hadron–hadron scattering, each hadron behaves as an extended object. They can be described by their size and the density of their constituents. The simplest models are demonstrated in Table 1.

Since long ago, it has been believed that hadrons contain some denser core surrounded by a meson (pion) cloud at their periphery. This idea was a cornerstone of the one-pion exchange model, which was first proposed in Ref. [121] to describe particle production in peripheral interactions. It evolved into the well-known multiperipheral and (multi)Reggeon exchange models (see, e.g., [122–124] for early review papers). They are rather successful in describing many features of multiparticle production processes. The multiperipheral approach developed, for instance, in the framework of the Bethe–Salpeter equation (see Ref. [124]) can be considered an attempt to account for the t-channel unitarity.

Nowadays, it is commonly believed that at very high energies, the total cross section is dominated by peripheral events. In modern parlance, this is related to the long-range nature of the field of 'perturbatively massless' gluons. The exchanged boson mass may mimic a nonperturbative mass gap in QCD with an 'effective' gluon mass of the order of 1 GeV and a gluon–gluon correlation length of about 0.3 fm. The pion mass scale is rather small, and more general 'boson' exchange is preferred. The weight factors of different mass scales take the impact parameter distribution of the particle opacity into account.

The role of inelastic channels in describing elastic scattering can be revealed by understanding the origin of and prescribing a definite shape to the overlap function $F(p, \theta)$ in the s-channel unitarity condition (18) or, equivalently, to its Fourier transform in the impact parameter picture. The scattering is mainly diffractive, i.e., it is due to the absorption of incoming waves in many open inelastic channels. Its quantitative field-theory treatment presents a serious unsolved problem.

The overlap function contains the sum of products of a matrix element of the inelastic process with a particular final state and the complex conjugate matrix element with the same final-particle content. However, their kinematical difference must be taken into account because the two final protons are scattered at an angle θ relative to the initial ones. Correspondingly, the overlap of the momentum distributions of the intermediate inelastic n-particle states is nontrivial kinematically and, what is especially important, the phases of these matrix elements become crucial. The phases are related to the position in space where particles are produced. It has been pointed out in many papers [125–128] that only the phase cancelation effect, which is closely related to particle correlations in inelastic processes, can lead to a realistic shape of the diffraction cone. The problem of properly accounting for them has not yet been solved.

At the same time, elastic scattering should be less peripheral because of a larger number of exchanged objects if regarded as an s-channel iteration of the overlap function. The great difficulty in transferring large momenta reveals itself already in the sharp shape of the forward diffraction peak. There have been numerous attempts to understand it in terms of the peripheral approach (see, e.g., [129–133]). Unfortunately, no framework for commonly accounting for both the s- and t-channel unitarity conditions has been developed.

In general, there have been many ideas proposed for describing elastic scattering processes, but no cogent theoretical arguments to justify the particular forms relying mainly on 'intuition' have been offered. The fact that they are very simple is usually the only advantage. Any strict interpretation is an idealization and as such should not be expected to be exactly true.

4.1.1 Geometry of the internal hadron structure. The key elements of the geometric approach are the use (a) of the impact parameter picture with Fourier–Bessel transformation (21), (22) from the transferred momenta amplitude to the spatial description, (b) of eikonal approximation (22), and (c) of unitarity condition (26). The S-matrix in the impact parameter picture is chosen in the exponential form

$$S(s, b) = \exp\left(-\Omega(s, b)\right), \tag{43}$$

and the convolution approximation for the real opacity Ω for elastic AB scattering is used:

$$\Omega(s, b) = K D_{\text{A}} \otimes D_{\text{B}}\,. \tag{44}$$

Here, $\otimes$ denotes the convolution of hadronic matter density distributions D for A and B, and K is an energy-dependent

factor. The assumptions about the validity of the eikonal approximation, the nearly imaginary character of the scattering amplitudes at low transferred momenta, the proportionality between the hadronic matter distribution and the electric charge distribution, the exponentiation of the S-matrix in b-space, and the validity of unitarity condition (26) are widely used.

The droplet model [134, 135] for elastic collisions was the first to fully exploit all the above elements. Particles were pictured as very much similar to nuclei. Correspondingly, the notion of the density distribution D inside a particle was introduced such that

$$\Omega(s,b) = \text{const} \int_{-\infty}^{+\infty} D\left((b^2 + x^2)^{1/2}\right) \mathrm{d}x\,. \tag{45}$$

In potential models, it corresponds to the WKB approximation. For the Gaussian shape of $h(s,b)$, it is possible to solve for D from (45), obtaining the function familiar in the theory of Bose–Einstein condensation of free particles [135]. In the droplet model, the disk properties are independent of the energy at sufficiently high energies. Many diffractive minima in the differential cross section have been predicted. The dipole form factors in the t-representation led to $\Omega(s,b)$ with a shape of the modified Bessel functions, which allowed fitting differential cross sections at ISR energies [136, 137]. The intuitive picture of high-energy hadron collisions as two extended objects breaking into fragments (and thus defining the overlap function!) has promoted the hypothesis of limiting fragmentation [138] inspired by the droplet model.

Models based on consideration of tower diagrams [50, 87] predict that the disk becomes larger and more absorptive as the energy increases. Both the black core and gray fringe expand with energy and become more absorptive.

The first estimates of the radii of protons, pions, and kaons from their form factors [139–141] showed that protons are larger than pions and kaons. This is not surprising, in view of the smaller cross sections of πp and Kp interactions than those of pp. The typical size is somewhat smaller than 1 fm. The proton hadronic matter distribution was fitted by a dipole form similar to the electric form factor but with the energy-dependent radius.

Other early attempts to consider the elastic scattering of hadrons also stemmed from the analogous simple geometric treatment of their internal structure [136, 137, 142–144]. Later, more complicated models were used. The main focus is, surely, on processes at small angles within the diffraction cone. They define the bulk contribution to the elastic scattering cross section due to the steep falloff of the distribution with increasing angles. Different models happen to fit the experimental data quite well in a wide energy range. But they fail outside the diffraction peak, as mentioned above. Large-angle scattering requires more central collisions with a lower impact parameter to probe the internal content of particles. Therefore, these regions of transferred momenta are discussed separately below.

Some ideas stemmed from regularities in inelastic processes. The multiplicity distributions of created particles are closely related to the purely geometric notion of the centrality of collisions. When the scaling of multiplicity distributions [146] was supported by experimental data, the proposal of the geometric scaling [145] for the elastic amplitude was promoted. The difficulties in accelerating the various parts of a nucleon without breaking it up had to be accounted for.

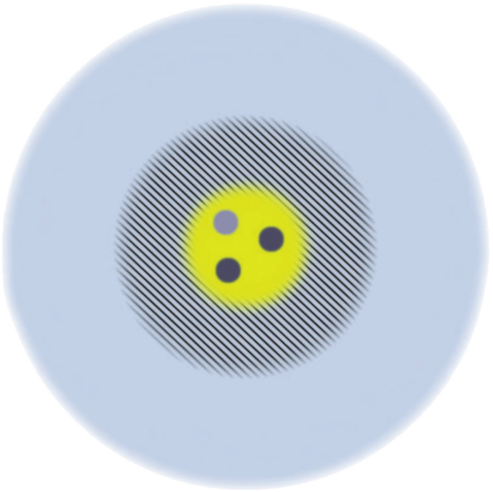


Figure 4. The nucleon structure according to the model in [63, 93, 94, 119]. The three regions of the internal structure are supposed to be directly responsible for the three regimes in the behavior of the differential cross section.

The basic idea of the geometric scaling is that at sufficiently high energies, the amplitude $A(s,t)$ depends on a single variable, the scaling parameter τ:

$$\tau = -\frac{t}{t_0} \ln^2 \frac{s}{s_0}\,. \tag{46}$$

This idea has led to several predictions at asymptotically high energies, and is still actively being debated. Such scaling was proved [147, 148] for cross sections increasing as $\ln^2(s/s_0)$ and for an infinitesimally small ratio of the real to imaginary part of the amplitude $\rho \to 0$ as $s \to \infty$. The latest results on $\rho(s,t)$ discussed in Section 4.2.5 do not support this assumption.

The purely geometric standpoint is adopted in Refs [63, 93, 94, 119]. The three regions in the behavior of the differential cross section are clearly reflected in the three spatial scales of the internal hadronic structure considered in [93, 94, 119,104]. The authors of this three-scale model claim that the nucleon has an outer cloud of the quark–antiquark condensate, an inner shell of the baryonic charge density, and a still smaller internal core of massless color-singlet valence 'quarks' surrounded by low-x gluon clouds about 0.3 fm in size. This picture is shown in Fig. 4.

The diffraction cone is described as a result of cloud–cloud interaction, represented by a class of potentials containing the sum of modified Bessel functions. The least massive exchanged quanta are the most important ones. At larger momentum transfers, the baryonic charge at intermediate distances is probed by the ω-exchange. The internal region filled in by the valence quarks starts playing its role in the presence of even larger transferred momenta.

The diffraction profile function, which defines the range of different densities and, correspondingly, different forces, is taken to be

$$\Gamma(s,b) = 1 - \Omega(s,b) = g(s)\left[\frac{1}{1+\exp\left[(b-r)/a\right]} + \frac{1}{1+\exp\left[(-b+r)/a\right]} - 1\right]. \tag{47}$$

The parameters r and a are energy dependent,

$$r = r_0 + r_1\left(\ln\left(\frac{s}{s_0}\right) - \frac{\mathrm{i}\pi}{2}\right), \qquad a = a_0 + a_1\left(\ln\left(\frac{s}{s_0}\right) - \frac{\mathrm{i}\pi}{2}\right), \tag{48}$$

and $g(s)$ is a coupling strength; $s_0 = 1\ \mathrm{GeV}^2$.

These functions render the shape of the differential cross section, similar to the Fraunhofer diffraction (see Section 4.1.2), with the form factor proportional to $\pi dq/\sinh(\pi dq)$ ($q^2 = -t$, d is an adjustable parameter) proposed a long time ago [148–153]. This form factor also extends somewhat to transferred momenta outside the diffraction cone. Unfortunately, the contemporary phenomenological analysis of experimental data is not able to determine the impact parameter profiles unambiguously.

The scattering due to ω-exchange is parameterized by the product of the ω-propagator and two form factors F directly in the (s,t)-representation:

$$A_\omega(s,t) \propto s\exp\left(\mathrm{i}\chi(s,0)\right)\frac{F^2(t)}{m_\omega^2 - t}\,. \tag{49}$$

The amplitude due to quark–quark scattering has two 'structure factors' G of valence quarks (different from the above form factors): the propagator with the black disc radius r_B of qq asymptotic scattering and s-dependent factors with the hard Pomeron intercept equal to $1 + \alpha_\mathrm{h}$:

$$A_\mathrm{qq}(s,t) \propto \mathrm{i}s\exp\left(\mathrm{i}\chi(s,0)\right)\left[s\exp\left(-\frac{\mathrm{i}\pi}{2}\right)\right]^{\alpha_\mathrm{h}}\frac{G^2(t)}{r_\mathrm{b}^{-2} + |t|}\,. \tag{50}$$

In total, there are seventeen adjustable parameters in the model.

As mentioned in [49], the fits according to this model predict too low a value of the slope B at $|t| = 0.4\ \mathrm{GeV}^2$ and strongly disagree with experiment at 7 TeV outside the diffraction peak (see Fig. 2). Formulas (49) and (50) are aimed to improve the fit just in this region, but they do not help.

In general, an internal region of the nucleon where the gluons cluster around the original valence quarks resembles the valon model [154, 155]. Similar pictures arise in the QCD-inspired models discussed below.

Surely, some care should be taken for any such model to be accepted and the geometric picture to be considered seriously, especially in view of its success or failure to describe experimental data in the whole range of transferred momenta at various energies.

4.1.2 The modified Fraunhofer diffraction. For a long time (see, e.g., Ref. [156]), the formulas of classical diffraction of light on a (black or grey) disk with the traditional Bessel functions have been used for hadronic reactions. Recently, an analogous expression for the elastic amplitude was considered in Ref. [157]:

$$A\left(s, |t| = q^2\right) = C\,\frac{dq}{\sinh(\pi dq)}\left(\mathrm{i}\,\frac{J_1(Rq)}{Rq} + \frac{\rho}{2}\,J_0(Rq)\right). \tag{51}$$

The free parameters in arbitrarily chosen analytic expression (51) are C, R, d, and ρ. The first term resembles the expression for the black disk, Eqn (36). The suppression at large transferred momenta is assumed to be approximated by the form factor in front of the Bessel functions. In the impact parameter representation, this shape corresponds to the ordinary Fermi profile used, e.g., in Refs [93, 94, 119, 158] and shown in Eqn (47):

$$h(b) \propto \frac{1}{1 + \exp\left[(b - R)/d\right]}\,. \tag{52}$$

The second term in brackets in (51) takes the contribution due to the real part of the amplitude into account. It should smooth the behavior of the differential cross section near zeros of the first term. This seems to be the only difference from the first component of the previously discussed model [93, 94, 119].

And, again, comparison with experimental data shows that the results of fits are satisfactory in the diffraction cone, but not outside it. The form factor in front of common Bessel functions does not fit the large-$|t|$ trends of experimental distributions.

Throughout these developments, modifications of early guesses have been found necessary, but the general spirit of the geometric description remains immutable and viable.

4.1.3 Electromagnetic analogies. The strongly interacting content of hadrons is often considered to be similar to their electromagnetic substructure [136, 137]. Similarly to the droplet model, the assumption of the proportionality between the hadronic matter distribution and the electric charge distribution is used in many models. However, in most of them, the electromagnetic form factors are used in combination with Reggeon exchanges because, considered alone, they do not reproduce the energy dependence of the main characteristics. But the assumption about the full congruence of these distributions is not necessarily valid, since gluons do not carry an electric charge even though they play an important (if not decisive) role in high-energy strong interactions. That is why the charge and matter distributions in some models are parameterized separately or some corrections are added.

Using the experience from calculations of tower diagrams in electrodynamics and the impact-parameter representation, it was proposed [92, 159–161] that the possibility of choosing the opacity $\Omega(s,\mathbf{b})$ in a factored form be considered:

$$\Omega(s,\mathbf{b}) = R(s)\,F(\mathbf{b}^2) + (\text{nonleading terms})\,, \tag{53}$$

where $R(s)$ is chosen to be crossing symmetric under $s \leftrightarrow u$ and to reproduce the energy dependence of the Pomeron, considered as a fixed Regge cut,

$$R(s) = \frac{s^c}{(\ln s)^{c'}} + \frac{u^c}{(\ln u)^{c'}}\,, \tag{54}$$

while $F(\mathbf{b}^2)$ is taken as the Bessel transform of

$$F(t) = f\left|G(t)\right|^2\frac{a^2 + t}{a^2 - t}\,. \tag{55}$$

Here, $G(t)$ stands for the proton 'nuclear form factor', parameterized like the electromagnetic form factor with two poles:

$$G(t) = \frac{1}{(1 - t/m_1^2)(1 - t/m_2^2)}\,. \tag{56}$$

Other factors with the parameter a are introduced 'by hand'. They can be treated just as a correction due to the different shapes of distributions of charge and matter. There are six adjustable parameters in total used at high energies if the Regge background is neglected. The noticeable t-dependence of the slope $B(t)$ in the diffraction cone is predicted. However, its values at 7 TeV are lower than experimental ones (about 18 GeV^{-2} instead of 20.1 GeV^{-2}) at $10^{-2} < |t| < 0.3$ GeV^2, slightly exceed them in the tiny interval near 0.35 GeV^2, and do not reach the value 23.6 GeV^{-2} mentioned above.

This model is close to the TOTEM data [48, 49] for the dip position and the exponential at very large $|t|$, but predicts values of the differential cross section in the Orear range about twice as large, $|t| \geqslant 0.36$ GeV^2 (see Fig. 2). In addition to the dip, some 'oscillations' at the transferred momenta of several GeV^2 are predicted (up to the energy of 6000 TeV) but not yet observed. In general, such structures appear as a byproduct of the eikonal approach and unitarization procedure (see, e.g., Ref. [95]). Their energy dependence is strongly determined by the parameters used in formula (54) to account for the crossing symmetry property of the amplitude.

The same parameters are crucial for the behavior of the real part of the amplitude. It is interesting that the model predicts the dominance of the imaginary part of the amplitude even at large transferred momenta. The real part becomes important only at zeros of the imaginary part. The dip and oscillations are noticeable precisely there. Near the cone, the model predicts two zeros of the real part of the full (Coulomb + nuclear) amplitude at $|t| = 0.0064$ GeV^2 and the nuclear amplitude alone at $|t| \geqslant 0.18$ GeV^2, as well as one zero of the nuclear imaginary part at $|t| = 0.5$ GeV^2. In the differential cross section, the last zero is partly compensated by the real part.

We note the difference between the power-like expression for $F(t)$ and its exponential behavior in the traditional Regge models. The exponentiation of this form of $F(t)$ leads to additional oscillations.

A similar but more complicated combination of the form factors has been used in Refs [73, 74, 162–165]. The authors consider t-dependent Mellin transforms of parton distributions and claim that the first moment G defines the form factor of the standard Pomeron, while the second moment H corresponds to interaction attributed to three nonperturbative gluons. Thus, the behavior of the differential cross section is determined by the electromagnetic form factors at small t and by the matter distribution at large t. The Born term of the elastic scattering amplitude is written as

$$A^{\text{Born}}(s,t) = h_1 G^2(t) F_a(s,t)\left(1 + \frac{r_1}{\hat{s}^{0.5}}\right) + h_2 H^2(t) F_b(s,t)\left(1 + \frac{r_2}{\hat{s}^{0.5}}\right), \qquad (57)$$

where

$$F_a(s,t) = \hat{s}^{\epsilon_1} \exp\left(B(s)t\right), \qquad F_b(s,t) = \hat{s}^{\epsilon_1} \exp\frac{B(s)t}{4}, \qquad (58)$$

$$G(t) = \frac{L_1^4}{(L_1^2 - t)^2}\,\frac{4m_\text{p}^2 - 2.793t}{4m_\text{p}^2 - t}, \qquad H(t) = \frac{L_2^4}{(L_2^2 - t)^2}, \qquad (59)$$

$L_1^2 = 0.71$ GeV^2, $L_2^2 = 2$ GeV^2, $\hat{s} = s\exp(-\text{i}\pi/2)/s_0$, $s_0 = 1$ GeV^2, $B(s) = \alpha' \ln(s/s_0)$, and $\alpha' = 0.24$ GeV^{-2}. We note that the slope of the second term is chosen as one fourth that of the first term. The final form of the amplitude is obtained after eikonalization of the Born contribution using the opacity

$$\Omega(s,b) = \frac{1}{2\pi}\int \text{d}^2 q\, \exp(\text{i}\mathbf{qb})\, A^{\text{Born}} \quad (s, q^2 = -t). \qquad (60)$$

The total cross section at 7 TeV was predicted to be equal to 95 mb. Authors demonstrate good fits of pp and $\text{p}\bar{\text{p}}$ differential cross sections, as well as of $\rho_0(s)$, in a wide energy range, including the TOTEM data. Only five (three for high energies and two for low energies) adjustable parameters are claimed to be used if all the above values are regarded as fixed. In fact, there are 10 such additional 'hidden' parameters in total if the hard Pomeron is also considered. Surely, the contribution from secondary Reggeons at LHC energies is negligible, i.e., smaller than the experimental errors.

The real part of the hadron amplitude is completely determined by the complex expression for $\hat{s}$. Its t-dependence appears just as a byproduct of the eikonalization procedure. As a function of t, it tends to zero at $|t| \approx 0.16$ GeV^2 at the energy of 7 TeV. The interesting predictions of the t-behavior of $\rho(s,t)$ at nonforward transferred momenta for different energies are presented. They are discussed in more detail in Section 4.2.4.

4.1.4 Reggeon exchanges. The Regge-pole model is beyond dispute one of the most explored. It will have already been noticed that the notion of Regge trajectories has been used in the preceding subsections as well. The only reason to discuss these models separately there was their stronger inclination to the use of nonexponential electromagnetic form factors and geometric pictures in the s-channel approach. At the same time, Reggeon models appeal mostly to the t-channel approach.

The amplidutes with Reggeon exchanges in the t-channel are natural candidates for explaining the exponential decrease in differential cross section (6) with the squared transferred momentum $|t|$ inside the diffraction cone. Just this shape is typical for them because it follows from the linearity of Regge trajectories. Moreover, they predict a logarithmic increase in the hadronic radii as the energy increases, i.e., a logarithmic increase in the cone slope B or the corresponding shrinkage of the width of the diffraction cone. This prediction is also supported by experiment. In the common Regge-pole models, the disk becomes larger and slightly more transparent as energy increases.

The standard Regge-type models [73, 74, 97, 166, 167] use the combination of contributions due to the exchange by the (multicomponent) Pomeron, Odderon, and secondary Reggeon trajectories corresponding to f and ω mesons with or without the form factors chosen in a simple exponential form or as power-like expressions resembling the electromagnetic structure of colliding partners discussed in the preceding subsection. The price to be paid is the increased number of adjustable parameters at each step of sophistication. To be more or less realistic, one has to use the knowledge about some of them from other (independent?) experimental results. But even under this condition, the ambiguity of their choice and sensitivity to fitted parameters leave some freedom in the conclusions.

The amplitudes of pp and $\text{p}\bar{\text{p}}$ scattering are approximated by the sum of terms corresponding to the leading (Pomeron and Odderon) and nonleading (f and ω meson) Regge

trajectories:

$$A(s,t)_{\mathrm{p\bar{p}}}^{\mathrm{pp}} = A_{\mathrm{P}}(s,t) + A_{\mathrm{f}}(s,t) \mp \left(A_{\omega} + A_{\mathrm{O}}(s,t)\right), \qquad (61)$$

where the labels P, f, O, and ω stand for the relevant contributions. The sign in the pp and p$\bar{\mathrm{p}}$ amplitudes differs for C-even and C-odd terms.

The contributions of the nonleading Regge poles are written as

$$A_{\mathrm{R}}(s,t) = a_{\mathrm{R}} \exp\left(-\frac{\mathrm{i}\pi\alpha_{\mathrm{R}}(t)}{2}\right) \exp\left(b_{\mathrm{R}} t\right) \left(\frac{s}{s_0}\right)^{\alpha_{\mathrm{R}}(t)}, \qquad (62)$$

with $\alpha_{\mathrm{R}}(t) = a_{\mathrm{R}} + b_{\mathrm{R}} t$.

While the secondary trajectories are usually chosen in a standard linear way, the Pomeron and Odderon contributions can be regarded, for example, as dipoles with nonlinear trajectories [166, 168–171]:

$$A_{\mathrm{P}}(s,t) = \mathrm{i}\,\frac{a_{\mathrm{P}} s}{b_{\mathrm{P}} s_0} \left\{ r_1^2(s) \exp\left[r_1^2(s)(\alpha_{\mathrm{P}} - 1)\right] - \epsilon_{\mathrm{P}} r_2^2(s) \exp\left[r_2^2(s)(\alpha_{\mathrm{P}} - 1)\right] \right\}, \qquad (63)$$

where $r_1^2(s) = b_{\mathrm{P}} + L - \mathrm{i}\pi/2$ and $r_2^2(s) = L - \mathrm{i}\pi/2$ with $L = \ln(s/s_0)$. The unknown Odderon contribution is assumed to be of the same form as that of the Pomeron. The parameters of the trajectories and of the absorption ϵ_P have to be adjusted. Their nonlinearity may be connected with the two-pion threshold following from the t-channel unitarity [168, 170, 172]. However, there could be double counting of the graphs with Pomerons attached on both sides to the pion loop. This is well known from old peripheral models of inelastic processes, where the self-consistent Bethe–Salpeter equation had to be used for the proper account of the pion–nucleon vertices. Different forms of nonlinear trajectories are in use. For instance, the Pomeron trajectory is chosen in [173] with four free parameters as

$$\alpha(t) = \alpha_0 - \gamma \ln\left(1 + \beta\sqrt{t_0 - t}\right). \qquad (64)$$

A more complicated nonlinearity was used in [172]. However, the use of the pion mass as a scale there is questionable in view of the above discussion.

The origin of the Pomeron and the parameterization of its trajectory are still being debated. There is no strict rule for choosing its shape. The dipole and even tripole forms of unitarized Pomerons have been attempted. They mimic cut contributions [174–177].

Moreover, there are arguments in favor of two Pomerons with different intercepts. Even the fits with three Pomerons are sometimes used [79, 97]. The soft Pomeron contributes a term with the energy dependence $s^{a_{\mathrm{s}}}$ ($a_{\mathrm{s}} \approx 0.08$) to hadron–hadron total cross sections, and the hard Pomeron makes a small contribution (at present energies) with a stronger energy dependence $s^{a_{\mathrm{h}}}$ ($a_{\mathrm{h}} \approx 0.4$). These values of the intercepts stem from the discussions of HERA data (see, e.g., [178]). Although the hard-Pomeron exchange was unnecessary for describing hadron–hadron total cross sections up to the energies $\sqrt{s}$ below 1 TeV, it may reveal itself at LHC energies, as argued in [179]. The model in [179] uses only two terms in the expansion for opacity:

$$\Omega(s,b) = \Omega_{\mathrm{s}}(s,b) - \frac{\lambda}{2}\,\Omega_{\mathrm{s}}^2(s,b), \qquad (65)$$

where Ω_{s} stands for the contribution from single exchanges of Reggeons (two Pomerons, f, and ω) with the adjustable parameter λ, as well as for the triple-gluon exchange of the form Cst^{-4} needed at larger values of $|t|$ and matched at some $t = t_0$ to exponential shapes of the diffraction peak and to the dip region. Certainly, adding such a term allows fitting the total cross section value at 7 TeV, but there is a suspicion that the sharp increase in the hard-Pomeron contribution will overpredict the cross sections at higher energies. The unitarization will become mandatory once again. The quality of the fit of the differential cross section beyond the diffraction peak is no better than of the fits shown in Fig. 2.

Several variant forms of Born amplitudes and different kinds of eikonalization have been attempted. There is no consensus on their choice.

A form of the eikonal similar to (65) is chosen in [180] with the exponential suppression

$$\Omega_{\mathrm{s}}(s,b) = A_0 \exp\left[-m(s)(r_0^2 + b^2)^{1/2}\right] \qquad (66)$$

for central interactions. The peripheral part of the Pomeron interaction with the meson cloud is parameterized [180] by a small term increasing with the energy and resulting in a $\sqrt{|t|}$ exponential fall-off of the differential cross section. The geometric picture corresponds to a black disk with a gray fringe, similarly to the above-described model [92].

In general, it is not easy to estimate the total number of adjustable parameters in different models. There are parameters related to either t- (b-) or s-dependence. In some papers, it is often assumed that part of them are known from fits of other characteristics of hadron or electromagnetic interactions at various energies and can therefore be considered known beforehand. For example, it is claimed that the model in [166] contains about 15 parameters. In this case, it is quite difficult to find the proper minima for the matrix of χ^2 values. It is well known how unstable the final results can be: one has to choose a step-by-step procedure for doing this and use some special computer codes.

There are 25 adjustable parameters shown in Table 1 in papers [79, 97]. However they include some values assumed to be *a priori* fixed in [166]. At the same time, additional form factors were inserted in the formulas, albeit with preliminarily 'fixed' parameters. They were used to fit 982 pp and p$\bar{\mathrm{p}}$ data points in a wide energy range. Besides the elastic differential cross sections, the total cross sections and the ratios ρ_0 were considered. The fit in the interference region of Coulomb and hadronic amplitudes with the same parameters helped in choosing among the different Coulomb phases proposed previously.

A similar situation is seen in the fits in Ref. [73, 74], where it is claimed that the number of parameters is much less (only 5!). However, there are many other, hidden parameters (in particular, those concerning the energy behavior and form factors). They are held fixed from the very beginning, as was discussed in Section 4.1.3.

As mentioned above, the simple exponential form of the differential cross section in the diffraction cone is quite well described. This becomes possible mainly due to the t-shape of the Pomeron trajectory in (63) and other Reggeons contributions in (62). The fits in this region at the different energies, shown, e.g., in [73, 74, 97, 166], are quite impressive. The evolution of the diffraction cone slope with energy is reproduced [as described by L in (63)]. Unfortunately, the variety of forms of Pomeron trajectories with different

intercepts, slopes, and shapes of residues unitarized in different ways and/or substituted by Regge cuts is so large that it is impossible to show all of them in this review due to the limited space.

The cuts with nonlinear trajectories mimic hard scattering [98]. A common problem appears in predicting them at larger angles. The fit according to the model in [79, 97] seems to be most successful in predicting the position of the dip and the shape at large $|t|$, but exceeds the absolute value approximately twofold. The model in [166] strongly underestimates it, with the wrong position of the dip and much slower decrease at $|t| > 1.5$ GeV^2. This is well demonstrated in Fig. 2 and is also discussed below.

We mention that all these papers follow the general approach proposed much earlier [170, 181–183]. They just deal with more detailed fits of newly available experimental data.

4.1.5 QCD-inspired models. Each incident particle consists of a superposition of Fock states with n partons [184], which are scattered instantaneously and simultaneously by the other particle. Some QCD-inspired models using this statement have been developed. The role of partons is played by quarks and gluons.

The two competing mechanisms of hadron interactions, the increase in the density (α in Table 1) and in the radius R, determine their specific features. In QCD, they can be respectively ascribed to the leading-order solution of the BFKL equation [185, 186] and to the long-range (Weizsäcker–Williams) nature of the field of massless gluons. The density increase due to the BFKL evolution leads to a power-law increase in the total cross section, which is nonunitary and violates Froissart bound (31). Therefore, at the critical density of the order $1/\alpha_s$, the density saturation must be taken into account [187]. The QCD evolution in all orders in the gluon density but in the leading logarithmic approximations is treated by the JIMWLK equations [188]. With account of multiple scattering effects, they can be simplified in the large-N_c limit to a single nonlinear BK equation for the gluon density [189, 190] when the induced field density is small.

The density growth effects are preasymptotic. According to [100], they are described by a hard Pomeron, while the growth of the size of the black saturated regions (the radius) is attributed to a soft Pomeron. The hard Pomeron manifests itself in small systems or in small subregions inside hadrons. The soft Pomeron appears in hadronic systems of a typical size and is related to an increased size in the impact parameter space. Only the increase due to the perturbative expansion in the transverse plane remains effective.

There is no consensus about this scenario proposed in Ref. [100]. The soft Pomeron is often used [80, 81, 191, 192] in attempts to explain the preasymptotic power-law growth of cross sections by an additional nonperturbative mechanism superimposed on the BFKL scenario of a hard Pomeron. It is ascribed mainly to the density growth of gluon clouds around quarks and not to the spatial scale of the interaction. Even though the size of gluon clouds increases, it is still limited by a short separation from their source. The proton looks like three valence quarks surrounded by gluon clouds or spots with mean sizes about 0.3 fm smaller than the proton radius, of the order of 1 fm. Radiation from any additional gluon in the cloud adds the factor $\ln(s/s_0)$ to the interaction cross section, and hence their sum gives a power-like term of the form

$$\sigma_t = \sigma_0 + \sigma_\Delta \left(\frac{s}{s_0}\right)^{\Delta}, \qquad \Delta = \frac{4\alpha_s}{3\pi} \approx 0.17, \tag{67}$$

with a large constant term σ_0 and small σ_Δ. Using the standard dipole form factors of protons and quasi-eikonal unitarization in the impact parameter space, the authors of this two-scale model [80, 81, 192] fit many distributions with 10 parameters for the t-dependence (subject to two additional constraints) and some parameters for the s-dependence. Such fits are, of course, aimed at high energies of colliding protons where the effects of secondary Regge trajectories die out. They are mainly successful in the diffraction cone and, consequently, in describing the energy dependence of the total and elastic cross sections.

Such a form of the total cross section with an energy-independent term σ_0 was proposed a long time ago [193–195] and actively developed later [196–198] in the framework of the parton model and semihard QCD, with the gluon–gluon interaction playing the main role.

The main role of gluons is also incorporated in [99, 110]. The profile is chosen in a form containing the gg, qq, and qg terms:

$$\Omega = \sigma_{gg} W(b;\mu_{gg}) + \Sigma_{gg}\left(C + C_R \frac{m_0}{s^{1/2}}\right) W(b;\mu_{qq}) + \Sigma_{gg} C_{qg} \ln\frac{s}{s_0}\, W\left(b;(\mu_{qq}\mu_{gg})^{1/2}\right), \tag{68}$$

where the impact parameter distribution functions are

$$W(b;\mu) = \frac{\mu^2(\mu b)^3 K_3(\mu b)}{96\pi}, \tag{69}$$

and the gluon–gluon interaction cross section is

$$\sigma_{gg} = C_{gg} \int \Sigma_{gg} \Theta(\tau s - m_0^2)\, F_{gg}(\tau)\, d\tau, \tag{70}$$

with $\Sigma_{gg} = 9\pi\alpha_s^2/m_0^2$, $F_{gg} = \int f_g(x_1) f_g(x_2) \delta(\tau - x_1 x_2)\, dx_1\, dx_2$, and $f_g(x) = N_g(1-x)^5/x^{1+\epsilon}$.

The Froissart bound for the total cross section is reproduced with

$$\sigma_t = 2\pi\left(\frac{\epsilon}{\mu_{gg}}\right)^2 \ln^2\frac{s}{s_0}. \tag{71}$$

The parameter μ_{gg} describes the area occupied by gluons in the colliding protons (the size effect), and ϵ is defined via their gluonic structure functions and therefore controls their soft gluon content (the density effect).

Again, being successful in the diffraction peak with its shape and normalization, the model in [99, 110] fails to predict the correct behavior of the differential spectrum outside it [48, 49]. Its prediction is more than three times larger than the experimental value at the dip and the subsequent maximum, while falling too steeply at ever higher $|t|$ above 1.5 GeV^2 (see Fig. 2).

Attempts to consider the semihard scattering of quarks and gluons can be found in Refs [197–201].

The traditional partonic description of the process is considered in a series of papers [172, 202–204]. The partonic approach with a hard BFKL Pomeron can be merged into the domain governed by the soft Pomeron. The transition from hard to soft is induced by absorptive multi-Pomeron effects in a limited energy range. The evolution produces parton

cascades, not strongly ordered in transverse momenta, with hot spots of a relatively small size in b-space. The saturation is driven by the enhanced multi-Pomeron graphs, also regulating the high-mass dissociation. The calculations are done with a 3-channel quasi-eikonal unitarization using the opacity formalism. They reproduce the shapes of the differential cross sections from ISR to LHC energies within the diffraction cone.

Another picture was considered in the framework of the functional integral approach in Refs [205–208] using the model of the stochastic vacuum and making the assumption that the proton has the quark–diquark structure of a color dipole, i.e., two quarks out of three are close together in transverse directions. A matrix cumulant expansion is used for vacuum expectation values of Wegner–Wilson loops [205] related to hadronic amplitudes. The QCD vacuum parameters (the gluon condensate or the string tension, the vacuum correlation length, and the parameter due to the non-Abelian tensor structure), as well as the hadron size, have been used. The imaginary part of the amplitude in the b-representation was calculated. Its contribution to experimentally measured quantities was shown to describe the ISR and Tevatron data in the diffraction peak reasonably well.

A more phenomenological approach to the quark–diquark model was attempted in Refs [209, 210]. As above, the correlated quark and diquark constituents are considered. According to the detailed analysis performed in [211], from ISR to LHC energies in the range $0.36 < |t| < 2.5$ GeV^2, the model is able to describe the data quite well, even outside the diffraction peak, except the narrow strip around the dip. But it shows a much stronger dip (by several orders of magnitude) there than the experimentally observed one. Moreover, similarly to the abovementioned calculations, the model ignores the contributions to the real part of the elastic scattering amplitude. As we saw previously, such contributions can smooth this dip. If so, their shape should drastically differ from that of the imaginary part, at least in this strip, as happened, for example, in the models with electromagnetic form factors [92, 165].

4.2 Intermediate angles: the dip and the Orear regime

As long ago as the 1960s, experiments on elastic pp and πp scattering at comparatively low energies between 6.8 and 19.2 GeV in laboratory system [2, 9–11] showed that the steep exponential fall-off of the differential cross section as a function of the squared transferred momentum $|t|$ is replaced by a slower dependence at larger $|t|$. They showed that just after the diffraction cone a shoulder was observed and, even more surprising, a behavior exponentially decreasing with the angle or with $\sqrt{|t|}$, which was called the Orear regime after its investigator [5, 6, 9]. A special session was devoted to these findings at the 1968 Rochester conference in Vienna. The shoulder evolved later into a minimum or dip at higher ISR energies. It has also been observed at the LHC, as seen in Figs 2 and 3.

It is interesting that at the FNAL–ISR energies $\sqrt{s} = 6$–60 GeV, the exponential fall-off with an increase in $\sqrt{|t|} \approx p_t$ was observed up to quite large values $|t| \approx 10$ GeV^2 [20, 28, 45], with the exponent in the range from 6.2 to 7 GeV^{-1} (see Table 7 in Ref. [45]). It is even larger at the LHC (about 8–9 GeV^{-1}). The region becomes narrower and shifts to lower values of $|t|$ from 0.5 to 1.5 GeV^2. The power-like regime already shows up at about $|t| \approx 2-2.5$ GeV^2 (see Fig. 3).

4.2.1 Gaussian fits. From the very beginning, it was noticed [5, 6] that it is possible to fit the differential cross sections at intermediate values of the momentum transfer by an exponential dependence on $\sqrt{|t|}$ (or θ) except the relatively small shoulder region. To take that into account as well, it was primarily proposed [3] to use fits with Gaussian functions with alternating signs of the coefficients directly in the expression for the amplitude. In this way, both the diffraction peak and larger-$|t|$ behavior could be described. No reference to any phenomenological model is given. From the geometrical standpoint, one can imagine an internal structure with envelopes of alternating density.

Such an empirical approach has been used [212–214] for fits of experimental data at ISR energies. The following parameterization of the amplitude is proposed in Ref. [214]:

$$A(s,t) = s\left[\left(\rho\sigma_t - \sum_{i=2}^{m} 4\pi a_i\right)\exp(b_1 t) + \sum_{i=2}^{m} 4\pi a_i \exp(b_i t) + i\left(\sigma_t - \sum_{j=2}^{n} 4\pi c_j\right)\exp(d_1 t) + i\sum_{j=2}^{n} 4\pi c_j \exp(d_j t)\right], \quad (72)$$

where $m < n$. The fits at different energies give information about the ratio of the real to imaginary part of the amplitude $\rho(t)$, besides the values of adjustable parameters a_i, b_i, c_j, and d_j. Two different methods were used. In total, there are 14 to 16 free parameters. The results of nonlinear fits are rather unstable, and the conclusions are somewhat controversial. In particular, the numbers of zeros of $\mathrm{Im}\, A(s,t)$ and $\mathrm{Re}\, A(s,t)$ differ in these methods. The dominance of the real part of the amplitude at intermediate values of the momentum transfer in one of the methods is not confirmed when the other method is used.

A similar fit was recently attempted and applied to the TOTEM data in Refs [111, 215]. The earlier proposal in Ref. [216] with phenomenologically chosen two t-exponentials and the relative interference phase responsible for the dip was applied to the TOTEM data. Using five parameters, it is possible to describe these data in the whole range of transferred momenta. We note that similarly to the model in [165], the slope of the second exponential term is chosen several times smaller than that of the main term. Moreover, when the electromagnetic form factors were tried in place of simple exponentials, the fit became worse.

Two exponentials without the interference term inside the diffraction peak and a Tsallis-type distribution outside it were used in [217]. It was possible, with the help of nine free parameters, to fit the data at energies from 19.4 GeV to 7 TeV.

In some way, this fit business with no reference to any theoretical model looks more like art than science, especially if no conclusions about the hadron structure are obtained. Such an approach will hardly be conclusive in the future.

4.2.2 Phenomenological models. Theoretical indications of the possibility of a new regime with an increase in transferred momentum were obtained even earlier [56, 122, 218]. It was treated as a consequence of the simple iteration of processes approximated by a Gaussian within the diffraction cone. The term I_2 in unitarity condition (18) with Gaussians inserted into the integrand gives rise to a Gaussian with a width that is twice as big, i.e., to a shape twice as wide as the diffraction

cone. Further iterations lead to further widening. Therefore, multiple exchanges were considered. However, the results did not fit new experimental findings. This failure was explained as resulting from the improper treatment of the unitarity requirements and incorrect choice of the overlap function.

The droplet-model relations between form factors and the elastic amplitude for hadronic scattering at infinite energy (see Eqns (1) and (2) in Ref. [219]) predict a series of kinks (or zeros) in the differential cross section, which could be related to dips. Dip position movement to lower $|t|$ with a growth of the total cross section was predicted in Ref. [220]. There is also an indication of several dips (or shoulders) at larger $|t|$ in the models in [92, 165, 197, 198] using the electromagnetic form factors with subsequent eikonalization (cf. Figs 2 and 11).

In accordance with the experimental data shown in Fig. 2, only one dip is predicted by other models. For example, it was described in Ref. [221] on the basis of a modified optical model [220]. In the framework of the geometric scaling approach [222], the numerical integration of the relation

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t}(s,t)\left(\frac{\mathrm{d}\sigma}{\mathrm{d}t}(s,0)\right)^{-1}=\frac{1}{1+\rho_0^2}\left[\phi^2(\tau)+\rho_0^2\left(\frac{\mathrm{d}\bigl(\tau\phi(\tau)\bigr)}{\mathrm{d}\tau}\right)^2\right], \tag{73}$$

where

$$\operatorname{Im}A(s,t)=s\sigma_{\mathrm{t}}\phi(\tau)\,,\qquad \phi(0)=1\,, \tag{74}$$

was performed with τ defined by Eqn (46).

It was predicted that the dip should even disappear at energies higher than $\sqrt{s}\approx 300$ GeV, but can reappear at ever higher energies. As we know now, it is clearly seen at 7 TeV. The imaginary part has been chosen in such a way that it has a zero at the dip. The absence of additional dips is explained as a deviation of the eikonal from a simple Gaussian with some flattening at small impact parameters (see Section 4.2.3). That shows strong sensitivity to the choice of tiny details of the phenomenological eikonal and also agrees with the properties of the overlap function to be discussed in more detail below. These results were confirmed and extended to $\mathrm{p\bar{p}}$ collisions in Ref. [223].

Processes described by diagrams with multiple exchanges by Pomerons are claimed to be responsible for the Orear regime at intermediate angles according to Refs [224, 225]. The differential cross section is predicted to have the form

$$\frac{1}{C}\frac{\mathrm{d}\sigma}{\mathrm{d}t}=\exp\left[-2\sqrt{2\pi\alpha'(0)|t|\xi\cot\frac{\phi}{2}}\,\varphi_1(\xi)\right]\times\left[1+\lambda\cos\left(2\sqrt{2\pi\alpha'(0)|t|\xi\tan\frac{\phi}{2}}+\varphi_0\right)\right], \tag{75}$$

where $\xi=\ln(s/4m^2)$, and C, ϕ, φ_0, φ_1, λ, and $\alpha'(0)$ are adjustable parameters. There are oscillations directly imposed on the exponential fall-off with the same exponential. They should be well pronounced. So far, no such oscillations have been observed.

A less strong statement about some saturation of the diffraction cone due to multiple Pomeron exchanges is made in Refs [226, 227].

4.2.3 Unitarity condition. A theoretical explanation based on the consequences of the unitarity condition in the s-channel has been proposed in Refs [57, 58]. The careful fit to experimental data showed good quantitative agreement with experiment [59]. Nowadays, the same approach helps explain the TOTEM findings [60] (see Fig. 5 below).

We consider the left-hand side and the integral term I_2 in unitarity condition (18) at the angles θ outside the diffraction peak. Because of the sharp fall-off of the amplitude with the angle, the leading contribution to the integral arises from a narrow region around the line $\theta_1+\theta_2\approx\theta$. Therefore, one of the amplitudes should be inserted at small angles within the cone as a Gaussian, while the other is kept at angles outside it. Integrating over one of the angles yields the linear integral equation

$$\operatorname{Im}A(p,\theta)=\frac{p\sigma_{\mathrm{t}}}{4\pi\sqrt{2\pi B}}\int_{-\infty}^{+\infty}\mathrm{d}\theta_1\,\exp\left[-\frac{Bp^2(\theta-\theta_1)^2}{2}\right]\times f_\rho\operatorname{Im}A(p,\theta_1)+F(p,\theta)\,, \tag{76}$$

where $f_\rho=1+\rho_0\,\rho(\theta_1)$. It can be solved analytically (see [57, 58] for more details) under two assumptions: that the role of the overlap function $F(p,\theta)$ is negligible outside the diffraction cone and the function f_ρ can be approximated by a constant, i.e., $\rho(\theta_1)=\rho_1=\mathrm{const}$. Both assumptions are discussed in the next subsections.

It is easy to verify that the eigensolution of this equation is

$$\operatorname{Im}A(p,\theta)=C_0\exp\left(-\sqrt{2B\ln\frac{Z}{f_\rho}}\,p\theta\right)+\sum_{n=1}^{\infty}C_n\exp\left[-(\operatorname{Re}b_n)\,p\theta\right]\cos\left(|\operatorname{Im}b_n|\,p\theta-\phi_n\right), \tag{77}$$

with

$$b_n\approx\sqrt{2\pi B|n|}\,(1+\mathrm{i}\,\mathrm{sign}\,n)\,,\qquad n=\pm1,\pm2,\ldots \tag{78}$$

This expression contains the term exponentially decreasing with θ (or $\sqrt{|t|}$) (Orear regime) with oscillations strongly damped by their own exponential factors. These oscillating terms are responsible for the dip. Just this formula was used in Refs [59, 60] for fits of experimental data in a wide energy range. The ratio ρ was approximated by its average values in and outside the diffraction cone, with $f_\rho=1+\rho_0\,\rho_1$, where ρ_1 is treated as the average value of ρ in the Orear region. The fits at comparatively low energies [59] are consistent with $f_\rho\approx 1$, i.e., with small values of ρ_l close to zero. The great surprise of the fit of TOTEM data in [60], shown in Fig. 5, was the necessity of using the negative value $\rho_1\approx -2.1$ large in modulus.

Being model-independent, this solution suffers from some limitations that are inherent for the unitarity relation in general and for unitarity equation (76) in particular. First, it predicts the dependence on transferred momenta $p\theta\approx\sqrt{|t|}$ but not the dependence on the collision energy. Second, it is applicable in a restricted (and not rigorously defined) range of angles in the Orear region.

The elastic scattering differential cross section outside the diffraction cone (in the Orear regime region) is

$$\frac{1}{p_1}\frac{\mathrm{d}\sigma}{\mathrm{d}t}=\left[\exp\left(-\sqrt{2B|t|\ln\frac{4\pi B}{\sigma_{\mathrm{t}}f_\rho}}\right)+p_2\exp\left(-\sqrt{2\pi B|t|}\,\right)\cos\left(\sqrt{2\pi B|t|}-\phi\right)\right]^2. \tag{79}$$

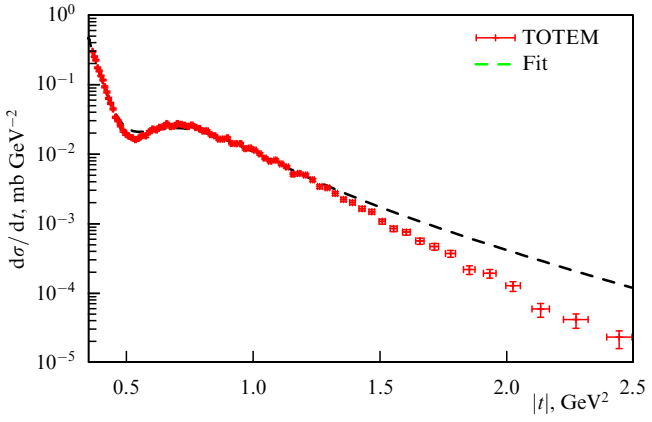


Figure 5. The fit of the differential cross section of elastic proton–proton scattering at $\sqrt{s} = 7$ TeV in the region beyond the diffraction peak according to the predictions of the unitarity condition [60]. Dots show experimental data, the line is the theoretical approximation.

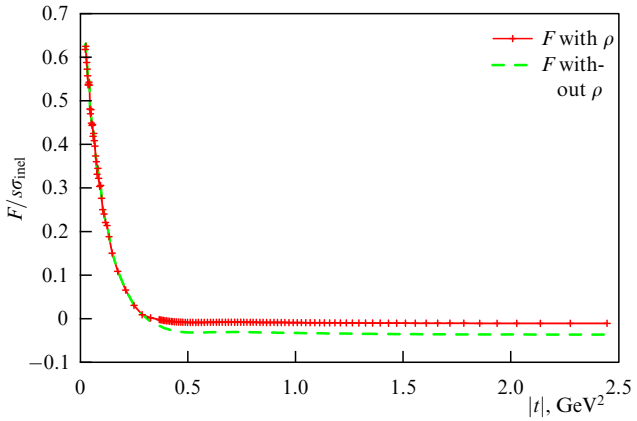


Figure 6. The overlap function at $\sqrt{s} = 7$ TeV obtained from the unitarity condition with substitution of experimental data on the differential cross section [60]. It is large in the diffraction cone and negligibly small outside it. The line nearest to the abscissa axis takes the real part of the amplitude into account, the farthest line is computed with $\rho = 0$.

It has been used in the fit in Fig. 5. Only the very first oscillating term in (77) is taken into account in this expression, because other terms are more strongly damped with $|t|$. It is important that the experimentally measured values of the diffraction cone slope B and the total cross section $\sigma_{\rm t}$ of the same experiment mostly determine the shape of the elastic differential cross section in the Orear region of transition from the diffraction peak to large-angle parton scattering. The value $Z = 4\pi B/\sigma_{\rm t}$ is so close to 1 at 7 TeV that the fit is extremely sensitive to f_ρ because $\ln(Z/f_\rho)$ in the first term determines the slope in this region. Therefore, it becomes possible for the first time to estimate the ratio $\rho_{\rm l}$ outside the diffraction cone directly from fits of experimental data.

Moreover, it was mentioned in footnote 2 in Ref. [58] that Eqn (76) is in fact an equation for $\theta^{1/2}{\rm Im}\,A(p,\theta)$. The factor $\theta^{1/2}$ was omitted in this review and all previous papers because it was assumed that "retaining it would exceed the accuracy of the derivation" of the equation. However, it will be worthwhile to take it into account in the future, multiplying the right-hand side of (79) by $|t|^{-1/4}$. This would slightly improve the fit in Fig. 5.

We note that this shape of differential cross section (79) differs from formula (75), first of all, because of the suppression of oscillations by the exponential factors in front of them, which decrease much more strongly than the leading exponential. In (75), the exponential is common for the main and oscillating terms, while in (79), the oscillations are strongly damped. They may give rise to the dip adjusted to the diffraction cone if their amplitude is sufficiently large. The small secondary damped oscillations at larger values of $|t|$ have been seen at comparatively low energies (see Ref. [59]) but have not yet been noticed at the LHC. We stress that fit (79) contains only three adjustable parameters: the overall normalization p_1, the amplitude of oscillations p_2, which determines the depth of the dip, and f_ρ, which helps find the ratio $\rho_{\rm l}$ outside the diffraction peak from the slope of the differential cross section there.

4.2.4 Overlap function and the eikonal. Both the overlap function and the eikonal are subject to the unitarization procedure, albeit in somewhat different approaches. Therefore, it is instructive to compare their different forms.

We discuss what shapes of the overlap function can be considered as suitable for further use. One of the assumptions used in solving the unitarity equation was the smallness of $F(p,\theta)$ in the Orear region. The results in [60, 228] give strong support to the validity of this assumption. The overlap function was calculated there directly from experimental data, by subtracting the elastic contribution I_2 from the left-hand side of the unitarity equation without appeal to any model. It is described by the formula

$$F(p,\theta) = 16p^2\left(\frac{\pi\,{\rm d}\sigma/{\rm d}t}{1+\rho^2}\right)^{1/2} - \frac{8p^4 f_\rho}{\pi}\int_{-1}^{1}{\rm d}z_2\int_{z_1^-}^{z_1^+}{\rm d}z_1\left(\frac{{\rm d}\sigma}{{\rm d}t_1}\frac{{\rm d}\sigma}{{\rm d}t_2}\right)^{1/2}K^{-1/2}(z,z_1,z_2)\,, \qquad (80)$$

where $z_i = \cos\theta_i$, $K(z,z_1,z_2) = 1 - z^2 - z_1^2 - z_2^2 + 2zz_1z_2$, and $z_1^{\pm} = zz_2 \pm [(1-z^2)(1-z_2^2)]^{1/2}$. The result at 7 TeV is shown in Fig. 6.

Certainly, the shadow of inelastic processes represented by the overlap function dominates within the diffraction peak. But it is extremely small outside. It is even smaller at the LHC energies [50] than at lower ones [228], where a similar behavior of the overlap function at large $|t|$ was observed previously. Hence, this assumption is well founded.

Moreover, it is quite understandable that $F(s,t)$ is very small at large $|t|$ in Fig. 6. This shows that its fit by the solution of the unitarity relation has been done by the proper eigenfunction (77) with the correct eigenvalues of the integral equation.

It is tempting to solve nonlinear inhomogeneous unitarity equation (18) by iterations. That has been attempted several times [56, 58, 122, 218]. The main problem is the choice of the overlap function. The simplest ansatz is the Gaussian form at all transferred momenta. The argument in favor of it is just that it plays the decisive role in the diffraction cone, where the elastic amplitude has a Gaussian shape. But the results fail to describe the Orear regime. This may be ascribed to the role of phases of inelastic processes that determine the genuine shape of the overlap function, and/or to the improper approximation of ρ by a constant outside the diffraction cone. Again, similarly to the situation in the b-representation, the tiny details of the shape prevent the proper outcome. No

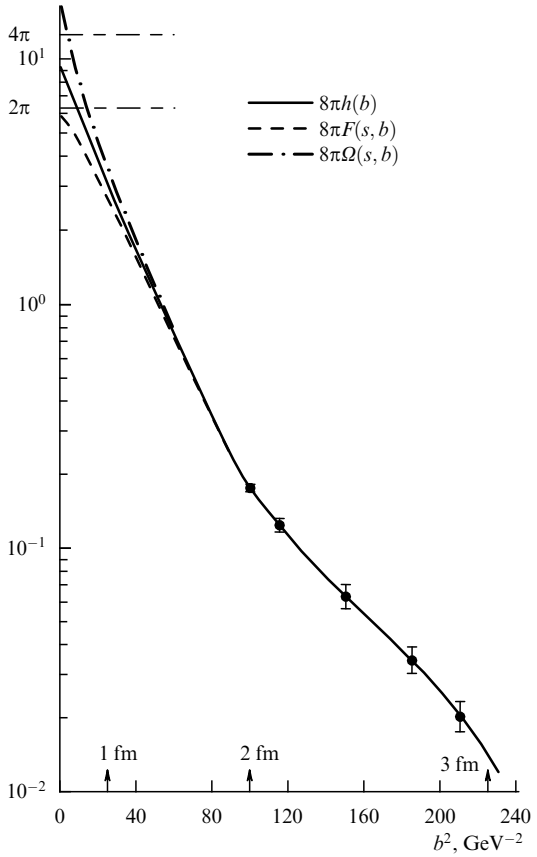


Figure 7. The shapes of the amplitude, overlap function, and eikonal extracted from experimental data at $\sqrt{s} = 52.8$ GeV as functions of the impact parameter squared (borrowed from [229]). In the notation of this review, the amplitude is $h(b) = \tilde{M}(b)/8\pi$, the overlap function is $F(s,b) = \tilde{O}(b)/8\pi$, and the eikonal is $\Omega(s,b) = \tilde{E}(b)/8\pi$. The corresponding spatial scales are shown on the abscissa axis.

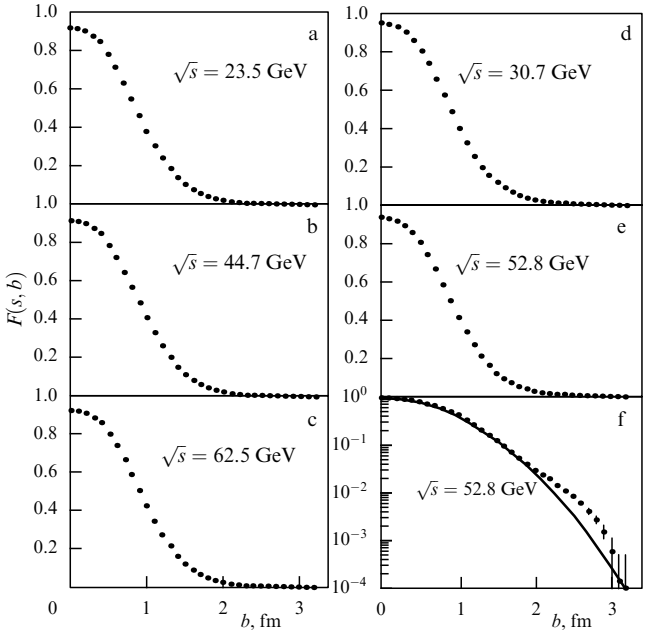


Figure 8. The overlap functions at ISR energies as functions of the impact parameter look similar (borrowed from [230]). The solid line at $\sqrt{s} = 52.8$ GeV is explained in the text.

approximations for the overlap function demonstrated in Fig. 6 have yet been proposed.

It is instructive to confront the shape of the overlap function $F(s,t)$ with results obtained in the impact parameter interpretation of proton–proton scattering. They were presented in Refs [229, 230] for ISR data and are demonstrated in Figs 7 and 8. The b-transformed amplitude $h(s,b)$, the overlap function $F(s,b)$, and the eikonal $\Omega(s,b)$ are shown in Fig. 7 at the energy $\sqrt{s} = 52.8$ GeV [229]. The transformed amplitude is almost Gaussian from the center to 2 fm with some flattening near the center. There is a tail beyond 2 fm with a much flatter slope. The flattening of the overlap function at the center is much stronger, while the eikonal is steeper there. Hence, we should not identify these three curves at small b, even though they almost coincide beyond 2 fm.

Similar features are seen in Fig. 8, taken from [230], where $F(s,b)$ at the same energy is displayed. The solid line on the logarithmic scale is a Gaussian adjusted to fit at $b = 0$ and $b = 1.6$ fm. A Gaussian adjusted between 0.6 and 1.6 fm would be higher at $b = 0$ and would require additional flattening. This flattening at small b corresponds directly to negative values of $F(s,t)$ at large $|t|$, as seen in Fig. 6. In the same way, slight variations of the eikonal $\Omega(s,b)$ at small b may lead to drastic disagreement of model fits with experimental data. Therefore, their success or failure at large $|t|$ depends on the accuracy of the chosen form of the eikonal at low b. A long tail above the solid line for large impact parameters is clearly seen in Fig. 8f. These figures demonstrate how accurate model formulas must be to correctly reproduce either the overlap function or the eikonal if the final goal is to describe the differential cross sections outside the diffraction peak.

A small 'edge' correction to the Gaussian shape of the eikonal has been claimed to be necessary for fits of experimental data at ISR energies on increasing total cross sections and structures of the differential cross sections in Refs [231–233]. For example, the correction factor k with some specific dependence on the impact parameter was introduced into the overlap function $F(s,b)$ in [233]. It changes the shape at small b and makes it similar to that shown in Fig. 7:

$$F(s,b) = P\exp\left(-\frac{b^2}{4B}\right) k\left(s, b\exp\left(-\frac{\gamma^2 b^2}{4B}\right)\right). \qquad (81)$$

It turns out that in the t-representation, the corresponding overlap function $F(s,t)$ has two zeros at $|t| = 0.645$ and 3.83 GeV^2 and becomes practically indistinguishable from zero already at $|t| > 3.5$ GeV^2. The last statement is in full agreement with the conclusions in Refs [59, 60].

Although the overlap functions in Fig. 8 look quite similar to each other, there is a slight difference, which was analyzed in [230, 233]. This difference reveals itself in a small increase at the level of 4% of the overlap function, with an energy increase at the impact parameters (radii) of about 1 fm, which implies the peripheral origin of this phenomenon. That was also discussed earlier [234]. Moreover, in Ref. [230], which deals with the direct analysis of experimental data at ISR, a shoulder of the overlap function at 2.3 fm was noticed. Its origin is unknown.

The overlap function in the b-representation is used in Ref. [235] to distinguish between the mechanisms of absorption and reflection with the help of the unitarity equation. In the latter case, the differential cross section at large momentum transfers is predicted to be 4 times larger.

The impact parameter picture used in almost all phenomenological models is very helpful for a qualitative

description of the process. But the forms of the eikonal in the b-representation turned out to be very approximate. In our opinion, their wide use in most papers dealing with extension to larger angles suffers from this deficiency. There are some arguments [88] that the eikonal approximation is only valid for sums of leading terms of the tower diagrams, but this is not correct in general. This is applicable to almost collinear processes only and does not properly take the separation due to transverse momenta into account. That is why quasi-eikonal models were developed where the intermediate states take inelastic diffraction processes into account, in addition to elastic ones. As a result, formulas like (39) and (40) were proposed. The eikonal does not properly reproduce the s-channel cuts of the scattering amplitude due to multiple scattering [236]. By itself, it does not guarantee precise unitarization. Moreover, the procedure of unitarity corrections is not well defined, because it can be implemented differently. The accuracy of unitarity relation (26) in the b-representation is not absolutely clear either, as discussed above, while its use is mandatory for interpretation of experimental data. That is why the model predictions shown in Fig. 2 fail to explain the data.

There is a drastic difference between the use of the Gaussian shape for the amplitude in the s-channel unitarity condition and the same shape for the overlap function, as well as its use directly in the b-representation. The exponential decrease [see Eqn (6)] of the differential cross section in the diffraction cone as a function of $|t|$ (or a Gaussian decrease with the angle) is an experimental observation. It can be used anywhere within its applicability range, as was done, for example, in solving Eqn (18). Hence, this solution is quite successful in fits of experimental data in the Orear region. The same shape cannot be used for the t-dependence of the overlap function, although it plays an important role in the formation of cone behavior.

It is often argued that the Fourier transform of the Gaussian is a Gaussian and therefore this shape can also be used in the b-representation. While the first part of the statement is correct, the second is wrong. The tails of differential cross sections are very sensitive to small b. Slight variations of this shape at small impact parameters lead to crucial changes in the behavior of the amplitude at large transferred momenta. Therefore, the predictions shown in Fig. 2, which use the impact parameter profiles close to Gaussian ones even in the vicinity of $b = 0$, are still successful inside the diffraction cone but completely fail outside it, where central collisions play an important role. It is very difficult in a particular model to guess the proper decline from the Gaussian shape at small impact parameters, which drastically influences the differential cross section at large transferred momenta.

Therefore, attempts to use non-Gaussian electromagnetic form factors were of some help in improving the situation, because they are closer in shape to the eikonal demonstrated in Figs 7 and 8. Further progress in this direction is necessary in order to understand the geometric content of the interaction region in ordinary space and time.

Nevertheless, it is hardly justified to blame the phenomenological model builders for their failure to predict the behavior of differential cross sections at large transferred momenta, where the differential cross section is many orders of magnitude lower than in the diffraction peak. The great and important task of fits of the energy behavior of total and elastic cross sections, the (s, t)-dependence of the differential cross section, and the ratio ρ in a wide range of energies and transferred momenta cannot be accomplished without free parameters and the physical intuition of model builders. The switch to higher energies allows eliminating corrections due to secondary Reggeons and improving the fits. There is hope of gaining clearer insight into the geometrical picture of hadron interactions.

4.2.5 Real part of the elastic scattering amplitude at nonzero transferred momenta. There are no reasonable arguments to neglect the t-dependence of the ratio $\rho(s, t)$ in (2) or of the phase ζ in (13): it seems to be important, even inside the diffraction cone. Using formula (10) and assuming that $\operatorname{Im} A(s, t)$ mainly determines the shape of the differential cross section in this region, we find that the real part must vanish at

$$t_0 = -2\,\frac{\mathrm{d}\ln\sigma_{\mathrm{t}}(s)/\mathrm{d}\ln s}{\mathrm{d}B(s)/\mathrm{d}\ln s}\,. \tag{82}$$

With the $\ln^2 s$-dependence of σ_{t} in (32) and $B(s)$ in (34) and using relation (42), we have

$$|t_0| = \frac{2}{B} = \frac{16\pi}{\sigma_{\mathrm{t}}}\,, \tag{83}$$

and hence $t_0 \to 0$ as $\sigma_{\mathrm{t}} \to \infty$. The estimates at LHC energies are $0.1 < |t| < 0.3\ \mathrm{GeV}^2$. Notably, they agree with the results obtained in the models in [165, 192].

There were several attempts to consider the behavior of $\rho(s, t)$ at larger transferred momenta in Refs [165, 230, 237–240]. The main efforts were spent on preventing differential cross sections from vanishing at those values of t where the imaginary part of the amplitude is zero in a particular model. The ratio $\rho(t)$ should be infinite, e.g., as in the models in Refs [165, 230]. The number of zeros of the imaginary part is sometimes greater than one. This is typical in the Fraunhofer diffraction or in models with electromagnetic form factors. Therefore, the singularities of $\rho(t)$ appear at different t in different models. The real part of the amplitude fills in these kinks, leaving some traces like shoulders or dips in the differential cross sections. For example, it is predicted in Ref. [165] that for pp scattering at 8 TeV, such traces appear at $|t| \approx 0.35\ \mathrm{GeV}^2$ and at $1.5\ \mathrm{GeV}^2$.

In Refs [237, 238], the dispersion relation between the phase and the modulus of the elastic amplitude considered in Refs [241, 242] was used with some input for the modulus fitted to the experimental data at laboratory energies above 100 GeV. The conclusion was that the real part exhibits a zero in the t-distribution above 200 GeV, which moves away from the forward direction as the energy increases.

In Ref. [239], the eikonal approximation was used following the proposal in Ref. [243]. Information about the interference region with a Coulomb amplitude similar to that in Eqn (30) was inserted into the total amplitude, with the result

$$\begin{aligned} A(s, t) = &-\frac{8\pi\alpha}{|t|}\, s f_1\bigl(|t|\bigr) f_2\bigl(|t|\bigr) \exp\left(\mathrm{i}\alpha\Phi\right) \\ &+ \mathrm{i}s\sigma_{\mathrm{t}} \exp\left(\frac{Bt}{2} - \mathrm{i}\zeta(s, t)\right)\left\{1 - \mathrm{i}\alpha\int_{-\infty}^{0} \mathrm{d}t'\, \ln\frac{t'}{t}\right. \\ &\left.\times\,\frac{\mathrm{d}}{\mathrm{d}t'}\left(f_1\bigl(|t'|\bigr) f_2\bigl(|t'|\bigr) \exp\left[\frac{Bt'}{2} - \mathrm{i}\bigl(\zeta(s, t') - \zeta(s, 0)\bigr)\right]\right)\right\}. \end{aligned} \tag{84}$$

The t-dependence of the phase was parameterized with the help of five parameters as

$$\zeta(t) = \zeta_0 + \zeta_1\left(\frac{t}{t_0}\right)^{\kappa} \exp\left(v|t|\right) + \zeta_2\left(\frac{t}{t_0}\right)^{\lambda}, \quad t_0 = -1 \text{ GeV}^2 . \tag{85}$$

The results showed that the phase [related to ρ by (14)] increases from values close to zero at $t = 0$ to about 0.5 in the interval $0.5 < |t| < 1$ GeV^2. This conclusion disagrees with the results in Refs [73, 74, 165] as well as with the arguments presented below.

A more general approach using the s-channel unitarity condition was developed in Ref. [240]. As explained above, the integral equation for the elastic amplitude is valid in the Orear region. Its analytic solution (77) was first obtained in the approximation where the values of ρ in f_ρ were replaced by their average values in the diffraction cone and in the Orear region. No zeros of the imaginary part of the amplitude were obtained. The dips at 7 TeV and lower energies were explained as resulting from damped oscillations. The necessity to introduce large negative values of ρ into the Orear region is the main outcome and surprise of the fit in Ref. [60]. In principle, this could happen if there were zeros of the imaginary part of the amplitude in this region, which would require very large values of $|\rho|$ near them. But there seem to be no such zeros there. We discuss this problem in more detail.

We first recall asymptotic predictions. It was shown in [244] that the ratio of real to imaginary parts of the amplitude can be calculated asymptotically at nonzero transferred momenta t as

$$\rho = \rho_0\left[1 + \frac{\tau\left(\mathrm{d}f(\tau)/\mathrm{d}\tau\right)}{f(\tau)}\right]. \tag{86}$$

We consider the leading term of solution (77). With the imaginary part of the amplitude in the Orear region represented as

$$\operatorname{Im} A(s,t) = C_0(s) f(\tau) , \tag{87}$$

it is possible to calculate ρ.

The very first approximation was to use the first term of solution (77) with average values of ρ both in the diffraction peak ($\rho_\mathrm{d} \approx \rho_0$) and in the Orear region (ρ_l) [240]. Then the following behavior of ρ was obtained:

$$\rho(s,t) = \rho_0\left(1 - \frac{a\sqrt{|t|}}{2}\right), \tag{88}$$

where

$$a = \sqrt{2B \ln \frac{Z}{1 + \rho_0\,\rho_\mathrm{l}}} . \tag{89}$$

We note that ρ passes through zero and changes sign at $|t| = 4/a^2 \approx 0.1$ GeV^2. This agrees with the general theorem on the change of sign of the real part of the high-energy scattering amplitude, which was first proved in Ref. [84]. A similar effect is discussed in Ref. [165]. But it is difficult to obtain $\rho_\mathrm{l} = -2.1$ as an average of (88) over the Orear region.

Moreover, this behavior of an unbounded decrease in ρ with $|t|$ does not look satisfactory. It can in fact be damped if instead of replacing ρ by ρ_l in the solution, we differentiate f according to (86), inserting there Eqn (87), i.e., the first term in (77). The following differential equation is then obtained:

$$\frac{\mathrm{d}v}{\mathrm{d}x} = -\frac{v}{x} - \frac{2}{x^2}\left[\frac{Z\exp\left(-v^2\right) - 1}{\rho_0^2} - 1\right]. \tag{90}$$

Here, $x = \sqrt{2B|t|}$ and $v = \sqrt{\ln\left(Z/f_\rho\right)}$. The dependence $\rho(t) = \left(Z\exp\left(-v^2\right) - 1\right)/\rho_0$ is shown in Fig. 9. It has a single zero at $|t| \approx 0.3$ GeV^2 and, which is indeed impressive, a large negative saturation value $\rho(|t| \to \infty) = -1/\rho_0$ for high transferred momenta $|t|$. We note that f_ρ tends to zero there. In the Orear region $0.3 < |t| < 1.4$ GeV^2, $\rho(t)$ steeply decreases. Nevertheless, taking the $\sqrt{|t|}$-exponential decrease in the distribution into account, the estimate of its average value -2.1 in this interval is not very bad at all, especially if we consider the result shown in Fig. 9 as another extreme approximation compared to Eqn (88).

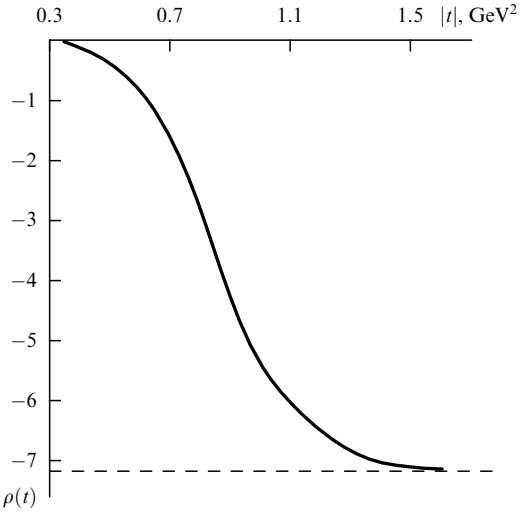


Figure 9. The ratio of the real to imaginary part of the amplitude obtained from the solution of Eqn (90) that follows from the unitarity condition [240].

The bold use of this procedure for derivation of Eqn (90) with $\rho(t)$ inserted directly into the solution is nevertheless not satisfactory, either. The two possibilities above should be regarded as two extremes for the shapes of $\rho(t)$.

Strictly speaking, the behavior of $\rho(t)$ should be taken into account primarily in the integrand. Then, inserting expression (86) in place of ρ_l in Eqn (76) and integrating by parts, we derive the linear integral equation

$$\operatorname{Im} A(x) = \frac{1}{Z\sqrt{\pi}}\int_{-\infty}^{+\infty} \mathrm{d}y \, \exp\left[-(x-y)^2\right] + \left[1 + 0.5\rho_0^2 + \rho_0^2(y-x)y\right] \operatorname{Im} A(y) , \tag{91}$$

with $F(p,\theta) = 0$ and new variables $x = \sqrt{B/2}\,p\theta$ and $y = \sqrt{B/2}\,p\theta_\mathrm{l}$. The kernel of this equation is not symmetric. Its solution has not yet been obtained, even numerically. However, some preliminary asymptotic estimates can be obtained from it [240].

In the preasymptotic energy region, we obtained [58] the Orear regime $\operatorname{Im} A \propto \exp\left(-ap\theta\right) \approx \exp\left(-ap_\mathrm{t}\right)$ with the exponential fall-off of the amplitude as a function of angles. We therefore seek a solution of Eqn (91) in the form $\operatorname{Im} A(x) = \exp\left(-ax\sqrt{2/B}\right)\phi(x)$. The Gaussian exponential

shifts to $x - y - a/\sqrt{2B}$. Replacing it with the δ-function of this argument, we obtain the equation in finite differences:

$$\phi(x) = Z^{-1} \exp\left(\frac{a^2}{2B}\right)\left[1 + 0.5\rho_0^2\left(1 + \frac{a^2}{B} - ap_t\right)\right] \times \phi\left(x - \frac{a}{\sqrt{2B}}\right). \quad (92)$$

Again, we cannot solve it directly, but reach an important conclusion about the zeros of the imaginary part of the amplitude. The expression in the square brackets is equal to zero at

$$p_{t0} = \frac{2}{a\rho_0^2}\left[1 + 0.5\rho_0^2\left(1 + \frac{a^2}{B}\right)\right] \approx \frac{2}{a\rho_0^2}. \quad (93)$$

With the present-day values of B, a, and ρ_0^2, this zero would appear at an extremely large $p_{t0} \approx 20$ GeV. However, zeros of the imaginary part of the amplitude in the Orear region just above the diffraction cone might appear as zeros of $\phi(x)$ itself. This result does not contradict the above statement about the absence of zeros in the case of small oscillatory terms in the solution of a homogeneous linear integral equation.

Moreover, the equation tells us that $\phi(x)$ and, consequently, the imaginary part of the amplitude can have zeros at $x_n = x_0 + a/\sqrt{2B}$. On the p_t axis, these zeros would be placed at rather short distances from one another.

In the black-disk limit, Z tends to 0.5. If ρ loses in the competition with $\ln(Z/f_\rho)$ and the argument of the logarithm becomes extremely close to 1 or even less, that would mean a drastic change of the regime in the Orear region [245]. What the outcome of the competition between decreasing Z and the negative values of ρ will be poses an interesting problem. Experimental data at higher energies will be able to give the answer.

As we see, the real part of the amplitude can dominate at large transferred momenta according to the unitarity condition. Compared to the imaginary part, it can be large and negative there. This conclusion contradicts, for example, the results of the models in [92, 165] with electromagnetic form factors, where the dominance of the imaginary part, on the contrary, is claimed everywhere except the tiny regions near its zeros placed in the Orear region, in particular. This disagrees with the above results. We must remember, however, that ρ is infinite at these zeros (see Fig. 10). A similar behavior of ρ in the case of a single zero was predicted in Ref. [230] at ISR energies, as shown in Fig. 10.

A similar shape of ρ is obtained in Ref. [165] at $|t| \approx 1.5$ GeV^2, but for the energy $\sqrt{s} = 8$ TeV. The real part decreases with $|t|$. Thus, the conclusions in different papers about the behavior of the real and imaginary parts of the elastic scattering amplitude are contradictory and require further theoretical studies and new experimental data.

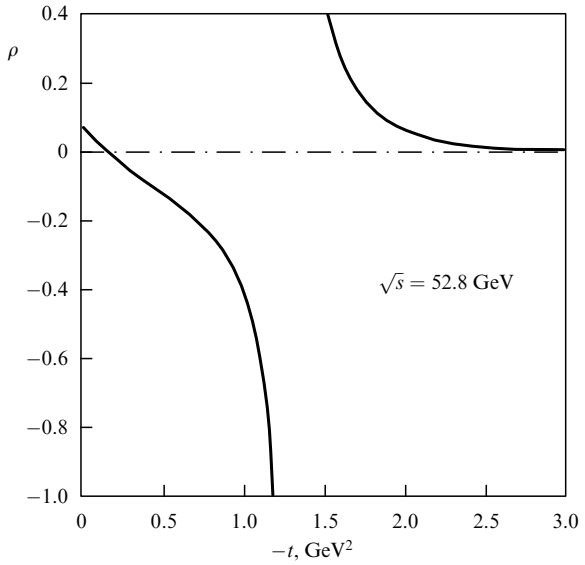


Figure 10. The dependence of the ratio of the real part to the imaginary part of the amplitude on transferred momenta obtained in a certain phenomenological fit [230] of experimental data at $\sqrt{s} = 52.8$ GeV. The singularities indicate the position where the imaginary part vanishes.

4.3 Scaling laws

We have written two formulas, (10) and (86), for the same function $\rho(s,t)$. Therefore, these two expressions must be identical. Equating them, we obtain [246] the partial differential equation

$$p - f(x)q = 1 + f(x), \quad (94)$$

where $p = \partial u/\partial x$, $q = \partial u/\partial y$, $u = \ln \operatorname{Im} A(s,t)$, $f(x) = 2\rho(s,0)/\pi$, $x = \ln s$, and $y = \ln|t|$. As usual, the variables s and $|t|$ should be regarded as scaled by the corresponding constant factors s_0^{-1} and $|t_0|^{-1}$.

Equation (94) can be rewritten as

$$\frac{\partial u}{\partial \ln \sigma_t} - \frac{\partial u}{\partial \ln t} = 1 + \frac{d \ln s}{d \ln \sigma_t}.$$

In the asymptotic black-disk limit $\sigma_t \propto \ln^2 s$ and $\rho(s,0) = \pi/\ln s$, we obtain

$$xp - 2q = x + 2, \quad (95)$$

and the solution is

$$u = \varphi_1\left(x \exp\frac{y}{2}\right) + x - y. \quad (96)$$

This yields the scaling law for

$$\frac{|t|}{s} \operatorname{Im} A(s,t) = \exp\left[\varphi_1\left(\sqrt{|t|}\ln s\right)\right] = \phi(z_1), \quad (97)$$

which implies a universal scaling dependence on a single variable $z_1 = \sqrt{|t|}\ln s$.

We temporarily neglect the contribution of the real part of the amplitude to the differential cross section. Then the asymptotic scaling law for the differential cross section times t^2 should be

$$t^2 \frac{d\sigma}{dt} = \phi_1^2\left(\sqrt{|t|}\ln s\right). \quad (98)$$

We note that the additional t^2-factor can be replaced by an s-dependence if absorbed in the argument of the scaling function ϕ. Then this formula coincides with that obtained in the geometric scaling approach [145, 247]. Thus, we have proved that the solution of partial differential equation (94) with the properly chosen $f(x)$ leads to the results known previously from geometric scaling.

At the same time, Eqn (94) is more general and can be used for different assumptions about $f(x)$. In particular, the

behavior of the total cross section at present energies is often approximated by formula (67) as a sum of a large constant term and another term that increases as some power of energy (see [192] for a recent reference). In this case, $\rho(s,0) = \pi\Delta/2$, and the equation is

$$p - \Delta q = 1 + \Delta\,. \tag{99}$$

Its solution is

$$u = \varphi_2(\Delta x + y) + x - y\,. \tag{100}$$

From here, we obtain another universal scaling dependence of the differential cross section on a single variable $z_2 = |t|s^{\Delta}$,

$$t^2\,\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \phi_2^2\left(|t|s^{\Delta}\right), \tag{101}$$

which could be valid at preasymptotic energies.

It follows from the above expressions that the energy dependence of the scaling variable is determined by the behavior of the total cross section, $|t|\sigma_{\mathrm{t}}$, if only the first term in Eqn (9) is used. If this scaling were valid, one would be able to predict the shape of the differential cross section at a higher energy once the total cross section is known there. The preliminary results of work with experimental data at energies from ISR to LHC have shown that just this dependence best reproduces the similarity of the shapes of the corresponding lines, even though their normalization differs somewhat. Further studies are necessary.

The above scaling laws must be satisfied for the imaginary part of the amplitude times the factor $|t|/s$ (see (97)). It follows from Eqn (10) that the real part satisfies an analogous scaling law, albeit with another factor, which differs in the two cases considered above. This would lead to the scaling violating terms when the contribution of the real part of the amplitude to the differential cross section is taken into account. The above scaling dependences of the differential cross section are modified as

$$t^2\,\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \phi_1^2(z_1) + 0.25\pi^2|t|\phi_1'^2(z_1)\,, \tag{102}$$

$$t^2\,\frac{\mathrm{d}\sigma}{\mathrm{d}t} = \phi_2^2(z_2) + 0.25\pi^2\Delta^2 z_2^2\phi_2'^2(z_2)\,. \tag{103}$$

The violation of scaling laws is different in these cases. The first law acquires a term with the coefficient depending only on the transferred momentum, while the second law acquires a term with the coefficient that depends both on energy and on the transferred momentum.

This violation of scaling laws must be negligible in the diffraction cone because the squared ratio of the real part to the imaginary part — which is crucial for the differential cross section — is extremely small there. It would be interesting to learn about the effect of these terms outside the cone, especially in the Orear region of transferred momenta.

We note that at small values of their arguments z_i, the scaling functions $\phi_i(z_i)$ must be respectively proportional to z_1^2 and to z_2, for the differential cross section to be equal to a constant at $t = 0$.

4.4 Hard scattering at large angles

4.4.1 Dimensional counting. The energy dependence of high-energy scattering processes at a fixed center-of-mass angle is of special interest. Dimensional scaling laws have been developed as an approach to understanding it. The large-angle scattering is determined by contributions due to central interactions of internal domains inside the colliding particles. The estimates according to perturbative QCD become justified due to the asymptotic freedom property. They depend on the number of constituent fields of the hadrons [248–250]. At large s and t and a fixed ratio s/t, we have

$$\left.\frac{\mathrm{d}\sigma}{\mathrm{d}t}\right|_{\mathrm{AB}\to\mathrm{CD}} \propto s^{-n+2} f\left(\frac{t}{s}\right), \tag{104}$$

where n is the total number of fields in A, B, C, D that carry a finite fraction of the momentum.

Assuming the existence of quark constituents, the $s \to \infty$, fixed-t/s prediction for pp scattering [249, 250] is $\mathrm{d}\sigma/\mathrm{d}t \propto s^{-10}$. For the elastic amplitude, it is

$$A_1(s,t) \propto \left(\frac{s_0}{s}\right)^{n/2-2} f_1\left(\frac{s}{t}\right). \tag{105}$$

This form can become more complicated for multiple scatterings. For example, the lowest-order graphs for m rescatterings [251] behave as

$$A_m(s,t) \propto \left(\frac{s_0}{s}\right)^{(n-m+1)/2-2} f_m\left(\frac{s}{t}\right), \tag{106}$$

and could become the leading ones. However, due to higher-order corrections, the resulting behavior could change not so drastically, and the result would be close to the initial estimate (105), as is shown in Ref. [252]. Further progress beyond the simple quark-counting rules was slowed down by complications in calculating the enormous number of Feynman diagrams.

4.4.2 Coherent scattering. In parallel, there were attempts to explain the $|t|^{-8}$ regime in pp scattering by dynamical mechanisms with the help of simple Feynman graphs. For protons (or their subregions) consisting of three valence quarks, we can assume coherent exchange by gluons [253–255] or by color-neutral pairs of gluons [256, 257] between them. The propagators of three gluons and their couplings produce an $\alpha_{\mathrm{s}}^6|t|^{-6}$ dependence, and two powers in the denominator are added by kinematical factors. The general problem of these approaches is the necessity to introduce additional factors in order to preserve both protons in their initial states in large-angle scattering. The corresponding powers of the QCD coupling constant should be included, of course, which leads to possible (strong?) modifications of the simple power law. Also, the exchange by three Pomerons instead of three pairs of gluons is possible. Because three colliding quarks share the total energy of the proton equally, their shares are smaller, and the whole process is farther from the asymptotic regime if treated at the parton level. None of these questions have been quantitatively resolved yet.

We note that the large-$|t|$ behavior of Reggeons composed of two Reggeized partons (quarks, gluons) can be calculated from the BFKL equation [258, 259].

The multi-Pomeron exchange for hadrons in a state with a minimum number of partons was considered in [260]. It was concluded that the differential cross section factors as a product of two $|S_0(s)|^2$, representing the probability of finding the initial and final particles in a 'bare' state, and

$d\hat{\sigma}(s,t)/dt$, describing the hard exchange interaction:

$$\frac{d\sigma}{dt} = \left|S_0(s)\right|^2 \frac{d\hat{\sigma}(s,t)}{dt} . \qquad (107)$$

The first factor describes the contribution of large transverse distances and the second represents the contribution of small ones. The hard exchange is determined by the Pomeron vertices, which are known semiclassically:

$$\frac{d\hat{\sigma}(s,t)}{dt} \propto g_1^2(t)\, g_2^2(t) \propto \frac{\left(\alpha_s(t)\right)^v}{|t|^N} \qquad (108)$$

with

$$v = n_1 + n_2 + |n_1 - n_2| ,$$
$$N = 0.5\left[3(n_1 + n_2) + |n_1 - n_2| - 1\right] , \qquad (109)$$

where n_i are the numbers of valence quarks in the colliding hadrons. This leads to a $|t|^{-8}$ behavior for pp and $|t|^{-7}$ for πp. The quantitative comparison with experimental data is more difficult because of much smaller values of the differential cross sections in this region and, correspondingly, larger error bars.

5. Discussion and conclusions

The new experimental data of the TOTEM collaboration at the LHC about elastic scattering of protons at an energy of 7 TeV has revived interest in these processes. The picture of very short-wavelength hadron collisions has become available, adding to our insight into the spatial structure of colliding particles and providing new intrinsic information pertaining to very short distance interactions. The total and elastic cross sections show a stable increase with energy. The share of elastic processes increases. The differential cross section has very intriguing properties. The exponential $|t|$-decrease persists at small transferred momenta, analogously to lower-energy data. But the diffraction cone slope is larger compared with low energies; it is stable up to transferred momenta $|t| \approx 0.3$ GeV^2, then this peak steepens and a dip appears at $|t| \approx 0.53$ GeV^2, with a subsequent maximum at $|t| \approx 0.7$ GeV^2. At somewhat larger angles, the exponential in the $\sqrt{|t|}$-regime prevails. It is replaced by the $|t|^{-8}$ behavior at ever larger transferred momenta, $|t| > 2$ GeV^2. At the same time, we are waiting for measurements at extremely small angles in the interference region of Coulomb and nuclear amplitudes to gain some knowledge about the real part of the forward scattering amplitude. It would be extremely interesting to learn its energy behavior and check our predictions from the dispersion relations.

The steeper slopes of the diffraction peak and of the Orear region at higher energies, and, correspondingly, their smaller extensions clearly demonstrate that it becomes more and more difficult for a high-energy particle to preserve its identity when scattering at larger transverse momenta.

This increase in the total cross section and, especially, in the share of the elastic cross section, as well as the peculiar change of regimes in the $|t|$-behavior of the differential cross section, require a theoretical interpretation. Short of a complete theory of hadron dynamics, we have to turn our attention to phenomenological models and some rare rigorous theoretical relations. The region of large transferred momenta became an Occam razor for them, as explained above.

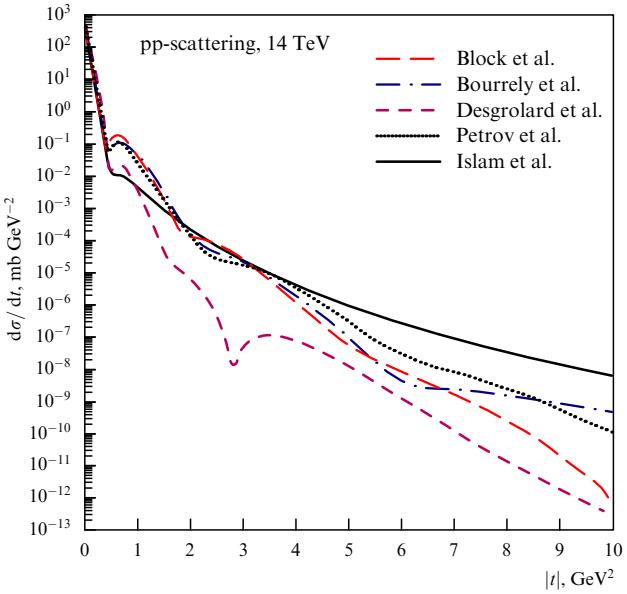


Figure 11. Model predictions for the behavior of the differential cross section of proton–proton scattering at $\sqrt{s} = 14$ TeV, presented in Ref. [119].

The geometric picture of the internal structure of protons and their collisions requires larger disk radii increasing with energy. Their blackness increases as well. Some separate subregions of different sizes and opacity are considered. The impact parameter approach is decisive in deciphering this structure. At ISR energies, the increase in the total cross section was attributed to some peripheral regions of nucleons. It is important to juxtapose these findings with the LHC data. The approach to the black-disk asymptotic limit has become very interesting. The proposal of geometric scaling reducing the number of independent variables is under investigation. At the same time, the scaling law may happen to be different from the geometric scaling.

There are many phenomenological models at our disposal, but it is still difficult to choose any particular one among them. Most of them are quite successful, albeit with many adjustable parameters, in describing the energy behavior of the cross sections and the main bulk of the elastic processes in the diffraction cone, but fail in their predictions outside it. The dynamical origin of many assumptions is still missing. The small details of the suspected break at small t, of the steepened slope, and of possible weak oscillations over a smooth exponential behavior of the diffraction peak are under investigation.

There are predictions of several dips and/or visible oscillatory behavior imposed on the trend of a generally decreasing dependence on the transferred momentum. As an example, in Fig. 11 borrowed from [119], the results of some model predictions for the differential cross section of proton–proton scattering at $\sqrt{s} = 14$ TeV are shown up to quite high values $|t| = 10$ GeV^2. They differ significantly, and further accurate experimental data expected to be obtained in 2015 – 2016 will surely be decisive in the choice of a model (if any!). The experience with unsuccessful predictions at 7 TeV in the region outside the diffraction cone is not very encouraging.

The problem of the behavior of the real part of the elastic scattering amplitude at nonforward transferred momenta is becoming very important. While the imaginary part of the

amplitude dominates at small angles in the diffraction cone, there are indications that just the real part prevails at high transferred momenta. The unitarity condition indicates some ways to solve this problem. However, there are other approaches with different conclusions.

Another important unsolved problem is the behavior of the overlap function. It certainly dominates in the diffraction cone, but seems to become negligibly small outside it. The phases of matrix elements of inelastic processes must play an important role in attempts to recover its shape. However, this presents the extremely difficult theoretical task of modeling them.

Unfortunately, there is little progress in understanding the regime of power counting for very hard constituent parton scattering, even though some recent attempts are quite promising.

To conclude, the aforementioned list of problems is not at all complete. Many other details should be clarified. Further experimental data will definitely shed light on ways to resolve them.

Acknowledgements

This work was supported by the RFBR grant 12-02-91504-CERN-a and by the RAS–CERN program.

References

1. Foley K J et al. *Phys. Rev. Lett.* **11** 425 (1963); *Phys. Rev. Lett.* **11** 503 (1963)
2. Cocconi G et al. *Phys. Rev. Lett.* **11** 499 (1963)
3. Krisch A D *Phys. Rev. Lett.* **11** 217 (1963)
4. Narayan D S, Sarma K V L *Phys. Lett.* **5** 365 (1963)
5. Orear J *Phys. Rev. Lett.* **12** 112 (1964)
6. Orear J *Phys. Lett.* **13** 190 (1964)
7. Baker W F et al. *Phys. Rev. Lett.* **12** 132 (1964)
8. Cocconi G et al. *Phys. Rev.* **138** B165 (1965)
9. Orear J et al. *Phys. Rev.* **152** 1162 (1966)
10. Allaby J V et al. *Phys. Lett. B* **27** 49 (1968)
11. Allaby J V et al. *Phys. Lett. B* **28** 67 (1968)
12. Akerlof C W et al. *Phys. Rev. D* **14** 2864 (1976)
13. Ayres D S et al. *Phys. Rev. D* **15** 3105 (1977)
14. Asa'd Z et al. *Nucl. Phys. B* **255** 273 (1985)
15. Amaldi U et al. *Phys. Lett. B* **44** 112 (1973)
16. Amendolia S R et al. *Phys. Lett. B* **44** 119 (1973)
17. Böhm A et al. *Phys. Lett. B* **49** 491 (1974)
18. Kwak N et al. *Phys. Lett. B* **58** 233 (1975)
19. Albrow M G et al. *Nucl. Phys. B* **108** 1 (1976)
20. Hartmann J L et al. *Phys. Rev. Lett.* **39** 975 (1977)
21. Conetti S et al. *Phys. Rev. Lett.* **41** 924 (1978)
22. Baksay L et al. *Nucl. Phys. B* **141** 1 (1978)
23. Nagy E et al. *Nucl. Phys. B* **150** 221 (1979)
24. Amaldi U, Schubert K R *Nucl. Phys. B* **166** 301 (1980)
25. Kuznetsov A A et al. *Sov. J. Nucl. Phys.* **33** 74 (1981) [*Yad. Fiz.* **33** 142 (1981)]
26. Schiz A et al. *Phys. Rev. D* **24** 26 (1981)
27. Fidecaro G et al. *Phys. Lett. B* **105** 309 (1981)
28. Faissler W et al. *Phys. Rev. D* **23** 33 (1981)
29. Rubinstein R et al. *Phys. Rev. D* **30** 1413 (1984)
30. Breakstone A et al. *Nucl. Phys. B* **248** 253 (1984)
31. Eisenhandler E et al. *Nucl. Phys. B* **113** 1 (1976)
32. Favart D et al. *Phys. Rev. Lett.* **47** 1191 (1981)
33. Amos N et al. *Nucl. Phys. B* **262** 689 (1985)
34. Amos N A et al. *Phys. Lett. B* **247** 127 (1990)
35. Ambrosio M et al. *Phys. Lett. B* **115** 495 (1982)
36. Sakamoto S et al. *Nucl. Phys. B* **195** 1 (1982)
37. Bozzo M et al. *Phys. Lett. B* **147** 385 (1984)
38. Bozzo M et al. *Phys. Lett. B* **155** 197 (1985)
39. Breakstone A et al. *Phys. Rev. Lett.* **54** 2180 (1985)
40. Bernard D et al. *Phys. Lett. B* **171** 142 (1986)
41. Bernard D et al. *Phys. Lett. B* **198** 583 (1987)
42. Augier C et al. *Phys. Lett. B* **316** 448 (1993)
43. Abe F et al. *Phys. Rev. D* **50** 5518 (1994)
44. Abazov V M et al. (D0 Collab.) *Phys. Rev. D* **86** 012009 (2012); arXiv:1206.0687
45. Zotov N P, Rusakov S V, Tsarev V A *Sov. J. Part. Nucl.* **11** 462 (1980) [*Fiz. Elem. Chastits At. Yadra* **11** 1160 (1980)]
46. Amaldi U, arXiv:1206.3954
47. Dataset of differential elastic cross sections, http://www.theo.phys.ulg.ac.be/~cudell/data/
48. Antchev G et al. (TOTEM Collab.) *Europhys. Lett.* **95** 41001 (2011)
49. Antchev G et al. (TOTEM Collab.) *Europhys. Lett.* **96** 21002 (2011)
50. Cheng H, Wu T T *Phys. Rev. D* **1** 1069 (1970); *Phys. Rev. D* **1** 1083 (1970)
51. Gribov V N, Migdal A A *Sov. J. Nucl. Phys.* **8** 583 (1969) [*Yad. Fiz.* **8** 1002 (1968)]
52. Bronzan J B, Kane G L, Sukhatme U P *Phys. Lett. B* **49** 272 (1974)
53. Fischer J, Kolář P *Phys. Lett. B* **64** 45 (1976)
54. Fischer J, Kolář P *Phys. Rev. D* **17** 2168 (1978)
55. MacDowell S W, Martin A *Phys. Rev.* **135** B960 (1964)
56. Van Hove L *Nuovo Cimento* **28** 798 (1963)
57. Andreev I V, Dremin I M *JETP Lett.* **6** 262 (1967) [*Pis'ma Zh. Eksp. Teor. Fiz.* **6** 810 (1967)]
58. Andreev I V, Dremin I M *Sov. J. Nucl. Phys.* **8** 473 (1968) [*Yad. Fiz.* **8** 814 (1968)]
59. Andreev I V, Dremin I M, Gramenitskii I M *Nucl. Phys. B* **10** 137 (1969)
60. Dremin I M, Nechitailo V A *Phys. Rev. D* **85** 074009 (2012)
61. Adachi T, Kotani T *Prog. Theor. Phys.* **35** 485 (1966)
62. Adachi T, Kotani T *Prog. Theor. Phys.* **39** 785 (1968)
63. Islam M M *Nucl. Phys. B* **104** 511 (1976)
64. Kundrát V, Lokajíček M, Krupa D *Phys. Lett. B* **544** 132 (2002)
65. Kundrát V, Kašpar J, Lokajíček M, arXiv:0912.1188
66. Bethe H A *Ann. Physics* **3** 190 (1958)
67. Rix J, Thaler R M *Phys. Rev.* **152** 1357 (1966)
68. West G B, Yennie D R *Phys. Rev.* **172** 1413 (1968)
69. Solov'ev L D *JETP* **22** 205 (1966) [*Zh. Eksp. Teor. Fiz.* **49** 292 (1965)]
70. Locher M P *Nucl. Phys. B* **2** 525 (1967)
71. West G B, Yennie D R *Phys. Rev.* **172** 1413 (1968)
72. Cahn R *Z. Phys. C* **15** 253 (1982)
73. Selyugin O V *Phys. Lett. B* **333** 245 (1994)
74. Selyugin O V *Phys. Rev. D* **60** 074028 (1999); arXiv:1201.4458
75. Block M M *Phys. Rev. D* **54** 4337 (1996)
76. Block M M *Phys. Rep.* **436** 71 (2006)
77. Kundrát V, Lokajíček M *Z. Phys. C* **63** 619 (1994)
78. Desgrolard P, Jenkovszky L, Struminsky B *Eur. Phys. J. C* **11** 145 (1999)
79. Petrov V A, Predazzi E, Prokudin A *Eur. Phys. J. C* **28** 525 (2003)
80. Kopeliovich B Z et al. *Phys. Rev. Lett.* **85** 507 (2000)
81. Kopeliovich B Z et al. *Phys. Rev. D* **63** 054001 (2001)
82. Froissart M *Phys. Rev.* **123** 1053 (1961)
83. Martin A *Nuovo Cimento A* **42** 930 (1966)
84. Martin A *Phys. Rev. D* **80** 065013 (2009)
85. Azimov Ya I *Phys. Rev. D* **84** 056012 (2011); arXiv:1204.0984; arXiv:1208.4304
86. Heisenberg W *Z. Phys.* **133** 65 (1952)
87. Cheng H, Wu T T *Phys. Rev. Lett.* **24** 1456 (1970)
88. Cheng H, Wu T T *Phys. Lett. B* **34** 647 (1971)
89. Block M M, Cahn R N *Rev. Mod. Phys.* **57** 563 (1985)
90. Kinoshita T *Phys. Rev. D* **2** 2346 (1970)
91. Chou T T, Yang C N *Phys. Lett. B* **128** 457 (1983)
92. Bourrely C, Soffer J, Wu T T *Eur. Phys. J. C* **28** 97 (2003)
93. Islam M M, Luddy R J, Prokudin A V *Phys. Lett. B* **605** 115 (2005)
94. Islam M M, Luddy R J, Prokudin A V *Int. J. Mod. Phys. A* **21** 1 (2006)
95. Desgrolard P et al. *Eur. Phys. J. C* **16** 499 (2000)
96. Golec-Biernat K, Wüsthoff M *Phys. Rev. D* **60** 114023 (1999)
97. Petrov V A, Prokudin A V *Eur. Phys. J. C* **23** 135 (2002)
98. Donnachie A, Landshoff P V *Nucl. Phys. B* **244** 322 (1984)
99. Block M M et al. *Phys. Rev. D* **60** 054024 (1999)
100. Kovner A, Wiedemann U A *Phys. Rev. D* **66** 034031 (2002)
101. Troshin S M, Tyurin N E *Phys. Lett. B* **707** 558 (2012); arXiv:1111.4454
102. Merino C, Shabelski Yu M, arXiv:1204.0769

103. Chao A W, Yang C N *Phys. Rev. D* **8** 2063 (1973)
104. Amos N A et al. *Phys. Lett. B* **247** 127 (1990)
105. Cheng H, Walker J K, Wu T T *Phys. Rev. D* **9** 749 (1974)
106. Fagundes D A, Menon M J, Silva P V R G, arXiv:1208.3456
107. Abreu P et al. (Pierre Auger Collab.), arXiv:1107.4804
108. Block M M, Halzen F *Phys. Rev. D* **72** 036006 (2005)
109. Block M M, Halzen F *Phys. Rev. D* **73** 054022 (2006)
110. Block M M, Halzen F *Phys. Rev. D* **83** 077901 (2011); arXiv: 1205.5514; arXiv:1208.4086
111. Grau A et al. *Phys. Lett. B* **714** 70 (2012)
112. Okorokov V A, arXiv:0907.0951
113. Campos S D, Okorokov V A *Int. J. Mod. Phys. A* **25** 5333 (2010)
114. Fagundes D A, Menon M J *Nucl. Phys. A* **880** 1 (2012); arXiv: 1208.0510
115. Dremin I M, Nazirov M T *JETP Lett.* **37** 198 (1983) [*Pis'ma Zh. Eksp. Teor. Fiz.* **37** 163 (1983)]
116. Henzi R, Valin P *Phys. Lett. B* **132** 443 (1983)
117. Dremin I M, Nechitailo V A *Phys. Rev. D* **70** 034005 (2004)
118. Dremin I M, Nechitailo V A *Phys. Rev. D* **84** 034026 (2011)
119. Islam M M, Kašpar J, Luddy R J *Mod. Phys. Lett. A* **24** 485 (2009)
120. Jenkovsky L, Lengyel A, Lontkovskyi D *Int. J. Mod. Phys. A* (2012), in press
121. Dremin I M, Chernavskii D S *Sov. Phys. JETP* **11** 167 (1960) [*Zh. Eksp. Teor. Fiz.* **38** 229 (1960)]
122. Amati D, Cini M, Stanghellini A *Nuovo Cimento 10* **30** 193 (1963)
123. Akimov V N et al. *Nucl. Phys. B* **14** 285 (1969)
124. Dremin I M, Dunaevskii A M *Phys. Rep.* **18** 159 (1975)
125. Fukuda H, Iso C *Nuovo Cimento A* **43** 43 (1966)
126. Zalewski K, Van Hove L *Nuovo Cimento A* **46** 806 (1966)
127. Michejda L *Nucl. Phys. B* **4** 113 (1967)
128. Giffon M, Hama Y, Predazzi E *Z. Phys. C* **19** 311 (1983)
129. Sopkovich N J *Nuovo Cimento* **26** 186 (1962)
130. Dar A et al. *Phys. Rev. Lett.* **12** 82 (1964)
131. Durand L (III), Chiu Y T *Phys. Rev. Lett.* **12** 399 (1964)
132. Arnold R C *Phys. Rev.* **136** B1388 (1964)
133. Jackson J D *Rev. Mod. Phys.* **37** 484 (1965)
134. Wu T T, Yang C N *Phys. Rev.* **137** B708 (1965)
135. Byers N, Yang C N *Phys. Rev.* **142** 976 (1966)
136. Chou T T, Yang C N *Phys. Rev.* **170** 1591 (1966)
137. Chou T T, Yang C N *Phys. Rev. Lett.* **20** 1213 (1968)
138. Benecke J et al. *Phys. Rev.* **188** 2159 (1969)
139. Chou T T *Phys. Rev. D* **11** 3145 (1975)
140. Chou T T *Phys. Rev. D* **19** 3327 (1979)
141. Chou T T, Yang C N *Phys. Lett. B* **244** 113 (1990)
142. Krisch A D *Phys. Rev.* **135** B1456 (1964)
143. Islam M M *Nuovo Cimento A* **48** 251 (1967)
144. Cheng H, Wu T T *Phys. Rev. Lett.* **22** 666 (1969)
145. Dias de Deus J *Nucl. Phys. B* **59** 231 (1973)
146. Koba Z, Nielsen H B, Olesen P *Nucl. Phys. B* **40** 317 (1972)
147. Auberson G, Kinoshita T, Martin A *Phys. Rev. D* **3** 3185 (1971)
148. Buras A J, Dias de Deus J *Nucl. Phys. B* **71** 481 (1974)
149. Frahn W E, Venter R H *Ann. Physics* **24** 243 (1963)
150. Frahn W E *Nucl. Phys. A* **302** 267 (1978)
151. Frahn W E *Nucl. Phys. A* **302** 281 (1978)
152. Sprung D W L, Martorell J *J. Phys. A Math. Gen.* **30** 6525 (1997)
153. Sprung D W L, Martorell J *J. Phys. A Math. Gen.* **31** 8973 (1998)
154. Hwa R C *Phys. Rev. D* **22** 759 (1980); *Phys. Rev. D* **22** 1593 (1980)
155. Hwa R C, Zahir M S *Phys. Rev. D* **23** 2539 (1981)
156. Singh V, Roy S M *Phys. Rev. D* **1** 2638 (1970)
157. Uzhinsky V, Galoyan A, arXiv:1111.4984
158. Brogueira P, Dias de Deus J *J. Phys. G Nucl. Part. Phys.* **37** 075006 (2010)
159. Bourrely C, Soffer J, Wu T T *Phys. Rev. D* **19** 3249 (1979)
160. Bourrely C, Soffer J, Wu T T *Phys. Rev. Lett.* **54** 757 (1985)
161. Bourrely C, Soffer J, Wu T T *Eur. Phys. J. C* **71** 1601 (2011)
162. Cudell J-R, Selyugin O V *Phys. Lett. B* **662** 417 (2008)
163. Selyugin O V, Teryaev O V *Phys. Rev. D* **79** 033003 (2009)
164. Selyugin O V, Teryaev O V *Found. Phys.* **40** 1042 (2010)
165. Selyugin O V, arXiv:1205.5867
166. Jenkovszky L L *Riv. Nuovo Cimento* **10** (12) 1 (1987)
167. Kontros J, Kontros K, Lengyel A, hep-ph/0006141
168. Fleming H, Predazzi E *Lett. Nuovo Cimento* **4** 556 (1970)
169. Giffon M, Hama Y, Predazzi E *Z. Phys. C* **25** 129 (1984)
170. Anselm A A, Gribov V N *Phys. Lett. B* **40** 487 (1972)
171. Fiore R et al. *Int. J. Mod. Phys. A* **24** 2551 (2009)
172. Khoze V A, Martin A D, Ryskin M G *Eur. Phys. J. C* **18** 167 (2000)
173. Bertini M et al. *Riv. Nuovo Cimento* **19** 1 (1996)
174. Lukaszuk L, Nicolescu B *Nuovo Cimento Lett.* **8** 405 (1973)
175. Jenkovszky L L, Martynov E S, Struminsky B V *Phys. Lett. B* **249** 535 (1990)
176. Cudell J R, Lengyel A, Martynov E *Phys. Rev. D* **73** 034008 (2006)
177. Martynov E *Phys. Rev. D* **76** 074030 (2007)
178. Donnachie A, Landshoff P V *Phys. Lett. B* **437** 408 (1998)
179. Donnachie A, Landshoff P V, arXiv:1112.2485
180. Goloskokov S V, Kuleshov S P, Selyugin O V *Z. Phys. C* **50** 455 (1991)
181. Donnachie A, Landshoff P V *Phys. Lett. B* **123** 345 (1983)
182. Donnachie A, Landshoff P V *Nucl. Phys. B* **267** 690 (1986)
183. Donnachie A, Landshoff P V *Phys. Lett. B* **185** 403 (1987)
184. Feynman R P, in *High Energy Collisions, Proc. of the Third Intern. Conf., Stony Brook, N.Y.* (Eds C N Yang et al.) (New York: Gordon and Breach, 1969) p. 237
185. Kuraev É A, Lipatov L N, Fadin V S *Sov. Phys. JETP* **45** 199 (1977) [*Zh. Eksp. Teor. Fiz.* **72** 377 (1977)]
186. Balitsky I I, Lipatov L N *Sov. J. Nucl. Phys.* **28** 822 (1978) [*Yad. Fiz.* **28** 1597 (1978)]
187. Gribov L V, Levin E M, Ryskin M G *Phys. Rep.* **100** 1 (1983)
188. Jalilian-Marian J et al. *Nucl. Phys. B* **504** 415 (1997)
189. Balitsky I *Nucl. Phys. B* **463** 99 (1996)
190. Kovchegov Yu V *Phys. Rev. D* **60** 034008 (1999)
191. Kharzeev D, Levin E *Nucl. Phys. B* **578** 351 (2000)
192. Kopeliovich B Z, Potashnikova I K, Povh B, arXiv:1208.5446
193. Cline D, Halzen F, Luthe J *Phys. Rev. Lett.* **31** 491 (1973)
194. Ellis S D, Kislinger M B *Phys. Rev. D* **9** 2027 (1974)
195. Gaisser T K, Halzen F *Phys. Rev. Lett.* **54** 1754 (1985)
196. Afek Y et al. *Phys. Rev. Lett.* **45** 85 (1980)
197. Durand L, Hong P *Phys. Rev. Lett.* **58** 303 (1987)
198. Durand L, Pi H *Phys. Rev. D* **40** 1436 (1989)
199. Margolis B et al. *Phys. Lett. B* **213** 221 (1988)
200. Block M M et al. *Phys. Rev. D* **58** 017503 (1998)
201. Kašpar J et al. *Nucl. Phys. B* **843** 84 (2011)
202. Martin A D et al. *Eur. Phys. J. C* **63** 189 (2009)
203. Ryskin M G, Martin A D, Khoze V A *Eur. Phys. J. C* **71** 1617 (2011); arXiv:1201.6298
204. Martin A D, Khoze V A, Ryskin M G, arXiv:1202.4966
205. Berger E R, Nachtmann O *Eur. Phys. J. C* **7** 459 (1999)
206. Berger E R *Nucl. Phys. B Proc. Suppl.* **74** 96 (1999)
207. Dosch H G, Simonov Yu A *Phys. Lett. B* **205** 339 (1988)
208. Dosch H G, Ferreira E, Krämer A *Phys. Rev. D* **50** 1992 (1994)
209. Bialas A, Bzdak A *Phys. Lett. B* **649** 263 (2007)
210. Bialas A, Bzdak A *Phys. Rev. C* **77** 034908 (2008)
211. Nemes F, Csörgö T, arXiv:1202.2438
212. Carvalho P A S, Menon M J *Phys. Rev. D* **56** 7321 (1997)
213. Ávila R F, Menon M J *Eur. Phys. J. C* **54** 555 (2008)
214. Fagundes D A, Menon M J, Silva G L P *Eur. Phys. J. C* **71** 1637 (2011)
215. Troshin S M, Tyurin N E *Mod. Phys. Lett. A* **27** 1250111 (2012); arXiv:1203.5137
216. Phillips R J N, Barger V *Phys. Lett. B* **46** 412 (1973)
217. Fagundes D A, Menon M J, Silva P V R G, arXiv:1204.5646
218. Cottingham W N, Peierls R F *Phys. Rev. B* **137** B147 (1965)
219. Chou T T, Yang C N *Phys. Rev. Lett.* **20** 1213 (1968)
220. Chou T T, Yang C N *Phys. Rev. Lett.* **46** 764 (1981)
221. Olsen S L *Z. Phys. C* **13** 215 (1982)
222. Dias de Deus J, Kroll P *Acta Phys. Polon. B* **9** 157 (1978)
223. Fischer J, Jakeš P, Novak M *Acta Phys. Polon. B* **14** 807 (1983)
224. Anselm A A, Dyatlov I T *Phys. Lett. B* **24** 479 (1967)
225. Anselm A A, Dyatlov I T *Sov. J. Nucl. Phys.* **6** 430 (1968) [*Yad. Fiz.* **6** 591 (1967)]
226. Khoze V A, Martin A D, Ryskin M G *Eur. Phys. J. C* **18** 167 (2000)
227. Gotsman E, Levin E M, Maor U *Phys. Lett. B* **353** 526 (1995)
228. Andreev I V, Dremin I M, Steinberg D N *Sov. J. Nucl. Phys.* **11** 261 (1970) [*Yad. Fiz.* **11** 468 (1970)]
229. Henyey F S, Tuan R H, Kane G L *Nucl. Phys. B* **70** 445 (1974)
230. Amaldi U, Schubert K R *Nucl. Phys. B* **166** 301 (1980)
231. Henzi R, Valin P *Phys. Lett. B* **48** 119 (1974)

232. Henzi R *Nucl. Phys. B* **104** 52 (1976)
233. Henzi R, Valin P *Nucl. Phys. B* **148** 513 (1979)
234. Henzi R, Margolis B, Valin P *Phys. Rev. Lett.* **32** 1077 (1974)
235. Troshin S M, Tyurin N E *Phys. Lett. B* **707** 558 (2012); arXiv: 1111.4454
236. Landshoff P V, Polkinghorne J C *Phys. Rev.* **181** 1989 (1969)
237. Kroll P *Nucl. Phys. B* **82** 510 (1974)
238. Grein W, Guigas R, Kroll P *Nucl. Phys. B* **89** 93 (1975)
239. Kundrát V, Lokajíek M, Krupa D *Phys. Rev. D* **35** 1719 (1987)
240. Dremin I M *JETP Lett.* **96** 277 (2012) [*Pis'ma Zh. Eksp. Teor. Fiz.* **96** 307 (2012)]
241. Odorico R *Nuovo Cimento A* **54** 96 (1968)
242. Alvarez-Estrada R F *Ann. Physics* **68** 196 (1971)
243. Cahn R *Z. Phys. C* **15** 253 (1982)
244. Martin A *Lett. Nuovo Cimento* **7** 811 (1973)
245. Dremin I M *Nucl. Phys. A* **888** 1 (2012)
246. Dremin I M, Radovskaya A A, arXiv:1209.1935
247. Dias de Deus J, Kroll P *J. Phys. G Nucl. Phys.* **9** L81 (1983)
248. Matveev V A, Muradyan R M, Tavkhelidze A N *Lett. Nuovo Cimento* **7** 719 (1973)
249. Brodsky S J, Farrar G R *Phys. Rev. Lett.* **31** 1153 (1973)
250. Brodsky S J, Farrar G R *Phys. Rev. D* **11** 1309 (1975)
251. Landshoff P V *Phys. Rev. D* **10** 1024 (1974)
252. Mueller A H *Phys. Rep.* **73** 237 (1981)
253. Donnachie A, Landshoff P V *Z. Phys. C* **2** 55 (1979)
254. Donnachie A, Landshoff P V *Z. Phys. C* **61** 139 (1994)
255. Dremin I M, Nazirov M T *Yad. Fiz.* **31** 1606 (1980)
256. Sotiropoulos M G, Sterman G *Nucl. Phys. B* **419** 59 (1994)
257. Sotiropoulos M G, Sterman G *Nucl. Phys. B* **425** 489 (1994)
258. Kwieciǹski J *Phys. Rev. D* **26** 3293 (1982)
259. Kirschner R, Lipatov L N *Z. Phys. C* **45** 477 (1990)
260. Kancheli O V, arXiv:1012.5385

Physics – Uspekhi **56** (1) 29 – 48 (2013)

REVIEWS OF TOPICAL PROBLEMS PACS numbers: 61.44.Fw, **62.25.–g**, 71.45.Lr, 72.20.Fr, 73.30.Mf, 78.70.Gq

High-frequency, 'quantum' and electromechanical effects in quasi-one-dimensional charge density wave conductors

V Ya Pokrovskii, S G Zybtsev, M V Nikitin, I G Gorlova,
V F Nasretdinova, S V Zaitsev-Zotov

DOI: 10.3367/UFNe.0183.201301b.0033

Contents

Abstract. Recent results (some previously unpublished) on the physics of charge density waves (CDWs) are reviewed. The synthesis conditions and unique properties of the quasi-one-dimensional compound NbS_3, with highly coherent room temperature CDWs, are described. A peculiar type of 'quantization' is discussed, which is observed in micro- and nanosamples of $K_{0.3}MoO_3$ and $NbSe_3$ due to the discrete nature of CDW wave vector values. The electric-field-induced torsional strain (TS) in quasi-one-dimensional conductors is considered. Research results on the TS of a noise character induced by sliding CDWs are presented, along with those on the inverse effect, the modulation of the voltage induced by externally driven TS. Results on the nonlinear conduction of TiS_3, a quasi-one-dimensional compound not belonging to the family of classical Peierls conductors, are also described.

V Ya Pokrovskii, S G Zybtsev, M V Nikitin, I G Gorlova, V F Nasretdinova, S V Zaitsev-Zotov Kotel'nikov Institute of Radioengineering and Electronics, Russian Academy of Sciences,
ul. Mokhovaya 11, korpus 7, 125009 Moscow, Russian Federation
Tel. +7 (495) 629 34 59, +7 (495) 629 36 56
Fax +7 (495) 629 36 78
E-mail: pok@cplire.ru

Received 16 December 2011, revised 18 July 2012
Uspekhi Fizicheskikh Nauk **183** (1) 33 – 54 (2013)
DOI: 10.3367/UFNr.0183.201301b.0033
Translated by Yu V Morozov; edited by A Radzig

1. Introduction

This review, based largely on the results obtained by the authors in the last 5–7 years, deals with electron transport and its relationship with the mechanical properties of quasi-one-dimensional charge density wave (CDW) conductors.

A CDW is a periodic self-consistent distortion of the crystal lattice and electron density observed at sufficiently low temperatures in metallic compounds composed of weakly bound conducting chains. There has been undiminishing interest in quasi-one-dimensional CDW conductors during the past 40 years. Although such conductors have not yet found practical application, their diverse and unique properties continue to attract the attention of researchers in different countries.

The history of CDWs dates back to Rudolf Peierls's prediction of instability of the crystal lattice in one-dimensional metals with respect to periodic distortions defined by double Fermi wave vector $2k_F$. Such distortion, known as the Peierls transition, can be described as the superposition of Fermi surfaces corresponding to k_F and $-k_F$. If the surfaces fit together perfectly, they are said to have the property of nesting. The prediction made by Peierls in the pre-war period and published in 1955 [1] was used by Fröhlich in 1954 [2] to propose a new collective mechanism of electron charge transport, and thereby to make an attempt to explain superconductivity. His explanation proved erroneous, but the proposed charge transfer mechanism was shown to actually operate 20 years later when conductors with a highly anisotropic crystalline and electronic structures were synthesized [3–7]. Reference [5] demonstrated an unusually high

conductivity in the microwave frequency range, besides nonlinear conduction, and thus gave direct evidence of the collective character of electron charge transport. Hence, although the basic ideas of the Peierls transition (confirmed by a number of experimental evidences) were formulated before 1975 [6], CDW sliding was first revealed in 1976. The main properties of CDWs are described in reviews [7, 8] and recent progress in this field is discussed in the proceedings of several workshops [9–11].

Thus, a characteristic pattern is observed in quasi-one-dimensional metals: the metal–insulator transition with lattice distortion occurring after a drop in temperature to the Peierls transition point. The period of distortion being defined by the wave vector $q = 2k_F$, a dielectric gap forms at the Fermi level, and the metal turns into a semiconductor. Then, a decrease in energy of the electron subsystem comes into play and this accounts for the observed transition. The lattice distortion is accompanied by modulation of electron density, i.e., CDW formation.

One-dimensionality, indispensable for Fermi surface nesting, is as important as interchain interaction; in other words, the conductors must be *quasi-one-dimensional*. Such interaction stabilizes phase transitions which are impossible in one-dimensional systems [12]. However, the temperature of the Peierls transition is much lower than predicted by the mean-field theory due to one-dimensional fluctuations; therefore, transition to the CDW state should rather be called three-dimensional ordering transition.

A change of CDW phase in time (sliding) results in the transfer of the charge of electrons condensed under the Peierls gap, i.e., all electrons that resided in the conduction band of the metallic state. This is just what is called the Fröhlich mode. CDW sliding is associated with the generation of narrow- and broad-band noises. The narrow-band noise frequency (also termed fundamental or washboard frequency) equals the reverse time of CDW movement by one period. It is directly proportional to the CDW velocity and may be used, in particular, to determine the charge density condensed in a CDW (if the current density is known). The fundamental frequency can also be found in experiments on CDW synchronization by an external high-frequency (HF) field (from the so-called Shapiro steps). Shapiro steps can be observed when dc bias V and alternating HF voltage are simultaneously applied to the sample [7, 8, 13]. When, at a certain value of V, the alternating voltage frequency coincides with the fundamental frequency of a sliding CDW (or one of its harmonics or subharmonics), the $I(V)$ dependence exhibits a voltage range characterized by constant CDW current, referred to as a Shapiro step. Accordingly, the dependence of $R_d \equiv dV/dI$ on V shows a peak. This effect is qualitatively explained as the action of an external HF field making a CDW creep with a given velocity in a certain voltage range (in a sense, frequency locking), due to which its differential resistance increases; it tends to infinity in the case of complete synchronization.

The infinite CDW conductivity predicted in Ref. [2] cannot be observed because of the interaction between CDWs and impurities, which enhances by virtue of the CDWs ability to deformation: the value of q changes with coordinate variation so that the CDW phase is adjusted to the impurity potential. The effect of impurities on CDWs (pinning) is characterized by coherence lengths at which phase gain is 2π due to a random phase interruption.

Pinning makes CDW sliding possible only in such electric fields, the strength of which is higher than the threshold value E_t. In strong electric fields, the CDW conductivity approximates the value that could be expected for the normal (metallic) state. Why conductivity of a CDW cannot exceed the normal conductivity of electrons condensed in it remains to be clarified. The limiting CDW creep velocity ('critical current'), whose overshoot must induce transition from the CDW state to the normal state, is not exactly known either. The experimental data suggesting that the critical current is reached and their possible interpretation are presented in Refs [14, 15], although they are not definitive. A separate problem is the limiting frequency of CDW generation or synchronization that may correspond to a current much lower than the 'critical' one.

The character of high-velocity CDW sliding is a topical issue discussed in this review because Section 2 includes the results of studies using nano-sized NbS_3 samples (phase II) with which record-breaking high CDW synchronization frequencies were obtained.

The properties of small-sized quasi-one-dimensional conductors constitute the subject matter of a special branch of CDW physics. Such samples exhibit a number of finite-size effects arising when certain dimensions of a sample become comparable to or smaller than CDW characteristic lengths. The properties of small samples were reviewed in Refs [16, 17]. One of the most interesting finite-size effects is responsible for the formation of discrete CDW states, i.e., sort of 'quantization'. If the lateral sizes of a sample are smaller than the respective coherence lengths and the sample is short enough, the creation or annihilation of even a single CDW period (~ 10 Å) may lead to experimentally observable effects. Thus, Ref. [18] reports conductivity steps in TaS_3 nanosamples related to a change in the number of CDW periods by ± 1 in single acts of CDW phase slippage. Stepwise variation of the q vector was directly observed in chromium crystals as well, where a two-dimensional spin density wave was generated [19]. We have recently observed such 'quantization' of the CDW vector in nanosamples of $K_{0.3}MoO_3$ (blue bronze) and $NbSe_3$, where CDWs proved to have a high degree of coherence. This effect is studied in Section 3.

CDW strain is closely associated with the existence of metastable states, i.e., nonequilibrium states in which CDW strain persists for a very long time. As known, metastable states in many compounds with CDWs may be created by thermal means using the temperature dependence of the wave vector. For example, a series of discrete metastable CDW states in TaS_3 or blue bronze nanosamples can be obtained at the same temperature based on their different temperature prehistories (see Section 3). Also, metastable states can be created by an electric field. The nature of such structurally nonuniform strain is fairly well known.

Apart from CDW strain, the elastic interaction between CDW, i.e., an electron crystal inside the host lattice, and the crystal itself is observed. Such interaction may be responsible for a decrease in Young's modulus if the CDW depins from impurities in the above-threshold electric fields [20]. It was later elucidated that CDW deformation results in the deformation of a crystal itself [21, 22]. Both uniform longitudinal [21] and nonuniform longitudinal [22] deformations were revealed. The accompanying effects are considered in Section 4, with special reference to torsional strain of quasi-one-dimensional CDW conductors in an electric field. The same section treats the inverse effect, i.e., voltage modulation induced by torsional strain.

Much attention has recently been attracted to compounds in which CDW generation is (possibly) unrelated to the Peierls transition, i.e., nesting [23, 24]. Interestingly, the very existence of CDWs in a such a well-known quasi-two-dimensional compound as $NbSe_2$ was questioned. The problem of a CDW nature may be just as well vital for NbS_3 (phase II) compound, where the CDW arises at $T = 150$ K and low concentration of electrons that depends in addition on the sample used (see Section 2). There are reports suggesting the existence of collective states (possibly CDWs) in the layered quasi-one-dimensional compound TiS_3, where electron concentration at room temperature is estimated to be 10^{18} cm^{-3}. If it is assumed that the lattice distortion in TiS_3 is described by usual nesting, the CDW period proves to be improbably long. The results of TiS_3 studies are considered in Section 5.

To recall, this review is not designed to comprehensively cover all research topics pertaining to quasi-one-dimensional CDW conductors, but focuses on the results recently obtained by the authors. Nevertheless, it contains a wealth of CDW physics data reported for the last 5–7 years. The study subjects included in the review are well-known compounds, such as TaS_3, $NbSe_3$, $K_{0.3}MoO_3$, $(TaSe_4)_2I$, rare NbS_3 (phase II) synthesized by the authors, and TiS_3 for which mechanisms of conductivity remain to be elucidated. Some new physical results were obtained in the course of this work, including certain formerly inaccessible values of selected parameters.

2. Synthesis of NbS_3 whiskers and the dynamic properties of a charge density wave in ultimately thin NbS_3 samples

To begin with, we shall consider the results of NbS_3 (phase II) research [7].

The quasi-one-dimensional NbS_3 compound is known to exist in two phases. Phase I is characterized by high resistance and some semiconducting properties; the crystal lattice is dimerized along the Nb chains. Phase II is remarkable for its CDWs incommensurate at room and higher temperatures; in this compound, CDWs can slide in an electric field. Apart from the high temperature of the Peierls transition (330–370 K), distinctive features of this compound include the strong anisotropy of its properties (as apparent from very weak coupling between the chains) and combination of high-frequency properties with high coherence of CDWs [25–27]. This compound was synthesized earlier by different authors [28], but they had problems with finding phase II whiskers (filamentary crystals), because they were outnumbered by phase I crystals in the material grown. The conditions for reproducible synthesis of phase II crystals have been described in the most recent publications [25–27].

CDW coherence in NbS_3 samples, combined with their high-frequency properties, implies the possibility of high-speed CDW sliding. We reached record-breaking synchronization frequencies with thin NbS_3 whiskers, i.e., extremely high velocities of CDW motion. The maximum current density in most compounds is limited by heating of the samples, while the observation of the generation or synchronization of HF oscillations is hampered by still lower currents. Heating effects can be weakened by selecting the thinnest samples and placing them on a substrate with high heat conduction. However, sample thinning is restricted by the size effect, due to which the threshold voltage V_t initially rises resulting in a decrease in CDW current at the same voltage, and then the CDW state undergoes transition to another dielectric state [16]. It was shown that phase II NbS_3 is an ideal compound for studying synchronization at ultimately high frequencies. Suffice it to say that a CDW retains its characteristic properties at least in samples as thin as 20 nm, due to its marked structural anisotropy. It needs to be emphasized that we conducted most CDW synchronization experiments at room temperature and did not use a cryostat; this facilitated matching the samples and microwave generators.

2.1 Synthesis

The determination of conditions for the reproducible growth of phase II NbS_3 whiskers made possible systematic studies of the properties of this compound, many of which proved unique. Therefore, we start by describing the conditions for phase II NbS_3 synthesis.

Three experiments on NbS_3 synthesis from a gaseous phase at the optimal temperature and its gradient were performed. The whiskers were grown in a three-zone Gero furnace (Germany) with a Modcon temperature control device. The temperature was set separately for each zone and, for instance, the appropriate values were 750, 700, and 650 °C to obtain the distribution shown by the bottom curve in Fig. 1.

An ampule was rinsed with an NaOH solution, then with distilled water, dried in a furnace at 800–900 °C for 48 hours, filled with reactants, and sealed under continuous pumping by a dry pump. The dried ampule was exposed to air for less than one hour. After the ampule was placed in the furnace, the temperature was gradually raised from 200 °C to the desired value at a rate of 20 °C/hr and remained constant thereafter during the entire period of whisker growth (usually 10 days), Variations of the growth period had no influence on the phase I to phase II ratio in finished NbS_3.

The mass of reacting Nb and S components remained unaltered in all experiments. These reactants were utilized in stoichiometric proportion with a small (10%) excess of sulfur on the base of 0.08 g of the end product per cm^3 of the ampule.

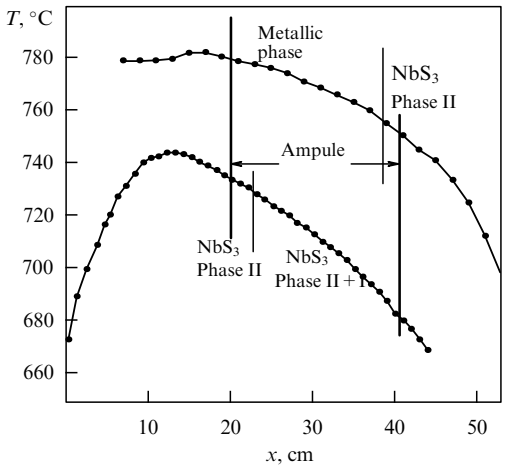


Figure 1. Two variants of temperature distribution over the furnace length (rough measurements in the absence of the ampule). Arrows indicate the positions of the ampule ends. Regions in which each of the two NbS_3 phases grows are shown along with the region where whiskers exhibiting metallic properties grow.

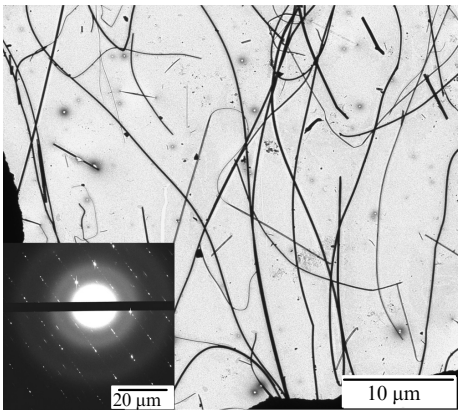


Figure 2. TEM images of NbS_3 whiskers. Inset: an electron diffraction pattern of one of the samples.

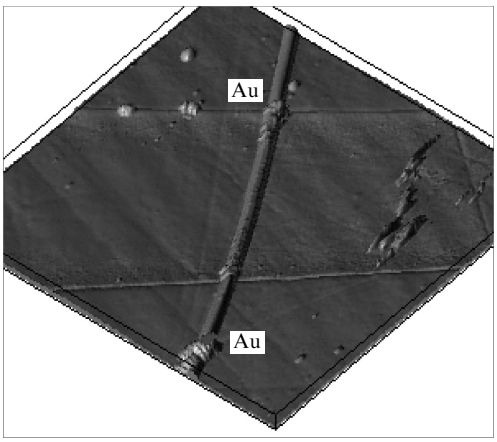


Figure 3. Atomic-force microscopy (AFM) image of an NbS_3 whisker (sample No. 11 from Table 1 in Section 2.4). The scan area is 13.56× 13.56 μm^2. Boundaries of two gold contacts made by the laser deposition method are seen in the image.

The sealing/unsealing procedures, as well as putting the ampule in the furnace, were the same in all experiments.

Two growth cycles were carried out with a temperature gradient of 50 °C for the ampule length (20 cm) (the bottom curve in Fig. 1). The temperature was below 720 °C in one cycle, and above 720 °C in the other. An additional cycle was conducted with a temperature gradient of 25 °C (the top curve in Fig. 1). A few repeated cycles confirmed the reproducibility of the results obtained.

The growth completed, the ampule was opened to remove the 'cocoon' (whisker wool). The transport properties of the whiskers were studied utilizing at least 10 filaments taken at selected points inside the cocoon, viz. at different distances from the cold end of the ampule. It this way, data were gathered on the physical properties of NbS_3 crystals, depending on temperature and its gradient. Figure 1 shows that the optimal temperature for phase II synthesis ranges $\approx 720-750$ °C.

The images of whiskers in a transmission electron microscope (TEM) are presented in Fig. 2; the inset shows the electron diffraction pattern for one of the whiskers at room temperature. Suprastructural reflexes resulting from the creation of two CDWs [26–28] can be seen. Bridge structures (Fig. 3) were made by laser-induced vacuum gold deposition through a micromask.

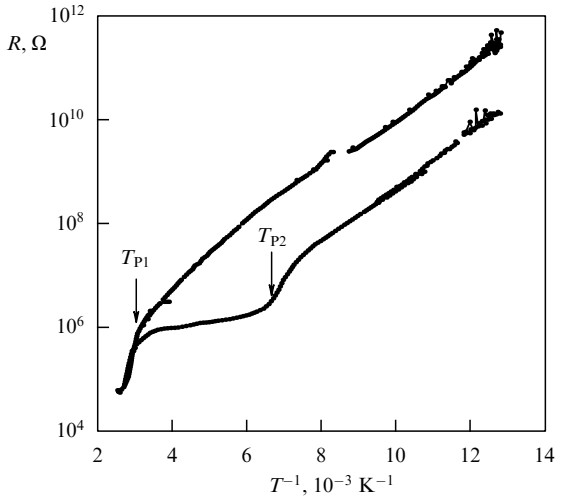


Figure 4. Temperature dependences of resistance for low-resistance (lower curve) and high-resistance (upper curve) NbS_3 whiskers. T_{P1} and T_{P2} are temperatures of Peierls transitions. The samples are 100 μm × 0.13 μm^2 and 90 μm × 0.2 μm^2 in area, respectively. The activation energy for both whiskers is ≈ 1200 K.

2.2 Properties of synthesized samples: nonlinear conduction at room temperature and temperature dependence of conductivity

Detailed measurements of phase II samples permitted distinguishing two groups of them. The samples synthesized at 670–700 °C have lower resistivity; their temperature dependences of resistance, $R(T)$, clearly exhibit two transitions: at $T \approx 360$ K, and $T \approx 150$ K. At synthesis temperatures 715–740 °C, the high-resistance (at room temperature) samples were grown with resistivity by almost an order of magnitude greater; their $R(T)$ dependences had no peculiarities near 150 K, as shown in Fig. 4 presenting characteristic $R(T)$ curves for low- and high-resistance samples. The tendency toward the disappearance of the peculiarity at $T = 150$ K is apparent in thinner samples, too.

The high-resistance samples at room temperature are characterized by a high (up to 80%) degree of CDW synchronization under microwave irradiation (Fig. 5). This suggests a high degree of CDW coherence in the crystal bulk and, therefore, their high quality. Low-resistance samples also exhibited Shapiro steps, but the relative decrease in their differential conductivity (degree of synchronization) was smaller.

At synthesis temperatures below 720 °C and a temperature gradient of 50 °C for 20 cm (bottom curve in Fig. 1), the output of phase 1 (semiconducting) samples shows an outstripping growth rate. Whiskers synthesized at temperatures $T > 760$ °C possess metallic properties (Fig. 6). Their resistivity is at least one order of magnitude lower than that of low-resistance phase II samples. Some dielectrization was observed only at temperatures below 150 K (see Fig. 6), which may be ascribed to 'residues' of the lower transition. There is a well-apparent tendency toward a decrease in lateral sizes of samples with increasing synthesis temperature.

Investigation on the crystals grown was undertaken by transmission electron microscopy at room temperature and confirmed that they fit into phase II samples, as also evidenced by the observation of two Peierls transitions near 365 and 150 K. At the same time, the intersample variation of properties gives evidence of a certain 'hidden' parameter that

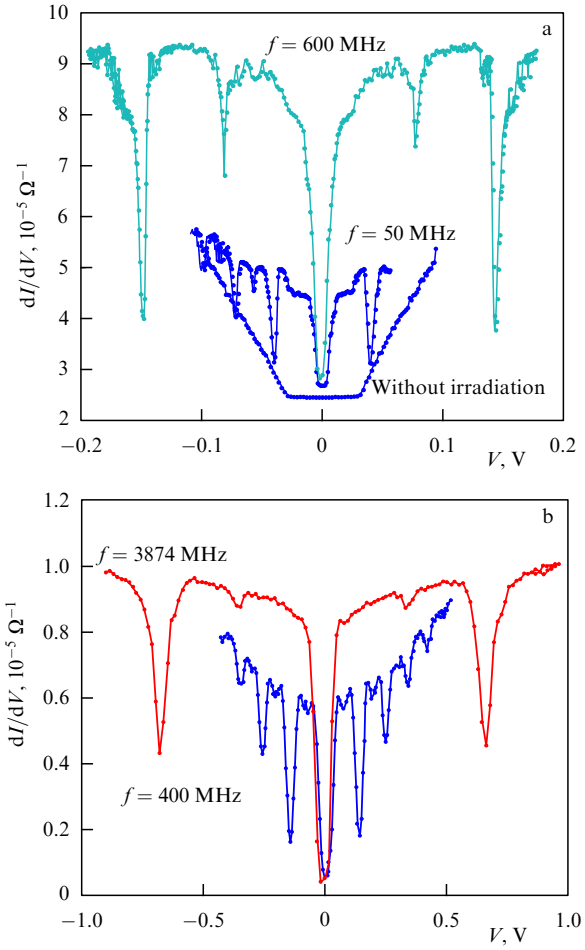


Figure 5. Dependences of differential conductivity of high-resistance NbS_3 samples at room temperature: (a) irradiation at 50 and 600 MHz, and in the absence of irradiation, sample area 7.5 μm × (8×10^{-2}) μm^2; (b) irradiation at 3874 and 400 MHz, sample area 6.5 μm × (7.6×10^{-3}) μm^2.

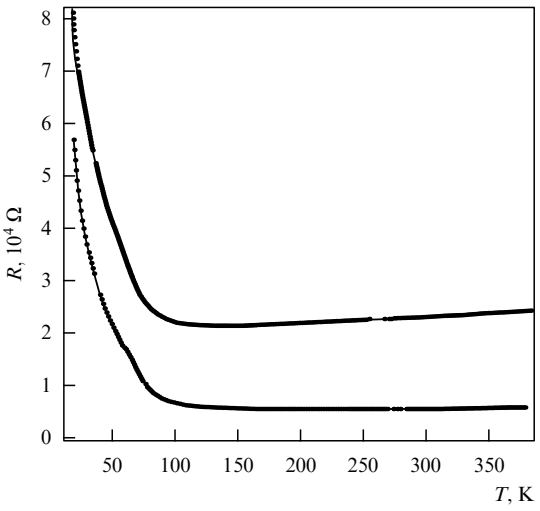


Figure 6. Temperature dependence of resistance for two samples grown at temperatures above 760 °C.

also varies depending on synthesis temperature. For example, a rise in temperature resulted in a marked jump of resistance at the upper transition and its fall at the low-temperature one. In doing so, resistivity at room temperature could show an

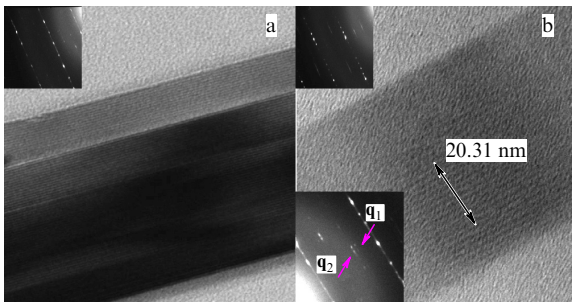


Figure 7. High-resolution electron micrographs of NbS_3 whiskers synthesized at (a) 670–700 °C, atomic layers spaced 9.9 Å apart; (b) 715–740 °C, atomic layers spaced 19.8 Å apart. Scale mark: 10 atomic layers (image scales are identical). Insets: microdiffraction images along direction [001]. Two superstructures are seen: $\mathbf{q}_1 = (1/2a^*, 0.297b^*, 0)$, and $\mathbf{q}_2 = (1/2a^*, 0.353b^*, 0)$.

order of magnitude rise simultaneously with a similar increase in the current–voltage characteristic (CVC) nonlinearity, i.e., a decrease in resistance at high voltages.

The results of structural research suggest that the 'hidden parameter' arises from period doubling along the crystallographic direction a in crystals grown at higher temperatures. Figure 7 presents high-resolution TEM images of samples from different groups. Both images demonstrate the structure of atomic layers extended along the b-axis. In a (low-resistance) sample grown at 670–700 °C (Fig. 7a), the period of this structure along the a-axis equals roughly 10 Å, whereas it is twice that in a (high-resistance) sample grown at 715–740 °C (Fig. 7b). A similar picture emerges from the electron diffraction patterns in the insets to Fig. 7: the period of diffraction pointing along direction a in Fig. 7a is twice that in Fig. 7b, which suggests the formation of a superlattice with the doubling of the period along the a-axis in high-resistance samples.

2.3 Nonlinear conduction at low temperatures

Nonlinear conduction of NbS_3 whiskers was investigated in a wide temperature range. Voltage dependences of differential resistance confirmed the formation of two CDWs in low-resistance samples: there are two temperature ranges in which the dependences take a threshold form [25]. The minimum threshold field E_t is around 2 V cm^{-1} for the upper CDW, and somewhat smaller for the lower one. The highly anisotropic structure of NbS_3 manifests itself in a short CDW coherence length across the chains, as apparent from the weak dependence of E_t on transverse dimensions of samples. Nevertheless, E_t exceeds 100 V cm^{-1} in the thinnest samples, in accordance with the finite-size effect known for other compounds with CDWs [16, 17].

One more peculiarity observed in some phase II samples is worthy of note. It is a sharp increase of resistance in response to a temperature drop in the $T < T_{P2}$ range. This rise may be characterized by an activation energy roughly equal to 2000 K (Fig. 8). Nonlinear conduction at temperatures $T < 150$ K has a threshold form (Fig. 9a) for the majority of the selected samples [25]. However, certain $R_d(V)$ dependences exhibit a sharp peak at $V = 0$ (Fig. 9b). In this case, near-maximum dependences can be roughly described by a power law. Dielectrization upon a drop in temperature and similar $R_d(V)$ dependences were observed earlier for very thin samples of the quasi-one-dimensional

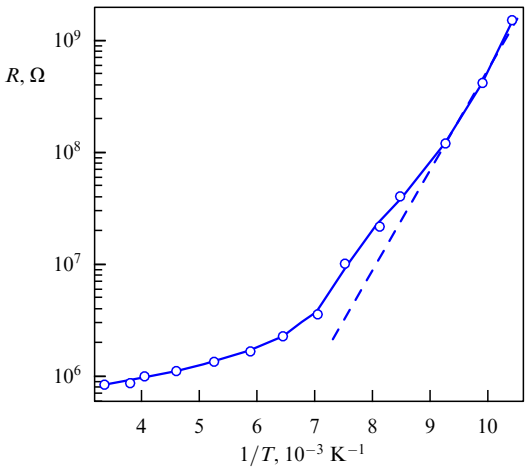


Figure 8. $1/T$ dependence of resistance for NbS_3 samples grown at 670–700 °C. All points are obtained from current–voltage characteristics (CVCs) shown in Fig. 9b. The activation energy at low temperatures is 2000 K (dashed line).

conductor $NbSe_3$ [29]: the peak at $V = 0$ seems to be superimposed on the usual threshold $R_d(V)$ dependence. The authors of Ref. [29] emphasize that such a peculiarity is characteristic of one-dimensional systems, but they failed to unambiguously identify its origin. It should be noted that similar dependences were also observed in $NbSe_3$ samples with artificially introduced defects [30, 31]. Therefore, the most plausible explanation for dielectrization of thin samples reduces to their defective structure. It is also worth mentioning a recent work [32] where power-law (pseudo-power-law, to be precise) dependences $R(V)$ and $R(T)$ in quasi-one-dimensional conductors, including $NbSe_3$, are explained by hopping conduction. This leads to the conclusion that $NbSe_3$ samples with temperature dependences of resistance similar to those presented in Figs 8 and 9b have more structural defects.

We shall turn back to the nature of such CVCs in Section 5, where the compound of a somewhat different type, TiS_3, is considered.

2.4 Limiting currents and synchronization frequencies of a charge density wave

The motion of CDWs in NbS_3 is confirmed by the observation of Shapiro steps (Fig. 5) for both upper and lower CDWs [25]. We prepared samples of the minimally possible small lateral sizes in order to obtain limiting currents (therefore, synchronization frequencies). Cross sections of about 10 such samples were first measured with an atomic force microscope. The study of Shapiro steps in *upper* CDWs of these samples showed that the ratio of CDW current density — corresponding to the main step — to the irradiation frequency was roughly equal for all samples within the scatter of the measurement (i.e., accuracy of cross section measurements) (see Table 1). This means that the cross section area of the samples can be found from the Shapiro steps, i.e., from the ratio of CDW current I_{CDW} to the irradiation frequency [29]. We used this approach in our further work.

The cross section area of the thinnest samples was about 5×10^{-4} μm^2, while the contacts were spaced ≈ 1 μm apart. Shapiro steps were identified in such samples at frequencies higher than 10 GHz. The limiting currents were constrained by the Joule heating of the samples, and limiting frequencies by the microwave power that was possible to feed into them. Better correlation was achieved between short samples and microwave radiation. Figure 10 shows typical $R_d(I)$ dependences at room temperature for the case of microwave irradiation (irradiation frequencies are indicated in the figure). The Shapiro steps, namely R_d maxima corresponding to the most complete synchronization of CDWs, are very apparent on the curves. The dependences of the CDW current

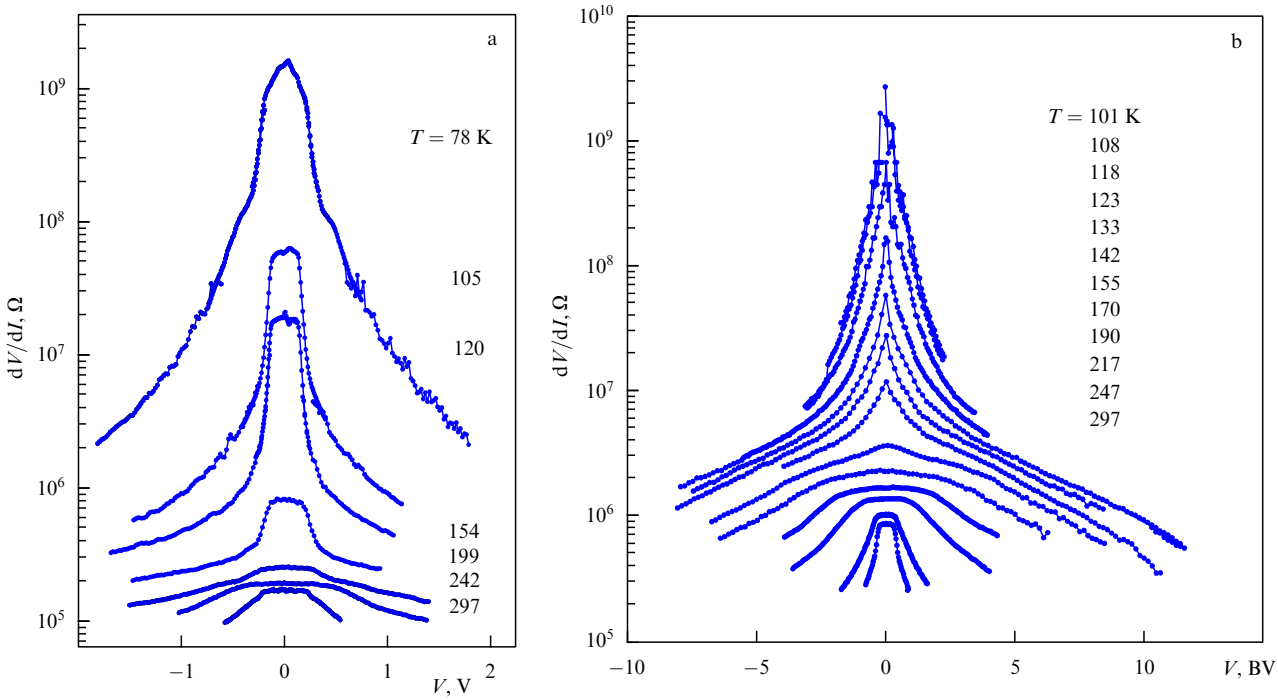


Figure 9. Voltage dependences of differential resistance for two samples grown at $T \approx 715$ °C: (a) low-resistance sample (50 μm × 0.08 μm^2) with two Peierls transitions, and (b) anomalous sample 60 μm × 0.07 μm^2 in size (see Fig. 8).

Table 1. Sample dimensions (width $w \times$ thickness $t \times$ length L) measured by AFM or an optical microscope; temperature T at which synchronization was studied; quantity $I_{CDW}/(2ef_{ex})$, i.e., the number of CDW chains in the sample; area s_0 per CDW chain, and number N of CDW chains per elementary cell, i.e., over an area of 180 Å^2 [7].

Sample	$w \times t \times L$, nm^2 × μm	T, K	$I_{CDW}/(2ef_{ex})$	s_0, per chain, Å^2	N, per elementary cell
1	140 × 50 × 3.3	295	2.1623×10^3 (< 1 GHz)	322	0.56
2	120 × 50 × 3.7	326	3.84×10^3	156	1.15
3	150 × 20 × 3.2	295	1.95×10^3 1.2806×10^3 (4 GHz)	154 234	1.2 0.78
4	750 × 23 × 3.5	295	10^4	171	1.06
5	160 × 40 × 5.5	295	1875	341	0.54
6	470 × 25 × 42	330 101	2250 1844	522 637	0.35 0.29
7	5000 × 1200	337	2×10^6	300	0.61
8	1000 × 400 × 500	337 131	1.64×10^5 1.05×10^4	244 3.8×10^3	0.75 0.0478
9	90 × 28 × 3	295	1367	184	1
10	44 × 14 × 1	295	304	202	0.9
11	425 × 180 × 7.3	295	3.98×10^4	191	0.94
12	430 × 26 × 8	295	6.585×10^3	169.8	1.06

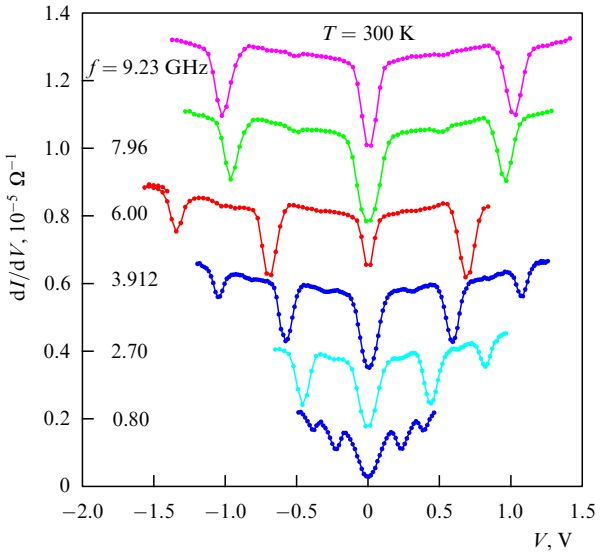


Figure 10. Current dependences of differential resistance for an NbS_3 nanosample 4 μm × (1.5×10^{-3}) μm^2 in size at different irradiation frequencies and room temperature.

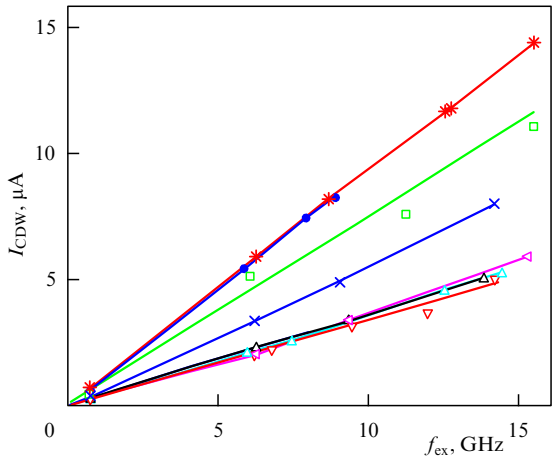


Figure 11. Dependences of CDW current, corresponding to the first Shapiro step, on the irradiation frequency for eight NbS_3 samples at room temperature.

I_{CDW}, i.e., nonlinear current at the main step, on the irradiation frequency f_{ex} are practically linear at frequencies below $f_{ex} = 16$ GHz (Fig. 11).

It is possible to determine the number of chains carrying CDW current (column 4 in Table 1) on the assumption that charge density transferred by a CDW equals $2e$ per chain for a single period, given the known I_{CDW}/f_{ex} ratio. If the sample's cross section area $s = wt$ is measured (column 2), the area s_0 per CDW chain (column 5) is known, too. According to Ref. [7], the area of an elementary cell in NbS_3 is 180 Å^2, with each cell encompassing eight non-equivalent niobium chains. This means that only one of the eight chains in any sample contributes to the upper CDW conduction (i.e., to the Fröhlich mode) (see column 6 in the Table 1).

The 'fundamental ratio' $I_{CDW}/(sf_{ex})$ for the upper CDW is roughly 18 A MHz^{-1} cm^{-2}, or much lower than for the known quasi-one-dimensional compounds. In particular, it is 25 A MHz^{-1} cm^{-2} for both upper and lower CDWs in $NbSe_3$, and 38 A MHz^{-1} cm^{-2} in NbS_3 [33]. The fundamental ratio for the lower CDW in NbS_3 is still smaller [25–27] and varies from sample to sample (see the table). This means that either only part of the elementary cells contribute to CDW conductivity or the charge transferred by the lower CDW is significantly smaller than $2e$ per period. Thus, the nature of the Peierls state at temperatures below 150 K remains unclear to a great extent (see also Ref. [27]).

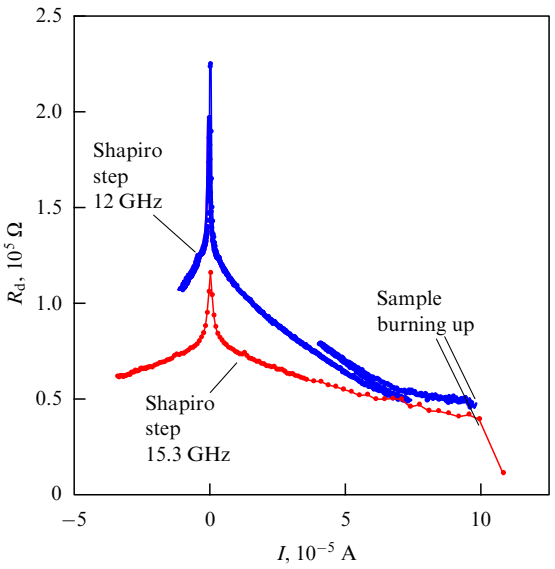


Figure 12. Current dependences of differential resistance for two NbS_3 nanosamples in the case of limiting currents. Sample sizes: 1 μm × (2×10^{-3}) μm^2 (upper curve), and 1 μm × (8.3×10^{-4}) μm^2 (lower curve). The sample undergoes degradation (its resistance grows) in the current range $(5-10) \times 10^{-5}$ A and thereafter burns up.

As far as the upper CDW is concerned, it should be noted that 16 GHz is the highest frequency on record at which CDW synchronization has ever been observed. It corresponds to the sliding CDW velocity on the order of 10 m s^{-1} in the absence of signs of approaching a limiting value associated with either CDW breaking (achievement of critical current) or the loss of its coherence. As mentioned in the Introduction, the indications of achieving the CDW critical current in TaS_3, $NbSe_3$, and blue bronze were reported in Refs [14, 15], but direct evidence of CDW breaking at high sliding velocities is still lacking. According to Refs [14, 15], the critical current corresponds to the fundamental frequency $f_0 \sim 10^{13}$ Hz. Thus far, we have not been able to match nanosamples to microwave radiation in excess of 16 GHz, i.e., feed into them a power sufficient for CDW synchronization. Nevertheless, we have estimated critical currents that can be passed through NbS_3 bridges. To this end, we have measured $R_d(I)$ dependences for a few samples, increasing the current till the sample burnt up or its degradation began (resistance increased).

Figure 12 presents such dependences for two samples. Shapiro steps were observed in them at the limiting irradiation frequencies of 12 and 15.3 GHz (see the corresponding peaks of $R_d(I)$ in the figure). It is seen from these dependences that the maximum current density is $\approx 6 \times 10^6$ A cm^{-2}, the corresponding CDW velocity ≈ 200 m s^{-1}, and the fundamental frequency ≈ 200 GHz. The last value may be regarded as an estimate of the maximum frequency that the samples can receive or generate. These parameters can possibly be enhanced by selecting still thinner samples, i.e., thinner than 10 nm. However, other experimental procedures are needed for this purpose.

The question is, what peculiarities of NbS_3 made it possible to obtain record-breaking CDW synchronization frequencies? Highest current densities on record are due to the aforementioned causes, such as strong structural anisotropy, allowing samples to be thinned to at least 20 nm without suppressing the CDW. Given that such samples fit tightly to the substrate, they can tolerate a current density of up to $\approx 6 \times 10^6$ A cm^{-2} (Fig. 12). Another peculiarity of NbS_3 is the low concentration of conducting chains (one of the eight) responsible for low bulk density of heat release. For a CDW to be synchronized by a microwave field, it must show 'high-frequency properties', besides high sliding velocity. A rigid CDW model and a strongly damped oscillator approximation [34, 35] permit associating such CDW properties of NbS_3 with high threshold fields (≈ 100 V cm^{-1} for nanosamples) and weak CDW friction $(1/\tau)$. In this model, the pinning frequency, i.e., the maximum frequency of CDW response to an alternating voltage, is estimated at $E_t e\tau/(\lambda m^*)$ [34, 35].

2.5 Unique properties of charge density waves in NbS_3

A unique property of NbS_3 phase II is the high temperature of transitions. As a matter of fact, the NbS_3 samples grown exhibit three Peierls transitions. TEM studies at room temperature demonstrated the existence of two superlattices (CDWs) with components along the b^* direction: $q_1 = 0.297b^*$, and $q_2 = 0.352b^*$ [28] (the data reported in Ref. [26] agree with those in paper [28]). Satellite reflexes corresponding to q_1 weaken appreciably for $T > T_{P1}$, i.e., starting from temperature 360 K [28]. As the temperature approaches T_{P1}, the threshold nonlinear conduction observed at near-room temperature disappears, too [25], thus meaning that this conduction is due to CDW sliding with q_1. On the other hand, one more CDW remains for $T > T_{P1}$. The study of its properties encounters difficulties because NbS_3 properties begin to degrade rapidly for $T > 400$ K. Nevertheless, fast measurements in the argon atmosphere allowed the temperature of the highest transition, T_{P0}, to be estimated at approximately 620 K. Threshold nonlinear conduction was also observed at temperatures close to 475 K.

The appearance of the third CDW with $T_{P2} = 150$ K is beyond doubt, but the lack of low-temperature diffraction research does not allow us to determine the magnitude of its q vector. Such studies are pending. They are of special interest because charge density carried by the lower CDW appears to be abnormally small and varies from sample to sample.

3. 'Quantization' of charge density wave states in thin samples of blue bronze and $NbSe_3$

Samples of quasi-one-dimensional conductors of submicron and nanometer sizes also possess unique properties in *below*-threshold fields, i.e., when the CDW is at rest. As mentioned in the Introduction, such samples exhibit conduction steps associated with single phase-slip (PS) events, i.e., a change in the number of CDW periods N by ± 1. If contacts create here stringent boundary conditions for the CDW phase, N must be an integer and quantity $q = 2\pi N/L$ must take discrete values. Such 'quantization' of q occurs at the intersection of quantum and classical physics: a CDW comprises a quantum condensate, and its period equals half the de Broglie wavelength of an electron at the Fermi level, $\lambda = \pi/k_F$. At the same time, discrete states of CDWs are similar to the states of a classical wave in a resonator.

In many quasi-one-dimensional conductors, q depends on temperature. Therefore, the CDW period can be altered by changing T. Because the phase slip of CDW requires that a certain energy barrier arising from the necessity of a local CDW break be overcome, $q(T)$ dependences display a hysteretic character (although only indirect evidence of q

hysteresis has so far been available, based on $R(T)$ dependences). This means that states with different q values can be obtained at the same temperature and, vice versa, states with $q = \text{const}$ can be studied in a certain temperature range.

To date, discrete CDW states have not been observable directly, i.e., using electron diffraction methods. (The sole exception is Ref. [19] reporting an X-ray study of thin single-crystal chromium films that yielded discrete q values of spin density waves, depending on the film thickness.) However, the 'quantization' of q is possible to observe, studying T-dependences of σ in micro- and nanosamples. In fields lower than E_t, conductivity is determined by quasiparticles excited over the Peierls gap. Each phase slip changes the charge of CDW and, therefore, the charge of the quasiparticles; moreover, it causes the chemical potential to shift [18]. Thus, jumplike changes of q must be accompanied by σ jumps.

Just such jumps were observed in a study of submicron TaS_3 whiskers [18]. Numerical calculations showed that the magnitude $\delta\sigma$ of jumps corresponds, *on the average*, to the creation and annihilation of a CDW period per chain. Repeated measurements of $\sigma(T)$ confirmed the possibility of observing reproducible discrete states in some cases. However, regular temperature switchovers between the states were never observed. Moreover, the jumps varied in magnitude. There were occasionally much smaller (as if fractional) steps among the majority of roughly equally high ones. Evidently, these deviations from the regular structure of the states are due to incomplete coherence of CDWs both along and across the sample, and to the absence of stringent boundary conditions at the contacts [36].

We conducted a similar study of thin $K_{0.3}MoO_3$ and $NbSe_3$ samples, in which conductivity jumps between discrete states corresponding to 'quantized' values of the wave vector were observed. An important difference from Ref. [18] was the method of contact manufacture, besides the selection and preparation of samples with a perfect structure. The contacts were made by the laser gold deposition technique, ensuring penetration of high-energy gold ions into the material to a depth of several angstroms and induction of radiation defects even farther from the surface. Evidently, such defects sharply enhance the value of E_t and, thereby, prevent the spread of PS-induced CDW strain outside the contact region. This means that the contacts create stringent boundary conditions for CDWs. We also note the low resistivity of such contacts (some 10^{-9} Ω cm^{-2}) and the 'metallic' behavior of their resistivity, suggesting the absence of a contact energy barrier [37].

We selected a few tenths of a micrometer-thick transparent lamellae of blue bronze extended along the b-axis, i.e., along the metal chains. The samples were a few dozen micrometers in length, and several-fold smaller in width [36]. The $NbSe_3$ whiskers were a few tenths of a micrometer in thickness and several dozen micrometers in length. Blue bronze and $NbSe_3$ samples with such dimensions retain the CDW properties characteristic of a bulk material. Detailed results of blue bronze studies are reported in papers [36, 38], and those of $NbSe_3$ in papers [37, 38].

Figure 13 depicts segments of temperature dependences of conductivity $\sigma(T)$ for three samples of blue bronze. The most clear picture of the states and changeover between them was obtained utilizing a $21 \times 5 \times 0.3$-μm^3 sample; it is depicted in the center of the figure showing the hysteresis loop typical of blue bronze. Both branches of the loop exhibit regular temperature jumps of conductivity roughly equal in height.

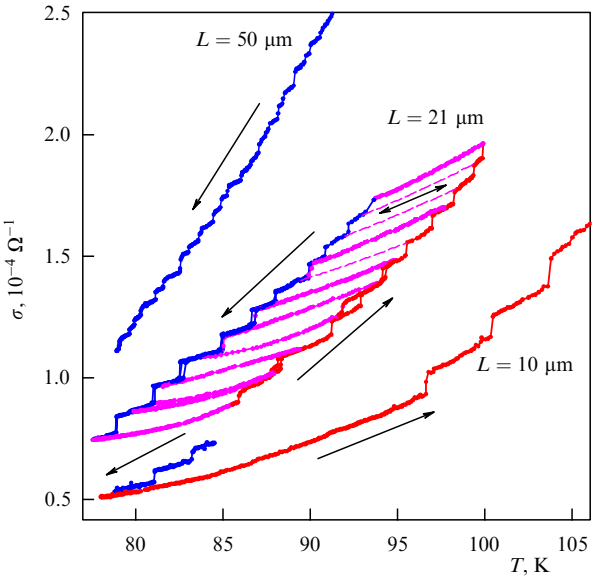


Figure 13. Segments of temperature dependences for three samples of blue bronze. The sample sizes are $21 \times 5 \times 0.3$ μm^3, $50 \times 7 \times 0.3$ μm^3 (conductivity is magnified 1.5 times), and 10 μm × 2 μm (the thickness was not measured, conductivity is divided by 1.5). Arrows indicate temperature scan direction. Dashed lines show omitted reversible segments of the dependence.

The magnitude of the jumps and its changes under the effect of varying temperature are quantitatively described by a semiconductor model. The segments of $\sigma(T)$ connecting the upper and lower branches were obtained by altering the scan direction of T. The $\sigma(T)$ dependences in these curves are reversible, and there are no 'allowed' states between them.

Because each jump corresponds to a $\pm 2\pi/L$ change in q, counting the number of jumps in a certain T range, it is possible to find how the q vector changes within this interval. Specifically, the change in q for $T < 130$ K was observed by this method in Ref. [36]. This observation is consistent with the result of X-ray studies [39], but the responsiveness of q to changes in our case proved much higher ($\delta q/q \sim 10^{-4}$). It was finally found that q continues to change even at temperatures below 70 K. Moreover, a hysteresis of $\delta q(T)$ dependence was revealed.

Because the concentration of electrons or holes (depending on which carriers play the key role) in each chain changes by $\pm 2/L$ in each jump, knowing the $\delta\sigma$ value: the mobility μ of the main carriers can be deduced

$$\mu = \frac{\delta\sigma L^2 s_0}{2es},$$

where s_0 is the area per CDW chain, and s is the sample cross section. The value of μ for blue bronze (in the present case, electron mobility) estimated in this way proved to be 10 cm^2 V^{-1} s^{-1}. This value practically coincides with Hall electron mobility of 13 cm^2 V^{-1} s^{-1} [40].

If the mobility is easy to determine in the case of blue bronze by measuring the Hall effect, it is much more difficult to do for semiconductors possessing a more complicated band structure, where different types of carriers make comparable contributions to conductivity. The Peierls' conductors (semimetals) with a complex structure are exemplified by $NbSe_3$. This compound has three types of Nb

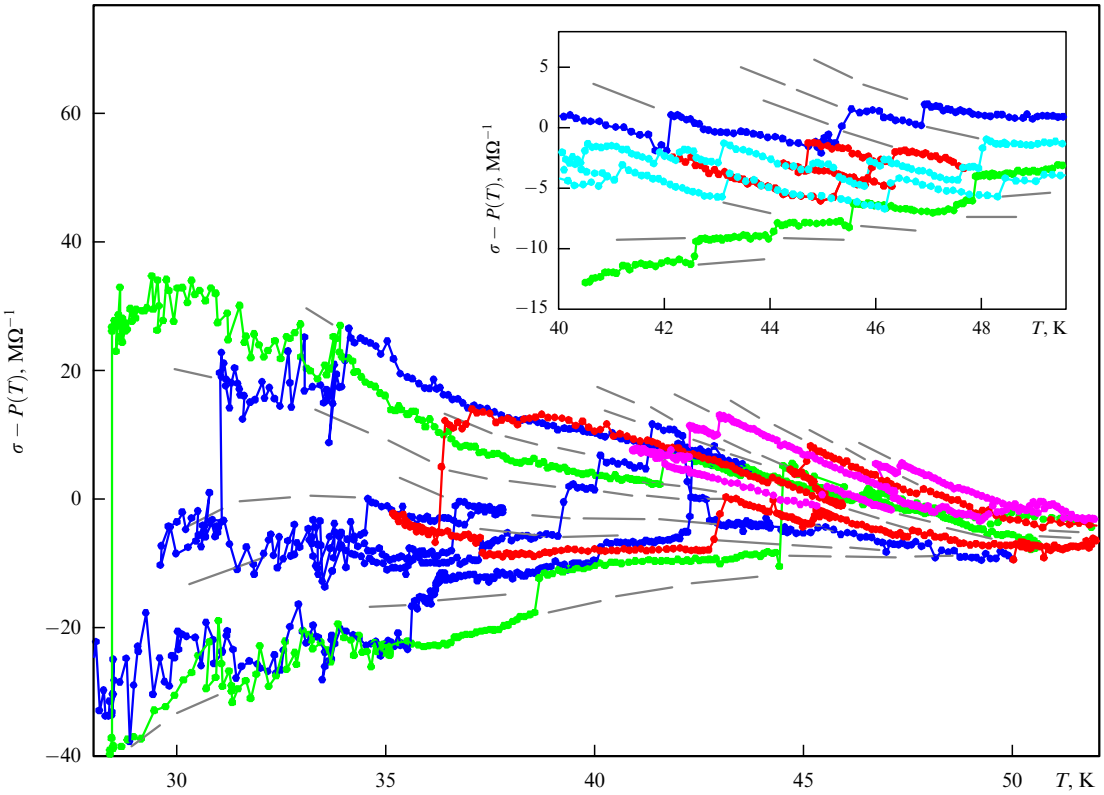


Figure 14. The $\sigma(T)$ dependence obtained after subtraction of polynomial approximation for an $NbSe_3$ sample of length $L = 33$ μm ($R_{300} = 3.2$ kΩ). Dashed lines roughly show omitted reversible fragments of the curves. Inset: a similar dependence for another sample ($L = 35$ μm, $R_{300} = 5.1$ kΩ) in a narrow temperature range.

chains in its structure, each associated with a specific quasiparticle carrier. Investigation of conductivity jumps due to CDW phase slips gives a unique possibility of determining the mobility of exactly those carriers that are associated with a given CDW.

The $NbSe_3$ compound gives evidence of two Peierls transitions: at $T_{P1} = 144$ K, and $T_{P2} = 59$ K. Incomplete dielectrization of the electron spectrum is responsible for a sharp rise in $NbSe_3$ conductivity with decreasing T. Therefore, we approximated the measured $\sigma(T)$ curves for each sample with a polynomial in order to distinguish small changes in conductivity due to single CDW phase-slip events. (The degree of the polynomial varied from 6 to 15.) Then, we subtracted this approximation from the $\sigma(T)$ dependences. The results of this procedure are presented in Fig. 14. The $\sigma(T)$ dependences for $T < T_{P2}$ for two submicron $NbSe_3$ samples were measured during repeated thermocycling. It can be seen from the figure that the curves obtained are actually sets of lines spreading out like a fan with lowering temperature. Each line corresponds to one of the discrete CDW states. The dashed lines help to identify the states that the sample did not occupy during measurements. The observed jumps between the states correspond to single — and sometimes double or triple — phase-slip events.

The most important information to be extracted from the figure is a rise in the magnitude of conductivity jumps $\delta\sigma$ with lowering temperature, which suggests an enhanced mobility of the carriers. The $\delta\sigma$ value was used to determine the temperature dependence of mobility [37]. The results for five samples are presented in Fig. 15. A drop in temperature T to 20 K increased the mobility — it approaches 10^5 cm^2 V^{-1} s^{-1}. It is worth noting that this value was obtained without using

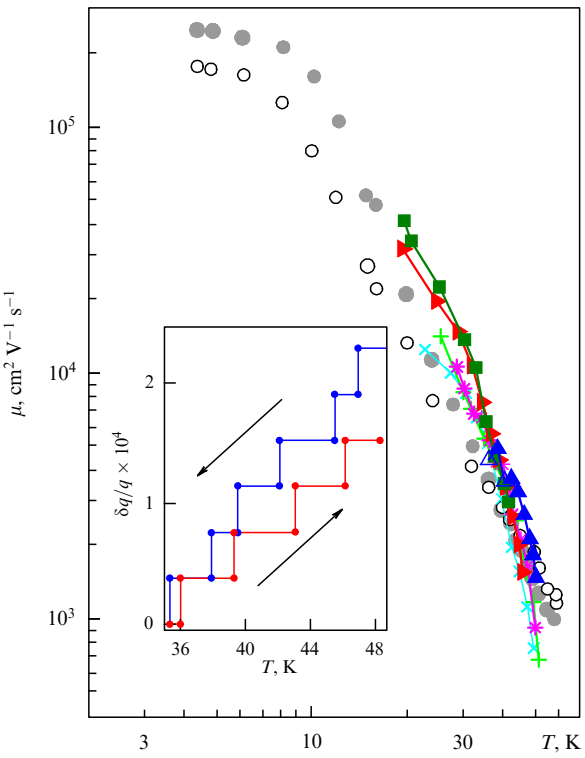


Figure 15. Temperature dependences of mobility obtained by the treatment of $\sigma(T)$ curves for five $NbSe_3$ samples 29–62 μm in length. Unlike symbols correspond to different samples. Full and open dots denote results of model calculation [41] for electron and hole quasiparticles, respectively. Inset: a segment of $q(T)$ dependence for one of the samples (see inset to Fig. 14) obtained on the assumption that each conduction jump corresponds to $\delta q = \pm 2\pi/L$.

any adjustable parameters. Because formation of the lower CDW is believed to be related to the appearance of a gap in the electron spectrum on type 1 chains, it is clear that we found mobilities of quasiparticles connected *precisely with these* chains.

Such a high mobility appears to be due to the presence of 'pockets', i.e., portions of the Fermi surface containing free holes. This unusual property of quasiparticles makes $NbSe_3$ especially interesting for researchers. Earlier authors calculated mobilities of carriers from measurements of the Hall effect, transverse magnetoresistance, and conduction anisotropy with the use of the so-called Ong two-band model. The results of calculation [41] for electrons and holes (principal carriers) are also shown in the inset to Fig. 15. Because six adjustable parameters and certain simplifications were employed in the calculation, its results appear unreliable. Later data suggesting small effective mass of the holes at low temperatures are purely qualitative [42]. Therefore, the rough coincidence between the result thus obtained (Fig. 15) and the results of calculation in Ref. [41] is of importance as the outcome of both the first direct measurement of mobility and the confirmation of the model [41]. Notice that the hole mobility increases even faster than it is expected from the Ong model (open circles in Fig. 15).

Conduction steps also observed in the temperature range $T_{P2} < T < T_{P1}$ are associated with 'quantization' of the states of the upper CDW with $T_P = 140$ K [37]. The mobility of carriers on type 3 chains proved to equal 400–600 $cm^2\,V^{-1}\,s^{-1}$ at $T = 85-135$ K [37], a rather high mobility for quasiparticle excitations of CDW. The steps in this temperature range can be related to the temperature dependence of the q vector of the upper CDW. $q_1(T)$, observed in an earlier study [43]. The $q_2(T)$ dependence has never been reported before for an obvious reason: the change in q_2 over the entire range 20 K $< T < T_{P2}$ is no more than $\sim 10^{-4}$ [37]. Such a change is rather difficult to observe by X-ray techniques.

To conclude, the observation of 'quantization' is not only interesting in itself, but it also allows determining with a high degree of accuracy the mobility of current-carrying quasiparticles. In the case of $NbSe_3$, it also revealed changes in the q-vector value and its hysteresis. The use of 'quantization' of the wave vector values makes it possible to study intricate effects arising from the reorganization of superstructure in different materials and to determine the mobility of carriers in Peierls semiconductors with a complicated structure of conducting bands.

4. Effects associated with deformation of quasi-one-dimensional charge density wave conductors

4.1. Nature of the deformation of quasi-one-dimensional conductors in an electric field

As noted in the Introduction, a CDW can be regarded as an electron crystal formed inside the host lattice and interacting with it. CDW depinning was shown to decrease Young's modulus Y in TaS_3 by 4% [20, 44–47], while the share modulus by the magnitude reaching 30% [20, 48, 49]. These findings were qualitatively explained on the assumption that the deformation of a crystal disturbs CDW equilibrium as well, i.e., results in producing CDW strain [44].

One can also envision the following situation: suppose that a CDW undergoes deformation under the effect of an external force. The interaction of two 'springs', the CDW and the lattice, causes the sizes of the crystal to change in such a way that the CDW approaches equilibrium, and thus the total elastic energy of the CDW and the lattice becomes minimal [22]. Variations of the length of quasi-one-dimensional conductors associated with CDW deformation were observed for the first time under the action of an electric field in Ref. [21]. It was shown that the length of TaS_3 samples may be different, even in a zero field, depending on the previously applied voltage, in connection with residual (metastable) CDW strain. The authors of Ref. [22] reported a change in the length one and a half orders of magnitude greater than in Ref. [21] ($\sim 5 \times 10^{-5}$), with the metastable CDW states in the TaS_3 sample being thermally created. These states are characterized by relatively uniform stretching or compression of the CDW over the sample length.

The semiempirical model proposed in Ref. [52] explains the thermally induced uniform deformation of the samples. Generally speaking, a change in the sample length as a result of CDW deformation can be expected only when the equilibrium CDW period λ nontrivially depends on the lattice parameters. For example, if a change in the lattice period a under the action of an external force alters λ/a, the sample length can be expected to change as a result of CDW deformation. Thus, the TaS_3 strain observed in Ref. [22] is associated with a decrease in the value of a CDW q vector during stretching of the sample, which was revealed in Refs [50, 51]. Model [52] predicts a markedly inhomogeneous longitudinal strain of the sample in an electric field. It is shown that such strain may be L_c/λ times the strain of piezoelectric ion crystals. The ratio of a CDW coherence length to a wave period, L_c/λ, may reach 10^6. However, nonuniform longitudinal strain of quasi-one-dimensional crystals has not yet been observed, presumably due to technical difficulties. An experimental approach with the use of AFM to study such deformation was proposed in Ref. [53].

At the same time, crystal strain need not be longitudinal. Bearing in mind the dependence of λ, not only on longitudinal but also on other components of sample's strain, other forms of inhomogeneous crystal strain in an electric field can be anticipated. Some of them are dealt with in Sections 4.2–4.7.

Torsional strain has been most thoroughly investigated due in part to the availability of relatively simple methods for the measurement of the torsion angle [54–63]. These methods are briefly discussed in Section 4.2.

4.2 Methods of torsion studies

To begin with, a crystal must be fixed so as to enable torsion. With this aim in view one end of the sample 1–5 mm in length and 10–30 μm in width was attached to the substrate by an indium pressure contact so that the sample was elevated above the substrate. The current was fed into the other end through a thin wire — a gold-plated whisker of high-temperature superconductor (HTSC) $Bi_2Sr_2CaCu_2O_x$, which did not really hinder sample twist. Different sample mounting schemes without resort to wires were also used, which is especially important for experiments with the thinnest samples. One scheme took advantage of the relationship between torsion and the presence of a polar axis in the sample that determined twist direction [54]. The sample was cut into halves transversely to conducting chains. One half was turned 180° and the former free end was glued to the cut surface of the other half, giving a structure of two pieces of one sample connected end to end with oppositely directed

polar axes. When the current was passed through this structure, the soldered ends of each piece turned to the same side [62].

Another configuration is much simpler as regards manufacturing technology [58, 62]. Both ends of a whisker were fixed on the substrate by two indium pressure contacts so that its middle part was elevated above the substrate (Fig. 19b in Section 4.5). In this case, half of the suspended part coated with a shunting gold layer served as a passive analog of the tension member, i.e. the wire in the preceding variant. The elasticity of the tension member made, obviously, the torsion angles roughly by half, which was taken into account in the treatment of the results.

In most experiments, torsion was studied by optical methods, in which a laser beam was focused on micromirrors glued to the samples ($Bi_2Sr_2CaCu_2O_x$ whiskers coated with a thin gold layer were also used). An American team studied torsion by a capacitance method based on the measurement of the spiral resonator frequency varying with the movement of the wire glued to the sample [59–61, 63].

4.3 Main results of torsion studies

Here are the main results of torsion research. It was revealed that the free end of the sample in the Peierls state turned as the current was passed through metal chains, with the direction of the turn determined by current direction [54]. Such an effect was observed in a number of CDW conductors: TaS_3, $(TaSe_4)_2I$, $K_{0.3}MoO_3$, and NbS_3 (phase II) [54, 56]. Dependences of the torsion angle on the electric field, $\varphi(E)$, measured at different E variation rates give evidence of two contributions [56]: (1) slow — threshold, hysteretic (see Section 4.4, Fig. 17b, dark curve), and (2) fast — practically linear in terms of voltage. Both contributions are related to CDW strain, probably due to surface pinning. In the first case, the strain was caused by the formation of metastable states; in the second case — by small reversible CDW deformations near an equilibrium or near a metastable state. In the case of 'slow' torsion, shear strain normalized to electric field amounts to 10^{-6} m V^{-1} (at least three orders of magnitude higher than the maximum values of piezomoduli in piezoelectrics). For TaS_3, the time of switchover between metastable strain states after the application of an electric field higher than the threshold one lies in the range $10^{-2}-10^{-1}$ s at liquid-nitrogen temperature and varies approximately as $\exp[900/(T\,[\mathrm{K}])]$ between 70 and 100 K [56]. Notice that the activation energy equal to 900 K is close to the half-width of the Peierls gap. A somewhat higher activation energy was obtained in Ref. [63], where it is shown that relaxation time decreases with increasing voltage and is invariably higher compared with the known CDW relaxation times found from the dynamics of the electromodulation of TaS_3 reflection index.

The second (fast) contribution [56] is studied less thoroughly. It is smaller than the first one by roughly 1–2 orders of magnitude and prevails in fields below threshold. In this case, CDW strain is small and metastable states do not form; nor does 'slow' strain develop. Moreover, the fast contribution predominates at relatively high frequencies, above 0.1–1 kHz for TaS_3 at $T \approx 80$ K, which distinguishes it from the slow contribution [56]. The upper frequency boundary for such response remains to be found. The TaS_3 torsion amplitude within the limits of experimental accuracy is independent of the alternating current frequency, at least till the first torsional resonance (1–10 kHz) occurs.

The general view of the dependence of torsional response on the amplitude and frequency is given in Fig. 16a [56], where the TaS_3 torsion amplitude is presented as a function of the alternating voltage amplitude in the form of symmetric meander at different alternating voltage frequencies. An analogous set of curves for blue bronze is displayed in Fig. 16b. These dependences suggest that the torsional response at low voltage modulation frequencies shows the threshold dependence on the amplitude. The threshold component of the response decreases with increasing frequency, whereas the linear one barely changes (in TaS_3), or decreases far less significantly (in blue bronze).

The resonant torsional vibrations under the action of an alternating electric field at frequencies up to 200 kHz and Q-factor above 10^4 are associated with 'fast' torsion [57]. In this case, the limiting frequency at which torsion occurs is

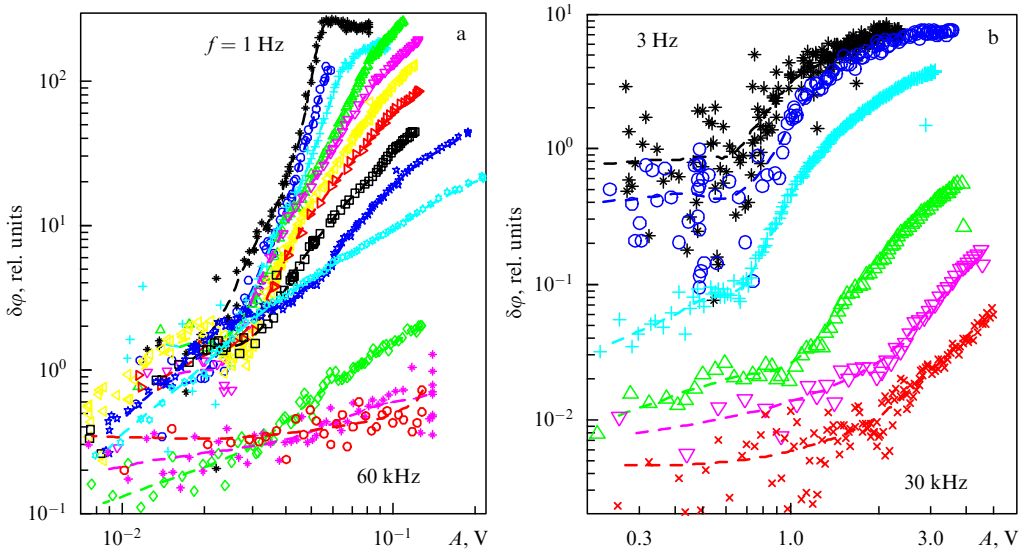


Figure 16. The amplitude of $\delta\varphi$ depending on amplitude A of symmetric rectangular alternating voltage: (a) for TaS_3 at frequencies (from top to bottom) $f = 1, 2, 4, 40, 110, 220, 440, 880, 2\times10^3, 9\times10^3, 15\times10^3, 30\times10^3, 60\times10^3$ Hz, and (b) for blue bronze at frequencies (from top to bottom) $f = 3, 30, 300, 5\times10^3, 10^4, 3\times10^4$ Hz. Dashed curves show smoothed results; the temperature $T \approx 82$ K.

restricted by measuring technique capabilities. The relationship between this contribution and the existence of CDW is well apparent from the behavior of temperature dependence, i.e., a sharp decrease in the amplitude in the Peierls transition region [56].

4.4 Torsional effects associated with the sliding of a charge density wave

The manifestation of CDW strain in the mechanical properties of the samples raised the question about the possible observation of torsional effects associated with CDW sliding, such as synchronization of CDW sliding in an electric field (Shapiro steps) and spatio-temporal nonuniformity of CDW motion apparent as broad- and narrow-band voltage fluctuations at a direct current exceeding the threshold one.

We begin with the Shapiro steps, the finding of which on CVCs suggests increasing *temporal* coherence of CDW motion: a CDW synphasically overcomes the periodic pinning potential at different points of the sample. At the same time, nothing is known about *spatial* coherence of CDW under synchronization conditions, i.e., about the effect of irradiation on the spatially inhomogeneous CDW strain. It is well known that a CDW can slide in a sample continuously without the formation of phase-slip centers in the bulk, as evidenced by the presence of near-contact CDW strain regions extending a few millimeters from the contacts [65]. This extended strain, like inhomogeneous CDW strain developed due to impurities, does not exclude, in principle, the possibility of synchronously overcoming pinning barriers at different points of a sample: the continuity of motion of a CDW means that its phase velocity (hence, fundamental frequency) is identical at all these points. For this reason, even full synchronization of CDW motion during microwave irradiation does not necessarily imply enhanced spatial coherence. Because torsion is associated with spatially inhomogeneous CDW strain, the manifestation of Shapiro steps may provide here information about spatial CDW coherence under synchronization.

Reference [58] reports the results of investigations into the 'slow' torsion of TaS_3 samples during microwave irradiation. Torsional strain was measured simultaneously with $R_d(I)$ dependences that exhibited Shapiro steps, i.e., $R_d(I)$ peaks. The $\varphi(I)$ dependences also showed peculiarities at the same current values at which Shapiro steps were apparent. Direct measurements of $\varphi(I)$ dependences (Fig. 17b, light curve) failed to reveal details of these peculiarities due to low-frequency noises produced by the photodiode and the amplifier. Therefore, we measured $d\varphi/dI$ derivatives depending on I, in analogy with CVC measurement by the synchronous detection method. In so doing, the current modulation frequency proved to be much lower than the frequency of the first torsional resonance (1–10 kHz). The I-dependence of $d\varphi/dI$ exhibited narrow peaks that crossed the $d\varphi/dI=0$ level (Fig. 17a). This means that under conditions of the fullest CDW synchronization, the sample begins to turn in the opposite direction, as also viewed in Fig. 17b presenting results of direct $\varphi(I)$ measurement (thin curve) together with the results of $d\varphi/dI$ integration. Evidently, the $\varphi(I)$ dependences measured by the two methods roughly coincide, except the hysteretic (slow) contribution, which is practically absent when measured by the differentiation method [58]. The dashed line in Fig. 17b fits approximate results of the $\varphi(I)$ extrapolation in the Shapiro step region in the absence of synchronization. It may be seen that, for the

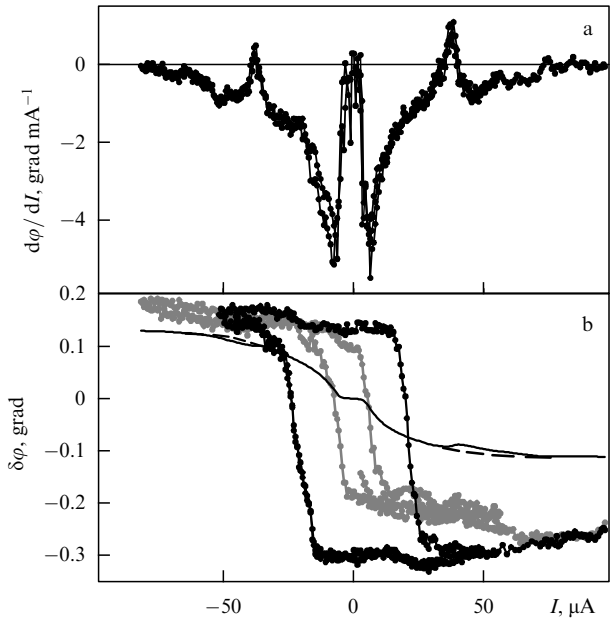


Figure 17. (a) Differential signal from a photodiode (rms value) measured by a synchronous detector during microwave irradiation at 500 kHz. Modulation frequency is 333 Hz (20 times lower than the frequency of first torsional resonance), amplitude $\delta I = 0.6$ μA, and $T = 121$ K. (b) $\varphi(I)$ dependences with (light curve with dots) and without (dark curve) irradiation at 500 kHz. The thin curve is the result of $d\varphi/dI$ integration, and the dashed curve fits approximate results of extrapolation of $\varphi(I)$ in the Shapiro step region in the absence of synchronization [58].

sample being studied, a change in the rotation angle, $10^{-2}{}^\circ$, is achieved due to CDW synchronization. The degree of synchronization is estimated at 10% from the $R_d(I)$ dependence [58], which is a good value for TaS_3. Thus, under full synchronization conditions, the torsion angle can be expected to deviate by 0.1° (in fact, zero deviation unless the hysteretic contribution is counted). The decrease in the sample's torsion angle implies that inhomogeneous CDW strain decreases, too, meaning that inhomogeneous CDW strain, at least its torsion-related part (most probably shear strain [54, 56]), must disappear under full synchronization. This finding suggests an enhancement of CDW *spatial* coherence in the presence of Shapiro steps. This conclusion is also consistent with the results of Ref. [66], where Young's modulus maxima and internal friction minima were observed in TaS_3 and $NbSe_3$ samples under synchronization conditions.

As also noted in Ref. [58], torsion measurements can be regarded as a highly sensitive method for studying CDW synchronization. This inference is illustrated by a comparison of peculiarities in the I-dependences of $d\varphi/dI$ and R_d for one of the TaS_3 samples singled out by more chaotic CDW motion. In this sample, the peaks corresponding to Shapiro steps are less pronounced, but peaks corresponding to fundamental frequency harmonics and subharmonics appear. It turned out that the fine structure of harmonics and subharmonics is more apparent in the I-dependence of $d\varphi/dI$ than in the I-dependence of R_d.

4.5 Sample vibrations during the flow of a direct charge density wave current

Shapiro steps research can be regarded as a tool for the study of CDW creep in the periodic pinning potential. This potential is directly manifested as the generation of narrow-

band electric noise during CDW creep. The generation of broad-band low-frequency noise is also associated with pinning and narrow-band noise [67–69]. It can be expected that the motion in the pinning potential must cause inhomogeneous dynamic, i.e., time-dependent, CDW strain.

We present below the results of a study on the dynamic strain of a sample through which direct CDW current is passed. The idea behind this experiment is as follows. If CDW strain results in sample's torsional strain, the dynamic inhomogeneous strain of a sliding CDW is likely to manifest itself as noise (torsional) vibrations (both stochastic and periodic) of the sample.

The first experiments were carried out on samples of $(TaSe_4)_2I$. CDW creep in this compound is distinguished by very low coherence, as apparent from the very indistinct peak of narrow-band noise ($\delta f/f \geqslant 1/10$) [70, 71] and relatively strong broad-band noise. Due to this, the amplitude of stochastic angular oscillations during CDW creep in this compound was higher than in TaS_3. However, difficulties encountered in the preparation of $(TaSe_4)_2I$-based torsional structures, which are related to the fragility of the resulting samples, precluded completion of these experiments. Improvement of the measuring scheme permitted reducing the level of sample vibrations and the amplifier noise; it made possible the study of torsional noises in the TaS_3 samples, as well.

When measuring noise, we ran a direct current I through a sample and recorded, as a rule, the oscillograms of the angle $\varphi(t)$ oscillations with a discretization frequency of 400 Hz for 40 s (16,000 points). Alternating (noise) voltage at the sample was simultaneously recorded in the oscillograph second channel. Discrete Fourier transform was used to find frequency dependences of spectral noise density for angle, $S_\varphi(f)$, and voltage, $S_V(f)$, in a range of 0.025–200 Hz. Measurements repeated at different currents yielded the dependences of spectral noise densities of φ and V on the frequency and current passed through the sample.

A set of $S_\varphi(f)$ dependences for one of the $(TaSe_4)_2I$ samples is presented in Fig. 18. Analogous data for TaS_3 at currents below and above the threshold are shown in Fig. 19b. Figure 19a depicts simultaneously measured dependences for *voltage* fluctuations $S_V(f)$. The following qualitative conclusions are relevant to both quasi-one-dimensional compounds: (1) current passage is accompanied by torsion angle fluctuations in excess of background vibrations (in a zero field); (2) these fluctuations are noticeable only at a voltage close to or above the threshold, and (3) the noise 'turns white', on the average, as V grows: if the difference $S_\varphi(f, V) - S_\varphi(f, 0)$ is approximated by the $1/f^\alpha$ dependence, the exponent α decreases from 2 to 1 for $(TaSe_4)_2I$, and from 1.5 to 0.5 for TaS_3 with increasing V. These observations indicate that vibrations qualitatively reproduce the main features of low-frequency noise associated with CDW sliding [72]. Noise 'whitening' with growing CDW current means that an increase in low-frequency noise precedes that of its high-frequency component. If the noise is regarded in a simplified representation as a result of superposition of two-level fluctuating systems ('fluctuators'), spectral changes related to growing currents suggest a rise in the mean frequency of their switchovers [73, 74].

A detailed comparison of spectral fluctuations of torsion angle and voltage for TaS_3 samples (see Fig. 19) also reveals some differences in their current-dependent evolution. For example, S_V starts to grow gradually even at below-threshold currents, whereas S_φ increases jumplike at a somewhat higher current. In both cases, a further increase in spectral density becomes saturated at currents exceeding roughly two times the threshold value.

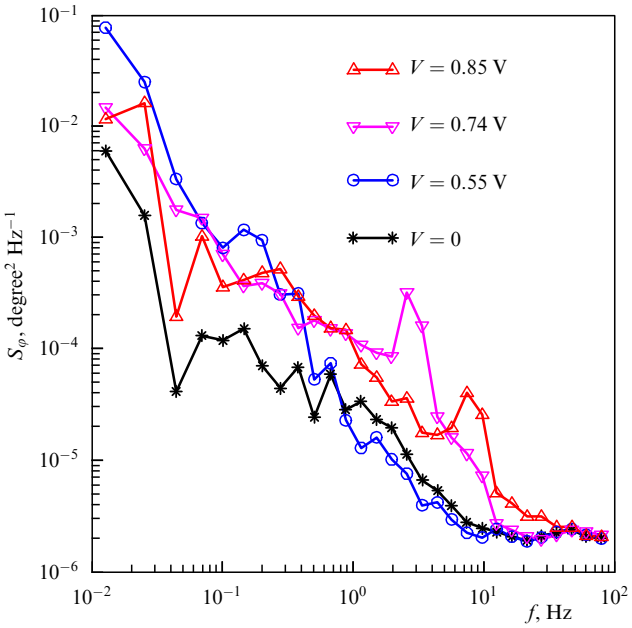


Figure 18. Spectral density of torsion angle fluctuations for a $(TaSe_4)_2I$ whisker during passage of the above-threshold direct current; $T = 147$ K, and discretization frequency is 200 Hz. Fluctuations surpass the noise level ($V = 0$) as the current increases. Threshold voltage is 0.5 V, sample length and width are 1.5 mm and 15 μm, respectively.

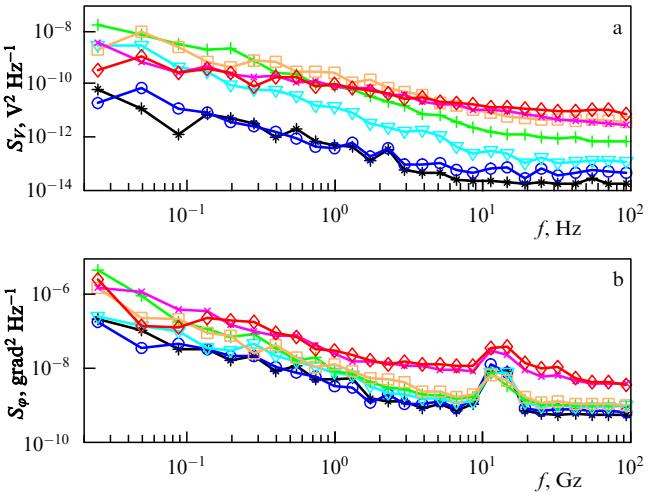


Figure 19. Spectral density of voltage fluctuations (a) and torsion angle fluctuations (b) in a TaS_3 whisker at constant current values 0 (∗), 1.20 μA (○), 2.24 μA (▽), 2.68 μA (+), 2.93 μA (□), 3.17 μA (×), and 3.41 μA (◇). Threshold current is $I_t = 2.8$ μA, length of the gold-free part is 3.5 mm, width 17 μm, and thickness 7 μm; the temperature $T = 83$ K.

Peculiarities in $S_\varphi(f, V)$ curves for $(TaSe_4)_2I$ samples (Fig. 18) need to be studied more thoroughly.

Detailed results of sample's vibration research, including those related to narrow-band noise, are presented in Ref. [75], where it is also shown that the cause of vibrations cannot be reduced to the action of noise voltage associated with CDW creep on the sample: the observed fluctuations of the torsion angle are 2–3 orders of magnitude greater than the angles due to electric noise effects.

Thus, it can be concluded that CDW sliding leads to two independent phenomena: generation of noise voltage, and onset of noise torsional vibrations. In other words, mechanical vibrations are due to the 'direct' mechanical effect of sliding CDWs.

We note that the relationship between torsion angle and voltage fluctuations is not trivial: angle fluctuations depend on inhomogeneous CDW strain, whereas voltage fluctuations may occur without CDW strains, e.g., in a rigid CDW model [13]. We did not find a significant correlation between simultaneously measured $V(t)$ and $\varphi(t)$ oscillograms.

4.6 Nature of torsion

All the main characteristics of torsion allow for the conclusion that it is associated with CDW strain. However, the kind of this strain remains unclear. It was shown in Ref. [54] that torsion cannot be induced by longitudinal CDW strain homogeneous throughout the volume. The relatively small magnitude of the torsion angle temperature hysteresis betokens such a conclusion [54]. The best studied variety of inhomogeneous CDW strain in an electric field, i.e., contraction at one contact and extension at the other, cannot also explain the observed torsion: the torsion angle increases monotonically and almost linearly [54] upon moving off the motionless contact. Therefore, we concluded that the sought-after CDW strain is roughly homogeneous over the sample length but inhomogeneous over its cross section.

It is argued in Refs [54, 56] that such strain may be due to CDW surface pinning and manifests itself in a CDW shear near a surface. As a crystal twists, the largest shear strain develops on the surface. It linearly decreases to zero far from the surface and vanishes at the axis of torsion. Therefore, it can logically be supposed that torsion relates precisely to the CDW strain near the surface. From the symmetry standpoint, torsion may occur in a crystal containing a polar axis. In the case of TaS_3 (rhombic syngony, point group 222 [7]), the polar axis determining torsional direction is absent. In principle, symmetry may be reduced in the Peierls transition, e.g., as a result of transformation of the rectangular cross section of the sample (or elementary cell) into a parallelogram. Such distortion would mean nonequivalence of the adjacent edges of a single-crystal parallel to the conducting chains, and nonequivalence of two directions along the chains. In this situation, 'torsional' surface pinning may occur if the pinning force near the adjacent edges is different.

Torsion may also result from structural defects of a sample. It was shown experimentally in Ref. [54] for three samples cut into halves across the conducting chains that the torsional directions for different pieces coincide if the direction of the chains is preserved. Thus, if torsion is associated with structural defects, such a structure extends over the total sample length. The linkage between structural defects and torsion was confirmed in Ref. [63], showing that the application of an external torque affects the torsion angle and can even cause inversion of the hysteresis loop (Fig. 17a), i.e. a change in the torsional direction at the same voltage sign.

It should be emphasized that the observation of torsion does not mean that a sample is free from developing other comparable strains. One of them, bending, was observed in TEM [56], although the overall strain pattern after the application of an electric field to the sample remains unclear. Specifically, inhomogeneous longitudinal strain expected from the data of Refs [22, 52] needs to be examined (see Section 4.1). Therefore, it cannot be concluded for the moment that shear strain manifested in torsion substantially predominates over other kinds of deformation.

4.7 Voltage modulation induced by torsional strain

Deformation of a sample in an electric field, i.e., an analog of the inverse piezoeffect, raises the question about the possibility of obtaining a signal analogous to the direct piezoeffect, i.e., the electromotive force (emf) resulting from the development of strains. The sample must be a dielectric if emf is to be observed. Resistivity of quasi-one-dimensional conductors at the liquid-nitrogen temperature (torsion is practically unexplored at lower temperatures) as a rule does not exceed $\rho = 1\ \Omega$ cm, which suggests very short charge relaxation time, i.e. Maxwell relaxation time $\tau = \varepsilon\varepsilon_0\rho$, ranging below 10^{-13} s at $\varepsilon = 1$, which excludes observation of the emf caused by developing sample's strain. However, it is worth searching for an analog of the direct piezoeffect at the liquid-helium temperature. The resistivity of TaS_3 samples in these conditions may be on the order of $10^{10}\ \Omega$ cm and higher.

At the same time, torsion (or other deformation) during current passage may produce a feedback signal related to torsion-induced modulation of resistance. Preliminary studies showed that such a signal does exist at above-threshold currents. Reference [57] reports detection of torsional resonances without using optical and capacitance methods. The TaS_3 whisker itself served as the receiver. Alternating current was fed into one part of the sample (actuator) and the signal was detected at another part (receiver) (Fig. 20a). When the alternating current frequency coincided with the resonance frequency, the frequency dependence of the signal had a maximum, suggesting an increase in the torsional vibration amplitude. A similar 'self-sensitive' actuator was designed to have no suspended contact joints (Fig. 20b) [76]. In this case, we modulated the resistance of the motionless part of the sample, contributing to the spread of CDW strain beyond the contact region.

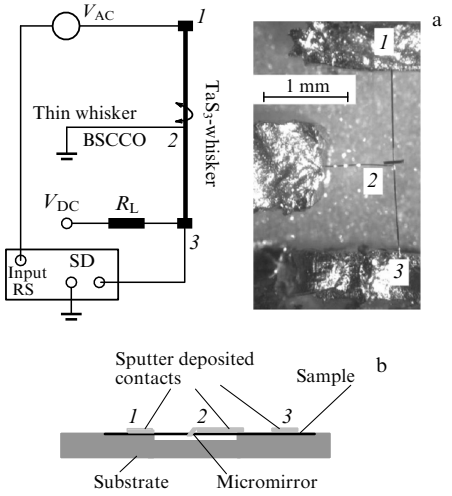


Figure 20. Variants of sample location in detecting resonant torsion without an optical scheme. (a) TaS_3 sample is elevated on two indium contacts [57]. (b) Sample without suspended contact joints: *1–3* — sputter deposited gold contacts [76]. In both cases, alternating voltage V_{AC} exciting vibrations is applied to contacts *1*, *2* and direct current (DC) flows through contacts *3*, *2*. At resonance, a synchronous detector (SD) registers torsional modulation of voltage between contacts *3* and *2*. RS labels reference signal.

The feedback signal was observed only if a direct current substantially higher (roughly an order of magnitude) than the threshold one was passed through the 'receiver'. By a rough estimate, the relative change in voltage in the sample at resonance compares with the surface shear strain of TaS_3: $(\delta V/V)/G \sim 1$. This result means that TaS_3 is not an unusually sensitive strain gauge, at least at liquid-nitrogen temperature.

The tensoelectric response of TaS_3 was more thoroughly investigated in Refs [59, 60], also at liquid-nitrogen temperature. Advantages of the device designed by the authors of Refs [59, 60] included the possibility of exciting torsional vibrations by an alternating magnetic field, although, unfortunately, the capacitance method proved unsuitable for exact calibration of the torsion angle in absolute units (this drawback was corrected in work [63]). It turned out that torsion-induced alternating voltage appeared only when the current was stronger than a certain threshold value in excess of the CDW depinning current. The feedback voltage sharply increased with voltage after this threshold was surpassed. Based on the results of papers [59, 60], the maximum surface shear of the samples was roughly estimated as $G \sim 2 \times 10^{-4}$. Hence, $\delta V/V \sim 10^{-3}$ and the estimate gives $(\delta V/V)/G \sim 5$, close to the above value.

We studied the tensoresistive response of TaS_3 samples at different temperatures. The results are presented below. The idea behind the experiment is that one of the contacts is placed on a motionless substrate, and the other on the bonding pad rotating about the axis roughly coincident with the sample axis, the pad which can be set in motion by a lever extending outward from the cryostat through the rubber gasket sealing and rigidly connected with it. The lever is jointed to the measuring mechanism of an X–Y recorder operating as a torsional actuator. In so doing, the torsion angle of the sample is equivalent to the displacement of the measuring mechanism divided by the lever length (as a rule, of order 1/10 rad). Above-threshold current was passed through the sample, and a 2–20 Hz frequency voltage was applied to the self-recorder. The alternating signal from the sample measured by the lock-in detection method constituted the sought-after tensoresistant or 'torsion-resistant' response.

Figure 21a exemplifies the current dependence of the sample's response. It can be seen that the response emerges at a current above the threshold value of I_t. The shape of this dependence is similar to that obtained in Refs [59, 60], but the signal saturates at currents above $\approx 2I_t$. Moreover, depinning thresholds for CDW (Fig. 21b) and for the appearance of torsional response are not significantly different, unlike those in Refs [59, 60]. At a modulation frequency of 3 Hz, the phase of response is practically unrelated to the current, i.e., the response does not fall behind torsion. It becomes delayed as the frequency increases, i.e., a quadrature signal appears, near the threshold field E_t, and this delay decreases with increasing current. The same conclusion is drawn from Fig. 3 of Refs [59, 60], showing that the quadrature signal at a frequency of 10 Hz first grows with a rise in V and becomes saturated thereafter.

The data presented in Fig. 21a allow the $(\delta V/V)/G$ ratio to be estimated. The value of $\delta V/V$ amounts to 5×10^{-4}, and $G = \delta\varphi\, w/2L \approx 0.07$ rad $\times$ 5 μm/3 mm $\sim 10^{-4}$, which gives the approximated ratio $(\delta V/V)/G \approx 5$, i.e., the same value as in Refs [59, 60]. Similar to the last experiments, the magnitude of the response and its current dependence varied appreciably from sample to sample. Sometimes, the $\delta V(I)$ maximum was

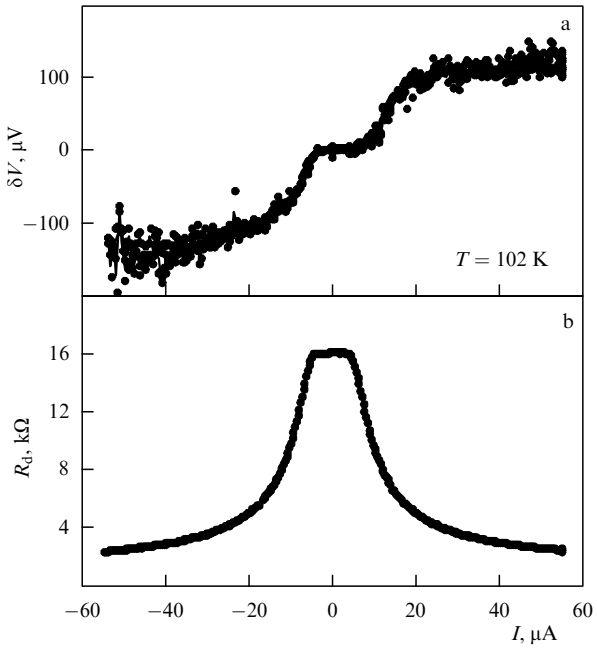


Figure 21. (a) Current-dependent alternating voltage across a TaS_3 sample recorded under torsion angle modulation; $f = 3$ Hz, modulation amplitude is 4°, and sample length 3 mm. (b) Current dependence of differential resistance in the same conditions.

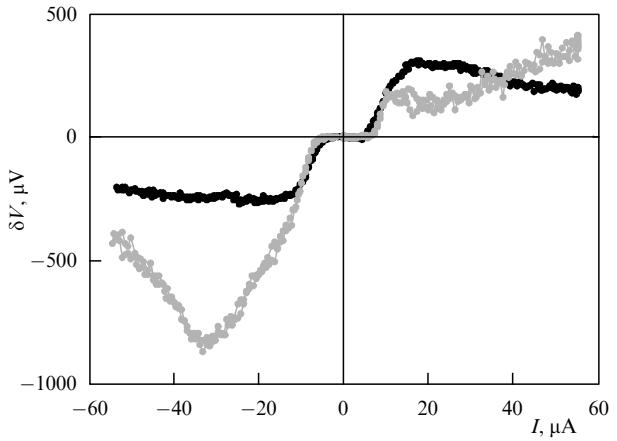


Figure 22. Current-dependent alternating voltage across a TaS_3 sample recorded under torsion angle modulation at $T = 89$ K (dark curve) and $T = 78$ K (light curve); $f = 7$ Hz, modulation amplitude is 3.5°, and sample length 3 mm.

observed instead of saturation (Fig. 22). The character of the dependence changed if a constant torsion angle of ≈ 1 rad ($G \sim 10^{-3}$) was added to the variable $\delta\varphi$.

The dependence was not always symmetric with respect to current reversal. As a rule, a drop in temperature to liquid-nitrogen temperature and below results in a symmetry breaking, as illustrated in Fig. 22 showing $\delta V(I)$ dependences at two temperatures (obtained for the same sample as in Fig. 21 on the next day).

Summarized results of tensoresistance measurements indicate that the $(\delta V/V)/G$ ratio for TaS_3 undergoing torsional strain may reach 5–10. It is a large enough value, bearing in mind that torsion causes only shear strain in which the volume of an elementary cell does not change in the first

approximation. References [77–80] make it possible to estimate (taking into account Young's modulus of 350 GPa [20]) the ratio of $\delta V/V$ to small *longitudinal* strain (tensoresistive coefficient) for TaS_3 as equaling at least 100 in the linear conduction regime (or somewhat below this value in the nonlinear regime). This parameter is comparable with that for p-silicon [81], traditionally employed as a tensoresistive sensor.

Studies into resistive feedback are also of interest from the scientific standpoint, as providing information about the effect of sample's strain on CDW dynamics. At temperatures above liquid-nitrogen temperature, one and the same strain roughly similarly changes the CDW current flowing in different directions. The model proposed in papers [59, 60] describes the modulation of CDW current by varying the threshold field during torsion. The CDW elastic modulus increases at lower temperatures (together with coherence length [82–84]) and the CDW dynamics is described by the motion of large domains. In this case, pinning of individual domains may lead to some asymmetry of the CVC. Accordingly, curves $\delta V(I)$ lose symmetry, too: one and the same torsional strain may have markedly different effects on CDW domains at I and $-I$, because their sizes are comparable to sample dimensions. It remains unclear how the torsional direction is related to the sign of δV. In principle, the relationship between them is feasible in the presence of the helical axis of symmetry that may be inherent in rhombic TaS_3 samples belonging to the point group 222 [7].

5. Nonlinear conduction in the quasi-one-dimensional conductor TiS_3

Sections 2–4 were concerned with group V transition metal trichalcogenides TaS_3, $NbSe_3$, NbS_3, which are typical quasi-one-dimensional conductors with CDWs. In the present section, we report the results of studies of a different sort of quasi-one-dimensional compound, exemplified by TiS_3; we recall that Ti is a group IV transition metal. Most trichalcogenides of such metals ($M^{IV}X_3^{VI}$) belong among diamagnetic semiconductors. Layered quasi-one-dimensional compounds $M^{IV}X_3^{VI}$ crystallize in the monoclinic phase. Metal chains parallel to the *b*-axis make up layers in the *ab* plane separated by double layers of chalcogen atoms. Up to now, the Peierls transition has been documented in the sole representative of this group, $ZrTe_3$. Surprisingly, CDWs originate in *a* and *c* directions perpendicular to the conducting chains [85], rather than along them as usual.

The electron concentration in TiS_3 samples at room temperature measured from the Hall effect [86] reaches $\sim 2 \times 10^{18}$ cm^{-3}, or 3–4 orders of magnitude lower than in known Peierls conductors. This suggests that TiS_3 may be a semiconductor or semimetal. At the same time, titanium trisulfide, like $ZrTe_3$, shows metallic properties at high temperatures: its resistance in the direction along the conducting chains decreases with decreasing temperature. At $T \approx 250$ K, the temperature dependence of resistance $R(T)$ has a minimum. For $T < 250$ K, resistance behaves in a dielectric manner [87, 88] and becomes frequency-dependent [87]. This might be attributed to CDW formation, but no structural phase transition in TiS_3 has thus far been observed [88], and the above result was explained as incidental to disordering. At the same time, nonlinear conduction has recently been reported to occur in TiS_3 at a temperature below 60 K and to increase as temperature continues to drop. Resistance R_d decreased by more than one order of magnitude at $T \approx 4.2$ K in an electric field of ~ 30 V cm^{-1} [80, 90].

Based on CVC measurements along the chains, the authors of Refs [89, 90] argue in favor of a collective mechanism of nonlinear conduction in TiS_3 samples. The dependences $R(T)$ have peculiarities at 120, 59, and 17 K in the form of the maxima of $|\mathrm{d} \ln R/\mathrm{d}T|$ derivatives in *a* and *b* directions, suggesting phase transitions to the condensed state, probably with the induction of a CDW. Temperature dependences of nonlinear conduction (at fixed voltage values) also exhibit peculiarities at the same temperatures as in the case of linear conduction. The growth in TiS_3 conduction anisotropy by two orders of magnitude with decreasing temperature down to $T \approx 50$ K is also a peculiar characteristic of Peierls quasi-one-dimensional conductors [91, 92].

Threshold nonlinearity typical of CDW depinning was also reported in papers [89, 90]. More detailed studies demonstrated the nonlinear behavior of $R_d(I)$ dependence at low currents as well. Conductivity began to increase practically at zero voltage, i.e., the threshold was absent (Fig. 23). As $I \to 0$, an increasingly sharper peak appeared on the $R_d(I)$ dependence as temperature decreased. CVCs had a power-law form at low temperatures (Fig. 24).

As mentioned in Section 2.3, peaks of $R_d(I)$ were observed in certain (seemingly most defective) NbS_3 samples (Fig. 9b), as well as in thin $NbSe_3$ samples with resistance above 1 kΩ μm^{-1} per unit length, i.e., with a cross section area of less than 3×10^{-3} μm^2, and in thick $NbSe_3$ samples with introduced defects [30, 31]. Threshold-free CVCs were also observed in TaS_3 at low temperatures [92].

The heightened interest in such dependences is due to the search for one-dimensional electron states of the Luttinger liquid type [29]. In this case, resistance must show the power-law dependence on temperature and applied voltage. Quasi-

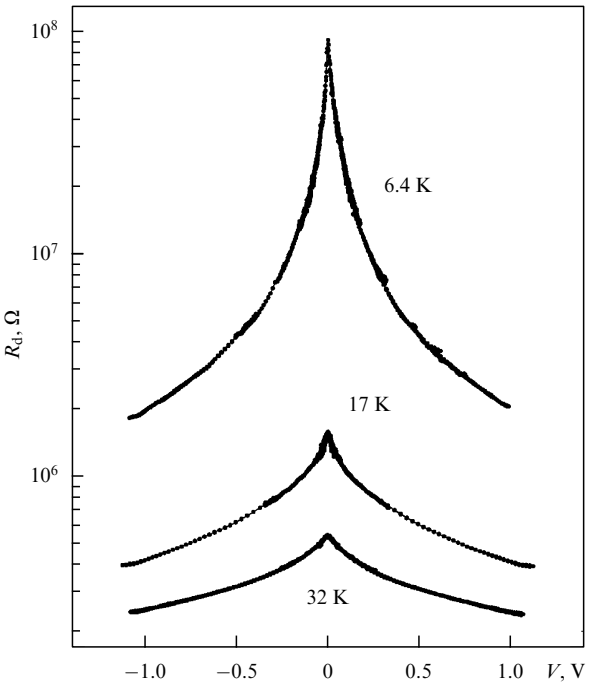


Figure 23. Voltage dependences of longitudinal differential resistance for a TiS_3 whisker at different temperatures. Distance between potential contacts measures 40 μm, and sample cross section area 60×2 μm^2.

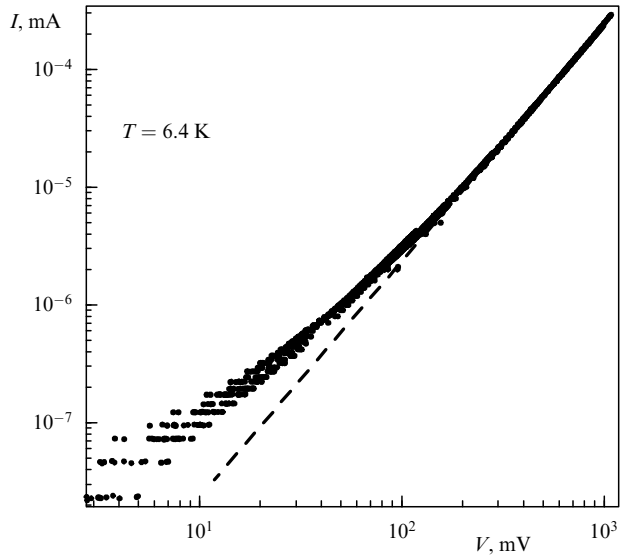


Figure 24. CVC of a TiS_3 whisker measured in the longitudinal direction at T = 6.4 K (corresponds to the upper curve in Fig. 23). For $V > 100$ mV, the dependence has the power-law form: $I \propto V^{\beta+1}$. The slope of the dashed straight line corresponds to exponent $\beta + 1 = 2$.

one-dimensional systems with some degree of disorder may also have an $R_d(I)$ maximum at zero current. Theoretical studies showed [32] that jumplike variations of conductivity in disordered quasi-one-dimensional compounds may be a cause of the power-law dependences of resistance: $R \propto T^{-\alpha}$, and $R \propto V^{-\beta}$. For a Luttinger liquid, one has $\alpha = \beta \approx 2$. In our TiS_3 whiskers, $\beta \approx 1$ near the liquid-helium temperature (Fig. 24). The exponent α found from the slope of $R(T)$ curves in a temperature range $10 < T < 40$ K for different samples proved close to the value of β, namely $\alpha \approx 1.0-1.5$ [93]. In all likelihood, the observed dependences are related to the quasi-one-dimensional character of conduction in combination with electron localization, rather than to the transition to one-dimensional conduction [32]. This reasoning seems especially rational for TiS_3, which shows neither high anisotropy of conduction nor a one-dimensional crystalline structure (crystal lattice parameters are as follows: $a = 0.50$ nm, $b = 0.34$ nm, and $c = 0.88$ nm) [90].

The form of a threshold-free nonlinear CVC in the above cases is determined by the spatially nonuniform potential for carrier motion. The current can be carried here by both one-electron excitations [29–32] and nonuniformly moving CDWs (CDW creep emerges or solitons execute motion) [91, 92, 94]. Nonlinear conduction in TiS_3 occurs for $T < 60$ K and is most likely associated with the phase transition at $T \approx 59$ K, apparent from $R(T)$ dependences [89, 90]. Therefore, it can be supposed that we observe in TiS_3 the nonuniform motion of electrons condensed in a collective state — that is, possibly, a CDW.

Recent studies have shown that the nonlinear conduction in TiS_3 samples also occurs in the transverse direction (along the *a*-axis) at temperatures below 120 K [93]. This nonlinearity appears to be related to CDW formation across the conducting chains in the manner it takes place in $ZrTe_3$ [85].

6. Conclusions

To sum up, the main results discussed in this review were possible to obtain owing to new approaches to the arrangement of experiments, synthesis of compounds (involving new ones) with a perfect structure and coherent CDWs, and application of original methods for preparing micro- and nanostructures based on these materials.

Elucidation of conditions for the synthesis of perfect phase II NbS_3 samples allowed CDW coherent motion to be observed. Synchronization of CDW motion (Shapiro steps) in NbS_3 was observed both at room temperature and T below 150 K. In the former case, a record-breaking 'washboard' frequency of CDWs was obtained with nanometer-thick samples and a CDW synchronization frequency up to 16 GHz. The limiting current density reached in such samples corresponds to the fundamental frequency of 100–200 GHz. Analysis of Shapiro steps gave reason to conclude that at room temperature the CDW current is carried by one of the eight chains contained in an elementary cell. The low density of such chains appears to be a cause behind the possibility of achieving CDW velocities on record in NbS_3 without sample overheating.

Conductivity measurement combined with electron microscopy of NbS_3 samples showed that phase II comprises two groups of samples with different properties determined by growth temperature: whiskers grown at higher temperatures are characterized by enhanced resistance at room temperature and period doubling along the crystallographic *a* direction.

The nature of a lower CDW in NbS_3 ($T < 150$ K) remains to be investigated. The current density carried by this CDW is much lower and strongly depends on sample characteristics. This means that either only part of the elementary cells contribute to CDW conductivity or that the charge carried by a lower CDW is considerably smaller than $2e$ per period. Structural studies of NbS_3 for $T < 150$ K are needed to clarify this issue; they may help to determine the period of CDWs and thereby lay the basis for further work.

In-depth conductivity studies of nanoscale samples of quasi-one-dimensional conductors with CDWs, such as blue bronze and $NbSe_3$, revealed a jumplike variation of resistance (observed earlier in TaS_3 [18]), suggesting specific 'quantization' of the CDW wave vector value. The high coherence of CDWs in these compounds made possible for the first time obtaining a set of discrete states equidistant in terms of conduction. Transition between the neighboring states means a change in the number of CDW periods between the contacts by a unity. Transitions between the states occur as regular temperature-dependent jumps of conductivity. Their temperature distributions were used to derive the temperature dependence of a CDW wave vector $q(T)$; sensitivity to q variations in our case proved substantially higher than that reached by X-ray methods. The mobility of quasiparticles was deduced from the jump magnitudes; it is of special importance for $NbSe_3$ having an intricate band structure. For the first time, we provided direct evidence of anomalously high mobility of residual carriers at low temperatures, suggesting the unique nature of this compound. Thus, q 'quantization' permits us to elucidate some microscopic characteristics of compounds with internal superstructure. In certain cases, this approach has no alternative.

One of the recently emerged lines of investigation into quasi-one-dimensional conductors runs through the research of abnormally large sample strain determined by CDW properties. Of special interest among other kinds of deformation is the huge torsional strain induced by an electric field. Torsion can be regarded as CDW strain transfer to the crystal

lattice, even if neither the form of this strain nor the mechanism of its transfer is known.

It has been shown that torsion contains two contributions: threshold hysteretic and linear. The largest contribution comes from static hysteretic of CDW strain, i.e., strain persisting for the most part after cessation of the application of electric voltage to a sample. At the same time, the dependences $\varphi(I)$ and $\varphi(t)$ exhibit peculiarities that reflect the properties of a sliding CDW. First, HF voltage applied gives rise to peculiarities analogous to Shapiro steps in the $\varphi(I)$ dependences. Their analysis demonstrated enhanced *spatial* coherence of CDWs under synchronization conditions. Second, the flow of a direct CDW current induces torsional vibrations of the samples, reflecting the presence of dynamic CDW strain.

A series of studies was designed to evaluate the influence of TaS_3 twist by an external actuator on the voltage being measured at such samples. No analog of the piezoelectric effect was observed, nor could it be observed owing to the very low resistance of known quasi-one-dimensional conductors with CDWs at temperatures above the liquid-nitrogen temperature. There is every reason to search for an analog of the direct piezoeffect at liquid-helium temperature, when the resistivity of such compounds as TaS_3, $(TaSe_4)_2I$ and $K_{0.3}MoO_3$ is high enough. Torsional strain at temperatures above the liquid-nitrogen temperature results in voltage modulation in TaS_3 samples through which the above-threshold direct current was passed. The maximum relative magnitude of $\delta V/V$ modulation factor normalized to the surface shear ranges 1–10. It is a large enough magnitude, bearing in mind that the volume of an elementary cell during torsion (shear strain) does not change in the linear approximation. This effect is attributed to the action of sample strain on CDW dissipation. For $T \lesssim 90$ K, the $\delta V(I)$ dependence becomes asymmetric, probably due to the rise in CDW coherence volume to a value comparable to the sample volume.

Detailed studies of the transport properties of TiS_3, a still poorly known compound, have been undertaken. It proved to show peculiarities on the temperature dependence of resistance and nonlinear conduction. These findings are explained by phase transitions to the collective electron state. Elucidation of the nature of these transitions (like the transition in NbS_3 at $T \approx 150$ K) is of special interest as occurring in a quasi-one-dimensional compound with a relatively low concentration of electrons. Structure research at low temperatures could elucidate this issue.

Acknowledgments

The authors thank S S Abramchuk for transmission electron microscopy of NbS_3, and R E Thorne and F Levý for the gift of high-quality samples. The work was supported by RFBR (grant No. 08-02-01303-a), the Human Potential Foundation, the Program No. 27 'Fundamentals of Basic Research on Nanotechnologies and Nanomaterials' of the Presidium of the Russian Academy of Sciences, the Program 'New Materials and Structures' of the Russian Academy of Sciences, and the Russian Federation Ministry of Education and Science (Agreement No. 8571). The study was carried out in the framework of the International European Laboratory 'Physical Properties of Coherent Electronic States in Solids' (with the participation of the Kotel'nikov Institute of Radioengineering and Electronics and the Department of Low-Temperature Properties of Solids) supported by CNRS (France), the Russian Academy of Sciences, and RFBR.

References

1. Peierls R *Quantum Theory of Solids* (Oxford: Clarendon Press, 1955) [Translated into Russian (Moscow: IL, 1956)]
2. Fröhlich H *Proc. R. Soc. Lond. A* **223** 296 (1954)
3. Devreese J T, Evrard R P, van Doren V E (Eds) *Highly Conducting One-Dimensional Solids* (New York: Plenum Press, 1979)
4. Fogle W, Perlstein J H *Phys. Rev. B* **6** 1402 (1972)
5. Monceau P et al. *Phys. Rev. Lett.* **37** 602 (1976)
6. Bulaevskii L N *Sov. Phys. Usp.* **18** 131 (1975) [*Usp. Fiz. Nauk* **115** 263 (1975)]
7. Monceau P, in *Electronic Properties of Inorganic Quasi-one-dimensional Compounds* Vol. 2 (Physics and Chemistry of Materials with Low-Dimensional Structures, Ser. B, Ed. P Monceau) (Dordrecht: D. Reidel, 1985) p. 139
8. Grüner G *Density Waves in Solids* (Frontiers in Physics, Vol. 89) (Reading, Mass.: Addison-Wesley Publ. Co., 1994)
9. Brazovskii S, Kirova N, Monceau P (Eds) *Proc. of the Intern. Workshops on Electronic Crystals, ECRYS 2008, Cargèse, France, 24–30 August 2008* (Physica B, Vol. 404, Nos 3–4) (Amsterdam: Elsevier, 2009); *Proc. of the Intern. Workshops on Electronic Crystals, ECRYS 2011, Cargèse, France, 15–27 August 2011* (Physica B, Vol. 407, No. 11) (Amsterdam: Elsevier, 2012)
10. Brazovskii S A, Kirova N N, Monceau P *Physica B* **404** xi (2009)
11. Brazovskii S A, Kirova N N, Monceau P *Physica B* **407** 1683 (2012)
12. Landau L D, Lifshitz E M *Statistical Physics* Vol. 1 (Oxford: Pergamon Press, 1980) [Translated from Russian: *Statisticheskaya Fizika* Pt. 1 (Moscow: Nauka, 1976)]
13. Brown S, Zettl A, in *Charge Density Waves in Solids* (Modern Problems in Condensed Matter Science, Vol. 25, Eds L P Gor'kov, G Grüner) (Amsterdam: Elsevier, 1989) p. 223
14. Sinchenko A A et al. *J. Phys. IV France* **12** Pr9-127 (2002); cond-mat/0101124
15. Sinchenko A A, Thesis for Doct. Phys.-Math. Sci. (Moscow: Moscow Inst. of Physics and Technology, 2010)
16. Zaitsev-Zotov S V *Phys. Usp.* **47** 533 (2004) [*Usp. Fiz. Nauk* **174** 585 (2004)]
17. McCarten J et al. *Phys. Rev. B* **46** 4456 (1992)
18. Borodin D V, Zaitsev-Zotov S V, Nad' F Ya *Sov. Phys. JETP* **66** 793 (1987) [*Zh. Eksp. Teor. Fiz.* **93** 1394 (1987)]
19. Kummamuru R K, Soh Y-A *Nature* **452** 859 (2008)
20. Brill J W, in *Handbook of Elastic Properties of Solids, Liquids, and Gases* (Eds-in-Chief M Levy, H E Bass, R R Stern) Vol. 2 (Ed. M Levy) (San Diego: Academic Press, 2001) p. 143
21. Hoen S et al. *Phys. Rev. B* **46** 1874 (1992)
22. Golovnya A V, Pokrovskii V Ya, Shadrin P M *Phys. Rev. Lett.* **88** 246401 (2002)
23. Weber F et al. *Phys. Rev. Lett.* **107** 107403 (2011)
24. Gor'kov L P *Phys. Rev. B* **85** 165142 (2012)
25. Zybtsev S G, Pokrovskii V Ya, Nasretdinova V F, Zaitsev-Zotov S V *Appl. Phys. Lett.* **94** 152112 (2009)
26. Zybtsev S G, Pokrovskii V Ya, Nasretdinova V F, Zaitsev-Zotov S V, in *VII Konf. 'Sil'no Korrelirovannye Elektronnye Sistemy i Kvantovye Kriticheskie Yavleniya' 18 Iyunya 2009 g., g. Troitsk, Moskovskaya Obl., Tezisy Dokladov* (Proc. VII Conf. 'Strongly Correlated Electron Systems and Quantum Critical Phenomena," June 18, 2009, Troitsk, Moscow region) (Troitsk: Inst. of the High Pressure Physics, Russ. Acad. of Sci., 2009) p. 16
27. Zybtsev S G, Pokrovskii V Ya, Nasretdinova V F, Zaitsev-Zotov S V *Physica B* **407** 1696 (2012)
28. Wang Z Z et al. *Phys. Rev. B* **40** 11589 (1989)
29. Slot E, Holst M A, van der Zant H S J, Zaitsev-Zotov S V *Phys. Rev. Lett.* **93** 176602 (2004)
30. Zaitsev-Zotov S V *Microelectron. Eng.* **69** 549 (2003)
31. Zaitsev-Zotov S V *JETP Lett.* **80** 445 (2004) [*Pis'ma Zh. Eksp. Teor. Fiz.* **80** 503 (2004)]
32. Rodin A S, Fogler M M *Phys. Rev. Lett.* **105** 106801 (2010)
33. Monceau P et al. *Phys. Rev. B* **28** 1646 (1983)
34. Grüner G, Zawadowski A, Chaikin P M *Phys. Rev. Lett.* **46** 511 (1981)
35. Monceau P, Richard J, Renard M *Phys. Rev. B* **25** 931 (1982)
36. Zybtsev S G, Pokrovskii V Ya, Zaitsev-Zotov S V *Nature Commun.* **1** 85 (2010)

37. Zybtsev S G, Pokrovskii V Ya *Phys. Rev. B* **84** 085139 (2011)
38. Zybtsev S G, Pokrovskii V Ya, Zaitsev-Zotov S V *Physica B* **407** 1810 (2012)
39. Girault S et al. *Phys. Rev. B* **38** 7980 (1988)
40. Forró L et al. *Phys. Rev. B* **34** 9047 (1986)
41. Ong N P *Phys. Rev. B* **18** 5272 (1978)
42. Sinchenko A A et al. *J. Phys. Condens. Matter* **21** 435601 (2009)
43. Moudden A H et al. *Phys. Rev. Lett.* **65** 223 (1990)
44. Mozurkewich G *Phys. Rev. B* **42** 11183 (1990)
45. Brill J W, Roark W *Phys. Rev. Lett.* **53** 846 (1984)
46. Jacobsen R L, Mozurkewich G *Phys. Rev. B* **42** 2778 (1990)
47. Xu Z G, Brill J W *Phys. Rev. B* **45** 3953 (1992)
48. Xiang X-D, Brill J W *Phys. Rev. B* **36** 2969 (1987)
49. Rivero A J et al. *Solid State Commun.* **106** 13 (1998)
50. Preobrazhenskii V B, Taldenkov A N, Kal'nova I Yu *JETP Lett.* **40** 944 (1984) [*Pis'ma Zh. Eksp. Teor. Fiz.* **40** 183 (1984)]
51. Preobrazhensky V B, Taldenkov A N, Shabanov S Yu *Solid State Commun.* **54** 399 (1985)
52. Pokrovskii V Ya *JETP Lett.* **86** 260 (2007) [*Pis'ma Zh. Eksp. Teor. Fiz.* **86** 290 (2007)]
53. Pokrovskii V Ya et al., in *Workshop on Recent Developments in Low Dimensional Charge Density Wave Conductors, Skradin, Croatia, June 29 – July 3, 2006*, p. 28
54. Pokrovskii V Ya, Zybtsev S G, Gorlova I G *Phys. Rev. Lett.* **98** 206404 (2007)
55. Gorlova I G, Zybtsev S G, Pokrovskii V Ya ''Aktyuator'' (''Actuator''), RF Patent No. 2375688. Application 2007141861. Priority of 14.11.2007. Reg. 10.12.2009
56. Pokrovskii V Ya et al. *Physica B* **404** 437 (2009)
57. Pokrovskii V Ya, Zybtsev S G, arXiv:0708.2694
58. Zybtsev S G, Nikitin M V, Pokrovskii V Ya *JETP Lett.* **92** 405 (2010) [*Pis'ma Zh. Eksp. Teor. Fiz.* **92** 448 (2010)]
59. Nichols J et al. *Phys. Rev. B* **79** 241110(R) (2009)
60. Nichols J et al. *Phys. Rev. B* **80** 039903(E) (2009)
61. Nichols J, Sandamali Weerasooriya C, Brill J W *J. Phys. Condens. Matter* **22** 334224 (2010)
62. Pokrovskii V Ya, Thesis for Doct. Phys.-Math. Sci. (Moscow: Kotel'nikov Inst. of Radioengineering and Electronics, Russ. Acad. of Sci., 2009)
63. Zhang H, Nichols J, Brill J W *Phys. Rev. B* **84** 125134 (2011)
64. Rai R C, Brill J W *Phys. Rev. B* **70** 235126 (2004)
65. Itkis M E, Nad' F Ya, Pokrovskii V Ya *JETP* **63** 177 (1986) [*Zh. Eksp. Teor. Fiz.* **90** 307 (1986)]
66. Bourne L C, Sherwin M S, Zettl A *Phys. Rev. Lett.* **56** 1952 (1986)
67. Bhattacharya S et al. *Phys. Rev. Lett.* **54** 2453 (1985)
68. Bhattacharya S et al. *Phys. Rev. Lett.* **59** 1849 (1987)
69. Bloom I, Marley A C, Weissman M B *Phys. Rev. B* **50** 12218 (1994)
70. Wang Z Z et al. *Solid State Commun.* **46** 325 (1983)
71. Mozurkewich G, Maki M, Grüner G *Solid State Commun.* **48** 453 (1983)
72. Bhattacharya S et al. *Phys. Rev. B* **40** 5826(R) (1989)
73. Pokrovskii V Ya, Zaitsev-Zotov S V, Monceau P, Nad' F Ya *J. Phys. IV France* **3** C2-189 (1993)
74. Bloom I, Marley A C, Weissman M B *Phys. Rev. B* **50** 5081 (1994)
75. Nikitin M V, Zybtsev S G, Pokrovskii V Ya *Phys. Rev. B* **86** 045104 (2012)
76. Zybtsev S G, Pokrovskii V Ya, in *INTERMATIC-2008. Materialy Mezhdunarod. Nauchno-Tekhnicheskoi Konf. 'Fundamental'nye Problemy Radioelektronnogo Priborostroeniya', 21 – 23 Oktyabrya 2008 g.* (Proceedings of Intern. Scientific and Technology Conf. 'Fundamental Problems of Radioelectronic Instruments'', 21 – 23 October 2008) (Ed. A S Sigov) (Moscow: Energoatomizdat, 2008) p. 18
77. Lear R S et al. *Phys. Rev. B* **29** 5656 (1984)
78. Preobrazhensky V B, Taldenkov A N, Kalnova I Ju, in *Charge Density Waves in Solids. Proc. of the Intern. Conf., Budapest, Hungary, September 3 – 7, 1984* (Lecture Notes in Physics, Vol. 217, Eds G Hutiray, J Sólyom) (Berlin: Springer, 1985) p. 357
79. Preobrazhensky V B, Taldenkov A N *Synthetic Met.* **19** 837 (1987)
80. Das K et al. *Phys. Rev. B* **52** 7915 (1995)
81. Klokova N P *Tenzorezistory* (Resistance Strain Gauges) (Moscow: Mashinostroenie, 1990); Laboratory study ''Primenenie tenzorezistorov dlya izmereniya usilii'' (''Application of resistance strain gauges for force measurement''), http://www.masters.donntu.edu.ua/2002/kita/gerchikov/lib/lib5.htm
82. Artemenko S N, Volkov A F, Kruglov A N *Fiz. Nizk. Temp.* **14** 356 (1988)
83. Pokrovskii V Ya, Zaitsev-Zotov S V *Synthetic Met.* **32** 321 (1989)
84. Hennion B, Pouget J P, Sato M *Phys. Rev. Lett.* **68** 2374 (1992)
85. Yomo R et al. *Phys. Rev. B* **71** 132508 (2005)
86. Gorochov O et al. *MRS Bull.* **18** 111 (1983)
87. Hsieh P-L, Jackson C M, Grüner G *Solid State Commun.* **46** 505 (1983)
88. Kikkawa S et al. *Phys. Status Solidi A* **61** K55 (1980)
89. Gorlova I G, Pokrovskii V Ya Gorlova I G, Pokrovskii V Ya *JETP Lett.* **90** 295 (2009) [*Pis'ma Zh. Eksp. Teor. Fiz.* **90** 320 (2009)]
90. Gorlova I G, Pokrovskii V Ya, Zybtsev S G *JETP* **111** 298 (2010) [*Zh. Eksp. Teor. Fiz.* **138** 335 (2010)]
91. Takoshima T et al. *Solid State Commun.* **35** 911 (1980)
92. Zhilinskii S K et al.*Sov. Phys. JETP* **58** 211 (1983) [*Zh. Eksp. Teor. Fiz.* **85** 362 (1983)]
93. Gorlova I G et al. *Physica B* **407** 1707 (2012)
94. Nattermann T, Giamarchi T, Le Doussal P *Phys. Rev. Lett.* **91** 056603 (2003)

Physics – Uspekhi **56** (1) 49 – 78 (2013)

REVIEWS OF TOPICAL PROBLEMS PACS numbers: **47.11. – j**, **47.27. – i**, 47.27.ek, 52.25.Dg

Nonequilibrium Kolmogorov-type particle distributions and their applications

V E Zakharov, V I Karas'

DOI: 10.3367/UFNe.0183.201301c.0055

Contents

Abstract. The paper reviews the current state of research on nonequilibrium (Kolmogorov type) stationary and nonstationary distributions of particles with statically screened Coulomb interaction that are exact solutions of the Boltzmann or Landau collision integral with a source and a sink ensuring the energy flow along the spectrum in momentum space. Analysis is made of the advantages of the new process (based on nonequilibrium distributions) of energy conversion and of the time-dependent nonequilibrium kinetics of an electron–phonon system of a crystal in a strong electric field (electroplastic effect).

V E Zakharov Lebedev Physical Institute,
Russian Academy of Sciences,
Leninskii prosp. 53, 119991 Moscow, Russian Federation,
L D Landau Institute for Theoretical Physics,
Russian Academy of Sciences,
142432 Chernogolovka, Moscow region, Russian Federation
Tel. +7 (499) 132 67 51
E-mail: zakharov@itp.ac.ru
V I Karas' National Scientific Center
'Kharkov Institute of Physics and Technology',
National Academy of Sciences of Ukraine,
ul. Akademicheskaya 1, 61108 Kharkov, Ukraine
Tel. +38 (057) 700 11 46. Fax +38 (057) 335 35 64
E-mail: karas@kipt.kharkov.ua

Received 23 December 2011
Uspekhi Fizicheskikh Nauk **183** (1) 55 – 85 (2013)
DOI: 10.3367/UFNr.0183.201301c.0055
Translated by S D Danilov; edited by A Radzig

1. Introduction

Interest in the nonequilibrium states of various physical systems is steadily growing at present, motivated by the development and extensive use of powerful particle and energy sources.

A universal (independent of the structure of a source and a sink) nonequilibrium stationary energy distribution over wavenumbers ε_k was first proposed by A N Kolmogoroff [1] in the theory of turbulence in an incompressible fluid for the interval of scales $2\pi/k$ intermediate between the scales of forced and efficiently damped motions. The well-known Kolmogorov spectrum of hydrodynamical turbulence has the form

$$\varepsilon_k = AP_1^{2/3}k^{-11/3}\,, \tag{1.1}$$

where A is a constant, P_1 is the spectral energy flux, and k is the wavenumber.

The derivation of formula (1.1) relies on the locality hypothesis in the turbulent motion, i.e., only comparable motion scales significantly interact with each other. This hypothesis remains unproven for the turbulence in an incompressible fluid (strong turbulence, etc.).

In physical systems allowing a description of the interaction between waves or particles in terms of the kinetic equations for waves, quasiparticles, or particles, the derivation of nonequilibrium stationary distributions reduces to solving kinetic equations. In this case, the locality of nonequilibrium stationary distribution corresponds to the convergence of the collision integral. Universal wave spectra that are solutions of the wave collision integral were first found by V E Zakharov [2] in the framework of the theory of weak turbulence of waves.

2. Theoretical studies of nonequilibrium stationary particle distribution functions with flux over spectrum

The Fermi–Dirac or Maxwell distribution function, representing the exact solutions of the respective quantum or classical Boltzmann collision integral, are the thermodynamically equilibrium distribution functions for electrons in degenerate or classical plasmas in the isotropic and spatially homogeneous case [3–7]. For a classical (nondegenerate) gas, the kinetic Boltzmann equation is written down as

$$\frac{\partial f(\mathbf{p})}{\partial t} = \int \mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{p}_2\,\mathrm{d}\mathbf{p}_3 W\left(\mathbf{p},\mathbf{p}_1|\,\mathbf{p}_2,\mathbf{p}_3\right) \times \left[f(\mathbf{p}_2)f(\mathbf{p}_3) - f(\mathbf{p})f(\mathbf{p}_1)\right] \times \delta(E+E_1-E_2-E_3)\,\delta(\mathbf{p}+\mathbf{p}_1-\mathbf{p}_2-\mathbf{p}_3)\,, \quad (2.1)$$

where $W\left(\mathbf{p},\mathbf{p}_1|\,\mathbf{p}_2,\mathbf{p}_3\right)$ is the transition probability due to collisions, $f(\mathbf{p})$ is the electron distribution function, $\mathbf{p}_l$ and E_l are the momentum and energy of an lth electron, and $\delta(x)$ is the Dirac delta function.

A distribution function which satisfies the condition

$$f(\mathbf{p}_2)f(\mathbf{p}_3) - f(\mathbf{p})f(\mathbf{p}_1) = 0 \quad (2.2)$$

is a stationary solution to equation (2.1).

It can readily be seen that functional equation (2.2), with account for energy and momentum conservation laws in particle collisions, leads to the thermodynamically equilibrium Maxwell distribution function.

The question concerning a nonequilibrium distribution function for a small portion of electrons that are relaxing on the equilibrium 'background' (the collision integral allows linearization in this case) in the presence of a uniform energy flux I_1 in momentum space, driven by ionization and recombination processes, has been addressed earlier (see, for example, monograph [8]). Closely related questions on the distribution function for neutrons in crystals were considered by A I Akhiezer and I Ya Pomeranchuk [9]. In these cases, the distribution function may noticeably deviate from the thermodynamically equilibrium solution, while its form turns out to be dependent on the structure of a source and a sink.

Universal nonequilibrium stationary power-law particle distributions ($f = Ap^{2s}$), which are exact solutions of the Boltzmann collision integral, were first obtained by A V Kats, V M Kontorovich, V E Novikov, and S S Moiseev [10, 11] by the group symmetry method. For such distributions to form, the source and sink of particles or energy must exist in momentum space, thus maintaining a constant spectral flux.

In order to determine the power-law exponent s for the transition probability, which is a homogeneous function of momenta of degree n, it is natural to use $\mathbf{p}_l/p$ variables in the integrand. In this case, equation (2.1) reduces to an integral that does not depend on p and the factor p^{4s+n+4}. Let us determine the particle (I_0) and energy (I_1) fluxes in momentum space. The fluxes are linked to the collision integral in the following way:

$$\mathrm{div}\left(j_i(p)\,\frac{\mathbf{p}}{p}\right) = -E^i\left(\frac{\partial f}{\partial t}\right)_{\mathrm{st}}, \quad (2.3)$$

where $I_i = 4\pi p^2 j_i$, and E is the particle energy. Solving equation (2.3) yields

$$I_i = A^2\alpha^{1-i}\,\frac{R(s,n)}{4s+n+9+2(i-1)}\,p^{4s+n+9+2(i-1)}, \quad (2.4)$$

where $\alpha = \mathrm{const}$, and n is the homogeneity index of transition probability.

From Eqn (2.4) it follows that for s_i satisfying the condition

$$\gamma_s = 4s_i + n + 9 + 2(i-1) = 0\,, \quad i = 0, 1\,, \quad (2.5)$$

the flux I_i is either constant in momentum space or zero if $R(s,n)$ has a zero of the first order at $s = s_i$ (the collision integral is then equal to zero). The distribution function Ap^{2s} corresponds to a nonequilibrium stationary case with a constant energy or particle flux. In this case, the flux direction is set by the sign of the derivative $\mathrm{d}R/\mathrm{d}\gamma_s$ at $\gamma_s = 0$, and A is defined by the expression

$$A^2 = I_i\,\alpha^{i-1}\lim_{\gamma_s\to 0}\left|\frac{\mathrm{d}R}{\mathrm{d}\gamma_s}\right|^{-1}. \quad (2.6)$$

Let us demonstrate through direct computations [12, 13] that for the Boltzmann and Landau collision integrals the function $R(s,n)$ satisfies the conditions formulated above in the case of nonequilibrium power-law distributions.

2.1 Exact solutions for the Landau collision integral

It is well known that in plasmas the collision integral describing the interaction between charged particles can be written in the Landau form (see, for example, book [3]):

$$\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{\mathrm{st}} = -\mathrm{div}\,\mathbf{j}_0,\ j_{0i} = \pi e^4\Lambda\int \mathrm{d}\mathbf{p}'\,\frac{u^2\delta_{ik}-u_iu_k}{u^3} \times\left[f(\mathbf{p})\,\frac{\partial f(\mathbf{p}')}{\partial p_k'} - f(\mathbf{p}')\,\frac{\partial f(\mathbf{p})}{\partial p_k}\right], \quad (2.7)$$

where $\mathbf{u} = (\mathbf{p}-\mathbf{p}')/m$, Λ is the Coulomb logarithm, and m and e are the electron mass and charge, respectively. Inserting the isotropic power-law distribution function Ap^{2s} into Eqn (2.7) and performing fairly simple calculations, one arrives at the following expression

$$\left(\frac{\partial f(\mathbf{p})}{\partial t}\right)_{\mathrm{st}} = 16\pi^2 m e^4\Lambda A^2 p^{4s}\,\frac{(4s+3)(4s+5)}{(s+1)(2s+3)(2s+5)} + \frac{16\pi^2 m e^4\Lambda A^2 p^{4s}}{3}\lim_{\substack{p_1\to 0\\ p_2\to\infty}}\left\{\frac{2s^2}{2s+3}\left(\frac{p_1}{p}\right)^{2s+3} + \frac{(2s+1)\,s}{2s+2}\left(\frac{p_2}{p}\right)^{2s+2} - \frac{2s+3}{2}\left(\frac{p_2}{p}\right)^{2s} - \frac{(2s-2)\,s}{2s+5}\left(\frac{p_1}{p}\right)^{2s+5}\right\}. \quad (2.8)$$

According to expression (2.8), the first term of the collision integral, i.e., the function $R(s,n)$, does indeed contain multipliers $4s+5$ and $4s+3$, and to the first power. This ensures, on the one hand, the constancy of energy flux, and, on the other hand, the disappearance of the collision integral for the power-law exponent $s_1=-5/4$. As concerns the exponent $s_0=-3/4$, it corresponds to a nonlocal distribution function, in which case the collision integral diverges (the second term under the limit sign becomes unbounded).

2.2 Exact solutions for the Boltzmann collision integral

Using expression (2.1) for the Boltzmann collision integral and substituting a power-law particle distribution function in the form Ap^{2s}, with the help of the δ-function expressing the momentum conservation law, we integrate expression (2.1) with respect to $\mathbf{p}_2$. Introducing new variables $\mathbf{p}_1$ and $\mathbf{q}$ instead of $\mathbf{p}_1$ and $\mathbf{p}_3$, we then write down the collision integral in the form

$$\frac{\partial f(\mathbf{p})}{\partial t}=-mA^2\int \mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{q}W\bigl(\mathbf{p},\mathbf{p}_1|\,\mathbf{p}+\mathbf{q},\mathbf{p}_1-\mathbf{q}\bigr)\times\left[|\mathbf{p}+\mathbf{q}|^{2s}|\mathbf{p}_1-\mathbf{q}|^{2s}-|p|^{2s}|p_1|^{2s}\right]\delta\bigl(\mathbf{q}(\mathbf{p}_1-\mathbf{p}-\mathbf{q})\bigr)\,, \quad (2.9)$$

where $\mathbf{q}=\mathbf{p}_1-\mathbf{p}_3$. The argument of the δ-function may become zero at $\mathbf{p}_1-\mathbf{p}-\mathbf{q}=0$ or $\mathbf{q}(\mathbf{p}_1-\mathbf{p}-\mathbf{q})=0$. The first case presents no interest, as it simply corresponds to permutation of particles upon their collisions, which sets the expression in square brackets to zero [i.e., it corresponds to condition (2.2)] and, consequently, the condition $(\partial f/\partial t)_{\mathrm{st}}=0$.

Let us introduce the angles θ and θ_1 between the vector $\mathbf{q}$ and vectors $\mathbf{p}$ and $\mathbf{p}_1$, respectively. Changing to spherical coordinates in Eqn (2.9) for $\mathbf{p}_1$ and $\mathbf{q}$, integrating over $\mathbf{p}_1$ with the help of the δ-function, and introducing a dimensionless variable $\tilde{q}$ $(\tilde{q}=q/p)$, we obtain

$$\frac{\partial f(\mathbf{p})}{\partial t}=-mA^2p^{4s+r+4}\int_0^8\tilde{q}\,\mathrm{d}\tilde{q}\int \mathrm{d}O\int \mathrm{d}O_1\tilde{W}\frac{1}{\cos\theta_1}\times\left(\frac{\cos\theta+\tilde{q}}{\cos\theta_1}\right)^2\left\{\left[|1+2\tilde{q}\cos\theta+\tilde{q}^2|^s\right]\times\left|\left(\frac{\cos\theta+\tilde{q}}{\cos\theta_1}\right)^2-2\tilde{q}(\cos\theta+\tilde{q})+\tilde{q}^2\right|^s-\left|\frac{\cos\theta+\tilde{q}}{\cos\theta_1}\right|^{2s}\right\}, \quad (2.10)$$

where r is the dimensionality of the transition probability W, and $\mathrm{d}O_i=\sin\theta_i\,\mathrm{d}\theta_i\,\mathrm{d}\varphi_i$.

Let us determine the particle (I_0) and energy (I_1) fluxes in momentum space, taking into account that in the case under consideration the fluxes can be expressed through the collision integral according to formula (2.3). For the probability W, being a homogeneous function of momenta to power n, $W=C_1q^n$, where C_1 is a constant, and n is any real number (in this case, $n=r$), the collision integral (2.10) can easily be integrated over θ_1, φ, and φ_1 and written down as

$$\left(\frac{\partial f}{\partial t}\right)_{\mathrm{st}}=\frac{2\pi^2mC_1A^2}{s+1}\,p^{4s+n+4}\left[J^{(1)}-J^{(2)}\right],$$

$$J^{(1)}=\int_{-1}^1\mathrm{d}x\int_0^\infty\mathrm{d}\tilde{q}\,\tilde{q}^{n+1}|1+2\tilde{q}x+\tilde{q}^2|^s\,|x|^{2s+2}\,,$$

$$J^{(2)}=\int_{-1}^1\mathrm{d}x\int_0^\infty\mathrm{d}\tilde{q}\,\tilde{q}^{n+1}|x+\tilde{q}|^{2s+2}\,,\quad x=\cos\theta\,.$$

The integral $J^{(2)}$ can be expressed in terms of beta functions $B(x,y)$ as

$$J^{(2)}=(2s+3)^{-1}\bigl[B(2s+4,n+2)-B(-2s-n-5,2s+4)+B(n+2,-2s-n-5)\bigr]\,.$$

To compute $J^{(1)}$, it is convenient first to carry out integration over $\tilde{q}$ (see, for example, handbook [14]), which gives

$$J^{(1)}=2^{-(2s+1)/2}\,\Gamma\left(\frac{1-2s}{2}\right)B(n+2,-n-2s-2)\times\int_0^1\mathrm{d}x\,x^{2s+2}(1-x^2)^{(2s+1)/4}\bigl[P^{s+1/2}_{n+s+3/2}(-x)+P^{s+1/2}_{n+s+3/2}(x)\bigr]\,,$$

where $P^{\mu}_{\nu}(x)$ is the spherical function. Making use of the property of spherical functions, namely

$$P^{\mu}_{\nu}(-x)=-\frac{\sin(\pi\nu)}{\sin(\pi\mu)}\,P^{\mu}_{\nu}(x)+\frac{\sin\bigl(\pi(\nu+\mu)\bigr)\,\Gamma(\nu+\mu+1)}{\sin(\pi\mu)\,\Gamma(\nu-\mu+1)}\,P^{-\mu}_{\nu}(x)\,,$$

and computing integrals over x, we arrive at

$$J^{(1)}=\frac{B(n+2,-n-2s-2)}{2s+3}\left[\left(1-\frac{\sin\bigl(\pi(s+n+3/2)\bigr)}{\sin\bigl(\pi(s+1/2)\bigr)}\right)\times{}_3F_2\left(\frac{n+2}{2},-\frac{n+2s+2}{2},1;\frac{1-2s}{2},\frac{2s+5}{2};1\right)+\frac{\Gamma(2s+n+3)\,\Gamma(1/2-s)\,\Gamma(s+5/2)\,\Gamma(s+2)}{2^{2s+1}\Gamma(n+2)\,\Gamma\bigl((4s+n+7)/2\bigr)\,\Gamma\bigl((2s-n+3)/2\bigr)}\times\frac{\sin\bigl(\pi(2s+n+2)\bigr)}{\sin\bigl(\pi(s+1/2)\bigr)}\right],$$

where ${}_pF_q(\alpha_1,\ldots,\alpha_p;\beta_1,\ldots,\beta_q;z)$ is the hypergeometric function. Thus, the expression for the function $R(s,n)$ entering into a particular solution for the flux I_i [see formula (2.4)], with account for the expressions obtained earlier for $J^{(1)}$ and $J^{(2)}$, takes the form

$$R(s,n)=\frac{-4\pi^3C_1}{(s+1)(2s+3)}\left\{B(2s+4,\,n+2)-B(-2s-n-5,\,2s+4)+B(n+2,\,-2s-n-5)-\left[B(n+2,\,-n-2s-2)\left(1-\frac{\sin\bigl(\pi(s+n+3/2)\bigr)}{\sin\bigl(\pi(s+1/2)\bigr)}\right)\times{}_3F_2\left(\frac{n+2}{2},-\frac{n+2s+2}{2},1;\frac{1-2s}{2},\frac{2s+5}{2};1\right)+\frac{(4s+n+7)(4s+n+9)\,\pi^2(2s+1)(2s+3)}{2^{2s+5}\Gamma(-2s)\,\Gamma\bigl((4s+n+11)/2\bigr)\,\Gamma\bigl((2s-n+3)/2\bigr)}\times\frac{\Gamma(s+2)}{\sin^2\bigl(\pi(s+1/2)\bigr)}\right]\right\}.$$

In Section 2.1 we considered a quadratic law for particle dispersion. The generalization to an arbitrary dispersion law $E=p^{\varsigma_1}/\alpha_1$ (where α_1 and ς_1 are some constants) does not lead to any principal complications, although the expression for $R(s,n,\varsigma_1)$ may prove to be more cumbersome. As for the power-law exponent s_i in the particle distribution function, it is defined for a general dispersion relation by the following

expression

$$s_i = -\frac{n+9+\varsigma_1(i-1)}{2\varsigma_1}\,. \tag{2.11}$$

It has been found that local nonequilibrium particle distribution functions (the collision integral converges for them) are associated with the power-law exponent s satisfying the conditions

$$-\frac{3}{2} < s_0 < -1\,, \quad -\frac{3}{2} < s_1 < -\frac{5}{4}\,. \tag{2.12}$$

Subscripts 0 (1) stand for $I_0\,(I_1) = \text{const}$.

According to condition (2.5), the following homogeneity indices for the transition probability correspond to the exponents from the ranges (2.12):

$$\begin{aligned} &-3 < n < -1\,, \quad I_0 = \text{const}\,,\\ &-4 < n < -3\,, \quad I_1 = \text{const}\,. \end{aligned} \tag{2.13}$$

2.3 Formation conditions for nonequilibrium stationary particle distribution functions in finite energy intervals

As follows from inequalities (2.13) formulated above, the collision integral diverges (the known singularity of W for small transferred momenta) in the case of the Coulomb interaction ($n = -4$). References [10, 11, 15] propose and Refs [12, 13] show that this divergence is alleviated by Debye screening. Consider the collision integral (2.10); for the transition probability, which corresponds to the screened Coulomb potential $W = 2e^4/(q^2 + a_1^2)^2$ (with q being the transferred momentum, and a_1 the Debye momentum), we get

$$\begin{aligned} \frac{\partial f(\mathbf{p})}{\partial t} = -mA^2 p^{4s} \int_0^\infty \tilde{q}\,\mathrm{d}\tilde{q} \int \mathrm{d}O \int \mathrm{d}O_1\, \frac{2e^4}{(\tilde{q}^2 + a^2)^2}\\ \times \frac{1}{\cos\theta_1}\left(\frac{\cos\theta + \tilde{q}}{\cos\theta_1}\right)^2 \Bigg\{ \left|1 + 2\tilde{q}\cos\theta + \tilde{q}^2\right|^s\\ \times \left| \left(\frac{\cos\theta + \tilde{q}}{\cos\theta_1}\right)^2 - 2\tilde{q}(\cos\theta + \tilde{q}) + \tilde{q}^2 \right|^s\\ - \left|\frac{\cos\theta + \tilde{q}}{\cos\theta_1}\right|^{2s} \Bigg\}, \end{aligned} \tag{2.14}$$

where $\tilde{q} = q/p$, and $a = a_1/p$.

We integrate over the angles θ_1, φ, φ_1, and θ, making use of standard integrals [14], and transform Eqn (2.14) to the sum of two integrals:

$$\begin{aligned} \frac{\partial f(\mathbf{p})}{\partial t} = \frac{-8\pi^2 m e^4 A^2}{(s+1)(2s+3)}\, p^{4s} \Bigg\{ \int_0^1 \frac{\mathrm{d}\tilde{q}\,\tilde{q}}{(\tilde{q}^2 + a^2)^2} \Bigg[(1+\tilde{q}^2)^s\\ \times \left({}_2F_1\left(-s, 2s+3; 2s+4; -\frac{2\tilde{q}}{1+\tilde{q}^2}\right) + {}_2F_1\left(-s, 2s+3; 2s+4; \frac{2\tilde{q}}{1+\tilde{q}^2}\right)\right)\\ - (1+\tilde{q})^{2s+3} - (1-\tilde{q})^{2s+3} \Bigg] + \int_0^1 \frac{\mathrm{d}\tilde{q}\,\tilde{q}^{-2s-2}}{(\tilde{q}^2 a^2 + 1)^2}\\ \times \Bigg[\tilde{q}^3 (1+\tilde{q}^2)^s \left({}_2F_1\left(-s, 2s+3; 2s+4; -\frac{2\tilde{q}}{1+\tilde{q}^2}\right) + {}_2F_1\left(-s, 2s+3; 2s+4; \frac{2\tilde{q}}{1+\tilde{q}^2}\right)\right)\\ - (1+\tilde{q})^{2s+3} + (1-\tilde{q})^{2s+3} \Bigg] \Bigg\}. \end{aligned} \tag{2.15}$$

To find the existence domains for power-law distribution functions which correspond to two different asymptotics of the transition probability W, the collision integral in two limiting cases: $a \ll 1$ and $a \gg 1$ is worth consideration. First, for small values of a ($a \ll 1$), we find the dependence of collision integral (2.15) on a. It can readily be seen that the main contribution in this case comes from the first integral in expression (2.15). Expanding the integrand in powers of $\tilde{q}$ and integrating the resulting series term by term, we obtain the expression for the collision integral:

$$\begin{aligned} \frac{\partial f(\mathbf{p})}{\partial t} = \frac{8\pi^2 m e^4 A^2}{(s+1)(2s+3)}\, p^{4s} \Bigg\{ \frac{(4s+3)(4s+5)}{2s+5} \Bigg[\ln a^2\\ + \frac{(2s+1)(2s+3)(2s+5)\,\pi^2 \Gamma(s+2)}{2^{2s+5}\Gamma(-2s)\,\Gamma((2s+7)/2)\,\Gamma((4s+7)/2)\,\sin^2(\pi(2s+1)/2)} \Bigg]\\ + K_1(s)\, a^2 \ln a^2 + \ldots \Bigg\}. \end{aligned}$$

Computing the energy flux according to formula (2.4), we find that the energy flux is negative and is determined by the logarithmic term only for $a^2 < 0.005$. Within the interval $0.005 < a^2 < 0.1$, the energy flux is directed oppositely (it is positive) to its direction for large momenta.

Thus, it is shown [12, 13] that in the region of momentum space $p \gg a_1$ the Debye screening, first, removes the Coulomb divergence and, second, does not change the exponent of the nonequilibrium stationary particle distribution function for constant energy flux in momentum space. The power-law exponent in this function corresponds to the asymptotics of W with the exponent $n = -4$. Additionally, it is established that within a certain domain in momentum space the direction of energy flux is opposite (positive) to its direction for large momenta. Then there is a local power-law distribution, the particle density in which is determined by the flux intensity. The conservative character of flux is ensured by the source and sink, the positions of which must agree with the flux direction found.

Numerous concrete physical tasks face the question of the formation, under the action of sources and sinks in momentum space, of power-law particle distributions in bounded energy intervals surrounded by domains where the particle distributions are in thermodynamic equilibrium. The electron–electron collision integral in a solid state plasma is computed in the approximation of the quadratic dispersion law. The divergence caused by the Coulomb interaction between particles is removed by the introduction, as above, of a matrix element describing the screened Coulomb interaction. The Boltzmann collision integral in the case of quantum statistics (see, for example, book [3]) can be represented in the form

$$\begin{aligned} \frac{\partial \eta(\mathbf{p})}{\partial t} = \frac{2}{(2\pi\hbar)^6} \int \mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{p}_2\,\mathrm{d}\mathbf{p}_3\, W(\mathbf{p}, \mathbf{p}_1 | \mathbf{p}_2, \mathbf{p}_3)\\ \times \left[\eta(\mathbf{p}_2)\,\eta(\mathbf{p}_3)\big(1-\eta(\mathbf{p})\big)\big(1-\eta(\mathbf{p}_1)\big) - \eta(\mathbf{p})\,\eta(\mathbf{p}_1)\right.\\ \left.\times \big(1-\eta(\mathbf{p}_2)\big)\big(1-\eta(\mathbf{p}_3)\big)\right]\\ \times \delta(E + E_1 - E_2 - E_3)\,\delta(\mathbf{p} + \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{p}_3)\,, \end{aligned} \tag{2.16}$$

where $W = (2\pi\hbar)^3 2e^4 \left(|\mathbf{p}_1 - \mathbf{p}_3|^2 + a_1^2\right)^{-2}$ is the matrix element describing the screened Coulomb interaction, and $\eta(\mathbf{p}_i)$ are the occupation numbers.

Let the electron distribution function be power-law in the inertial range (between the source and sink), and thermo-

dynamically equilibrium (Fermi) outside it, namely

$$\eta_i = \eta_s(\mathbf{p}_i), \text{ if } p' \leqslant |\mathbf{p}_i| \leqslant p'' \left(\eta_s = \frac{(2\pi\hbar)^3}{2}\,\alpha |I_1|^{1/2} p_i^{2s}\right),$$

$$\eta_i = \eta_{\mathrm{F}}(\mathbf{p}_i), \text{ if } p' > |\mathbf{p}_i|, |\mathbf{p}_i| > p'' \left(\eta_{\mathrm{F}} = \left[1 + \exp\frac{p_i^2 - p_{\mathrm{F}}^2}{2mk_{\mathrm{B}}T}\right]^{-1}\right),$$

where α is the proportionality factor, I_1 is the energy flux, p_{F} is the Fermi momentum, k_{B} is the Boltzmann constant, and T is the equilibrium temperature of electron gas. The collision integral is computed with the goal to determine the extent of the inertial range for the power-law distributions η_s; therefore, the momentum p over which there is no integration in Eqn (2.16) should lie in this interval: $\eta = \eta_s(p)$. It has been shown above that the nonequilibrium distribution function may have the form $\eta_s = A_1 p^{2s}$ for metals in the region of momentum space where $a_1 < p$, i.e., the source and sink must be arranged so that the conditions $p', p'' > p_{\mathrm{F}}$ are satisfied. When computing the integral in Eqn (2.16), we will neglect thermal smearing of the Fermi distribution function, because it will only lead to corrections which are insignificant by virtue of the condition $T \ll E_{\mathrm{F}}$ (E_{F} is the Fermi energy). Thus, a step function of unity amplitude can be taken for η_{F}:

$$\eta_{\mathrm{F}} = \theta(p_{\mathrm{F}}^2 - p_i^2)\,,$$

and the unit function is defined hereinafter as

$$\theta(x) = \begin{cases} 0, & x < 0\,, \\ 1, & x \geqslant 0\,. \end{cases}$$

Suppose that $\eta_s \ll 1$ in the entire inertial range. We will evaluate expression (2.16) with an accuracy up to terms logarithmic in a_1, similarly as for the Landau collision integral. This implies that we will account for interactions with a small momentum transfer in collisions. It has been shown above that for power-law distribution functions the nonlogarithmic terms in the Boltzmann collision integral, related to large momentum transfer in collisions, turn out to be important if momenta lie in a certain range characteristic of a solid state plasma in conductors. Computation of nonlogarithmic terms in this case faces considerable difficulties, so the conditions to be derived below take into account only weak momentum exchange and will be valid for the plasma in semiconductors, whereas in the plasma of conductors they are more stringent than necessary. Bearing in mind the remarks and clarifications made above, the collision integral can be represented as

$$\begin{aligned}\frac{\partial\eta(\mathbf{p})}{\partial t} = {} & \frac{2}{(2\pi\hbar)^6}\int_{0\leqslant|p_1|\leqslant p'} \mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{p}_2\,\mathrm{d}\mathbf{p}_3 W \\ & \times \left[\eta_2\eta_3\theta(p_1^2 - p_{\mathrm{F}}^2) - \eta_s\eta_1(1-\eta_2)(1-\eta_3)\right] \\ & \times \delta(E + E_1 - E_2 - E_3)\,\delta(\mathbf{p} + \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{p}_3) \\ & + \frac{2}{(2\pi\hbar)^6}\int_{p'\leqslant|p_1|\leqslant p''} \mathrm{d}\mathbf{p}_1\,\mathrm{d}\mathbf{p}_2\,\mathrm{d}\mathbf{p}_3 W \\ & \times \left[\eta_2\eta_3 - \eta_s\eta_{s1}(1-\eta_2)(1-\eta_3)\right] \\ & \times \delta(E + E_1 - E_2 - E_3)\,\delta(\mathbf{p} + \mathbf{p}_1 - \mathbf{p}_2 - \mathbf{p}_3)\,. \end{aligned} \quad (2.17)$$

We begin with the second integral on the right-hand side of Eqn (2.17), and perform the same manipulations as in the derivation of Eqn (2.10). Integrating over φ and φ_1 and substituting W, which corresponds to the screened Coulomb interaction, we obtain

$$\left(\frac{\partial\eta(p)}{\partial t}\right)_{p'}^{p''} = \int_{0\leqslant(x+q)\,y\leqslant p''} \frac{q\,\mathrm{d}q}{(q^2+a^2)^2}\int \mathrm{d}x \int \mathrm{d}y F(y,x,q) - \int_{0\leqslant(x+q)\,y\leqslant p'} \frac{q\,\mathrm{d}q}{(q^2+a^2)^2}\int \mathrm{d}x \int \mathrm{d}y\, F(y,x,q)\,, \quad (2.18)$$

where

$$F(y,x,q) = \frac{16\pi^2 e^4 m A^2 p^{4s}}{(2\pi\hbar)^3}\,|y|(x+q)^2 \times \left\{(1+2qx+q^2)^s\left[y^2(x+q)^2 - 2qx + q^2\right]^s - |y|^{2s}|x+q|^{2s}\right\},$$

$$q = \frac{|\mathbf{p}_1 - \mathbf{p}_3|}{|\mathbf{p}_1|}\,,\ x = \cos\theta\,,\ y = (\cos\theta_1)^{-1}\,,\ \cos\theta = \frac{|\mathbf{p}\,\mathbf{q}|}{|\mathbf{p}|\,|\mathbf{q}|}\,,$$

$$\cos\theta_1 = \frac{|\mathbf{p}_1\,\mathbf{q}|}{|\mathbf{p}_1||\mathbf{q}|}\,,\ A = \frac{(2\pi\hbar)^3}{2}\,\alpha|I_1|^{1/2}\,.$$

In order to write down the integration limits in Eqn (2.18), one needs to find the domains of existence for the inequalities $0 \leqslant (x+q)\,y \leqslant p''$ and $0 \leqslant (x+q)\,y \leqslant p'$, taking into account that $p'' > 1$, $p' < 1$, $|x| < 1$, and $1 \leqslant |y| < \infty$.

Since the function $F(y,x,q)$ is even with respect to y, after carrying out the change of variable $z = y^2(x+q)^2$, formula (2.18) can be reduced to the form

$$\begin{aligned}\left(\frac{\partial\eta(p)}{\partial t}\right)_{p'}^{p''} = {} & \int_0^{p''-1}\mathrm{d}q\int_{-1}^{1}\mathrm{d}x\int_{(x+q)^2}^{p''^2}\mathrm{d}z\,F_1(z,x,q) \\ & + \int_{p''-1}^{p''+1}\mathrm{d}q\int_{-1}^{p''-q}\mathrm{d}x\int_{(x+q)^2}^{p''^2}\mathrm{d}z\,F_1(z,x,q) \\ & - \int_0^{1-p'}\mathrm{d}q\int_{-p'-q}^{p'-q}\mathrm{d}x\int_{(x+q)^2}^{p'^2}\mathrm{d}z\,F_1(z,x,q) \\ & - \int_{1-p'}^{1+p'}\mathrm{d}q\int_{-1}^{p'-q}\mathrm{d}x\int_{(x+q)^2}^{p'^2}\mathrm{d}z\,F_1(z,x,q)\,, \end{aligned} \quad (2.19)$$

where

$$F_1(z,x,q) = \frac{8\pi^2 e^4 m A^2 p^{4s}}{(2\pi\hbar)^3}\,\frac{q}{(q^2+a^2)^2} \times \left\{(1+2qx+q^2)^s(z-2qx+q^2)^s - z^s\right\}.$$

We will not resolve the behavior of the power-law distribution function near the energy source and sink, i.e., at $p' \approx 1$ and $p'' \approx 1$. Relatedly, to maintain accuracy up to the terms logarithmic in a, it suffices to compute contributions from the first and third terms on the right-hand side of Eqn (2.19). Having performed integration over z and x, we need to expand the integrand in series in q and, since the main contribution comes from $q \leqslant a$, retain only the terms up to those proportional to q^3.

Integration of the first and third terms in Eqn (2.19) results in the following expression

$$\left(\frac{\partial\eta(p)}{\partial t}\right)_{p'}^{p''} = \frac{16\pi^2 e^4 m A^2 p^{4s}}{(2\pi\hbar)^3} \times \left\{\left[\frac{(4s+5)(4s+3)}{(2s+3)(2s+5)(s+1)} + \frac{s(2s+1)}{3(s+1)}\,(p''+1)^{s+1} \times {}_3F_2\left(-s-1, s+\frac{3}{2}, s+1; s+\frac{1}{2}, s; \frac{1}{p''^2+1}\right)\right] \times \int_0^{\min(1,p''-1)} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2} + \left[p'^{2s+1} - \frac{2s(-s+1)}{(2s+5)(s+1)} \times p'^{2s+5} - \frac{s(4s+5)}{(2s+3)(s+1)}\,p'^{2s+3} - \frac{2s+3}{3}\,p'^{2s+1}(1+p'^2)^s \times {}_3F_2\left(-s, s+\frac{5}{2}, s+2; s+\frac{3}{2}, s+1; \frac{p'^2}{p''^2+1}\right)\right] \times \int_0^{\min(p', 1-p')} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2}\right\}.$$

Let us evaluate the first integral in Eqn (2.17) carrying out the same manipulations as above, with subsequent integration by the formula

$$\int_\alpha^\beta f(x) \prod_{k=1}^n \theta\bigl(\varphi_k(x)\bigr)\,\mathrm{d}x = F(\beta) \prod_{k=1}^n \theta\bigl(\varphi_k(\beta)\bigr) - F(\alpha) \prod_{k=1}^n \theta\bigl(\varphi_k(\alpha)\bigr) - \sum_{l=1}^n \sum_{i=1}^{m_l} F(a_i^l)\,\frac{\varphi_l'(a_i^l)}{|\varphi_l'(a_i^l)|} \times \prod_{\substack{k=1\\k\neq l}}^n \theta\bigl(\varphi_k(a_i^l)\bigr)\bigl[\theta(\beta-a_i^l)\,\theta(a_i^l-\alpha) - \theta(a_i^l-\beta)\,\theta(\alpha-a_i^l)\bigr].$$

The expression for the first integral in Eqn (2.17) will take the form

$$\left(\frac{\partial\eta(p)}{\partial t}\right)_0^{p'} = \frac{32\pi^2 e^4 m A p^{2s}}{(2\pi\hbar)^3}\,p_\mathrm{F}^3\,\frac{s}{3}\int_0^{p'} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2}\,.$$

The final expression for the energy flux in momentum space, which is defined through the collision integral according to relationship (2.3), becomes

$$I_1 = -\frac{32\pi^3 e^4 A^2 p^{4s+5}}{(2\pi\hbar)^3}\left\{\frac{4s+3}{(2s+5)(2s+3)(s+1)} \times \int_0^{\min(1,p''-1)} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2} - \frac{s}{3(s+2)}\,p'^{2s+1}\int_0^{p'} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2}\right\} - \frac{32\pi^3 e^4 A p^{2s+5}}{(2\pi\hbar)^3}\,\frac{s}{3(2s+5)}\,p_\mathrm{F}^3\int_0^{p'} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2}\,. \tag{2.20}$$

From the last relationship, one can derive a condition to be imposed on the electron density in the vicinity of a sink (the density which governs the source intensity) to keep the flux I_1 constant and, consequently, the distribution η_s close to the universal one in the inertial range (the interval between the source and sink). This condition will be satisfied if the first term in Eqn (2.20) is much larger that the other two:

$$\eta_s(p) \gg p_\mathrm{F}^3\left|\frac{s(2s+3)(s+1)}{3(4s+3)}\right| \int_0^{p''} \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2} \times \left(\int_0^1 \frac{q^3\,\mathrm{d}q}{(q^2+a^2)^2}\right)^{-1}.$$

Thus, an explicit expression was found for the quantum Boltzmann collision integral as a function of the momentum p, exponent s, momentum p' corresponding to the energy sink, and momentum p'' corresponding to the energy source. It may be concluded that in the absence of particles outside the energy range which is located between the source and sink, the exponent s differs from the universal one by no more than 10% in the limits of the inertial range:

$$|p''-p'| \approx (5-6)\,p_\mathrm{ch}\,, \quad p' \approx p_\mathrm{ch} = 2a_1\,, \tag{2.21a}$$

References [16, 17] consider a case more often encountered in solid state plasmas, where the electron distribution function has a power-law form in the interval between the energy source and sink in momentum space and follows the thermodynamically equilibrium Fermi–Dirac distribution outside it. It is shown that a nonequilibrium stationary distribution of electrons is close to the universal one ($s = -5/4$ for an unbounded inertial range) if the location of source and sink, as well as their intensities, satisfies the conditions specified below. So, the power-law exponent s in the distribution η_s (η_s are the occupation numbers) will differ from $-5/4$ by less than 10% if the following conditions are met:

$$|p''-p'| \approx (5-6)\,p_\mathrm{ch}\,,\ \eta_s(p) \gg 10^{-3}\,,\ p' \approx p_\mathrm{ch} = (2-3)\,a_1\,. \tag{2.21b}$$

Thus, the nonequilibrium universal electron distribution is possible even if occupation numbers are significantly less (by one or two orders of magnitude) than the equilibrium ones.

3. Theoretical studies of nonequilibrium nonstationary particle distribution functions with flux over spectrum

3.1 Numerical modeling of the formation of particle distribution functions for Landau–Fokker–Planck type equations

3.1.1 Fully conservative difference schemes for Landau–Fokker–Planck type equations. The nonlinear kinetic Boltzmann equation [3–8] describes a system of many particles interacting by the laws of classical mechanics, and serves as the basic equation in models of rarefied gas dynamics. In its general form, the equation for the particle distribution function f_α, depending on spatial coordinates $\mathbf{r}$, velocities $\mathbf{v}$, and time t, can be written down as

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v}\,\frac{\partial f_\alpha}{\partial \mathbf{r}} + \frac{\mathbf{F}_\alpha}{m_\alpha}\,\frac{\partial f_\alpha}{\partial \mathbf{v}} = \left(\frac{\partial f_\alpha}{\partial t}\right)_\mathrm{st} + S_\alpha\,, \tag{3.1}$$

where m_α is the mass of α particles, F_α is the force acting on an α particle, S_α are the sources (sinks) of particles and energy (open systems), and $(\partial f_\alpha/\partial t)_\mathrm{st}$ is the collision integral. It stands as the fundamental equation in applications requiring a mathematical description of the dynamics of rarefied gases and plasmas. Models of kinetic processes governed by Coulomb collisions occupy a prominent place in applications related to laboratory and magnetospheric plasmas, plasma-chemistry, and solid state plasma.

In plasmas, the Coulomb collisions involving small-angle scattering (with small transferred momentum) are more important than collisions with a substantial change in

velocity (momentum). The collision integral for charged particles was first derived by L D Landau (see, for example, book [3]) from the Boltzmann collision integral, with account for the smallness of momentum transferred in Coulomb collisions and the effect of particle charge screening by other particles outside a sphere with Debye radius. A Landau type equation is also obtained by accounting for small-angle scattering, but for arbitrary potentials of interaction between particles [18–28]. The most pertinent object for modeling with the help of the Landau type collision integral are gases with power-law potentials and an infinite radius of interaction between particles.

The most widely used method of numerical modeling is (regular) finite difference method (Monte Carlo type methods suffer from certain drawbacks for long-range forces). However, a combined approach should be employed, ensuring approximation of equations describing actual dynamics and adequate representation of physical laws underlying the physical model. This approach, based on conservative and fully conservative finite-difference schemes, is related to the work done by A N Tikhonov, A A Samarskii, Yu P Popov, A V Bobylev, I F Potapenko, V A Chuyanov, and others [18–28].

Such an approach, together with high accuracy, gave the capability of obtaining asymptotic solutions in certain cases and testing analytical approaches and other modeling methods.

In a spatially homogeneous case, the Boltzmann collision integral for the distribution function $f(\mathbf{v}, t)$ was written in the form (2.1), where the transition probability W due to collisions is defined by the differential scattering cross section $\sigma(u, \mu)$, which is a function of the relative velocity $u > 0$ and the parameter $\mu = \cos\theta$ ($0 \leqslant \theta \leqslant \pi$ is the scattering angle). Finding the scattering cross section $\sigma(u, \mu)$ for a given interaction potential $U(r)$ is a well-known task in classical mechanics (see, for example, Refs [22–28]).

The collision integral in the Landau form for particles of the same kind carrying the charge e in a spatially-homogeneous case was given in Section 2.1 [see Eqn (2.7)] and may be represented in the form

$$\frac{1}{\Gamma}\left(\frac{\partial f(\mathbf{v})}{\partial t}\right)_{\mathrm{st,\,L}} = \frac{\partial}{\partial v_i}\left\{\int \mathrm{d}\mathbf{w} U_{ij}\left(\frac{\partial}{\partial v_j} - \frac{\partial}{\partial w_j}\right) f(\mathbf{v}) f(\mathbf{w})\right\},$$

$$U_{ij} = \frac{u^2\delta_{ij} - u_i u_j}{u^3}, \qquad (3.2)$$

where $\Gamma = 2\pi e^4 \Lambda / m^2$, and the symmetric kernel U_{ij} is a function of particle relative velocity $\mathbf{u} = \mathbf{v} - \mathbf{w}$. A rough condition for the applicability of equation (3.2) is furnished by the inequality $e^2 n^{1/3} \ll T$, implying that the mean energy of Coulomb interaction is small compared to the mean kinetic energy (n is the particle number density, and T is the temperature expressed in energy units).

3.1.2 Numerical modeling of relaxation of the particle distribution function for the Landau–Fokker–Planck collision integral. Let us consider the relaxation of the initial distribution for gases of particles with power-law interaction potentials in the case of isotropic distribution function $f(\mathbf{v}, t) = f(|\mathbf{v}|, t)$, which obeys the Landau–Fokker–Planck equation in the symmetric form

$$\frac{\partial f}{\partial t} = \frac{1}{v^2}\frac{\partial}{\partial v}\left\{\frac{1}{v}\int_0^\infty \mathrm{d}w\, Q(v, w)\left[wf(w)\frac{\partial f(v)}{\partial v} - vf(v)\frac{\partial f(w)}{\partial w}\right]\right\}, \qquad (3.3)$$

where $Q(v, w)$ is the symmetric kernel, namely

$$Q(v, w) = \frac{\pi}{8} v^3 w^3 \int_{-1}^{1} \mathrm{d}\mu (1 - \mu^2)\, u\sigma(u),$$

$$u^2 = v^2 + w^2 - 2vw\mu.$$

It can be verified that in the absence of sources and sinks the conservation laws for the particle density and energy follow from equation (3.3):

$$n = 4\pi \int_0^\infty \mathrm{d}v\, v^2 f(v, t) = \mathrm{const},$$

$$k_{\mathrm{B}} T = \frac{4\pi m}{3n}\int_0^\infty \mathrm{d}v\, v^4 f(v, t) = \mathrm{const}, \quad t \geqslant 0.$$

The only equilibrium stationary solution to the kinetic Boltzmann equation is furnished by the Maxwell distribution. Exploring the formation of the particle distribution function at energies essentially exceeding the mean energy (the tails of the particle distribution function), apart from its academic interest, can be helpful in tasks related to wave–particle interactions, electron acceleration by a field and tail formation by runaway electrons, and for the problem of thermonuclear fusion (electron cyclotron heating, lower-hybrid resonance, nuclear reactions at the tails of the ion distribution function, and so on). For the Coulomb potential, such research has been conducted in Refs [21–28].

The equilibrium solution takes on the form

$$f_{\mathrm{M}} = \frac{4}{\pi^{1/2}}\left(\frac{3}{2}\right)^{3/2} \exp\left(-\frac{3}{2}v^2\right), \quad v_{\mathrm{th}} = 1,$$

where v_{th} is the thermal velocity. We obtain the conservation laws from equation (3.3) by integrating both its sides over the velocity with an appropriate weight:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{J}{v}\bigg|_0^\infty = 0, \quad \frac{\mathrm{d}E}{\mathrm{d}t} = (vJ)\big|_0^\infty - 2\int_0^\infty J\,\mathrm{d}v = 0.$$

By applying the integro-differential method to equation (3.3), we obtain the following (implicit) difference scheme:

$$\frac{f_i^k - f_i^{k-1}}{\tau} = \frac{1}{v_i^2 h_{i+1/2}}\left[\frac{J_{i+1/2}^k}{v_{i+1/2}} - \frac{J_{i-1/2}^k}{v_{i-1/2}}\right], \quad i = 2, \ldots, M. \qquad (3.4)$$

The scheme is set forth as the local conservation law for the particle number, whereas the approximation of function J_i must ensure the implementation of energy conservation law. We replace the upper integration limit in Eqn (3.3) by the value of velocity at the last point. Approximating the integrals with the trapezoidal rule, and derivatives with centered differences, yield

$$J_{i+1/2}^k = \sum_{m=1}^{M} Q_{i+1/2,\,m+1/2}\, h_{m+1} \times \left(\frac{f_{i+1}^k - f_i^k}{h_{i+1}}\,\frac{f_m^k v_m + f_{m+1}^k v_{m+1}}{2} - \frac{f_{m+1}^k - f_m^k}{h_{m+1}}\,\frac{f_i^k v_i + f_{i+1}^k v_{i+1}}{2}\right),$$

where $Q_{i+1/2,\,m+1/2} = Q_{m+1/2,\,i+1/2}$. The particle number density is written out as

$$n=\sum_{i=1}^{M-1}\frac{f_i v_i^2+f_{i+1}v_{i+1}^2}{2}\,h_{i+1}$$
$$=\sum_{i=1}^{M-1}\frac{f_i v_i^2}{2}\,h_{i+1}+\sum_{i=2}^{M}\frac{f_i v_i^2}{2}\,h_i=\sum_{i=2}^{M-1}\frac{f_i v_i^2}{2}\,h_i\,,$$

since it is assumed that $f_{M+1}=0$ and that $v_1=0$. Accordingly, the first point does not enter into the difference equation. Let us compute the change in particle number over a time step with the aid of difference equation (3.4) by summing both its parts with an appropriate weight:

$$\Delta n=\sum_{i=2}^{M-1}h_i v_i^2\left(f_i^k-f_i^{k-1}\right)=\tau\left[\frac{J_{M+1/2}^k}{v_{M+1/2}}-\frac{J_{3/2}^k}{v_{3/2}}\right].$$

Assume that $J_{M+1/2}=0$, since $f_M=0$ at the tail, and formally set $J_{3/2}=0$ (we will learn below about its implications), then $\Delta n=0$. The change in energy is given by

$$\Delta E=\sum_{i=2}^{M-1}h_i v_i^4\left(f_i^k-f_i^{k-1}\right)$$
$$=\tau\left[\frac{v_{M+1}^2}{v_{M+1/2}}\,J_{M+1/2}-\frac{v_2^2}{v_{3/2}}\,J_{3/2}-2\sum_{i=2}^{M-1}h_{i+1}J_{i+1/2}\right].$$

If we take into account that the particle number is conserved, and that the relationship

$$\sum_{i=1}^{M-1}h_{i+1}J_{i+1/2}=0$$

is valid, then $\Delta E\approx 0$. Difference equation (3.4) is written out at point $i=2$ in the following way:

$$\frac{f_i^k-f_i^{k-1}}{\tau}=\frac{1}{v_i^2 h_{i+1/2}}\,\frac{J_{i+1/2}^k}{v_{i+1/2}}\,,\quad i=2\,. \tag{3.4a}$$

We need to know the boundary condition for the distribution function at the point $i=1$, and we derive it from the condition $J_{3/2}=0$.

For a numerical solution, the finite-difference scheme is rearranged into a system of nonlinear difference equations solved at each time step and, accordingly, at each iteration by the sweep method.

Let us rewrite the scheme in a form convenient for numerical computations. To do so, the following notation is introduced:

$$\Sigma_{i+1}=\sum_{m=1}^{M-1}h_{m+1}Q_{i+1/2,m+1/2}(v_m f_m+v_{m+1}f_{m+1})\,,$$

$$\Theta_{i+1}=\sum_{m=1}^{M-1}Q_{i+1/2,m+1/2}(f_{m+1}-f_m)\,,$$

$$A_i=\frac{1}{v_{i-1/2}\,h_i}\,(\Sigma_i-v_i h_i\Theta_i)\,,\quad B_i=\frac{1}{v_{i-1/2}h_i}\,(\Sigma_i+v_i h_i\Theta_i)\,,$$

$$\alpha_i=\frac{\tau}{v_i^2 h_i}\,A_{i+1}\,,\quad \beta_i=\frac{\tau}{v_i^2 h_i}\,(A_i+B_{i+1})\,,$$

$$\gamma_i=\frac{\tau}{v_i^2 h_i}\,B_i\,,\quad \psi_i=-f_i^{k-1}\,,\quad i=2,\ldots,M-1\,.$$

The set of difference equations is then written out as

$$\alpha_i f_{i+1}-(1+\beta_i)f_i+\gamma_i f_{i-1}=\psi_i\,,\ i=2,\ldots,M-1\,, \tag{3.5}$$

with the first difference equation involving

$$\alpha_2=\frac{\tau}{v_2^2 h_i}\,A_3\,,\quad \beta_2=\frac{\tau}{v_2^2 h_i}\,B_3\,,\quad \gamma_2=0\,.$$

From the equality $J_{3/2}=0$, it follows that the functions at the initial points are linked to each other at any moment of time by the following relationship

$$f_1=f_2\left(1-v_2^2\,\frac{N_2}{E_2}\right). \tag{3.6}$$

If the boundary condition of the second kind (equaling zero of the derivative) is approximated based on the formulation of an exact problem with the first- or second-order approximation, the conservation law is not necessarily warranted. What does the approximation imply if it is derived formally from the requirement that a difference analog of the conservation law be implemented? The boundary condition arising formally proves to be quite reasonable. Consider, for definiteness, charged particles and the kernel $Q(v,w)$ corresponding to their Coulomb interactions. Suppose the distribution function is quasi-Maxwellian: $f_i^{\mathrm{M}}=C\exp\left(-v_i^2/v_{\mathrm{th}}^2\right)$. Then, estimating expression (3.4a) and neglecting terms on the order of $O(h^3)$, we find

$$f_1^{\mathrm{M}}=f_2^{\mathrm{M}}\left(1+\frac{v_2^2}{v_{\mathrm{th}}^2}\right).$$

We only note that, initially conceived as just a formality, the requirement of full conservatism leads to a more natural approximation of exact conditions in a subsequent analysis of exact initial data. We revisit this question when constructing a scheme for the Landau–Fokker–Planck equation.

Since the difference schemes are chosen implicit, the time step τ is determined by the required solution accuracy and the distribution character. The scheme is nonmonotonic; it can be made monotonic but then we have to sacrifice the second-order spatial approximation.

Since the distribution function should be nonnegative for stability and by its physical sense, the coefficients of difference equation (3.5) should satisfy the conditions α, $\gamma>0$, $1+\beta>0$, which requires $h_i=v_i-v_{i-1}$, $h_i<E_i/n_i v_i$. The Landau–Fokker–Planck equation is a parabolic one, and as $v\to\infty$ it degenerates into an almost hyperbolic equation; therefore, to account for the exponentially decreasing functions, the step should even be reduced in order to ensure accuracy. A rough discretization step estimate based on velocity, which can be over or under for particular tasks, takes the form $h<v_{\mathrm{th}}^2/nL$.

For power-law interaction potentials $U=a/r^{\beta}$, where $1<\beta<4$, the symmetric kernel $Q(v,w)$ is represented in the form [24, 25]

$$Q(v,w)=\frac{a(v,w)(v+w)^{n_1+4}+b(v,w)|v-w|^{n_1+4}}{(n_1+2)(n_1+4)(n_1+6)}\,, \tag{3.7}$$

$$a(v,w)=(n_1+4)\,vw-(v^2+w^2)\,,$$

$$b(v,w)=(n_1+4)\,vw+(v^2+w^2)\,,\quad n_1=\frac{\beta-4}{\beta}\,.$$

The negative values of n_1 correspond to soft potentials ($1 \leqslant \beta \leqslant 4$). For charged particles, one has $\beta = 1$ ($n_1 = -3$).

To illustrate the performance of the difference scheme, the initial distribution is chosen as a δ-function which is approximated in the following manner:

$$f(v_i, 0) = \begin{cases} \dfrac{2}{v_{i+1} - v_{i-1}}, & v_i = 1, \\ 0, & v_i \neq 1. \end{cases}$$

Such an approximation models the difference delta function; relatedly, both the number density and energy of particles become equal to unity. We limit ourselves to only presenting results of simulation for this type of initial distribution.

A relaxation of initial distribution to the equilibrium state belongs to classical problems of plasma physics and is among the test problems for any model of collisional plasma.

Moving along, we note that the particle number density is preserved with a machine accuracy (random error), while the energy depends on the accuracy of δ iterations because of the nonlinearity of the equation. For $\delta = 10^{-3}$, the relative error makes $\Delta\varepsilon = 10^{-2}\%$, and for $\delta = 10^{-7}$, it is $\Delta\varepsilon = 10^{-5}\%$, i.e., the energy error does not exceed 10^{-7}. Such a level of accuracy is needed, for example, in exploring asymptotic problems.

Finite initial conditions are considered at $t = 0$, with the initial function $f_0(v) = f(v, 0)$ being confined to the thermal range $v \approx v_{\text{th}}$.

At long times, when the relaxation in the thermal domain is completed, the equation can be considered as a linear one in the hot domain for $v \gg v_{\text{th}}$:

$$\frac{\partial f}{\partial t} = \frac{n}{8} \frac{1}{v^2} \frac{\partial}{\partial v} \left[v^3 \sigma(v) \left(\frac{T}{m} \frac{\partial f}{\partial v} + vf \right) \right].$$

Here, we took into account that in the hot domain the kernel $Q(v, w) \to \pi w^3 v^4 \sigma(v)/6$ as $v \to \infty$, and used the scattering cross section $\sigma(u, \mu) = g_\beta(\mu)\, u^{-4/\beta}$ [22–25] for the power-law interaction potential. Substituting the expression for $Q(v, w)$ into Eqn (3.3), for $v \gg v_{\text{th}}$ we arrive at the final form of the last equation:

$$\frac{\partial f}{\partial t} = \frac{n g_\beta}{8} \frac{1}{v^2} \frac{\partial}{\partial v} \left[v^{3-4/\beta} \left(\frac{T}{m} \frac{\partial f}{\partial v} + vf \right) \right], \tag{3.8}$$

where the following notation was introduced:

$$g_\beta = 2\pi \int_{-1}^{1} \mathrm{d}\mu\, g_\beta(\mu)(1 - \mu^2). \tag{3.9}$$

For example, formula (3.9) for the Coulomb potential ($\beta = 1$) gives $g_1 = 32\pi e^4 L/m$. In this case, Eqn (3.8) reduces to the well-known linear Landau–Fokker–Planck equation for plasmas.

Further, let us consider equation (3.8) for arbitrary potentials $U = \alpha/r^\beta$, $1 \leqslant \beta < 4$. The finite-difference scheme has already been presented above. For the chosen variables, the conservation laws for the particle number density and energy look like

$$n = \int_0^\infty v^2\, \mathrm{d}v f(v, t) = 1, \quad E = \int_0^\infty v^4\, \mathrm{d}v f(v, t), \quad t \geqslant 0. \tag{3.10}$$

The thermal velocity $v_{\text{th}} = 1$, and the equilibrium solution is written down as

$$f_{\text{M}}(v) = \frac{4}{\pi^{1/2}} \left(\frac{3}{2} \right)^{3/2} \exp\left(-\frac{3}{2} v^2 \right).$$

At the initial moment of time, the function is the delta function $f(v, 0) = \delta(v - 1)/v^2$ confined in the thermal domain. When choosing the boundary L for the velocity interval, hot particles are estimated from the Maxwell distribution, which yield $L \approx (7-8)\, v_{\text{th}}$. The difference scheme expressed by formula (3.5) is used. The initial function is approximated as mentioned above. For this approximation, the number of particles equals one, and the kinetic energy $v_i^2 = 1$.

We present below the results of modeling. The function smears over the thermal domain ($0 < v < 2$) rather rapidly and then becomes monotonic, attaining its maximum at zero at a certain moment t_0 which corresponds to the so-called collisional time. This characteristic time t_0 is only weakly dependent on the exponent β in the expression for the particle interaction potential $U = \alpha/r^\beta$. In this domain, the distribution functions for different exponents stay rather close to each other during the entire relaxation process for various potentials. The distinction is more prominent in the hotter domain with $v > 2$ and at distribution tails. To make the presentation of results more explicit, the function normalized on the Maxwell distribution, $g(v, t) = f(v, t)/f_{\text{M}}(v)$, is introduced. Figures 1–3 plot the graphs of function $f(v, t)$ in the velocity interval $2 < v/v_{\text{th}} < 5$ for different time moments and values of the parameter β. All curves demonstrate the wave character of propagation of $g(v, t)$ toward the domain of high velocities. For the Coulomb interaction ($\beta = 1$, $n_1 = -3$, $U = \alpha/r$), the function $g(v, t)$ is shown in Fig. 1. The solution has the character of a wave propagating with a stable (stiff) front. For the particle interaction potentials with $1 < \beta < 2$, the tail relaxation proceeds more slowly than the relaxation of the distribution core. Beginning with $2 < \beta < 3$, the tail evolution gradually loses the wave character. Figure 2 demonstrates that the wave front slowly rises with time ($\beta = 2$, $n_1 = -1$, $U = \alpha/r^2$). For stiff potentials ($\beta > 4$), the characteristic relaxation times for the distribution core and

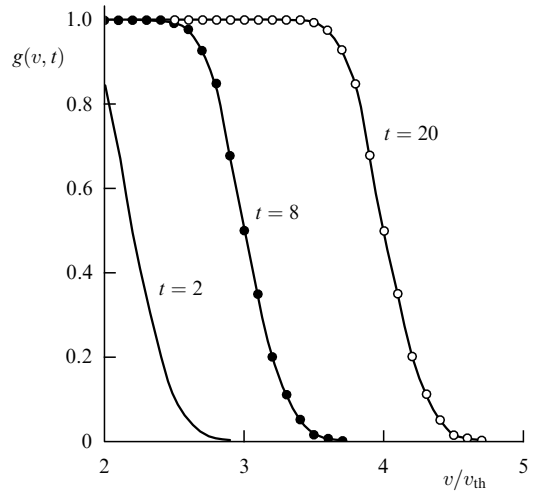


Figure 1. Plots of the distribution function normalized to the Maxwell distribution, $g(v, t) = f(v, t)/f_{\text{M}}(v)$, for velocities in the range $2 < v/v_{\text{th}} < 5$ for several moments of time t (in arbitrary units). The time is normalized on the time of electron–electron collisions (3.11). Plots correspond to the case of Coulomb interactions: $n_1 = -3$, $\beta = 1$, $U = \alpha/r$.

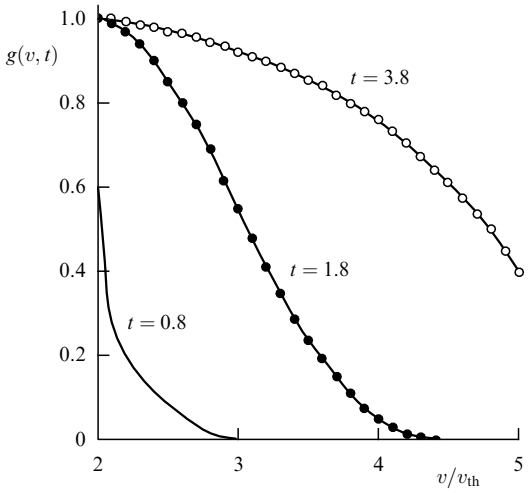


Figure 2. Plots of the distribution function normalized to the Maxwell distribution, $g(v,t) = f(v,t)/f_{\rm M}(v)$, for velocities in the range $2 < v/v_{\rm th} < 5$ for several moments of time t (in arbitrary units). The time is normalized on the time of electron–electron collisions (3.11). Plots correspond to the case of dipole interactions: $n_1 = -1$, $\beta = 2$, $U = \alpha/r^2$.

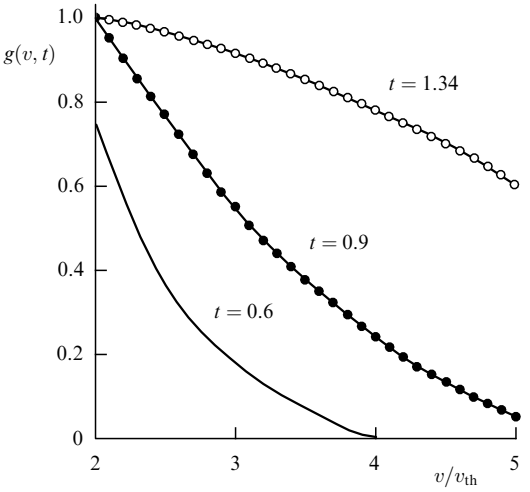


Figure 3. Plots of the distribution function normalized to the Maxwell distribution, $g(v,t) = f(v,t)/f_{\rm M}(v)$, for velocities in the range $2 < v/v_{\rm th} < 5$ for several moments of time t (in arbitrary units). The time is normalized on the time of electron–electron collisions (3.11). Plots correspond to the case of interactions of Maxwell molecules: $n_1 = 0$, $\beta = 4$, $U = \alpha/r^4$.

tail are practically indistinguishable (times of 1–2). As a consequence, the velocities of propagation and smearing are hardly distinguishable (Fig. 3), and the relaxation in thermal and hotter domains occurs simultaneously.

As can be concluded, the front width $\Delta_{\rm f}(t)$ is essentially dependent on the power-law exponent β in the interaction potential. Indeed, $\Delta_{\rm f}(t)$ takes a constant value for $1 < \beta < 2$, i.e., the propagation of $g(v,t)$ in this case has the character of a wave with a stable profile, which does not smear with time for $t > 0$. This fact was discovered numerically at $\beta = 1$ in Refs [22–25], and the respective analytical solution was developed. Beginning from the exponent $\beta = 2$, the front starts to slightly smear out, following a weak logarithmic dependence on time: $\Delta_{\rm f}(t) \propto \sqrt{\ln t}$. The solution should preserve its wave character. The propagation speed $v_{\rm f}(t)$ of a front is higher than the rate of its smearing.

To facilitate the analysis, equation (3.8) is recast in a convenient form by introducing dimensionless variables for the velocity, time, and distribution function:

$$x = \left[\frac{v}{v_{\rm th}}\right]^{(4+\beta)/2\beta} X^{(4+\beta)/2\beta}, \quad \tau = t\,\frac{\rho g_\beta}{8}\,\frac{4-\beta}{\beta}\left[\frac{X}{v_{\rm th}}\right]^{(4-\beta)/2\beta}, \tag{3.11}$$

$$f(v,t) = f_{\rm M}(v)\,u(x,\tau)\,,$$

where $X = 2\beta(4-\beta)/(4+\beta)^2$, and $f_{\rm M}(v)$ is the Maxwell distribution function. Substituting representation (3.11) into Eqn (3.8), we obtain the equation for the distribution function $u(x,\tau)$ in dimensionless variables:

$$\frac{\partial u}{\partial \tau} + \frac{x^{1-p}}{p}\,\frac{\partial u}{\partial x} = \frac{1}{2}\,\frac{\partial^2 u}{\partial x^2}\,, \quad p = 2\,\frac{4-\beta}{4+\beta}\,. \tag{3.12}$$

The type of the last equation gives an idea about the wave character of its solution. Indeed, it is worth mentioning that we are interested in the hot domain $x \gg 1$ and the times at which $u(x,\tau) \approx 1$ and slow establishment of the equilibrium solution $u_{\rm M}(x) = 1$ occurs at tails, while $u(x,\tau) \to 0$ as $x \to \infty$. The condition $x \gg 1$ will be taken into account with the aid of new variables $\tilde{x} = x/x_0$ and $\tilde{\tau} = \tau/x_0^p$, where $x_0 \gg 1$ is some characteristic scale. In this case, equation (3.12) acquires a small parameter $x_0^{p-2} \ll 1$ of the higher derivative. Hence, it follows that the Landau equation changes its type in $x \to \infty$ tails and becomes a transport operator. Taking this into account, we introduce a formal parameter ε of the second derivative. We obtain, as a result, that

$$\frac{\partial u}{\partial \tau} + \frac{x^{1-p}}{p}\,\frac{\partial u}{\partial x} = \frac{\varepsilon}{2}\,\frac{\partial^2 u}{\partial x^2}\,. \tag{3.13}$$

At $\varepsilon \approx 0$, the last equation reduces to a first-order equation, and the equilibrium solution $u_{\rm M}(x)$ is simply carried along its characteristics. The typical solution of equation (3.13) has the form of a step function:

$$u(x,\tau) \approx \theta(\tau^{1/p} - x)\,, \quad p = 2\,\frac{4-\beta}{4+\beta}\,. \tag{3.14}$$

Approximate solution (3.14) properly reflects the asymptotic law followed by the wave front, $x_{\rm f}(\tau) \propto \tau^{1/p}$, but does not describe the front structure. In order to analyze it, equation (3.13) is transformed with due regard for the information on the wave front behavior obtained earlier. Introducing new variables

$$z = \frac{x - \tau^{1/p}}{\sqrt{\varepsilon}}\,, \quad u(x,\tau) = \phi(z,\tau)\,,$$

we rewrite equation (3.13) for the new function $\phi(z,\tau)$ as

$$\frac{\partial \phi}{\partial \tau} = \frac{1}{2}\left\{\frac{\partial^2 \phi}{\partial z^2} - \frac{2}{p}\,\frac{\tau^{1/p} - 1}{\sqrt{\varepsilon}}\left[\left(1 + \sqrt{\varepsilon}\,\frac{z}{\tau^{1/p}}\right)^{1-p} - 1\right]\frac{\partial \phi}{\partial z}\right\}. \tag{3.15}$$

Suppose ε is small but retain z finite. In this conditional limit, formula (3.15) reduces to

$$\frac{\partial \phi}{\partial \tau} = \frac{1}{2}\left(\frac{\partial^2 \phi}{\partial z^2} - \gamma\,\frac{z}{\tau}\,\frac{\partial \phi}{\partial z}\right), \quad \gamma = 2\,\frac{1-p}{p}\,. \tag{3.16}$$

The boundary conditions for the function $\phi(z,\tau)$ are as follows: $\phi \to 1$ when $z \to -\infty$ (behind the front), and $\phi \to 0$ when $z \to 0$ (ahead the front). The function $\phi(z,\tau)$ is constructed as a solution to the initial value problem for time $\tau > 1$ with the respective initial condition $\phi(z,1) = \theta(-z)$ [cf. function (3.14)]. By an appropriate variable change, equation (3.16) reduces to the heat conduction equation [24, 25]. As a result, a self-similar solution to equation (3.16) is obtained:

$$\phi(z,\tau) = \Phi\left[z\sqrt{\frac{1-\gamma}{2(\tau-\tau^{\gamma})}}\right], \quad \tau > 1,$$
$$\Phi(z) = \frac{1}{\sqrt{\pi}}\int_z^{\infty} \mathrm{d}y \, \exp(-y^2). \tag{3.17}$$

Returning to the variable x and recalling that $\varepsilon = 1$, we find the quasistationary solution to equation (3.17):

$$u(x,\tau) \approx \Phi\left[\frac{x-\tau^{1/p}}{\sqrt{2}}\sqrt{\frac{1-\gamma}{\tau-\tau^{\gamma}}}\right], \quad \tau > 1, \quad \gamma = 2\,\frac{1-p}{p}. \tag{3.18}$$

In order to learn where solution (3.18) is valid, let us formulate all the assumptions made en route. The kinetic equation is considered for $x \gg 1$ and a large time (larger than the Coulomb collision time) $\tau \gg 1$ close to the wave front $x_{\mathrm{f}}(\tau) \propto \tau^{1/p}$, $x - \tau^{1/p} \ll \tau^{1/p}$. The last inequality justifies the passage from Eqn (3.15) to Eqn (3.16), subject to the condition $\varepsilon = 1$, and leads to the following constrains. The derived solution (3.18) is inapplicable, first, in the interval $0 < x \ll x_{\mathrm{f}}$, and, second, for particles outside the region $x > 2x_{\mathrm{f}}$. Neither case is interesting from a practical viewpoint because, with a good accuracy, it can be assumed that $u(x,\tau) \approx 1$ in the first region $x \ll x_{\mathrm{f}}(t)$, and $u(x,\tau) \approx 0$ in the second region $x \geqslant 2x_{\mathrm{f}}$. Formula (3.18) can be simplified with account for the condition $\tau \gg 1$. Since the final result depends on γ, three cases may be considered for the expression under the radical sign in formula (3.18). The expression $(1-\gamma)(\tau-\tau^{\gamma})^{-1}$ for $\gamma < 1$ tends to $(1-\gamma)\,\tau^{-1}$, and for $\gamma > 1$ it tends to $(\gamma-1)\,\tau^{-\gamma}$. The case of $\gamma = 1$ is a boundary one and is different from the other cases. Notice that $\gamma = 1$ corresponds to $\beta = 2$, i.e., the dipole interaction between particles. In this limiting case, the expression is rewritten as

$$(1-\gamma)(\tau-\tau^{\gamma})^{-1} = (1-\gamma)\left(\tau^{\gamma}\sum_{i=1}^{\infty}\frac{\left[(1-\gamma)\,\ln\tau\right]^i}{i!}\right)^{-1}.$$

The last expression contains only the first term at $\gamma = 1$, because the coefficients of the sum disappear for all $i > 2$.

Thus, simplified expressions for the asymptotic solution $u(x,\tau)$, which is a function of the exponent γ, are arrived at under the condition $\tau \gg 1$:

$$u(x,\tau) = \Phi\left\{\frac{x-\tau^{1/p}}{\sqrt{2}}\sqrt{\frac{1-\gamma}{\tau}}\right\}, \quad \gamma < 1;$$
$$u(x,\tau) = \Phi\left\{\frac{x-\tau^{1/p}}{\sqrt{2}}\sqrt{\frac{1}{\tau\ln\tau}}\right\}, \quad \gamma = 1; \tag{3.19}$$
$$u(x,\tau) = \Phi\left\{\frac{x-\tau^{1/p}}{\sqrt{2}}\sqrt{\frac{\gamma-1}{\tau}}\right\}, \quad \gamma > 1.$$

Finally, we reformulate the results obtained above in terms of original variables v, t, and for the distribution function $f_{\beta}(v,t)$:

$$f_{\beta}(v,t) \approx n\left(\frac{m}{2\pi T}\right)^{3/2}\exp\left(-\frac{mv^2}{2T}\right) \times u\left(v\sqrt{\frac{m}{T}},\; t\,\frac{\rho g_{\beta}}{8}\left(\frac{m}{T}\right)^{(4-\beta)/2\beta}\right). \tag{3.20}$$

The constant factor g_{β} is defined by formula (3.9). The function $u(v,t)$ has the form of a propagating wave, the front of which travels according to the law

$$v_{\mathrm{f}}(t) = \left(\frac{4-\beta}{\beta}\right)^{\beta/(4-\beta)}.$$

The function $u(v,t)$ is described by three expressions depending on the value of the exponent β:

$$u(v,t) = \Phi\left\{2\,\frac{\sqrt{\beta(2-\beta)}}{4+\beta}\,v_{\mathrm{f}}V^{(4+\beta)/2\beta}\right\}, \quad 1 \leqslant \beta < 2;$$
$$u(v,t) = \Phi\left\{\frac{1}{3}\,(\ln v_{\mathrm{f}})^{-1/2}v_{\mathrm{f}}V^{3/2}\right\}, \quad \beta = 2; \tag{3.21}$$
$$u(v,t) = \Phi\left\{\sqrt{\frac{2(\beta-2)}{4-\beta}}\left[\frac{2\beta(4-\beta)}{4+\beta}\right]^{(4-\beta)/2\beta} \times v_{\mathrm{f}}^{(4+\beta)/\beta}V^{(4+\beta)/2\beta}\right\}, \quad 2 < \beta < 4,$$

where $V = (v - v_{\mathrm{f}}(t))/v_{\mathrm{f}}(t)$.

The usability conditions for solution (3.21) have the form of strong inequalities:

$$v \gg 1, \quad v_{\mathrm{f}}(t) \gg 1, \quad \left|v - v_{\mathrm{f}}(t)\right| \ll v_{\mathrm{f}}(t).$$

Notice that $u\bigl(v_{\mathrm{f}}(t),t\bigr) = 1/2$. As usual, we define the front width in the following form

$$\Delta_{\mathrm{f}}(t) = 2\left|\frac{u(v,t)}{\partial u(v,t)/\partial v}\right|_{v=v_{\mathrm{f}}} = \left|\frac{\partial u(v,t)}{\partial v}\right|^{-1}_{v=v_{\mathrm{f}}}.$$

Consequently, for the cases considered above we obtain

$$\Delta_{\mathrm{f}}(t) = \sqrt{\frac{\pi\beta}{2-\beta}}, \quad 1 \leqslant \beta < 2;$$
$$\Delta_{\mathrm{f}}(t) = \sqrt{\pi\ln v_{\mathrm{f}}(t)}, \quad \beta = 2; \tag{3.22}$$
$$\Delta_{\mathrm{f}}(t) = \sqrt{\frac{\pi\beta}{2-\beta}\,\frac{2\beta(4-\beta)}{(4+\beta)^2}}\left[v_{\mathrm{f}}(t)\right]^{2(\beta-2)/\beta}, \quad 2 < \beta < 4.$$

It can be seen that the front width $\Delta_{\mathrm{f}}(t)$ essentially depends on the exponent β in the interaction potential. Thus, $\Delta_{\mathrm{f}}(t)$ takes a constant value for $1 < \beta < 2$. The propagation of $u(v,t)$, in this case, has the character of a wave with a stable profile which is left unsmeared with time for $t > 0$. For $\beta \geqslant 2$, the front gradually smears out, showing a weak logarithmic dependence $\Delta_{\mathrm{f}} \propto \sqrt{\ln t}$. The solution behavior should still preserve its wave character. The front propagation speed exceeds the rate at which it is smeared out:

$$\frac{\partial\Delta_{\mathrm{f}}/\partial t}{\partial v_{\mathrm{f}}/\partial t} = \frac{\mathrm{d}\Delta_{\mathrm{f}}}{\mathrm{d}v_{\mathrm{f}}}.$$

3.2 Numerical modeling of the formation of nonequilibrium particle distribution functions for stationary self-consistent sources and sinks

In this section, we analyze the formation of nonequilibrium quasistationary distribution functions in a spatially homogeneous isotropic plasma composed of one kind of particles, in the presence of localized sources (sinks) of particles (energy) in velocity space. The existence of nonequilibrium distribution functions assumes the presence of sources and sinks in momentum space. An energy (particle) source and sink can be maintained by ion beams, powerful laser radiation, emission current, beams of charged particles emitted in fusion or fission reactions, etc. The analysis is based on the Landau–Fokker–Planck equation, which is a model of the Boltzmann equation for arbitrary interaction potentials of particles [21–26].

We concentrate now on a numerical study of the evolution of nonequilibrium distribution functions and their dependence on various input parameters — the intensity of sources, the degree of their localization in velocity space, and so forth. Section 4 compares results simulated numerically with those obtained in a particular experiment dealing with the irradiation of a semiconductor by fast ions [29–32]. Fully conservative difference schemes [22–28] are applied for numerical modeling.

Numerical implementation of the problem solved here faces a fundamental difficulty rooted in the nonlinearity of the collision integral. As has already been mentioned, two conservation laws should be obeyed in the absence of external sources; otherwise, the dissipative properties of the difference scheme may distort the result through the influence of implicit sinks or sources. For this reason, numerical schemes capable of appropriately handling the nonlinearity of the modelled equation are employed [26].

Nonlinear operator (3.3) with the symmetric kernel $Q(v, w)$ for power-law interaction potentials (3.7) is utilized in numerical modeling. For charged particles $\beta = 1$ ($n_1 = -3$), so then $Q(v, w) = (2/3)\, w^3$ for $w < v$, and $Q(v, w) = (2/3)\, v^3$ for $w > v$.

Following a common procedure of making equations dimensionless, we change to the variables

$$v' = \frac{v}{v_{\rm th}}\,; \quad t' = \frac{\tilde{t}}{t_\beta}\,, \quad t_\beta = \frac{32\pi v_{\rm th}^3}{n g_\beta}\, v_{\rm th}^{4(\beta-1)/\beta};$$
$$f' = \frac{4\pi v_{\rm th}^3}{n} f; \quad S' = \frac{4\pi v_{\rm th}^3 t_S S}{n_S}\,. \tag{3.23}$$

To implement the difference scheme, the infinite interval in velocity space is replaced by a finite interval $[0, v_{\max}]$ selected so as to account for high-energy particles, and the boundary condition for the distribution function is taken as $f(v_{\max}, t) = 0$. The sources S_+ and sinks S_- in most cases are chosen as δ-functions:

$$S_\pm = \frac{I_\pm\, \delta(v - v_\pm)}{v^2}\,. \tag{3.24}$$

If the intensities of a source and sink satisfy the relationship $I_+ = I_- v_-^2 / v_+^2$, the energy acquired by the system from outside equals zero, but the particle density in the system decreases (if the source is associated with larger velocities than the sink), i.e., in this situation an analog of constant energy flux in momentum space is realized with nonconservation of particle density in the system. Since we are dealing with charged particles, as the electron density drops in a certain region, thermal electrons from its surroundings tend to replenish it, driven by the arising electric field. In the framework of the spatially homogeneous model considered here, this can be taken into account by introducing yet another source with an intensity $I_{\rm th}$ so as to compensate for the decrease in particle number and, respectively, the generation of the electric field. In this manner, we can formulate a consistent model with two sources of intensities I_+ and $I_{\rm th}$ and one sink of intensity I_-, in the framework of which neither the energy nor the particle density will change. This will be achieved if the intensities of sources and sinks satisfy the two relationships

$$I_{\rm th} - I_- + I_+ = 0\,, \quad I_{\rm th} v_{\rm th}^2 - I_- v_-^2 + I_+ v_+^2 = 0\,. \tag{3.25}$$

They can be rewritten as the expressions for the intensities of sources in terms of the sink intensity I_-:

$$I_+ = I_- \frac{v_-^2 - v_{\rm th}^2}{v_+^2 - v_{\rm th}^2}\,, \quad I_{\rm th} = I_- \frac{v_+^2 - v_-^2}{v_+^2 - v_{\rm th}^2}\,. \tag{3.26}$$

Additionally, in numerical computations we will sometimes consider sources (sinks) which are distributed exponentially over velocities:

$$S_\pm \propto I_\pm \exp\left[-b(v - v_\pm)^2\right]. \tag{3.27}$$

Such an expression for the source or sink proves to be convenient for exploring the dependence of nonequilibrium distribution functions on the source shape. We remark that, strictly speaking, the introduction of sinks which are independent of the distribution function may face severe problems. In this arrangement, one may specify such initial conditions that solutions will become negative in the vicinity of the sink as time progresses. For this reason, the sources (sinks) are frequently modelled by terms that are proportional to the distribution function sought after, namely

$$S_\pm = I_\pm \frac{\delta(v - v_\pm)}{v^2}\, f(v, t)\,. \tag{3.28}$$

It will be recalled that in the discrete case the function $\delta(v - v_1)$ differs from zero only at $v = v_1$. The initial distribution is chosen as Maxwellian or δ-functional. It should be noted that results are practically insensitive to the shape of the initial distribution function, except for the very initial stage. Iterations are carried out on each time step, the particle number is preserved up to the machine precision, and the energy is conserved up to 7–8 significant figures.

Later, we discuss the main results of numerical simulations in the presence of particle or energy fluxes in momentum space [29–31]. There are the source S_+ and sink S_- on the right-hand side of kinetic equations (3.3) to maintain the flux in momentum space. We begin with the case when the source and sink conform in momentum space with the direction of flux transferred as a result of collisions. Notice that analytical consideration of equations for a localized source and sink gives the correct flux direction — from high to low velocities. It can be seen from Fig. 4 that, as time progresses, a nonequilibrium stationary (Kolmogorov type) particle distribution is established in the interval between the source and sink, in agreement with the presence of energy flux in momentum space, while outside this interval the distribution function remains a thermodynamically equilibrium one. To convince ourselves once again that the location of source and

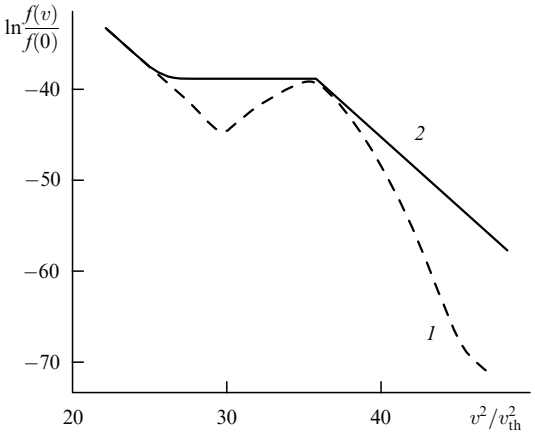


Figure 4. Dependence of the logarithm of the nonequilibrium stationary distribution function normalized to its value at $v = 0$ on the dimensionless velocity squared. The computations use the initial Maxwell distribution function and the source function with $I = 10^{-16}$, $v_- = 5$, $v_+ = 6$, and are carried out for the Landau–Fokker–Planck equation at the instants of time $t_2 = 100$ (solid line) and $t_1 = 25$ (dashed line).

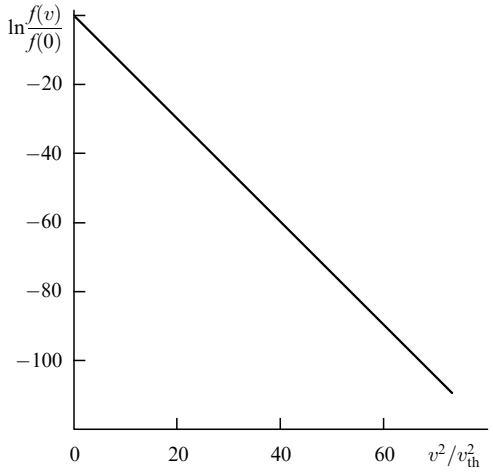


Figure 5. Dependence of stationary (equilibrium) distribution function for nonconforming location of the source and sink, $v_+ = 5$, $v_- = 7$, obtained from computations based on the Fokker–Planck and also the Landau equations for $\beta = 1, 2, 3$.

sink need to conform with the flux direction in momentum space, we carried out computations with opposite locations of the source and sink in energy space. Figure 5 displays the logarithm of distribution function versus (dimensionless) velocity squared for the incorrect arrangement of the source and sink. Apparently, the particle distribution function stays in thermodynamic equilibrium as the flux intensity is varied over several orders of magnitude — this reinforces the importance of conforming the placement of a source and sink.

For the variant presented in Fig. 6, the functional dependence of the source and sink on velocity is exponential. The source S_+ occupies a 'narrow' domain in the vicinity of energies that corresponds to seven thermal velocities, while the sink S_- is also sufficiently local in the region of four thermal velocities. The localization regions of the source and sink are controlled by the value of coefficient α_1 in the exponential function. In the case considered, the coefficient α_1 is fairly large (it equals 100), which ensures their strong localization. We explore the dependence of the electron distribution function on the degree of source and sink localization in energy space. With this aim, the magnitude of coefficient α_1 was reduced in the next simulations by one order to $\alpha_1 = 10$. From the comparison of results presented in Figs 6 and 7, it can be concluded that the character of the nonequilibrium stationary distribution in the main domain between the source and sink does not depend on the degree of source (sink) localization, which witnesses in favor of the locality of the distribution function.

Figure 8 plots the dependence of distribution functions on the square of the dimensionless velocity for various flux intensities. It is found that for small intensities of source I_+ (sink I_-), the universal nonequilibrium distribution is formed in the vicinity of velocities $v \leqslant v_+$, which is caused, first, by a reduction in the Coulomb scattering cross section as velocity is increased ($\sim v^{-3}$) and, second, by the always existing flux of energy and particles (because of the Coulomb diffusion) into the region of the basic, 'background' equilibrium distribution function. For this reason, as the intensity is increased, the universal nonequilibrium particle distribution forms, tending to occupy an ever larger region between the source and sink, which is related to the reduction in the relative role of the flux

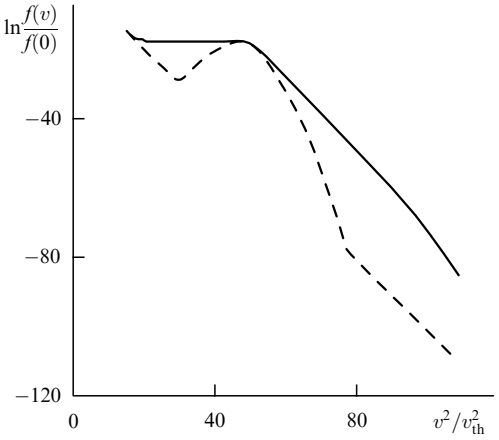


Figure 6. Distribution function computed from the Fokker–Planck equation for the source (sink) $S_\pm \sim I_\pm \exp\{-\alpha_1(v - v_\pm)^2\}$, $\alpha_1 = 100$, $v_- = 4$, and $v_+ = 7$. The dashed and solid lines correspond to time moments $t = 25$ and $t = 100$, respectively.

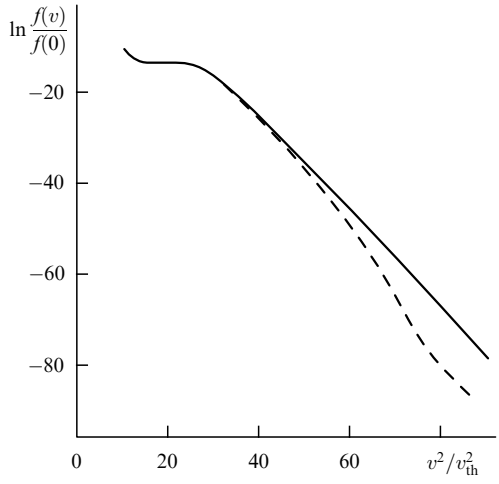


Figure 7. Distribution function computed from the Fokker–Planck equation for the source (sink) $S_\pm \sim I_\pm \exp\{-\alpha_1(v - v_\pm)^2\}$, $\alpha_1 = 10$, $v_- = 3$, and $v_+ = 5$. The dashed and solid lines correspond to time moments $t = 25$ and $t = 100$, respectively.

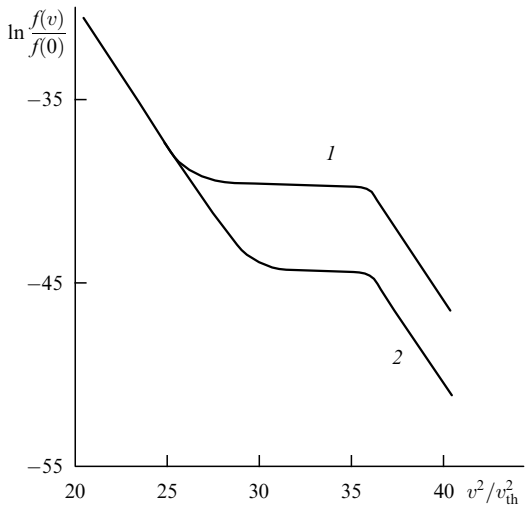


Figure 8. Stationary distribution function obtained from the Landau equation at $\beta = 2$ for a source (sink) in the form of δ-function (3.24) for various flux intensities $I = 0.01$ (*1*), 0.001 (*2*) and $v_- = 4$, $v_+ = 6$.

leaving for (transferred to) the 'background' plasma. It is worth noting that the magnitude of the nonequilibrium distribution function grows together with the intensity, since it is proportional to the magnitude of flux (2.6). Some numerical values of the nonequilibrium stationary distribution function obtained by solving the Landau equation are listed in Tables 1 and 2. The results present a detailed study of the dependence of the solution on the flux intensity, which varies in wide limits in momentum space. Function (3.24) was chosen to describe the source (sink) of particles: the sink was located at the point $v_- = 4$, and the source at $v_+ = 8$, and $\Delta E = 0$. The last equality implies that the energy flux remained constant in momentum space; however, since additional sources in the thermal domain were not involved, the particle density did not stay constant.

From Table 1 it can be seen that the amplitude of the distribution function increases with the intensity I of the flux produced by the source and sink. At low intensities (up to 0.1), the values of the distribution function grow proportionally to I, because a large contribution comes from the interaction of nonequilibrium particles (i.e., particles from the interval between the source and sink) with 'background' particles which are described by the thermodynamically equilibrium distribution function. For intermediate intensities (from 0.1 to 20), the distribution function is the universal one throughout the interval between the source and sink and is proportional to the square root of flux intensity, in agreement with expression (2.6). At large intensities, the distribution function ceases to exhibit this proportionality, because the sink intensity itself depends on the distribution function, in agreement with the chosen model of sink (3.28).

Consider the formation of the nonequilibrium particle distribution function, which corresponds to a constant energy ($\Delta E = 0$, $I_+ = I_- v_-^2/v_+^2$) or particle ($\Delta N = 0$, $I_+ = I_-$) flux. To facilitate the comparison of results, it is convenient to plot the distribution function normalized on its value at zero (Fig. 9). It can be seen that for fluxes of both particles and energy, gradually decaying distributions are formed, characterized by close exponents s.

It is interesting to learn about the form of the distribution function for various laws of interaction between particles.

Table 1.

I	$f(3.95)$	$f(7.95)$	$f(8)$
10	0.393×10^{-9}	0.417×10^{-10}	0.417×10^{-10}
1	0.475×10^{-9}	0.144×10^{-10}	0.144×10^{-10}
0.1	0.508×10^{-9}	0.189×10^{-11}	0.189×10^{-11}
0.01	0.517×10^{-9}	0.197×10^{-12}	0.197×10^{-12}
0.001	0.518×10^{-9}	0.198×10^{-13}	0.198×10^{-13}
0.0001	0.519×10^{-9}	0.196×10^{-14}	0.196×10^{-14}

Table 2.

β	1	1.5	2	3	4
$f(8)$	0.197×10^{-12}	0.123×10^{-13}	0.266×10^{-14}	0.771×10^{-15}	0.384×10^{-15}

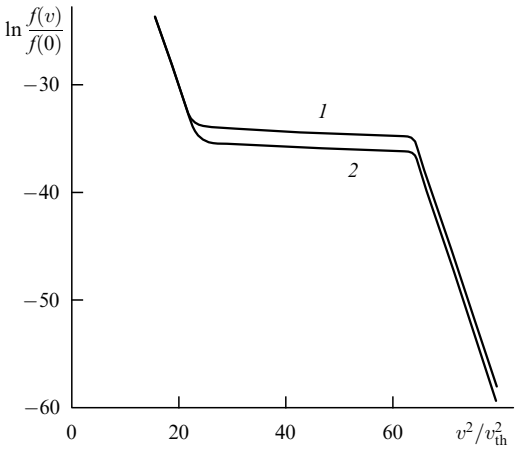


Figure 9. Stationary distribution function derived from the Landau equation at $\beta = 1$ for the source (sink) in the form of the δ-function (3.24); $|I_\mp| = 0.01$, $v_- = 4$, and $v_+ = 8$. Curves *1* and *2* correspond to constant energy ($\Delta E = 0$) and particle ($\Delta N = 0$) fluxes, respectively.

The values of the power-law exponents from the interval $1 \leqslant \beta \leqslant 4$ are considered. It should be kept in mind that $\beta = 1$ corresponds to the Coulomb interaction potential, $\beta = 2$ corresponds to the dipole interaction, and $\beta = 4$ describes the interaction of the so-called Maxwell molecules.

Table 2 presents the dependence of the distribution function $f(v_+)$ on the exponent β of interparticle interaction potential for $I = 0.01$. It is evident that for equal intensities of the source and sink, the values of the distribution function at the same value of velocity decrease, as the parameter β is increased, by almost three orders of magnitude. Figure 10 displays nonequilibrium distribution functions for the constant energy flux of intensity $I = 0.001$ (the sink and source are of the form (3.28), and are located at $v_- = 4$ and $v_+ = 8$ points, respectively), and the exponents $\beta = 1$, 2, 3. It is apparent that nonequilibrium distribution functions with close values of power-law exponents s are formed for all these values of β, which agrees with the conclusions of Ref. [30]. The absolute value of the distribution function in the nonequilibrium region drops as the exponent β is increased. These results are in qualitative agreement with the analytical results presented above.

Next, consider the evolution of the distribution function in the framework of the above-described self-consistent, spatially homogeneous model with two sources of intensities I_- and $I_{\rm th}$ and one sink with intensity I_-, in which the energy

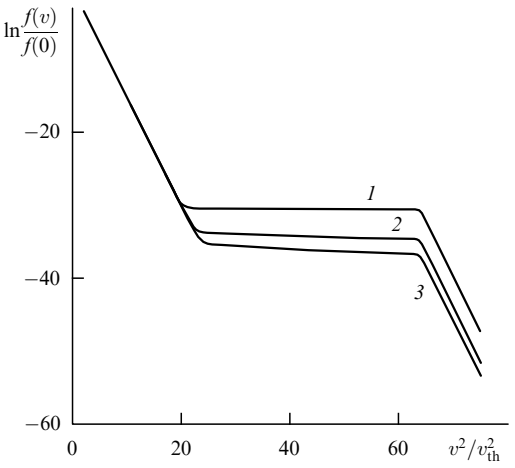


Figure 10. Dependence of nonequilibrium stationary distribution functions on velocity squared for a constant energy flux with the intensity $I = -0.01$, for the source (sink) in the form (3.28) and $v_- = 4$, $v_+ = 8$. The curves are computed from the Landau equation for the exponent $\beta = 1$ (*1*), 2 (*2*), and 4 (*3*).

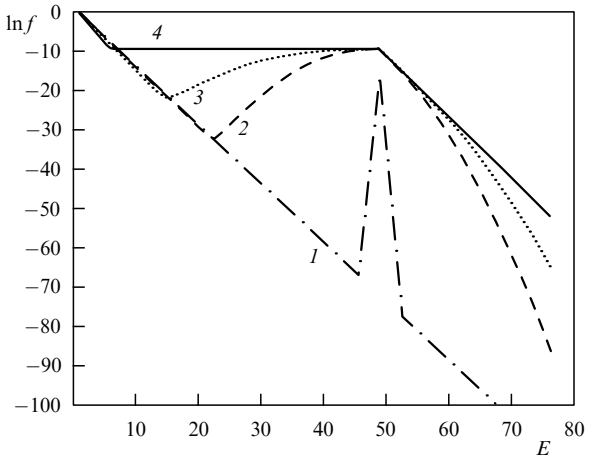


Figure 12. Energy dependence of the logarithm of the nonequilibrium distribution function obtained from the Landau equation at $\beta = 1$ for the source and sink in the form of the δ-function (3.24), $I_- = -0.002$, $v_- = 2$, $v_{1+} = 1$, $v_{2+} = 7$ at various time moments $t = 0.001$ (*1*), 10 (*2*), 20 (*3*), and 200 (*4*).

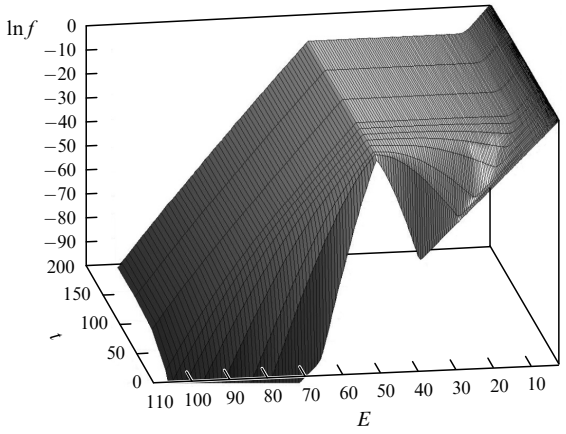


Figure 11. Time evolution of nonequilibrium distribution function computed from the Landau equation at $\beta = 1$ for the source and sink in the form of the δ-function (3.24); $I_- = -0.001$, $v_- = 2$, $v_{1+} = 1$, $v_{2+} = 7$.

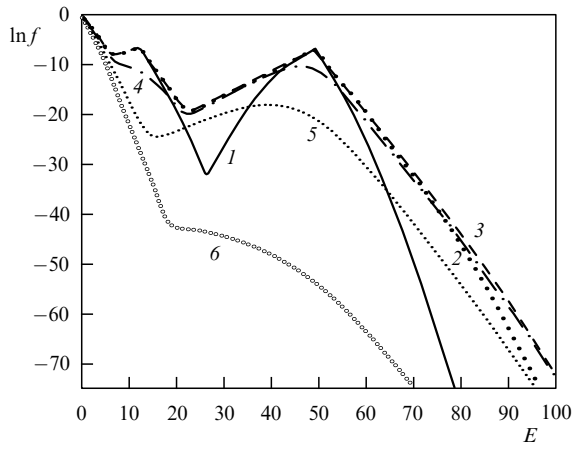


Figure 13. Evolution of the distribution function under the action of nonstationary sources and a stationary distributed sink at several time moments $t = 5$ (*1*), 20 (*2*), 30 (*3*), 35 (*4*), 50 (*5*), and 100 (*6*). Time moment $t = 30$ corresponds to switching the sources off. The sources with equal intensities $I_{1,2} = 0.01$ are localized at velocities $v_{1+} = 3.5$ and $v_{2+} = 7.0$; the sink is proportional to the distribution function and operates in the range $v_- \geqslant 2$.

and particle density do not vary with time. The intensities of sources expressed in terms of the sink intensity I_- satisfy relationships (3.26):

$$I_+ = I_- \frac{v_-^2 - v_{\rm th}^2}{v_+^2 - v_{\rm th}^2}\,, \quad I_{\rm th} = I_- \frac{v_+^2 - v_-^2}{v_+^2 - v_{\rm th}^2}\,.$$

Figure 11 presents the time evolution of the nonequilibrium distribution function obtained in the framework of the self-consistent model from the Landau equation at $\beta = 1$ and the source and sink in the form of δ-function (3.24) for a sink of intensity $I_- = -0.0001$ located at the point $v_- = 2$, and two sources located at points $v_{1+} = 1$ and $v_{2+} = 7$, respectively, with the intensities given above. It can be seen that the distribution function stays in thermodynamic equilibrium outside the inertial range, with the temperature coinciding with the initial one.

In the inertial range (corresponding to a constant energy flux between the source and sink), a distribution function that barely varies with velocity (a plateau) is set. The distribution function attains its stationary limit most rapidly in the vicinity of a source, while its formation in the vicinity of a sink takes several hundred dimensionless time units (Figs 12, 13). The increase in flux intensity (see Fig. 13) leads to the increase in the magnitude of the distribution function but leaves the temperature of the thermodynamically equilibrium distribution function without any changes.

3.3 Numerical modeling of the formation of nonequilibrium particle distribution functions for nonstationary, nonconforming sources and sinks

In the numerical simulations discussed in Section 3.2, we dealt with either sources and sinks of limited intensity or with the self-consistent model in which the sources and sinks are, first, localized in momentum space and, second, agree in intensities so as to provide a pure situation with constant energy (particle) flux in momentum space. In this case, nonequilibrium stationary distribution functions formed in the inertial range as time elapsed.

As shown in Section 4, in actual experimental practice (for example, when a solid state plasma is irradiated by a beam of high-energy ions), one is dealing not only with intensity-unmatched but also with nonstationary sources and sinks, i.e., with sources acting during a finite time interval which is much shorter than the time the sinks are active. Moreover, the sinks and sometimes the sources can be distributed over almost the entire region of momentum space. As we shall see in the subsequent exposition [29–31], the distribution functions in this case will be quasistationary or nonstationary.

Let us analyze the evolution of the distribution function for $\beta_{1,2,3}=1,2,4$ under the action of only sources with low intensity. The formation of nonequilibrium distribution can be subdivided into three stages. During the first, short stage, the system still 'remembers' its initial conditions. The duration of this stage does not vary too much for different exponents β and makes approximately $t\approx 1$. The formation of the main part of the distribution happens during the second stage. Its duration essentially depends on the source position v_+, but not on its intensity, provided it is low. The distribution function acquires the shape of a plateau or gently decays between the source and the cold region, depending on the source intensity. The establishment of quasistationary distribution ends by the tail formation. Its duration depends essentially on the exponent β. The evolution of the main distribution part ends in dimensionless units for $t\approx 50$ at $\beta=1$, for $t\approx 2$ at $\beta=2$, and for $t\approx 1$ at $\beta=4$.

3.4 Formation mechanism of electron distribution function for solid state plasma interacting with beams of electromagnetic radiation or fast charged particles

In this section, we intend to draw attention to specific features of conductive and emissive properties that emerge when high-energy particle beams or laser radiation act on a semiconductor plasma.

(1) Let us compare electron energy relaxation times due to electron–electron and electron–phonon collisions. Because of ionization caused by beams of intense electromagnetic radiation, electrons with energies $E\geqslant\hbar\omega$, where ω is the radiation frequency, are produced. In the case of irradiation by particle beams, the energy spectrum of released electrons embrace the interval from tens to tens of thousands of electron-volts. According to Ref. [33], the frequency of electron–electron collisions for electrons of sufficiently high energy E ($E\gg k_{\rm B}T$) at high temperatures $T>T_{\rm D}$ ($T_{\rm D}$ is the Debye temperature) is given by the expression

$$\gamma^{\rm ee}(E,T)=\gamma_0^{\rm ee}(T)\left[1+\left(\frac{E}{k_{\rm B}T}\right)^2\right], \tag{3.29}$$

where $\gamma_0^{\rm ee}(T)$ is the classical high-temperature frequency for collisions between electrons, which is proportional to T^2. The frequency of electron–phonon collisions under the same conditions is expressed as [34]

$$\gamma^{\rm ef}=\frac{f(T_{\rm D})\,T}{T_{\rm D}}, \tag{3.30}$$

where $f(T_{\rm D})$ is the classical high-temperature frequency of collisions between electrons and phonons at $T=T_{\rm D}$. For the processes discussed here, the conditions needed for formulas (3.29) and (3.30) to be valid are satisfied, since $E>10$ eV, $T=300$ K, and $T_{\rm D}=200-300$ K. Notice that quantities reciprocal to collision frequencies as given by Eqns (3.29) and (3.30) do not coincide in the general case with electron energy relaxation time, since one has to take into account the number of collisions it takes a particle to lose its energy E, i.e., the factor $\xi=E/E_1$ (E_1 is the energy lost by an electron per collision). According to data from Ref. [34], in the case considered one has

$$\gamma^{\rm ef}\approx\gamma_0^{\rm ee},\quad \gamma^{\rm ee}\gg\gamma_0^{\rm ee}, \tag{3.31}$$

whereas the factor ξ for electron–electron collisions can be on the order of unity, while for electron–phonon collisions $E/k_{\rm B}T_{\rm D}>3\times 10^2$. Thus, the relaxation time due to electron–electron collisions is substantially shorter in this case than the relaxation time owing to electron–phonon processes.

The comparison of characteristic ionization time with the relaxation times indicates that the electron distribution function will be quasistationary in our case, and will be largely determined by the electron–electron collisions. Relatedly, it can be found from the condition that the Boltzmann (Landau–Fokker–Planck) collision integral becomes zero.

From the analysis presented above, it follows that for a semiconductor plasma in the energy interval $E-E_{\rm F}>E_{\rm F}$ a power-law distribution may exist, which corresponds to a constant flux of energy or particles in momentum space. In this case, the particle distribution will be formed by collisions with electrons having energies satisfying the condition $E-E_{\rm F}>E_{\rm F}$, as well as with (equilibrium) electrons in the basic background.

It was shown that the nonequilibrium electron distribution function is close to a universal distribution if the intensity of flux created by sources and sinks in momentum space is sufficiently large.

(2) Let us consider, as an example, the irradiation of solid state plasma by a beam of fast ions (with velocities exceeding the velocities of electrons in atoms). We concentrate on a typical situation along a track of an ion. Let the energy of a helium ion $\varepsilon_{\rm i}\approx 5$ MeV, the excitation potential $\Phi\approx 100$ eV, and the ion range in matter $R_{\rm tr}\approx 10^{-3}$ cm. Then the helium ion creates 10^4-10^5 particles on its path owing to ionization, the radius of the ionization track is commensurable with the mean free path of released electrons ($R_{\rm e}\approx 10^{-6}$ cm), the electron number density created by a single helium ion in its track $n_{\rm en}\approx 10^{19}$–$10^{20}$ cm^{-3}, while the equilibrium density $n_{\rm e}\approx 10^{22}$ cm^{-3}. For this ratio of electron densities, the intensity of the source (sink), as can be concluded from the preceding consideration, is sufficient for the nonequilibrium distribution function to form.

We dwell on the energy loss channels for a fast ion in a solid state plasma in more detail. Notice that even for parameters characteristic of inertial thermonuclear fusion on ion beams, the interaction of ion beams with solid state plasma lacks the 'beam density effect' (cf. Ref. [35]), i.e., the dependence of ion energy loss on particle density in the beam. Bearing this circumstance in mind, we estimate the energy loss by a single ion by the Bethe–Bloch formula. According to monographs [36, 37], the energy lost by a fast nonrelativistic particle is transferred to matter in two ways. A part of the energy is spent to excite collective oscillations of the wake charge density, while the other part is transferred to individual electrons, leading to their ionization. The first part corresponds to macroscopic energy losses in distant collisions and weak momentum transfer. The second part corresponds to collisions with large transferred momenta. A relatively large part of the energy lost by the particle is spent

to excite collective oscillations. The energy $\Delta\varepsilon_k$ pertaining to the oscillations of wake charge density can be presented in the form [36, 37]

$$\frac{\Delta\varepsilon_k}{\Delta\varepsilon} = \frac{\ln(v/10v_0)}{2\ln(v/v_0)}, \quad (3.32)$$

where $\Delta\varepsilon$ is the total particle energy loss, and v_0 is the electron velocity in the ground state of a hydrogen atom. Expression (3.32) indicates that the energy $\Delta\varepsilon_k$ of wake charge density oscillations is comparable to an order of magnitude with the total particle energy transferred to the matter.

A fast particle traversing matter may create slow electrons by two equally probable mechanisms — the avalanche ionization, and the ionization through plasma oscillations. The main features intrinsic to the ionization electron formation through plasma oscillations are linked to the fairly long lifetime of wake charge oscillations, as well as to their substantial extension in space. Because of the long lifetime of the wake charge, the secondary ionization inside the beam persists long after the particle's passage. A substantial number of slow electrons in the cascade ionization are produced at the beginning of the cascade, being triggered by the secondary electron with large energy. Since the range of such an electron in matter is large, the dominant number of slow electrons are produced in cascade ionization at distances on the order of the electron range. For this reason, the ionization by the wake potential shows up as the main process that determines the distribution of ionization electrons near the axis of the particle track, while the cascade ionization governs the distribution of ionization electrons at distances on the order of the electron range from the track axis. The ionization by the field of the charged beam particle proper occurs only at the moments when the particle is flying past, while the wake oscillations of charge density play the role of the linear source of secondary electrons, which is preserved long after the passage of the charged particle and therefore markedly determine the behavior of the ionization pattern with time.

As mentioned above, despite the ion travel time along its track being small, the characteristic time of avalanche ionization by the wake charge density is rather large, reaching approximately 10^{-13} s.

If the irradiation frequency satisfies the condition $\hbar\omega \gg k_B T$, the interaction of intense electromagnetic radiation with solid state plasma results in the liberation of a large number of high-energy electrons which form, in agreement with the consideration above, the nonequilibrium stationary electron distribution function. Thus, in both cases of irradiation — by beams of intense electromagnetic radiation and by beams of fast particles — we are dealing with a nonequilibrium electron distribution function, which is formed in the inertial range as a result of electron–electron collisions described by the Boltzmann or Landau–Fokker–Planck collision integrals, and which essentially differs from the equilibrium distribution function through a large number of high-energy electrons.

4. Experimental studies of nonequilibrium particle distribution functions

Section 3.4 showed that local nonequilibrium isotropic stationary particle distributions may exist in collisional plasmas. The existence of such distribution functions hinges on the presence of sources or sinks of particles or energy in momentum space. This takes place when beams of charged particles, laser radiation, or microwave radiation interact with dense plasmas, when nuclear or thermonuclear reactions are maintained in plasma, etc. Earlier [12, 13, 15, 38], attention was drawn to a set of important consequences which, for one thing, stem from so radical a change in the energy distribution of particles and, for another, have an essential bearing on applications (influence on the Landau damping, the Lawson criterion in tasks of controlled fusion, application in astrophysics, and others). However, gas plasma constitutes a highly unstable medium in which collective processes may play a particularly important role and 'conceal' the collision phenomena between particles. From this viewpoint, solid state plasma, allowing one to control the departure from equilibrium in a stable regime, seems to be more attractive. Here, the source or sink of energy (particles) can be furnished by ion beams, powerful laser radiation, an emission current, beams of charged particles produced in fusion or fission reactions, and so forth.

In this section, our goal is to draw attention to specific features of conductive and emissive properties of metals and semiconductors subject to the action of intense beams of particles or laser radiation. In this connection, Sections 4 and 5 consider anomalies in the emissive properties of metals, enabling one to create new sources of current or converters of the radiation energy into electric energy, promising substantial advantages over thermionic ones (see, for example, Ref. [39]) with respect to their efficiency and emission current [16].

4.1 Experimental studies of nonequilibrium electron distribution functions in emission induced by laser radiation

Paper [12] called attention to the possibility of an anomalous increase in the photoconductivity of a semiconductor exposed to light of frequency ω insufficient to trigger the transition between its bands, $\hbar\omega < V_g$, where V_g is the band gap width, by creating a nonequilibrium distribution of electrons and holes. Based on experimental research, the authors of Ref. [40] pointed to the significant modification of the conductive properties of semiconductors irradiated by α-particles.

It is known [41, 42] that two peaks in the emission current are observed on illuminating metal foils by a nanosecond pulse of a powerful laser, $Q = 10^{14}$ erg $(\text{cm}^2\,\text{s})^{-1}$. The first peak, almost synchronous with the laser pulse, contains a large number of 'fast electrons' (the maximum energy for tungsten is 14.5 eV). The second peak, lagging $\tau \sim 10^{-7}$–10^{-8} s behind the first one, contains electrons with energies that do not exceed 2 eV. A satisfactory explanation for the appearance of fast electrons as being due to the Maxwell distribution function is impossible [43], because the experimental results of Refs [41, 42] would correspond to the temperature $T_e = 30{,}000$ K, which is an order of magnitude higher than the tungsten melting temperature. As concerns the emission current, two mechanisms of its production are well known: the multiquantum photoeffect and thermionic emission, both giving emission currents which are smaller by many orders of magnitude.

Turning to the mechanism based on a nonequilibrium electron distribution function forming under these conditions [13, 38], one gets plausible estimates for the magnitude of emission current and its dependence (fast peak) on the retarding potential. As for the slow peak, in all probability,

over its initial part the emission current is contributed not only by the equilibrium distribution (thermionic emission) but also by a nonequilibrium nonstationary component linked to the 'breakup' of power-law distribution.

If the intensity of laser beams is very large, a plasma layer builds up near the solid body surface, and the appearance of high-energy electrons may be explained by a soliton formation under resonance pumping [44]. This mechanism, however, has only a limited validity domain and is unsuitable for explaining the experimental results obtained by Knecht [41, 42].

Thus, it is shown that a series of experimental data on the magnitude of emission current from metals, induced by laser irradiation, and the dependence of current on the retarding potential [41, 42] cannot be explained in the framework of equilibrium distribution function [43], but gains a satisfactory explanation with the help of mechanism [13, 38] that hinges on the presence of a nonequilibrium situation.

4.2 Experimental studies of nonequilibrium electron distribution functions in emission induced by beams of fast ions

To describe kinetic electron emission induced by ions, one resorts to theories proposed in Ref. [45] for low ion energies, and in Ref. [46] for the range of high energies. According to the mechanism of secondary emission proposed by Sternglass [46], secondary electrons are formed because of ionization by fast ions, as described by the Bohr–Bethe theory, then diffuse to the surface and exit into a vacuum. The secondary emission coefficient Δ_e for this mechanism is proportional to the specific ionization losses and does not depend on the work function φ, conductivity, or other basic properties of the substance. Notice that, for the thermal mechanism of emission, the coefficient Δ_e is proportional to the square of specific ionization losses and essentially (exponentially) depends on the electron work function.

In reviews and experimental studies [47–75] and in the literature cited therein, it is proven that the secondary emission coefficient Δ_e is proportional to the energy loss by fast particles, i.e., available data confirm the mechanism described in Ref. [46]. It is noteworthy that the proportionality coefficient κ in this dependence is practically independent of the incident ion energy, but depends on the target material and may change severalfold if one material is replaced by another. Prior to research involving nonequilibrium particle distributions, both theoretical and empirical expressions linking the proportionality coefficient κ with the target parameters and incident ion energy were absent.

Theoretical consideration of secondary electron emission from aluminum, induced by protons and α particles and carried out in Refs [65, 75, 76], is not able to predict the values of coefficient κ observed experimentally or explain the broad energy spectrum (especially for forward emission) of secondary electrons. Note that most experimental research is concerned with backward emission. However, as shown in Ref. [54], the energy spectrum of secondary (backward) electrons is not universal if the energy of the impinging particle is varied. In this case, even such a rough characteristic of the process as the ratio of the forward secondary emission coefficient to the backward one may vary (see Refs [54, 70]).

In our opinion, the expression for the secondary electron emission coefficient Δ_e that is most physically transparent is offered by the formulas

$$\Delta_e = \kappa \frac{d\varepsilon}{dx}, \quad \kappa = \frac{PL_e}{\varepsilon_e}, \tag{4.1}$$

proposed in accordance with Refs [46] and [59] and used in Ref. [55], where $d\varepsilon/dx$ is the energy loss by an impinging particle, P is the probability of the event that the surface barrier will be surmounted by an internal secondary electron, L_e is the depth of the layer from which the emission electrons emerge, and ε_e is the energy spent for every internal secondary electron formation. Reference [59] suggests taking ε_e equal to that of the respective substance, but in a gaseous phase under normal conditions.

We suppose that the secondary emission coefficient Δ_e should be proportional to the energy loss per atom, i.e., be dependent on the number density of atoms N_a as $N_a^{-1/3}$, and be inversely proportional to the excitation potential Φ taking into account the collective character of the interaction of a charged particle with the electron subsystem of the target material. In accordance with Refs [77–79], the Sternglass formula is modified as follows:

$$\Delta_e = \frac{\zeta Z_{1\mathrm{eff}}^2}{\Phi N_a^{1/3}} \left(-\frac{d\varepsilon}{dx}\right)_p, \tag{4.2}$$

where $(-d\varepsilon/dx)_p$ is the energy loss by a proton moving with the velocity of the impinging ion, $Z_{1\mathrm{eff}}$ is the effective ion charge in the target, and ζ is some constant. In this case, it is reasonable to use an expression for $(-d\varepsilon/dx)_p$ proposed in Ref. [80] and valid in the beam energy range from several keV to 50 MeV:

$$\left(-\frac{d\varepsilon}{dx}\right)_p = \frac{ay + by^2}{0.01y^{2.55} + c_1}, \tag{4.3}$$

where y is the velocity ($\times 10^8$ cm s^{-1}), and a, b, c_1 are coefficients characterizing the material of the target and having a rather pronounced periodic dependence on the atomic number Z_2 of the target element. Thus, the coefficient a varies with a period close to $Z_2 = 18$. The values of coefficients a, b, and c_1 for aluminum, beryllium, graphite, and nickel [80], which are used in Section 4.3, are collected in Table 3.

The energy dependence of secondary emission coefficient Δ_e is explored rather thoroughly, yet the description of the distribution of emitted electron over energies emerges as a rather tough problem. Results of experimental research on the energy spectrum of secondary electrons emitted when protons or α particles traverse thin foils are reported in Refs [47–79, 81–84]. However, in Refs [81–83] the spectrum was explored in a narrow energy range (0–10 eV), and it is only mentioned that the spectrum is of a nonthermal nature, and in Refs [54, 55, 79], although their measurements cover a

Table 3. Coefficients characterizing the target material.

Material	a	b	c_1
Al	2.4	0.018	0.36
Be	2.42	0.001	0.37
Ci	2.92	0.018	0.4
Ni	6.8	0.01	0.77

wide secondary-electron energy range (0–100 eV), the distribution function can be judged only by its integral characteristic because of imperfections in the experimental technique utilizing a nonspherical analyzer, which is not fully satisfactory. Reference [54] proved that for a broad range of proton energies (20–250 keV) the energy spectrum of secondary electrons in the forward direction is defined by a universal power-law function, whereas for the backward emission such a universal dependence is absent. Additionally, it was shown that the forward secondary emission coefficient is almost twice as large as the backward one. It can readily be seen that, by studying the energy spectrum of secondary electron emission (SEE) with a spherical analyzer for a point source of SEE, it is possible to retrieve the electron distribution function in a metal. The emission current is determined as

$$I_{\mathrm{em}} = B_1 \int_{\varphi+E_{\mathrm{F}}+q_{\mathrm{e}}U}^{E_{\max}} Ef(E)\,\mathrm{d}E\,, \tag{4.4}$$

where U is the retarding potential, and B_1 is a constant; therefore, the derivative of current over the retarding potential is proportional to the electron distribution function:

$$\frac{\mathrm{d}I_{\mathrm{em}}}{\mathrm{d}U} = B_2(\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U)f(\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U)\,. \tag{4.5}$$

When the electron distribution function has a power-law form, plotting the dependence of $\mathrm{d}I_{\mathrm{em}}/\mathrm{d}U$ on $\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U$ in logarithmic coordinates enables one to easily find the power-law exponent from the slope of the curve.

To verify the theoretical ideas pertaining to the mechanism of electron distribution function formation in interactions of ion beams with a solid state plasma, formulated in Sections 2 and 3, an analysis was carried out of the experimental secondary electron energy spectrum and the dependence of the secondary emission coefficient Δ_{e} in the forward direction, not only on energy losses of α particles and protons in matter, but also on the excitation potential Φ.

The comparison of current–voltage characteristics for different targets with due account of thermal electrons invites the conclusion that the mechanism of secondary emission is not thermal and that secondary electrons knocked out the target by α particles are distributed according to a law different from the exponential one.

4.2.1 Study of kinetic electron emission from metals. We turn now to presenting experimental results on the exploration of the secondary electron energy spectrum with the aid of a spherical three-grid analyzer. Such measurements allow one to retrieve the power-law exponent s of the electron energy distribution function with the help of a single differentiation of the current–voltage characteristic because, in this case, in the domain where the distribution function follows the power law, we have

$$\frac{\mathrm{d}I_{\mathrm{em}}}{\mathrm{d}U} = B_3(\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U)^{s+1}\,, \tag{4.6}$$

where B_3 is a constant.

Accordingly, dependence (4.6) is a straight line on the logarithmic scale with the tangent of the slope angle equal to $s+1$.

The experimental points for the dependences of $\lg(\Delta I_{\mathrm{em}}/\Delta U)$ on $\lg(\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U)$ for aluminum and beryllium targets are well fit by three straight lines, which corresponds to different power-law exponents in the energy intervals 0–10 eV, 10–40 eV, and 40–100 eV. The exponents s for aluminum and beryllium are only different in the range of small energies.

Departing from expression (4.3) for the particle energy loss, and taking into account the values of coefficients a, b, and c, it is possible to explain the dependence of secondary emission coefficient Δ_{e} for aluminum and nickel on the energy ε of impinging particle in the form $\Delta_{\mathrm{e}} \sim \varepsilon^{-0.73}$ in a wide energy range, which was mentioned in Refs [64, 65]. Since $\Phi\,[\mathrm{eV}] \approx 13.5Z_2$ [85, 86], in the energy interval from 1 to 10 MeV the range of an α particle in matter is determined from the empirical formula [86, 87]

$$R_{\mathrm{tr}} = 0.174 \times 10^{-3} A_{\mathrm{m}}^{1/3} \rho_{\mathrm{m}}^{-1} \varepsilon^{3/2}\,, \tag{4.7}$$

where A_{m} and ρ_{m} are the atomic weight and the density of matter, respectively. For estimates, it can approximately be assumed that

$$\Delta_{\mathrm{e}} \sim Z_{\mathrm{leff}}^2 \frac{\rho_{\mathrm{m}}^{2/3}}{Z_2} V_{\mathrm{b}}^{-1}\,, \tag{4.8}$$

where V_{b} is the velocity of impinging particle. Formulas (4.7) and (4.8) provide the correct relationship between the secondary emission coefficient Δ_{e} for aluminum, beryllium, and graphite. High absolute values of Δ_{e} obtained in Refs [77, 79] are explained by a substantial contribution to the electron emission from α particles flying at an angle to the normal to the film and experiencing higher energy losses than the particles moving in the direction of the normal.

In order to alleviate the drawback of Ref. [79] caused by averaging the secondary electron spectrum over energies of impinging particles and the take-off angle of the secondary electrons, the experimental research dealing with the energy spectra of secondary electron emission induced by a proton beam traversing a target was carried out for different electron take-off angles for Al, Cu, and Be [78, 88]. The dependence of secondary emission coefficient in the forward direction was studied not only with respect to the energy loss of protons with an energy of 1 MeV in matter, but also to other macroscopic characteristics of the target. The energies of secondary electrons knocked out of the target were analyzed by the retarding potential method with the aid of a narrow-aperture (4×10^{-4} steradian), three-grid analyzer mounted at different angles (30°, 45°, and 75°) to the beam direction. The secondary emission coefficient was determined by the ratio of the total secondary electron current I_{e} to the current of protons I_{p}. The experiments utilized targets with a thickness of 5.6 μm for aluminum, 9.7 μm for beryllium, and 1.2 μm for copper. The measured dependences of $I_{\mathrm{e}}/I_{\mathrm{p}}$ on U allow one to determine the power-law exponent s by once differentiating these dependences because, in this case, in the range where the distribution function follows the power law, we have

$$\frac{\mathrm{d}}{\mathrm{d}U}\frac{I_{\mathrm{e}}}{I_{\mathrm{p}}} = B_4(\varphi + E_{\mathrm{F}} + q_{\mathrm{e}}U)^{s+1}\,, \tag{4.9}$$

where B_4 is a constant.

Consequently, function (4.9) represents a straight line on the logarithmic axes, the tangent of its slope being $s+1$. It was shown that the experimental points gather around three straight lines that correspond to different power-law exponents on the intervals 0–10 eV, 10–40 eV, and 40–100 eV.

These exponents coincide rather accurately ($\sim 10\,\%$) with those found with the aid of a spherical analyzer in the experiments on bombarding the same targets by α particles (see above and Ref. [77]). The power-law exponents were found to differ for different targets only in the range of small energies 0–10 eV. Together with the energy spectrum, the values of the integral characteristic — the secondary emission coefficient Δ_e for aluminum, copper, and beryllium targets — were determined to be equal to 2.5, 1.6, and 4.6, respectively. As we have already mentioned, the coefficient Δ_e is proportional to the energy loss of a fast particle:

$$\Delta_e = \kappa \frac{d\varepsilon}{dx}, \tag{4.10}$$

where κ is the proportionality coefficient, which may vary several times between various materials.

The experimental results for the three targets considered here agree with the data for aluminum and beryllium discussed earlier and the data of Ref. [70] for graphite. The coefficients Δ_e for different targets relate as the reciprocal of the excitation potential Φ multiplied by $N_a^{1/3}$. We note that the data for copper [88] are somewhat worse, but, possibly, this stems from the imperfect character of the surface layer for the copper foil used.

Thus, the experimental studies of the energy spectrum of secondary electrons demonstrated that the electron energy distribution function is essentially nonequilibrium and decays by a power law as energy is increased, with the same exponents for energies in excess 10 eV for different target materials. They confirmed the proposed theoretical dependence of the secondary emission coefficient on the excitation potential Φ and the number density of atoms N_a.

Electron distributions over energies have been studied in the case of ion–electron emission induced by beams of H^+ ions with energies from 0.75 to 3.0 MeV, and H_2^+ ions with energies from 1 to 2.5 MeV. The ion beam current comprised 0.1–0.4 μA, while the beam diameter on target was 3 mm in all experiments. Thin foils made of silver, copper, nickel, or titanium were used as targets, and their thickness was less than the range of H^+ and H_2^+ ions with the given energy in the target material. The residual gas pressure in the vacuum camera reached 10^{-6} Torr. The energy distributions of secondary electrons were measured in a spherical analyzer with a retarding field. The latter was created between the target and two hemispheres. To eliminate the electron emission under the action of a stream of striking particles, the Faraday cylinder was under the antidynatron potential (~ 20 V) created by a DC power source. The current of the Faraday cylinder was amplified by an electrometric amplifier. The signal from the amplifier was passed to a D3-28 computer. The current of electrons overcoming the retarding field of the analyzer was collected by a forward hemisphere and, after amplification, passed to the computer, too. The emission current I_e of secondary electrons was measured on two intervals of electron energy U: 5–50 eV with a step of 1 eV, and 35–200 eV with a step of 5 eV. The operations of preliminary processing were performed automatically with the assistance of specially developed software. Expression (4.6) was rewritten in the form

$$y = (s+1)\,x + a, \tag{4.11}$$

where $y = \lg(dI_e/dU)$, $x = \lg(U + E_F + \varphi)$, and $a = \lg B_3$, and then the exponent s and parameter a were determined.

Table 4. Absolute values of power-law exponent s as a function of proton energy.

E_p, MeV	Titanium	Nickel	Copper	Silver
0.75	5.4 2.28	5.6 1.88	7.74 1.81	6.54 2.7
1.0	5.15 2.0	4.66 1.45	6.36 1.64	6.52 2.54
1.5	5.95 2.64	5.38 2.32	5.8 0.86	6.42 1.98
2.0	7.16 2.93	5.53 2.78	5.77 2.85	4.38 2.08
2.5	6.06 2.12	4.58 1.64	5.01 2.11	5.25 1.96
3.0	5.66 1.88	4.32 1.66	5.73 2.67	6.34 2.7

Table 5. Absolute values of power-law exponents s as a function of the H_2^+ ion energy.

$E_{H_2^+}$, MeV	Titanium	Nickel	Copper	Silver
1.0	6.04 1.98	5.34 1.87	2.81 2.43	
1.5	4.66 1.82	3.40 1.18	5.36 2.36	6.73 2.0
2.0	4.56 1.49	4.09 1.57	3.66 2.16	6.11 2.54
2.5	4.31 1.28	8.73 2.62	5.73 2.67	5.48 2.7

The fit skill was characterized by the parameter

$$q_{fit} = \frac{1}{N_y} \sum_{i=1}^{N} \left(1 - \frac{y_{exp}}{y_{theor}}\right)^2. \tag{4.12}$$

Here, N_y is the number of y values used in determining the exponent s:

$$y_{exp} = \lg\left(\frac{1}{U}\frac{dI}{dU}\right), \quad y_{theor} = \lg\left(\frac{1}{U}\frac{dI_{theor}}{dU}\right), \tag{4.13}$$

where the function dI_{theor}/dU is computed by formula (4.6) using the values of s and B_3 found.

Typical plots of the secondary electron spectrum were approximated by two straight lines on the intervals from 5 to 30 eV, and from 30 to 250 eV, with different values for the power-law exponent.

Table 4 lists the values of power-law exponent s for the two regions (the upper and lower values) as a function of the energy of the impinging beam of H^+ ions for different target materials. Table 5 displays the same quantities, but for a beam of H_2^+ ions.

As can be seen from Tables 4 and 5, in most experiments the power-law exponent in the first region (upper values) increases together with the ion energy loss in the matter. In the second region (lower values), an apparent dependence escapes detection.

4.2 2 Studies of kinetic electron emission from semiconductors. The information available in the literature on emissive properties of materials irradiated by beams of fast ions largely pertains to metals. The apparent lack of data for effective electron emitters widely used in photoemission and

electronic devices cannot escape attention. The effective secondary electron emitters based on antimony and cesium compounds are disseminated most widely. Owing to their high secondary photoemission and electron emission coefficients, which are commonly attributed to low heights of the potential barrier at the interface between the sample surface and a vacuum, materials of that type are actively used as photocathodes and dynodes in photoelectron multipliers and other devices [89, 90]. Indeed, the value of the secondary electron emission coefficient for antimony–cesium cathodes ranges $\Delta_e = 3-4$ for a low energy of primary electrons $E_e = 100$ eV, and the maximum value amounting to $\Delta_{e\,max} = 8-10$ for the SEE coefficient is achieved at energies $E_e = 500-600$ eV [89, 90]. The rather high secondary emission coefficient is the consequence of not only the low work function for this material but, arguably, also the formation of nonequilibrium distribution functions of the power-law form.

The study of the distribution functions of electrons formed in the solid state plasma of an antimony–cesium cathode hit by beams of fast light ions was carried out on a setup described in detail in Ref. [32]. An electrostatic Van de Graaff ion accelerator, used as a source of primary particles, made it possible to produce beams of hydrogen (H^+) and helium (He^+) ions. Measurements of the energy spectra of electrons of secondary ion–electron emission (SIEE) were carried out for beams of H^+ ions with energies from 1.00 to 2.25 MeV, and He^+ with energies from 1.75 to 2.25 MeV with a step of 0.25 MeV. The explored cathode, utilized as a target, consisted of an antimony-cesium layer with a thickness exceeding the range of impinging ions in the material, fitted to a massive nickel substrate. The target 10 mm in diameter was fixed in a copper casing attached to a moving holder. The ion beam collimated with the aid of a system of diaphragms hit the target causing SIEE from its surface in the backward direction. The plane of the target was installed perpendicularly to the beam axis. The beam diameter on the target measured 3 mm. The ion current density at the target did not exceed 30 μA cm^{-2}. The chamber was pumped out with an NMD-0,4-1 magnetic-discharge pump and NVPR-16D forevacuum pump with a nitrogen trap. In all experiments conducted, the residual vacuum in the chamber was at least 10^{-6} Torr. Electrons emitted from the target surface were collected at a spherical collector made of two hemispheres of radius 100 mm. The target on the holder was placed inside the collector. The gap between the hemispheres equaled 15 mm. The entrance window of each hemisphere was 10 mm in diameter. Simultaneously with the collector current I_C, the target current I_T was measured, too. The target current represents the sum of the ion beam current I_B and the current of secondary electrons that reached the collector: $I_T = |I_C| + I_B$. The measured currents of collector I_C and target I_T, amplified by electrometric amplifiers, were passed to a personal computer via an analog-to-digital converter. To calibrate the measuring system, the Faraday cylinder was located behind the rear hemisphere, enabling direct registration of the ion beam current I_{FC} when the beam was not traversing the target. The Faraday cylinder was 20 mm in diameter and had the length $l = 130$ mm. The Faraday cylinder current I_{FC} was measured with the help of an F303 current instrument. The SIEE coefficient was determined from the formula

$$\Delta_{ie} = \frac{|I_C|}{|I_C| - I_T}\,. \tag{4.14}$$

Studying the energy spectrum of electrons produced by SIEE with the aid of a spherical analyzer for a pointwise emission source, one can reconstruct the explicit shape of the electron distribution function inside the solid substance [16, 32, 78]. When the distribution function exhibits a power-law shape, the derivative of the emission current over electron energy, dI/dU, can be written out as

$$\frac{dI}{dU} = B(E_F + \varphi + eU)^{s+1}\,, \tag{4.15}$$

where B is a constant. Consequently, on the logarithmic axes, dependence (4.15) represents a straight line inclined at an angle with tangent equal to $s + 1$.

The energy distributions of secondary electrons in backward emission were measured with the help of a spherical collector in the energy analyzer mode with a retarding field on the interval from 0 to 100 V with a step of 1 V. The retarding electric field was created between the target and the two hemispheres. Since the radius of the energy analyzer was substantially larger than the target size, the field configuration was close to a spherically symmetric one. A ceramic tube covered from the outside with a resistive layer and measuring 5 mm in diameter was used as the target holder. The resistivity of the layer R_l was varied nonlinearly along the tube so that the holder potential did not disturb the field inside the analyzer. The target had a galvanic contact with one end of the resistive layer, while its other end was grounded. The retarding potential was fed to the target inside the ceramic tube from a source of saw-tooth voltage controlled through a PC. Accordingly, the current flowing along the resistive layer created the required potential distribution along the length of the holder. In the experiment, secondary electrons reached the collector while moving radially. When the retarding potential was applied to the target, only electrons with energies sufficient to overcome the retarding field reached the collector. The software controling the experiment enabled collecting statistics of 100 measurements of the electron emission current during 7 s for every magnitude of retarding potential. It then carried averaging over these 100 experimental points, writing the result to computer memory. Differentiating the measured dependences of the collector current on the retarding potential (the retardation curves), one may retrieve the energy spectrum of SIEE electrons and then reconstruct their distribution function.

The experimental research of the energy spectrum of SIEE electrons carried out in this manner indicates that, for all explored energies of ions, the electron distribution function formed under nonequilibrium conditions in the plasma of antimony–cesium cathode has a power-law shape.

Figure 14 displays a typical distribution function for nonequilibrium electrons for the explored sample hit by He^+ ions with an energy of 1.75 MeV.

The experimental points are well fit by two straight lines corresponding to different power-law exponents in the energy ranges 5–30 eV and 30–100 eV. By processing the experimental data, the respective exponents were recovered. Table 6 lists the values of exponents $-s_{E1}$ and $-s_{E2}$ for the two portions of the distribution function, which correspond to the energy intervals specified above, as a function of the energy of impinging H^+ and He^+ ions.

In our opinion, the power-law exponent of the secondary electron distribution function may preserve generally the dependence on the energy (specific ionization losses) of fast ions, since the intensity of the source of additional particles in

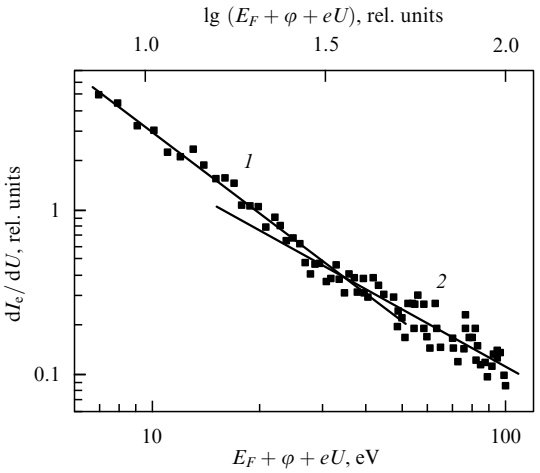


Figure 14. Typical dependence of $\lg(dI_e/dU)$ on $\lg(E_F + \varphi + eU)$ for an antimony–cesium cathode bombarded by He^+ ions with the energy of 1.75 MeV. Section *1* of the distribution function (the energy range 5–30 eV) is described by the power-law with exponent $s_{E1} = -2.9$, and section *2* (30–100 eV) with $s_{E2} = -2.5$.

Table 6.

Ion	Energy, MeV	$-s_{E1}$	$-s_{E2}$
H^+	1.25	2.9	2.5
	1.5	3.0	2.5
	1.75	2.9	2.5
	2.0	3.0	2.4
	2.26	3.0	2.6
He^+	1.75	2.9	2.5
	2.0	2.8	2.2
	2.26	2.8	2.3

momentum space is defined precisely by the ion's specific ionization losses. As was pointed out in Refs [13, 16, 17], the power-law exponent is independent of the structure of sources and sinks only under certain special conditions. In such cases, one speaks about the universal electron distribution function with the exponent $-5/4$ [16]. In experiments conducted earlier with the He^+ ion beam and thin metallic films, the exponents s were measured, and it was shown that the absolute value of power-law exponent s_{E1} of the distribution function on the first energy interval, corresponding to the range of slow electrons ($E < 35$ eV), decreases with the growth of the ion's specific ionization losses in metals [91]. The authors of Ref. [57] point out that a fraction of the fast electrons increases together with the energy of impinging ions. As is seen from Table 6, the exponents s_{E1} differ but slightly for different energies of impinging ions and, consequently, for different specific ionization losses of ions in an antimony–cesium sample, although the absolute value of the power-law exponent grows (decays) for protons with an increase in energy (specific ionization losses). Such behavior is not observed for helium ions. Noteworthy is the fact that the variation of the power-law exponent is within 10%, so that additional research is needed to reliably establish its dependence on energy losses.

Figure 15 plots the dependence of electron emission yield Δ_e on the energy of incident H^+ and He^+ ions for the antimony–cesium cathode. As can be concluded from the plot, the values of electron emission yield Δ_e for the explored

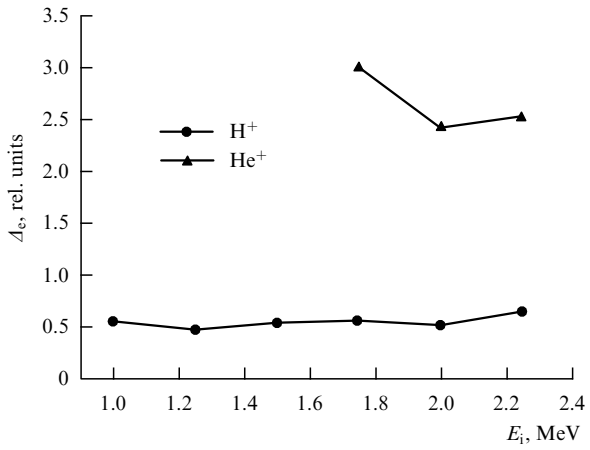


Figure 15. Dependence of electron emission yield Δ_e on the energy of impinging H^+ and He^+ ions for an antimony–cesium cathode.

antimony–cesium compound exceed those for many metallic samples [59]. This may have the following cause. As has already been mentioned, a part of the nonequilibrium electrons, formed in a solid state plasma bombarded by beams of fast charged particles, diffuses to the surface and are ejected into the vacuum. The electron emission occurs from a subsurface layer substantially thinner than the ion penetration depth and is governed by laws defining the motion of these electrons to the surface.

The results of experimental studies presented in Ref. [32] also point to the dependence of the electron emission yield Δ_e induced by He^+ ions from a germanium sample on specific ionization losses dE/dx of He^+ ions, which is approximated well by a straight line, thus confirming the proportionality between these quantities. It should be mentioned that ion beam currents considered in Ref. [32] do not exceed 10 μA. In this case, the emission current varies noticeably with time, reflecting the nonstationary behavior of sources and sinks in energy space. For every ion track, the electron distribution function has sufficient time to pass all stages of its evolution, so that the dependence of emission current on the retarding potential observed in the experiments stemmed from the superposition of currents produced during all time intervals when the nonequilibrium electron distribution function existed. Figure 16 presents the dependence of emission current on retarding potential [29–32]. The main result of comparing theoretical and experimental data is the conclusion that the account for source nonstationarity is an essential factor in explaining the dependence of current on retarding potential, when the experimental technique of collecting the charge escaping the entire film surface over a sufficiently long time interval (several seconds) is utilized. Figure 17 compares electron number densities $N(E) \sim f(v, t)\, v$ for nonstationary electron distribution functions simulated numerically at various time instants [30, 31] with those observed in the experiment [84] for secondary emission electrons induced by 1-keV electrons from polycrystalline aluminum (A1). The energy of the volume plasmon amounts to 15.5 eV, and the A1 work function equals 5.25 eV [31, 84].

4.3 Direct transformation of particle kinetic energy into electric energy based on nonequilibrium particle distributions

Based on the research findings discussed in Sections 4.1 and 4.2, a secondary-emission radioisotope current source has

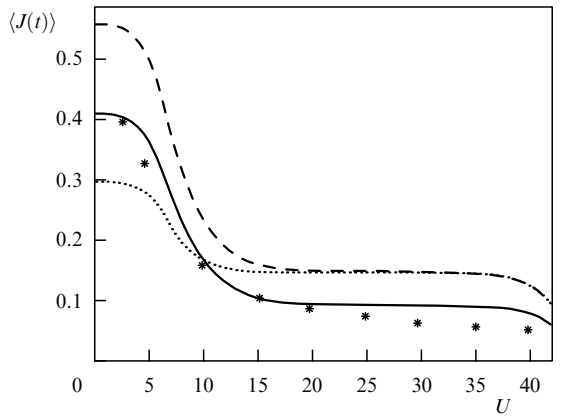


Figure 16. Dependence of ion–electron emission current on the retarding potential for gallium arsenide (GaAs) bombarded by H^+ ions with an energy of 1.25 MeV. The curves correspond to the current averaged over time $t = 10$ (dotted line), 20 (dashed line), and 100 (solid line); the stars display the experimental results [32].

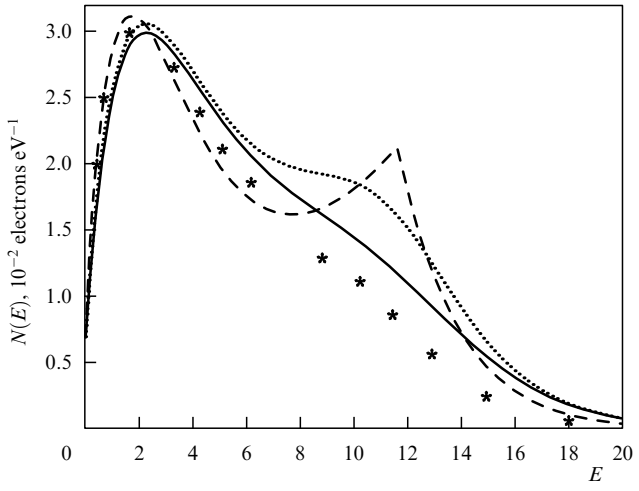


Figure 17. Dependence of the number density of electrons with energy E, $N(E) = f(v)\,v$, on the energy of emission electrons produced by 1-keV electrons bombarding polycrystalline aluminum. The solid, dashed, and dotted curves correspond to the results of numerical modeling, and the stars plot the experimental data [84].

been developed [92–100], containing a radioisotope layer in a hermetic enclosure under conditions of sufficiently high vacuum. This layer was sandwiched between metal emitters whose thickness does not exceed the range of charged particles emitted by the radioisotope in the respective metal. Each of the emitters is fabricated as a sequence of alternating layers of two distinct metals with different secondary electron emission coefficients, insulated electrically by vacuum gaps. The efficiency of a radioisotope source of this type is determined by the fact that secondary electrons are produced along the entire path of a charged particle in metals, i.e., the energy of charged particle is directly converted into the energy of electrons, the number and mean energy of which are incommensurately higher than in the thermionic emission. It was established that making use of heavy particles in sources of electric current leads to a high secondary electron emission coefficient owing to a negligibly small scattering of these particles, since their motion is approximately rectilinear. The secondary electron distribution function is nonequilibrium, with the mean energy of emitted electrons exceeding 10 eV. The source efficiency coefficient increases as a result of augmented secondary emission under the action of δ-electrons (see Refs [95–97]). This source, therefore, has high energy indices which are directly proportional to the number of emitter layers. Since the full thickness of the emitter does not exceed the range of a charged particle emitted by the radioisotope, the increase in layer number, and hence emitter efficiency, is only possible by making layers thinner. However, with a reduction in thickness the layers lose their constructive rigidity, and as a result their electric isolation may be destroyed, for example, because of sagging when the vacuum gap between the layers gets thinner. Subsequently, a secondary emission source of current possessing sufficient constructive rigidity for layers of reduced thickness was offered [98–100]. Its emitter consists of alternating, electrically isolated layers of two unlike materials with different secondary electron emission coefficients. The emitter layers are separated by dielectric grids which electrically insulate the layers and increase the stiffness of the emitter's construction. In so doing, the dielectric grids are thicker than the working layers of the emitter. To achieve the best results when implementing this invention, the grid from a dielectric material (ceramic or plastic) should be superimposed directly on one of the emitter layers. To boost the energy efficiency of the proposed current source, it is desirable to make one of the alternating layers in the emitter from a metal, and the other one from a semiconductor material with a higher secondary electron emission coefficient than for the metal of the first layer.

5. Kinetics of an electron–phonon system of a crystal in a strong electric field

In the 1960s, the phenomenon of a sharp reduction in the plastic resistance of metals was discovered, whereby the conduction electron subsystem of metals is excited either by irradiating them by fast charged particles or by passing high-density electric current $j = 10^8 - 10^9$ A m^{-2}. It could not be reduced to a trivial thermal action (in a macroscopic manifestation) of current, so an assumption has been put forward on the existence of electron–dislocation interaction influencing the mechanical properties of crystals [101, 102].

It was proposed to call the phenomenon the electroplastic effect (EPE). The mechanism of the EPE was associated with an increase in mobility of dislocations in the domain with sources and, hence, with the intensification of source operation. In its purest form, the EPE was explored in metallic single crystals of Zn, Cd, Sn, and Pb [101]. If samples of these materials in a process of deformation are subject to current pulses $10^2 - 10^3$ A mm^{-2} in magnitude and 10^{-4} s in duration or are irradiated with accelerated electrons in the slip direction, a reduction in strength is observed, manifested in a jumplike drop in deforming stress. Oscillations of stress are linked to jumps in the plastic strain of objects. It was discovered that, synchronously with the passage of current pulses and the reduction in the deforming force, groups of slip bands are formed, and also that the jumps in deforming force are much smaller in the range of quasielastic deformations than after the yield point. These oscillations are anomalously high in the vicinity of the material yield point. Jumps in deforming force in diagrams decrease in tests carried out in the stress relaxation mode. A characteristic feature of the EPE in single crystals is the absence of temperature dependence in a broad interval from 77 to 300 K.

It was shown that as the electron energy is increased past the knock-out threshold for atoms ($E_{tr} \approx 0.7$ MeV for zinc), the augmented plasticity of a metal on irradiation is overlain by the effect of its hardening on irradiation, caused by the appearance of additional stoppers for dislocations in the form of point defects and their ensembles. If the electron density in a pulse is increased, the plasticity on irradiation first increases and then weakens. Weakening of the effect is explained by the influence of possible partial degeneration of electron gas in a metal on motion and interaction of dislocations [102].

It is shown that the activation volume does not change considerably on irradiation of a metal by electrons and that an increase in creep rate is explained by the decrease in the time scale (increase in the frequency) of the process of thermally activated passage of barriers by dislocations [102].

5.1 Kinetic description of the electroplastic deformation effect

The plastic deformation of crystals under applying an external load occurs in most cases through dislocation slip [103]. The pertinent basic equation (the Orowan relationship) governing the kinetics of plastic deformation in a slip plain has the form

$$\dot{\mathcal{E}}_d = b\rho_d V_d(\sigma^*, T)\,, \tag{5.1}$$

where $\dot{\mathcal{E}}_d$ is the plastic deformation rate, ρ_d is the density of mobile dislocations, b is the magnitude of their Burgers vector, and $V_d(\sigma^*, T)$ is the mean velocity of long-range dislocation displacement, which depends on the effective deforming shear stress $\sigma^* = \sigma - \sigma_i$ and the temperature T, where σ_i are the internal shear stresses in the slip plane (the analog of dry friction). In the region of sufficiently weak σ^*, a moving dislocation colliding with local stoppers (impurity atoms or other defects of a crystalline lattice) may linger there over a long time. It is believed that the mechanism by which a dislocation segment surmounts barriers involves thermal fluctuations, provided temperatures are not too low. Then, one has

$$V_d(\sigma^*, T) = lv(\sigma^*, T) = lv_0 \exp\left(-\frac{H(\sigma^*)}{k_B T}\right), \tag{5.2}$$

where l is the mean distance between stoppers, and $v(\sigma^*, T)$ is the frequency they are passed through. The explicit form of function $H(\sigma^*)$ depends on the model of the potential barrier (its amplitude and shape) retarding the dislocation slipping, and also on the barrier distribution along the dislocation line. In experimental data processing, it is routinely assumed that

$$H(\sigma^*) = H_0\left(1 - \left(\frac{\sigma^*}{\sigma_{cr}}\right)^{p_0}\right)^{q_0}. \tag{5.3}$$

If the role of stoppers is played by impurity atoms, then H_0 is the energy parameter of dislocation–impurity interaction, σ_{cr} is the critical stress for dislocation motion through a grid of impurity barriers in the slip plane without activation, and p_0, q_0 are some power-law increments determined from the analysis of experimental curves.

5.2 Mathematical model

To quantitatively describe the dynamics of an electron–phonon system in a metallic film, Ref. [104] resorted to an important simplifying assumption that the isotropic part of the electron distribution function is a Fermi type with a time-dependent electron temperature. The authors of Ref. [104] note that, although the introduction of electron temperature is equivalent to a frequently used assumption on instantaneous thermalization of an electron subsystem, it cannot always be strictly justified. So, in the region of very low temperatures $T_e < T^*$ (the temperature $T^* \approx T_D^2/E_F$), where electron–electron collisions dominate electron–phonon collisions, the electron distribution function is thermalized on the time scale of electron–electron interaction, τ_{ee}. In ordinary, relatively pure metals, $T^* \sim 1$ K, while in specially contaminated films, where the electron–electron interaction is intensified through the occurrence of weak localization effects, T^* can reach the value of order 10 K. For temperatures $T_e > T^*$ (but $T_e < T_D$), the electron relaxation in relatively thick films does not involve direct electron–electron interaction, but occurs indirectly, through the exchange by phonons. It was shown earlier by one of the authors of Ref. [104] that the electron distribution function, resembling a Fermi one in shape, is also formed in relatively thin films (from which nonequilibrium phonons may escape to the substrate without being re-absorbed by electrons) solely as a result of phonons being emitted by 'hot' electrons. The consideration relied on a rather strong assumption that all the energy acquired from the electric field is converted to the electron temperature, which is unjustified and, as will be seen from the further analysis, does not correspond to the actual solution to the problem if the electric field strength is not very small. In both cases, the characteristic thermalization time for electrons is that of electron–phonon collisions, τ_{ep}. We also note that in optically thick films the uniformity of electron temperature over the film thickness is furnished by the fast withdrawal of electrons from the skin-layer and high electron heat conductivity compared to the phonon heat conductivity [103]. Because of an additional diffusive reduction in the density of hot electrons, the electron subsystem is thermalized substantially faster, so that for optically thick films the approximation of instantaneous thermalization leads to good agreement between theoretical and experimental results. Reference [104] dealt with the case of weak 'heating', but we, while considering the EPE, are dealing with very strong heating and need, therefore, to carry out a consistent kinetic consideration of both electron and phonon subsystems, which comprises the main content of Refs [30, 105, 106].

In the kinetic description, the behavior of electrons obeys the Boltzmann equation for the electron distribution function $f(\mathbf{r}, \mathbf{p}, t)$ with related collision integrals:

$$\frac{\partial f}{\partial t} + \mathbf{v}\,\frac{\partial f}{\partial \mathbf{r}} + \frac{d\mathbf{p}}{dt}\,\frac{\partial f}{\partial \mathbf{p}} = I_{ee} + I_{ep} + I_{ed}\,,$$

$$\frac{d\mathbf{p}}{dt} = e\left[\boldsymbol{\mathcal{E}}(\mathbf{r}, t) + \mathbf{v} \times \mathbf{B}(\mathbf{r}, t)\right],$$

where I_{ee}, I_{ep}, and I_{ed} are the respective collision integrals of electrons with electrons, electrons with phonons, and electrons with impurities and lattice defects, $\mathbf{v}$ is the velocity, $\mathbf{r}$ is the radius vector, $\mathbf{p}$ is the electron momentum, t is the time, $\mathcal{E}$ is the electric field strength, and $\mathbf{B}$ is the magnetic field induction. In what follows, we assume that the magnetic field is absent.

The phonon distribution function also obeys the kinetic equation with collision integrals, namely

$$\frac{\partial N_p(\mathbf{q}_p)}{\partial t} + \mathbf{v}_q\,\frac{\partial N_p(\mathbf{q}_p)}{\partial \mathbf{r}} = I_{pe} + I_{pp} + I_{pd}\,, \tag{5.4}$$

where I_{pe}, I_{pp}, and I_{pd} are the respective collision integrals of phonons with electrons, phonons with phonons, and phonons with impurities and lattice defects, $\mathbf{v}_q = \hbar\partial\Omega/\partial\mathbf{q}_p$ is the phonon speed, $\mathbf{q}_p$ is its momentum, and $N_p(\mathbf{q}_p)$ is the phonon distribution function. Recognizing that collisions of electrons with impurities, phonons, and defects lead to the isotropization of the electron distribution function, we seek it in the form

$$f(\mathbf{p},t) = f\big(E(p)\big) + \mathbf{f}_1\big(E(p)\big)\,\frac{\mathbf{p}}{p}\,. \tag{5.5}$$

Taking into account that energy transfer in collisions of electrons with phonons is very small, we simplify the electron–photon collision integrals, namely, we expand the isotropic part of the electron distribution function in series in small energy transfer, retaining the quadratic terms, viz.

$$\begin{aligned} f(\mathbf{p}\pm\mathbf{q}_p) &\equiv f\big(E(\mathbf{p})\pm\hbar\Omega(\mathbf{q}_p\big) \\ &= f\big(E(p)\big) \pm \frac{\partial f\big(E(p)\big)}{\partial E}\,\hbar\Omega + \frac{\partial^2 f\big(E(p)\big)}{\partial E^2}\,\frac{(\hbar\Omega)^2}{2}\,, \end{aligned}$$

and substitute this expansion into the collision integrals. Let us take into account the specific form of transition probability $w(q_p)$ and also the frequency of phonon–phonon collisions $\nu_{pp}(q_p)$:

$$w(q_p) = w_0 q_p\,, \quad w_0 = \frac{4\pi^2 m^2 \varepsilon_{1A}^2}{2(2\pi\hbar)^3\rho}\,,$$

$$\nu_{pp}(q_p) = \nu_{pp0} q_p^2\,, \quad \nu_{pp0} = \frac{T_{gr}^3 c_s}{a_{gr} T_D M_c}\,,$$

where $\hbar\Omega(q_p) = c_s q_p$, ε_{1A} is the constant of the deformation potential, T_{gr} is the lattice temperature, M_c is the net mass of two atoms, c_s is the speed of sound, and a_{gr} is the lattice parameter.

The distribution functions for electrons $f(\varepsilon)$, as well as for phonons $N_p(q_p)$, are dimensionless and satisfy the following normalization conditions

$$\frac{1}{2\pi^2}\left(\frac{2m}{\hbar^2}\right)^{3/2}\int_0^\infty E^{1/2} f(E)\,\mathrm{d}E = n\,, \tag{5.6}$$

where n is the number density of electrons in the valence band (in metals it is also the conduction band, since it is only partially filled), and

$$\frac{1}{2\pi^2}\left(\frac{1}{\hbar^3}\right)\int_0^{q_D} q_p^2 N_p(q_p)\,\mathrm{d}q_p < \infty\,, \tag{5.7}$$

where q_D is the Debye phonon momentum defined by the identity $k_B T_D = c_s q_D$. Hereinafter, all quantities are listed for nickel (in which case computational and observational results can be compared [108]): $c_s = 5\times10^5$ cm s^{-1}, $T_D = 375$ K; hence, the maximum phonon momentum is 103×10^{-19} g cm s^{-1}, the electron number density in the valence band $n = 2.5\times10^{22}$ cm^{-3}, the nickel density $\rho = 8.9$ g cm^{-3}, and the lattice constant (the distance between neighboring atoms) $a = 3.5\times10^{-8}$ cm.

In a thermodynamically equilibrium state, the electron distribution function $f(E)$ is given by the Fermi–Dirac distribution

$$f(E) = \left[\exp\left(\frac{E-E_F}{k_B T_e}\right)+1\right]^{-1}, \tag{5.8}$$

where $E_F = 5\times10^{-12}$ erg, and T_e is the temperature of the electron component [in experiments, it reached 20 K and 80 K, coinciding with the lattice (phonon) temperature initially (before the electric field $\mathcal{E} = 0.31$ CGSE units $= 94$ V cm^{-1} was switched on)]. Based on the residual nickel resistivity, $\rho_{cur} = 3\times10^{-6}\ \Omega$ cm, we determine the frequency of electron collisions with impurities and lattice defects: $\nu_{ed} = 3\times10^{13}$ s^{-1}. The frequency of electron–phonon collisions depends on the phonon momentum as $\nu_{ep} = \nu_{ep0}\,q/q_D$, and $\nu_{ep0} = 2\times10^{10}$ s^{-1}. Since numerical integration of the electron–electron collision integral presents considerable difficulties due to its nonlinearity, and since its role consists in redistributing the energy supplied by the outer electric field between electrons (quasithermalization), integrating the coupled system for isotropic electron and phonon distribution functions, we limit ourselves to retaining only the electron–phonon and phonon–electron collision integrals, but on the time interval t_{ee} during which the contribution from electron–electron collisions can be omitted. The time t_{ee} is estimated by its lower bound from the condition that the energy released in the conductor upon the passage of electric current heats (it is assumed that electrons have relaxed over this time to a thermodynamically equilibrium state) the electron subsystem to a temperature comparable to the initial one, i.e., one has

$$\frac{\mathcal{E}^2}{\rho_{cur}}\,t_{ee} = c_p\,\rho\,T_e\,, \tag{5.9}$$

where $c_p = 25$ J kg^{-1} K^{-1} is the specific heat for nickel. From Eqn (5.9) we find the expression for the time interval t_{ee}:

$$t_{ee} = c_p\,\rho\,T_e\,\frac{\rho_{cur}}{\mathcal{E}^2}\,. \tag{5.10}$$

For the electric field strength $\mathcal{E} = 0.31$ CGSE units $= 94$ V cm^{-1}, the time $t_{ee} = 45\tau_{ep0}$. We will measure time in units inverse to the electron–phonon collision frequency in both equations for electron and phonon distribution functions. The energies of electrons E and phonons $c_s q_p$ are put into a dimensionless form by dividing their by $k_B T_e$. In this case, the system of equations is written out as

$$\begin{aligned} \frac{\partial f(\tilde p)}{\partial\tilde t} - \Delta\tilde\varepsilon\,\frac{1}{\tilde p^2}\,\frac{\partial}{\partial\tilde p}\left[\tilde p^2\,\frac{\partial f(\tilde p)}{\partial\tilde p}\right] &= \frac{1}{\tilde p^2} \\ &\times \frac{\partial}{\partial\tilde p}\left\{\frac{1}{\tilde p}\,\frac{\partial f(\tilde p)}{\partial\tilde p}\int_0^{2\tilde p}\mathrm{d}\tilde q_p\,\tilde q_p^4\left[N_p(\tilde q_p)+\frac12\right]\right. \\ &\left. + f(\tilde p)\big(1-f(\tilde p)\big)\sqrt{\frac{1}{\alpha}}\int_0^{2\tilde p}\mathrm{d}\tilde q_p\,\tilde q_p^3\right\}, \end{aligned} \tag{5.11}$$

$$\begin{aligned} \frac{\partial N_p(\tilde q_p)}{\partial\tilde t} = \int_{q_p/2}^\infty \mathrm{d}\tilde p\,\tilde p\left\{\frac{1}{\alpha}\,f(\tilde p)\big(1-f(\tilde p)\big) + \big[N_p(\tilde q_p)+1-f(\tilde p)\big]\right. \\ \left.\times\left[\frac{1}{\sqrt\alpha}\,\tilde q_p\,\frac{\partial f(\tilde p)}{\tilde p\,\partial\tilde p} + \tilde q_p^2\,\frac{\partial}{2\tilde p\,\partial\tilde p}\,\frac{\partial f(\tilde p)}{\tilde p\,\partial\tilde p}\right]\right\}, \end{aligned} \tag{5.12}$$

where

$$\tilde t = t\nu_{ep0}\,, \quad \Delta\tilde\varepsilon = \frac{e^2\mathcal{E}^2}{6m\nu_{ep0}\nu_{ed}k_B T_e}\,, \quad \tilde p = \frac{p}{\sqrt{2mk_B T_e}}\,,$$

$$\tilde q_p = \frac{q_p}{\sqrt{2mk_B T_e}}\,,$$

$$\alpha = \frac{mc_s^2}{2k_B T_e}\,, \quad 0 \leqslant \tilde q_p \leqslant 42.5\,.$$

Integration of systems (5.11) and (5.12) was carried out with the help of fully conservative difference schemes. The conservative character of the scheme is a necessary requirement, since only it ensures that errors do not accumulate in computations over long time intervals (see, for example, Ref. [107]). The essential point consists in satisfying several conservation laws; in the case considered, they are the laws of conservation of energy and particles.

5.3 Results of numerical modeling and their discussion

As a result of numerical simulations, nonstationary distribution functions for electrons $f(p)$ and phonons $N_{\rm p}(q_{\rm p})$ over momenta have been found. Figure 18 plots the dependence of the electron distribution function on dimensionless momentum at different instants of time. The leftmost curve corresponds to the thermodynamically equilibrium function, which is also the initial one for the solution of systems of equations (5.11) and (5.12). As can be seen from Fig. 18 (curves shift to the right with time), the electron distribution function deviates more and more from the equilibrium function as time progresses.

It is thus established that the energy received by the electron subsystem from the external electric field is partially transferred to the phonon subsystem as a result of electron–phonon collisions (a small portion is involved because of the quasielastic character of electron–phonon collisions); its largest part does not contribute to the formations of the thermodynamically equilibrium electron distribution function (contrary to what is frequently assumed; see, for example, Ref. [104]), but is channelled in the formation of intense high-energy tails. Such a drastic modification of the electron distribution function causes the formation of a phonon distribution function enriched very strongly by phonons with an energy close to the Debye one. Our results deviate significantly from the results of Ref. [107], where the phonon distribution function practically coincides with the Bose–Einstein distribution in this range of momenta, but possesses a temperature corresponding to that of the electron subsystem. As follows from our numerical simulations, the 'temperature' (more precisely, the mean energy of electrons, since the electron distribution function strongly departs from the thermodynamically equilibrium form) of the electron distribution function varies insignificantly, i.e., the thermalization of the energy obtained from the electric field does not happen, but high-energy tails are formed, which are responsible for this cardinal change in the phonon distribution function. High-energy tails with ever growing intensity develop with time in the phonon distribution function, because the momentum transfer in electron–photon collisions implies a rather small energy transfer; many phonons are born at the Debye energy, i.e., their distribution function is enriched with the Debye phonons. Further, we analyze the behavior of the product of the phonon distribution function and the cubed dimensionless momentum for the thermodynamically equilibrium case (Bose–Einstein distribution) and the phonon distribution function at various moments of time after the electric field starts to act (Fig. 19).

Thus, with the help of numerical modeling of the electron–phonon system in a strong pulsed electric field, which relies on fully conservative schemes, nonequilibrium distribution functions of electrons and phonons have been found, and it has been shown that:

— the isotropization of the electron distribution function occurs because of collisions with lattice defects;

— the electron distribution function does not become a thermodynamically equilibrium one because the electron–electron collisions contribute essentially less in this situation than the electron–phonon collisions, and collisions with the 'alien' subsystem do not result in thermalization;

— the distribution functions for electrons and phonons contain high-energy tails because the momentum transfer in electron–phonon collisions implies a rather small energy transfer;

— many phonons are born in the vicinity of the Debye energy, i.e., the phonon distribution function is enriched with the Debye phonons.

By way of illustration, Fig. 20a presents the dependences of the phonon distribution function multiplied by the cubed dimensionless momentum for the thermodynamically equilibrium Bose–Einstein function corresponding to the temperature of the substrate (dashed curve), the stationary phonon distribution function taken from Ref. [107] (solid curve), and the nonequilibrium phonon distribution function from Ref. [105] (the dotted curve corresponding to the time moment $t_{\rm ee} = 50\tau_{\rm ep0}$, which is a characteristic energy relaxation time due to collisions between electrons) on the

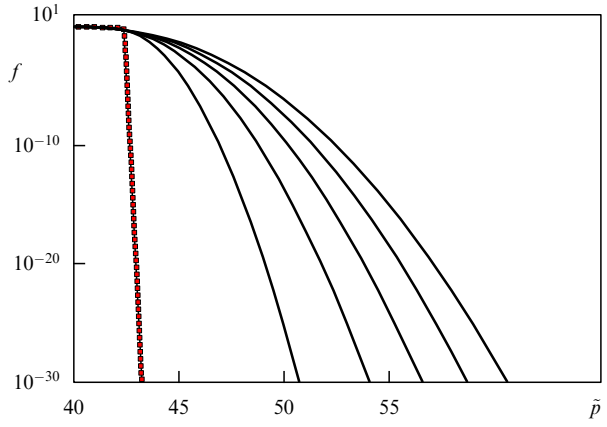


Figure 18. Dependence of the electron distribution function on the dimensionless momentum at different time moments ($t = 0$, 10, 20, 30, 40, and 50); the curves shift to the right with increasing time $\tilde{t}$.

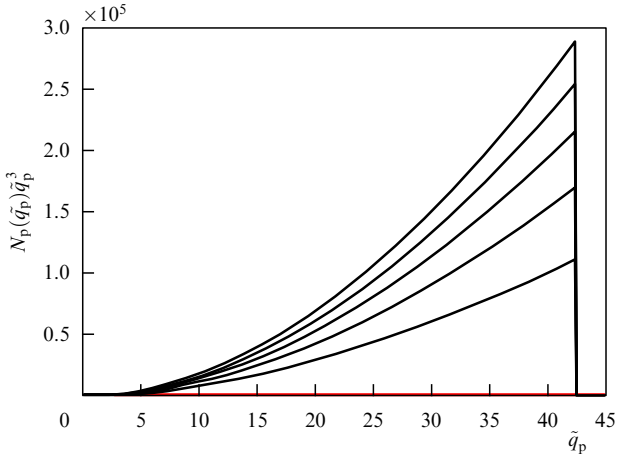


Figure 19. Dependence of the phonon distribution function multiplied by the cubed momentum (at time moments $\tilde{t} = 0$, 10, 20, 30, 40, and 50) on the dimensionless momentum; the curves go up with increasing time $\tilde{t}$.

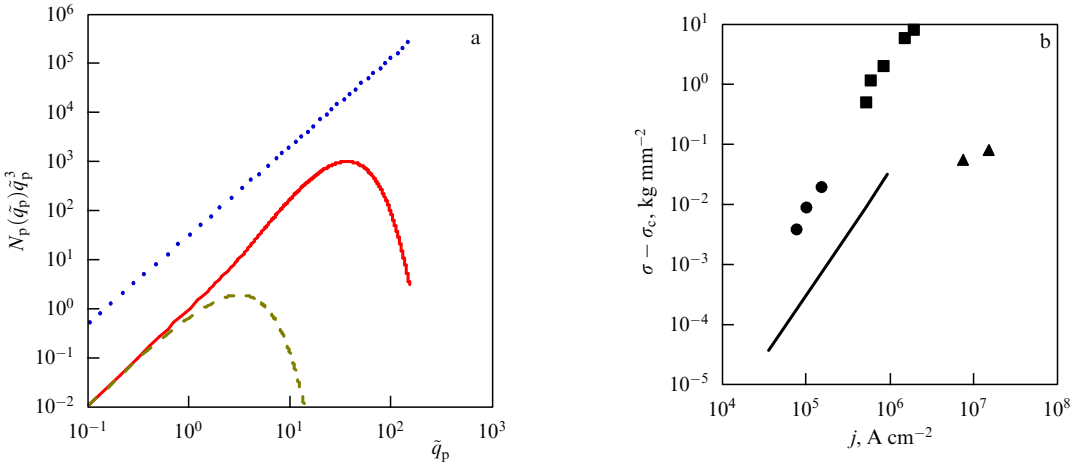


Figure 20. (a) Dependences of the phonon distribution function multiplied by the cubed dimensionless momentum for the thermodynamically equilibrium Bose–Einstein function corresponding to the temperature of the substrate (lower dashed line), the stationary phonon distribution function from Ref. [107] (solid line), and the nonequilibrium phonon distribution function from Ref. [105] (dotted curve) on the dimensionless momentum $\tilde{q}_p$ under the action of electric field $\mathcal{E} = 96$ V cm^{-1}. (b) The drop in load at constant deformation rate as a function of current density. The solid line corresponds to Joule heating; triangles correspond to theoretical nonequilibrium consideration; dots plot the data of Refs [27, 28] at $T_{start} = 78$ K, and squares plot the experimental data obtained by V P Lebedev at $T_{start} = 293$ K for a constant deformation rate of $\dot{\varepsilon}_d = 2.7 \times 10^{-4}$ s^{-1} and the following parameters $\rho_d = 10^8$ cm^{-2}, $l = 10^3 b$, $b = 3.52 \times 10^{-8}$ cm, $x_c = 2b$, $U_0 = 8 \times 10^{-13}$ erg, $L = 10^{-4}$ cm, and $\mu = 1.2 \times 10^{12}$ dyne cm^{-2}.

dimensionless momentum. The dependences are given for the following parameters: the acting electric field $\mathcal{E} = 96$ V cm^{-1}, the substrate temperature $T_b = 4.2$ K, the thickness of nickel film $d = 10^{-4}$ cm, and the electron temperature $T_e = 41.68$ K, which is established according to results of Ref. [107].

From Fig. 20, it is seen that under the action of electric field $\mathcal{E} = 96$ V cm^{-1} the curve for the product of the phonon distribution function and the cubed dimensionless momentum for the thermodynamically equilibrium Bose–Einstein function, corresponding to the temperature of the substrate, practically coincides with the result for the stationary (in partly nonequilibrium approximation) phonon distribution function from Ref. [107] for phonons with small momenta, and differs rather significantly (by almost two orders of magnitude) from the curve for phonons with large momenta. But, as indicated by Fig. 20, this does not lead to a substantial growth in the effective temperature serving actually as a quantity controling the effects of electroplastic deformation. The nonequilibrium phonon distribution function from work [105] provides the product of the distribution function and the cubed dimensionless momentum, which is more than two orders of magnitude larger than the relevant product for the partly nonequilibrium situation over the entire range of phonon momenta. The reduction in the load stress computed for the nonequilibrium case is in satisfactory agreement with experimental data.

Let us compare the reduction in the load stress as a function of the density of electric current passing through a thin metallic sample under the condition of stationary strain rate (the experimental dependence was obtained by V P Lebedev) with the theoretical prediction obtained in Section 5.3. By considering the Landau–Gofman model [108] and formulas (5.1)–(5.3), the following expression for the plastic strain rate was derived:

$$\dot{\mathcal{E}}_d = b\rho_d l v_0(T) \exp\left[-\frac{bL^2\sigma_c^2}{4S\mu k_B T_{eff}(T)}\left(1-\frac{\sigma}{\sigma_c}\right)^2\right].$$

Since the strain rate was stationary in the experiments, we obtain the relationship between jumps in the load and the effective temperature, the expression for which is given in Ref. [105]:

$$\delta\sigma = \sigma(j=0) - \sigma(j) = 10^{-8}\sigma_c\left[\sqrt{a(j)\,T_{eff}(j)} - \sqrt{a(0)\,T_{eff}(0)}\right],$$

$$a(0) = a(T_{start}), \quad T_{eff}(0) = T_{eff}(T_{start}), \quad \sigma_c = \frac{2}{bL}\frac{U_0}{x_c},$$

$$a(T) = \ln\left(\frac{\dot{\mathcal{E}}_0(T)}{\dot{\mathcal{E}}_d}\right)\frac{4\mu S k_B}{bL^2\sigma_c^2}, \quad \dot{\mathcal{E}}_0(T) = b\rho_d l v_0(T),$$

where $T(j)$ is expressed in terms of Joule heating in an equilibrium case or through the respective effective temperature, which is defined by the nonequilibrium phonon distribution function. Substituting the parameters corresponding to the experiment into this relationship yields the theoretical dependence plotted together with the experimental data in Fig. 20b. From the analysis carried out and Fig. 20, it can be seen that the dependence of load stress reduction on the current density under a constant strain rate, which would satisfactorily describe the experimental dependence, cannot be obtained (the difference is 2–3 orders of magnitude) either from the thermodynamically equilibrium approach or from the partly nonequilibrium approach proposed in Ref. [107]. Relatedly, a more promising framework for explaining anomalous electroplastic properties of metals and semiconductors observed experimentally involves computing a nonequilibrium phonon distribution function which is the solution to the above-considered two-component electron–phonon system of equations [105, 106].

6. Conclusions

This review discusses the current state of research pertaining to stationary and nonstationary nonequilibrium electron distributions with flux along the spectrum in solid state plasmas and their application to a new radioisotope current source design.

By analyzing the Boltzmann collision integral, it is shown that in a homogeneous and isotropic medium for a source and

sink localized in momentum space there is a local stationary nonequilibrium distribution for nonrelativistic charged particles interacting by the Coulomb law with account for static screening. This distribution corresponds to a constant energy flux I_1: $f^{(i)}(p) = A_i I_i p^{2s_i}$, $s_i = -(3D - 4/\beta + 2(i-1))/4$, where D is the space dimension, β is the power-law exponent in the particle interaction law ($\beta = 1$ for the Coulomb interaction), and $s_{i=1} = -5/4$. This stationary nonequilibrium distribution is also an exact solution of the collision integral in the Landau form. For electrons in a solid state plasma, the interaction between which is described by the screened Coulomb potential, a local nonequilibrium distribution function may form which corresponds to a constant energy flux in momentum space. Analytical consideration was performed for a stationary source and sink localized in momentum space. Numerical modeling in the framework of the Landau–Fokker–Planck collision integral with the help of fully conservative difference schemes allows one to find nonequilibrium particle distribution functions for sources and sinks which are nonlocal, nonstationary, and nonconforming in momentum space (as is the case, for example, in ionization by direct collisions and wake waves). The existence conditions and the intensity of energy flux in momentum space are found, which allow the formation of the nonequilibrium distribution function for electrons with energy exceeding the Fermi energy. It is shown how the results obtained can be used to predict the behavior of semiconductors with their intrinsic and impurity conductivities upon irradiating them by beams of fast ions or laser radiation. The existing and presented results of experimental research on energy and angular distributions of secondary emission electrons, induced by ions, including molecular ions, witness in favor of the importance of ionization by wake fields excited by ions. The experimental results are in good agreement with theoretical predictions. Based on theoretical research into nonequilibrium electron distributions induced by ion beams in a solid state plasma, a new method is proposed for transforming fission energy into electric energy. A detailed comparison of the current source based on this principle with existing ones is carried out. The advantages and disadvantages of each existing type of radioisotope current sources are considered. The advantages of the proposed source are described and how to implement it is outlined.

The mechanism of the EPE in the framework of the model of a dislocation string passing stoppers as a result of its excitation by phonons is studied and substantiated. With the aid of numerical modeling of an electron–phonon system placed in a strong pulsed electric field, relying on fully conservative schemes, nonequilibrium distribution functions are found for electrons and phonons, and it is shown that the electron and phonon distribution functions exhibit high-energy tails, i.e., many phonons are born in the vicinity of the Debye energy and the distribution function is enriched with Debye phonons.

Based on the random impact model (in the framework of the Langevin approach), which is due to phonons in our problem, it is demonstrated that

— the thermodynamic approach cannot describe the electroplastic deformation effect;

— based on the actual phonon distribution function found as the solution of the two-component electron–phonon system of equations for the case involving the action of a strong impulse electric field on a metal, it is possible to explain anomalous electroplastic properties of metals and semiconductors observed experimentally, i.e., the EPE.

The authors regret that, because of limitations on the review size, they had to omit certain interesting and important results. To partly compensate for this gap, the list of references includes several reviews [109–114].

References

1. Kolmogoroff A N *C. R.* (*Dokl.*) *Acad. Sci. USSR* **30** 301 (1941) [*Dokl. Akad. Nauk SSSR* **30** 299 (1941)]
2. Zakharov V E *J. Appl. Mech. Tech. Phys.* **6** (4) 22 (1965) [*Zh. Prikl. Mekh. Tekh. Fiz.* (4) 35 (1965)]
3. Silin V P *Vvedenie v Kineticheskuyu Teoriyu Gazov* (Introduction into the Kinetic Gas Theory) (Moscow: Nauka, 1971)
4. Rosenbluth M N, MacDonald W M, Judd D L *Phys. Rev.* **107** 1 (1957)
5. Trubnikov B A, in *Reviews of Plasma Physics* Vol. 1 (Ed. M A Leontovich) (New York: Consultants Bureau, 1965) p. 105 [Translated from Russian: *Voprosy Teorii Plazmy* (Problems in the Plasma Theory) Issue 1 (Ed. M A Leontovich) (Moscow: Gosatomizdat, 1963) p. 98]
6. Gurov K P *Osnovaniya Kineticheskoi Teorii*: *Method N N Bogolyubova* (Foundations of Kinetic Theory: N N Bogoliubov Method) (Moscow: Nauka, 1966)
7. Gurov K P, in *Nonequilibrium Phenomena I*: *the Boltzmann Equation* (Eds J L Leibowitz, E W Montroll) (Amsterdam: North-Holland, 1983) [Translated into Russian (Moscow: Mir, 1986)]
8. Chapman S, Cowling T G *The Mathematical Theory of Non-uniform Gases; An Account of the Kinetic Theory of Viscosity, Thermal Conduction, and Diffusion in Gases* (Cambridge: Univ. Press, 1952) [Translated into Russian (Moscow: IL, 1960)]
9. Akhiezer A I, Pomeranchuk I Ya *Nekotorye Voprosy Teorii Yadra* (Some Questions from the Theory of Nucleus) (Moscow – Leningrad: Gostekhizdat, 1948)
10. Kats A V et al., Preprint No. 42 (Khar'kov: Inst. of Radiophysics and Electronics, Acad. of Sci. of Ukrainian SSR, 1974)
11. Kats A V et al. *JETP Lett.* **21** 5 (1975) [*Pis'ma Zh. Eksp. Teor. Fiz.* **21** 13 (1975)]
12. Karas' V I *Sov. Tech. Phys. Lett.* **1** 438 (1975) [*Pis'ma Zh. Tekh. Fiz.* **1** 1020 (1975)]
13. Karas' V I, Moiseev S S, Novikov V E *Sov. Phys. JETP* **44** 744 (1976) [*Zh. Eksp. Teor. Fiz.* **71** 1421 (1976)]
14. Gradshteyn I S, Ryzhik I M *Tables of Integrals, Series, and Products* (New York: Academic Press, 1980) [Translated from Russian: *Tablitsy Integralov, Summ, Ryadov i Proizvedenii* (Moscow: Nauka, 1971)]
15. Kats A V et al. *Sov. Phys. JETP* **44** 93 (1976) [*Zh. Eksp. Teor. Fiz.* **71** 177 (1976)]
16. Karas' V I, Abstract of Doct. Phys. Math. Sci. Thesis (Kharkov: Gorky Kharkov State Univ., 1988)
17. Karas' V I, Moiseev S S, Shuklin A P *Ukr. Fiz. Zh.* **25** 820 (1980)
18. Samarskii A A *The Theory of Difference Schemes* (New York: Marcel Dekker, 2001) [Translated from Russian: *Teoriya Raznostnykh Skhem* (Moscow: Nauka, 1977)]
19. Samarskii A A, Popov Yu P *Raznostnye Skhemy Gazovoi Dinamiki* (The Difference Schemes of Gas Dynamics) (Moscow: Nauka, 1975)
20. Bobylev A V *Sov. Sci. Rev. Sect. C Math. Phys. Rev.* **7** 111 (1988)
21. Bobylev A V, Chuyanov V A *USSR Comput. Math. Math. Phys.* **16** (2) 121 (1976) [*Zh. Vychisl. Mat. Mat. Fiz.* **16** 407 (1976)]
22. Potapenko I F, Chuyanov V A *USSR Comput. Math. Math. Phys.* **19** (2) 192 (1979) [*Zh. Vychisl. Mat. Mat. Fiz.* **19** 458 (1979)]
23. Potapenko I F, Chuyanov V A *USSR Comput. Math. Math. Phys.* **20** (2) 249 (1980) [*Zh. Vychisl. Mat. Mat. Fiz.* **20** 513 (1980)]
24. Bobylev A V, Potapenko I F, Chuyanov V A *USSR Comput. Math. Math. Phys.* **20** (4) 190 (1980) [*Zh. Vychisl. Mat. Mat. Fiz.* **20** 993 (1980)]
25. Bobylev A V, Potapenko I F, Chuyanov V A *Sov. Phys. Dokl.* **25** 994 (1980) [*Dokl. Akad. Nauk SSSR* **255** 1348 (1980)]
26. Potapenko I F, Chuyanov V A *USSR Comput. Math. Math. Phys.* **22** (3) 269 (1982) [*Zh. Vychisl. Mat. Mat. Fiz.* **22** 751 (1982)]
27. Potapenko I F et al. *Phys. Rev. E* **56** 7159 (1997)

28. Potapenko I F, de Azevedo C A *J. Comput. Appl. Math.* **103** 115 (1999)
29. Karas' V I, Potapenko I F *Plasma Phys. Rep.* **28** 837 (2002) [*Fiz. Plazmy* **28** 908 (2002)]
30. Karas' V I, Potapenko I F *Comput. Math. Math. Phys.* **46** (2) 294 (2006) [*Zh. Vychisl. Mat. Mat. Fiz.* **46** (2) 307 (2006)]
31. Potapenko I F, Bornatici M, Karas V I *J. Plasma Phys.* **71** 859 (2005)
32. Kononenko S I et al. *Plasma Phys. Rep.* **30** 761 (2004) [*Fiz. Plazmy* **30** 722 (2004)]
33. Gurzhi R N *Sov. Phys. JETP* **6** 506 (1958) [*Zh. Eksp. Teor. Fiz.* **35** 965 (1958)]
34. Ziman J M *Electrons and Phonons* (Oxford: Clarendon Press, 1960) [Translated into Russian (Moscow: IL, 1962)]
35. Basko M M, Sokolovskii M V, Preprint No. 177 (Moscow: ITEP, 1980)
36. Kalashnikov N P, Remizovich V S, Ryazanov M I *Collisions of Fast Charged Particles in Solids* (New York: Gordon and Breach, 1985) [Translated from Russian: *Stolknoveniya Bystrykh Zaryazhennykh Chastits v Tverdykh Telakh* (Moscow: Atomizdat, 1980)]
37. Otsuki Y-H *Charged Beam Interactions with Solids* (New York: Taylor and Francis, 1983) [Translated into Russian (Moscow: Mir, 1985)]
38. Karas' V I, Moiseev S S, Novikov V E *JETP Lett.* **21** 245 (1975) [*Pis'ma Zh. Eksp. Teor. Fiz.* **21** 525 (1975)]
39. Hansen L K (Ed.) *Thermionic Converters and Low-Temperature Plasma* (Oak Ridge, TN: Technical Information Center, U.S. Dept. of Energy, 1978); Moizhes B Ya, Pikus G E (Eds) *Termoemissionnye Preobrazovateli i Nizkotemperaturnaya Plazma* (Moscow: Nauka, 1973)
40. Aseevskaya A S, Ivkin E B, Kolomiets B G *Sov. Tech. Phys. Lett.* **2** 117 (1976) [*Pis'ma Zh. Tekh. Fiz.* **2** 305 (1976)]
41. Knecht W L *Appl. Phys. Lett.* **6** 99 (1965)
42. Knecht W L *Appl. Phys. Lett.* **8** 254 (1966)
43. Anisimov S I et al. *Deistvie Izlucheniya Bol'shoi Moshchnosti na Metally* (The Action of High-Power Radiation on Metals) (Eds A M Bonch-Bruevich, M A El'yashevich) (Moscow: Nauka, 1970)
44. Anisimov S I, Ivanov M F *Sov. Phys. Dokl.* **20** 758 (1975) [*Dokl. Akad. Nauk SSSR* **225** 280 (1975)]
45. Parilis E S, Kishinevskii L M *Sov. Phys. Solid State* **3** 885 (1961) [*Fiz. Tverd. Tela* **3** 1219 (1961)]
46. Sternglass E J *Phys. Rev.* **108** 1 (1957)
47. Kaminsky M *Atomic and Ionic Impact Phenomena on Metal Surfaces* (New York: Academic Press, 1965) [Translated into Russian (Moscow: Mir, 1967)]
48. Arifov U A *Interaction of Atomic Particles with a Solid Surface* (New York: Consultants Bureau, 1969) [Translated from Russian: *Vzaimodeistvie Atomnykh Chastits s Poverkhnost'yu Tverdogo Tela* (Moscow: Nauka, 1968)]
49. Medved D B, Strausser Y E *Adv. Electron. Electron. Phys.* **21** 101 (1965) [Abridged Russian Translation: *Usp. Fiz. Nauk* **91** 485 (1967)]
50. Frischkorn H J et al. *Nucl. Instrum. Meth. Phys. Res.* **214** 123 (1983)
51. Petrov N N, Abroyan I A *Diagnostika Poverkhnosti s Pomoshch'yu Elektronnykh Puchkov* (Surface Diagnostics with the Help of Ion Beams) (Leningrad: Izd. LGU, 1977)
52. Abroyan I A, Eremeev M A, Petrov N N *Sov. Phys. Usp.* **10** 332 (1967) [*Usp. Fiz. Nauk* **92** 105 (1967)]
53. Dorozhkin A A, Petrov N N, Petrov A A *Izv. Akad. Nauk SSSR Ser. Fiz.* **40** 1687 (1976)
54. Mechbach W, Braunstein G, Arista N *J. Phys. B At. Mol. Phys.* **8** L344 (1975)
55. Hasselkamp D, Scharmann A *Surf. Sci.* **119** L388 (1982)
56. Hasselkamp D, Scharmann A *Phys. Lett. A* **96** 259 (1983)
57. Hasselkamp D, Hippler S, Scharmann A *Nucl. Instrum. Meth. Phys. Res. B* **2** 475 (1984)
58. Krebs K H *Vacuum* **33** 555 (1983)
59. Hasselkamp D et al. *Nucl. Instrum. Meth.* **180** 349 (1981)
60. Baragiola R A, Alonso E V, Oliva Florio A *Phys. Rev. B* **19** 121 (1979)
61. Alonso E V et al. *Phys. Rev. B* **22** 80 (1980)
62. Hasselkamp D, Scharmann A *Phys. Status Solidi A* **79** K197 (1983)
63. Holmén G, Svensson B, Burén A *Nucl. Instrum. Meth. Phys. Res.* **185** 523 (1981)
64. Koyama A, Yagi E, Sakairi H *Jpn. J. Appl. Phys.* **15** 1811 (1976)
65. Koyama A, Shikata T, Sakairi H *Jpn. J. Appl. Phys.* **20** 65 (1981)
66. Mironov E S, Nemenov L M *Sov. Phys. JETP* **5** 188 (1957) [*Zh. Eksp. Teor. Fiz.* **32** 269 (1957)]
67. Svensson B, Holmén G *Phys. Rev. B* **25** 3056 (1982)
68. Frischkorn H J et al. *Phys. Rev. Lett.* **49** 1671 (1982)
69. Sternglass E J *Phys. Rev.* **95** 345 (1954)
70. Garnir H P, Dumont P D, Baudinet-Robinet Y *Nucl. Instrum. Meth. Phys. Res.* **202** 187 (1982)
71. Rothard H et al. *Phys. Rev. A* **51** 3066 (1995)
72. Mischler J et al. *Surf. Sci.* **136** 532 (1984)
73. Anno J N *J. Appl. Phys.* **34** 3495 (1963)
74. Koyama A et al. *Jpn. J. Appl. Phys.* **21** 1216 (1982)
75. Svensson B, Holmén G *J. Appl. Phys.* **52** 6928 (1981)
76. Schou J *Phys. Rev. B* **22** 2141 (1980)
77. Batrakin E N et al. *Poverkhnost'* (12) 82 (1986)
78. Batrakin E N et al. *Sov. Phys. JETP* **62** 633 (1985) [*Zh. Eksp. Teor. Fiz.* **89** 1098 (1985)]
79. Babenko V A et al. *Sov. Phys. Tech. Phys.* **25** 508 (1980) [*Zh. Tekh. Fiz.* **50** 848 (1980)]
80. Gott Yu V *Vzaimodeistvie Chastits s Veshchestvom v Plazmennykh Issledovaniyakh* (Interaction of Particles with Matter in Plasma Research) (Moscow: Atomizdat, 1978)
81. Agranovich V M et al. *Sov. Phys. JETP* **30** 220 (1970) [*Zh. Eksp. Teor. Fiz.* **57** 401 (1969)]
82. Agranovich V M et al. *Sov. Phys. JETP* **34** 805 (1972) [*Zh. Eksp. Teor. Fiz.* **61** 1511 (1971)]
83. Daukeev D K, Lebedev S Ya *Sov. Phys. Solid State* **12** 507 (1970) [*Fiz. Tverd. Tela* **12** 655 (1970)]
84. Baragiola R A, Dukes C A, Riccardi P *Nucl. Instrum. Meth. Phys. Res. B* **182** 73 (2001)
85. Bronshtein I M, Fraiman B S *Vtorichnaya Elektronnaya Emissiya* (Secondary Electron Emission) (Moscow: Nauka, 1969)
86. Siegbahn K (Ed.) *Alpha- Beta- and Gamma-Ray Spectroscopy* Vol. 1 (Amsterdam: North-Holland Publ. Co., 1965) [Translated into Russian (Moscow: Atomizdat, 1969)]
87. Price W J *Nuclear Radiation Detection* (New York: McGraw-Hill, 1958) [Translated into Russian (Moscow: IL, 1960)]
88. Kononenko S I *Dopovidi Nats. Akad. Nauk Ukrainy* (1) 87 (2001)
89. Soboleva N A, Melamid A E *Fotoelektronnye Pribory* (Photoelectron Instruments) (Moscow: Vysshaya Shkola, 1974)
90. Fomenko V S, Podchernyaeva I A *Emissionnye i Absorbtsionnye Svoistva Veshchestv i Materialov* (Emission and Absorption Properties of Matter and Materials) (Moscow: Atomizdat, 1975)
91. Rosebury F *Handbook of Electron Tube and Vacuum Techniques* (Reading, Mass.: Addison-Wesley, 1965) [Translated into Russian (Moscow: Energiya, 1972)]
92. Zhurenko V P et al. *Plasma Phys. Rep.* **29** 130 (2003) [*Fiz. Plazmy* **29** 150 (2003)]
93. Karas' V I, Moiseev S S, Preprint No. 77-24 (Kharkov: Kharkov Inst. of Physics and Technology, Acad. Sci. of the Ukrainian SSR, 1977)
94. Karas' V I, Moiseev S S *Ukr. Fiz. Zh.* **24** 1724 (1979)
95. Balebanov V M et al. *Plasma Phys. Rep.* **24** 732 (1998) [*Fiz. Plazmy* **24** 789 (1998)]
96. Balebanov V M et al. *Atomic Energy* **84** 324 (1998) [*Atom. Energ.* **84** 398 (1998)]
97. Balebanov V M et al., Author's Certificate No. 1737559; *Byull. Izobret.* (20) (1992)
98. Balebanov V M et al., Patent RF No. 2050626; *Byull. Izobret.* (35) (1995)
99. Balebanov V M et al., Patent RF No. 2054742; *Byull. Izobret.* (5) (1995)
100. Balebanov V M et al., Patent RF No. 2050625; *Byull. Izobret.* (35) (1995)
101. Gromov V E, Tsellermaier B Ya, Bazaikin V I *Elektrostimulirovannoe Volochenie: Analiz Protsessa i Mikrostruktura* (Electrosimulated Drawing: the Process Analysis and Microstructure) (Moscow: Nedra, 1996)
102. Spitsyn V I, Troitskii O A *Metallofizika* **51** 18 (1974)
103. Kosevich A M *Teoriya Kristallicheskoi Reshetki* (*Fizicheskaya Mekhanika Kristallov*) (The Theory of Crystal Lattice (Physical

Mechanics of Crystals Ser.)) (Kharkov: Vishcha Shkola, Izd. pri Khark. Gos. Univ., 1988)
104. Bezuglyi A I, Shklovskii V A *JETP* **84** 1149 (1997) [*Zh. Eksp. Teor. Fiz.* **111** 2106 (1997)]
105. Karas' V I, Potapenko I F *Vopr. Atom. Nauki Tekh. Ser. Fiz. Radiats. Povrezhd. Radiats. Mater.* (4 – 2) 42 (2009)
106. Zakharov V E, Karas' V I *Vopr. Atom. Nauki Tekh. Ser. Yad.-Fiz. Issled.* (2(66)) 204 (2010)
107. Perrin N, Budd H *Phys. Rev. Lett.* **28** 1701 (1972)
108. Landau A I, Gofman Yu I *Sov. Phys. Solid State* **16** 2220 (1975) [*Fiz. Tverd. Tela* **16** 3427 (1974)]
109. Zakharov V, Dias F, Pushkarev A *Phys. Rep.* **398** 1 (2004)
110. Zakharov V E (Ed.) *Nonlinear Waves and Weak Turbulence* (American Mathematical Society Translations, Ser. 2, Vol. 182) (Providence, RI: American Mathematical Society, 1998)
111. Musher S L, Rubenchik A M, Zakharov V E *Phys. Rep.* **252** 177 (1995)
112. Zakharov V E, L'vov V S, Falkovich G *Kolmogorov Spectra of Turbulence I. Wave Turbulence* (Berlin: Springer-Verlag, 1992)
113. Zakharov V E, in *Basic Plasma Physics* Vol. 2 (Handbook of Plasma Physics, Vol. 2, Eds A A Galeev, R N Sudan) (Amsterdam: North-Holland, 1984) p. 3 [Translated into Russian (Moscow: Energo-atomizdat, 1984) p. 48]
114. Kontorovich V M *Radiofiz. Radioastron.* **11** (1) 5 (2006)

Physics – Uspekhi **56** (1) 79 – 92 (2013)

PHYSICS OF OUR DAYS

PACS numbers: 47.15.G –, 47.57.J –, 47.63.Gd, 64.75.Yz, 81.05Xj

Active colloids

I S Aranson

DOI: 10.3367/UFNe.0183.201301e.0087

Contents

Abstract. A colloidal suspension is a heterogeneous fluid containing solid microscopic particles. Colloids play an important role in our everyday life, from food and pharmaceutical industries to medicine and nanotechnology. It is useful to distinguish two major classes of colloidal suspensions: equilibrium and active, i.e., maintained out of thermodynamic equilibrium by external electric or magnetic fields, light, chemical reactions, or hydrodynamic shear flow. While the properties of equilibrium colloidal suspensions are fairly well understood, active colloids pose a formidable challenge, and the research is in its early exploratory stage. One of the most remarkable properties of active colloids is the possibility of dynamic self-assembly, a natural tendency of simple building blocks to organize into complex functional architectures. Examples range from tunable, self-healing colloidal crystals and membranes to self-assembled microswimmers and robots. Active colloidal suspensions may exhibit material properties not present in their equilibrium counterparts, e.g., reduced viscosity and enhanced self-diffusivity, etc. This study surveys the most recent developments in the physics of active colloids, both in synthetic and living systems, with the aim of elucidation of the fundamental physical mechanisms governing self-assembly and collective behavior.

I S Aranson Materials Science Division, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA;
Department of Engineering Sciences and Applied Mathematics, Northwestern University, 2145 Sheridan Rd, Evanston, IL 60208, USA
E-mail: aronson@anl.gov

Received 17 July 2012, revised 2 October 2012
Uspekhi Fizicheskikh Nauk **183** (1) 87 – 102 (2013)
DOI: 10.3367/UFNr.0183.201301e.0087
Translated by the author; edited by M S Aksent'eva

1. Introduction

Colloids are substances consisting of macroscopic particles larger than atoms or molecules, typically from 10 nm to 100 μm, dispersed in another continuum phase, such as liquid or gas [1]. Colloids play an important role in our everyday life (e.g., milk, pigmented ink, blood) and are critical to many industries, from food, pharmaceutical, and medicine to nanotechnology [2] and electronics (colloidal pigment particles are used in low power consumption electrophoretic displays for e-readers [3]). The cross-disciplinary field of colloidal suspensions is an active area of research. Colloidal physics is an important part of 'soft matter', a rapidly expanding field of contemporary science dealing with physical states easily deformable by thermal stresses, fluctuations, or external forces. The scope of soft matter includes colloids, polymers, foams, gels, granular materials, and some biological materials, like suspensions of motile microorganisms (for recent reviews, see [4–9]).

A multitude of forces governs interactions between colloidal particles, like steric repulsion, electrostatic and magnetic forces (for magnetic colloids), van der Waals forces due to electric dipole moments of colloidal particles, gravity, entropic forces, and hydrodynamic and other forces due to the gradients of the surfactant, etc. [9]. Under various experimental conditions, interacting colloidal particles may form a variety of steady states, from colloidal glasses [10] and gels [11] to highly ordered colloidal crystals [12]. Ordered colloidal suspensions find application as optical materials with a photonic band gap [13, 14] (see Fig. 1 in Section 2).

It is important to emphasize the connections between colloids and the related concept of metamaterials. The 20th century saw remarkable progress based on our understanding of hard materials at the atomic level. Only recently did the concept of metamaterials emerge: human-made materials with artificial building units. Metamaterials were conceived, for example, to control the propagation of waves [15] or vary opto-mechanical properties [16]. New functionalities of

metamaterials, such as a negative index of refraction or optical band gap, result from preprogrammed features of artificial meta-atoms. Functionalized colloidal particles, playing the role of meta-atoms, offer a promising platform for the design of metamaterials via controlled and flexible bottom-up assembly.

There is a significant body of work dedicated to various aspects of mostly equilibrium colloidal structures obtained as a result of static self-assembly [17–30]. Some examples of ordered structures, like photonic band gap crystals and ionic colloidal crystals, obtained by equilibrium self-assembly, are shown in Fig. 1a–d in Section 2.

The purpose of this study is to survey recent progress in active (or driven) colloidal systems, both living and synthetic, where formation of dynamic structures occurs under out-of-equilibrium conditions. In particular, actively self-assembled colloidal structures offer new functionalities not available under equilibrium conditions: an ability of self-propulsion [31–33], self-repair [34, 35], manipulation of cargo particles [36], and other functions usually associated with living systems [37]. While active colloids is a rapidly expanding field, there are only a few reviews on this subject [38–40], especially from the point of view of physics.

It is natural to classify collective motion in active colloidal systems according to the way how the energy is injected. One large class is represented by systems driven by external fields, such as electric and magnetic fields, or by hydrodynamic flows. In this situation, the external field exerts forces or torques on the colloidal particles, either in the bulk or at the liquid/solid or liquid/liquid interface. In the second case, the particles are driven internally, for example, by chemical reactions, ultraviolet light, etc. (see, e.g., [41]). Consequently, no net external force is applied to the particle, and the particle is propelled due to the generation of local force dipoles [42]. A large group of self-propelled particles, such as bimetallic chemical microswimmers, asymmetric Janus particles (named after the two-faced Roman god), and even a majority of swimming microorganisms, belongs to this class [43]. It was recently shown by the example of motile bacteria that a suspension of active swimmers may exhibit material properties different from their equilibrium counterpart: a sevenfold reduction of viscosity [44], a dramatic increase in diffusivity [45, 46], etc. It remains a major challenge, however, to obtain similar effects with suspensions of synthetic swimmers.

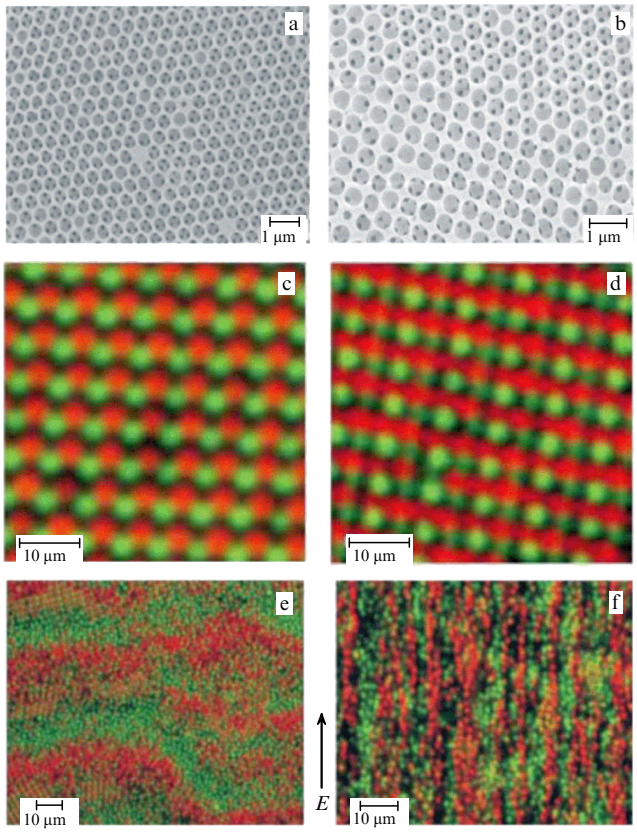


Figure 1. (Color online.) Illustrations of self-assembled colloidal structures. (a) Highly ordered regions with hexagonal symmetry extending over 10 μm assembled from functionalized (patchy) colloidal silica particles. The colloidal structure exhibits a photonic band gap in the visible light range. (b) Regions showing co-existing hexagonal and square symmetry, from [13]. (c, d) Confocal microscope images illustrating highly-ordered self-assembled structures in 'ionic' colloidal crystals of oppositely charged particles assembled from positive (red, 1.08 μm) and negative (green, 0.99 μm) polymethylmethacrylate spheres. (e, f) Electric field E induced structures in the same system: stationary bands perpendicular to the field directions (e) and lanes of oppositely moving particles parallel to the field direction, emerging for larger amplitudes of the applied electric field (f). (Reproduced with the kind permission of the authors of [20, 47] and the Nature Publishing Group).

2. Collective behavior of externally-driven colloidal systems

In this section, we consider large-scale collective states emerging when the colloidal systems are driven by external fields: electric or magnetic. Figures 1–5 illustrate representative self-assembled states observed in this broad class of systems, from nonequilibrium steady-state patterns, like separated bands, lanes, and self-healing membranes [35, 47] to dynamic structures: rotating binary vortices, pulsating rings [48], and self-assembled microswimmers [33].

2.1 Electrostatically-driven colloidal systems

It appears rather challenging to classify patterns and collective behavior emerging in externally driven colloidal systems. For example, in a colloidal system energized by an electric field, the outcome of self-assembly is very sensitive to the amplitude, frequency of the electric field, properties of the suspending fluid (e.g., viscosity, conductivity), and the size, composition, and material properties of the colloidal particles. A system of oppositely-charged colloidal particles in an aqueous solution subject to a low frequency ac electric field segregates into bands perpendicular to the applied field direction [47] (Fig. 1e). The bands disintegrate due to the particles' Brownian diffusion once the field is switched off. The authors of [47] argue that the band formation is caused by collisions between particles moving in opposite directions.

In a similar system of oppositely charged colloids, but under slightly different conditions (a dc electric field of higher magnitude), instead of the static bands formed perpendicular to the field, a phenomenon of the formation of lanes is observed (Fig. 1f). Lanes are stripes formed by particles of opposite polarity and moving in opposite directions, like cars on a two-way highway. In contrast to the static bands, lanes are aligned parallel to the applied field [18]. The authors [18] suggest that the dynamic mechanism responsible for the formation of the lanes is an enhanced lateral mobility of particles induced by collisions with particles driven in the opposite direction. Once the lanes are formed, a particle's

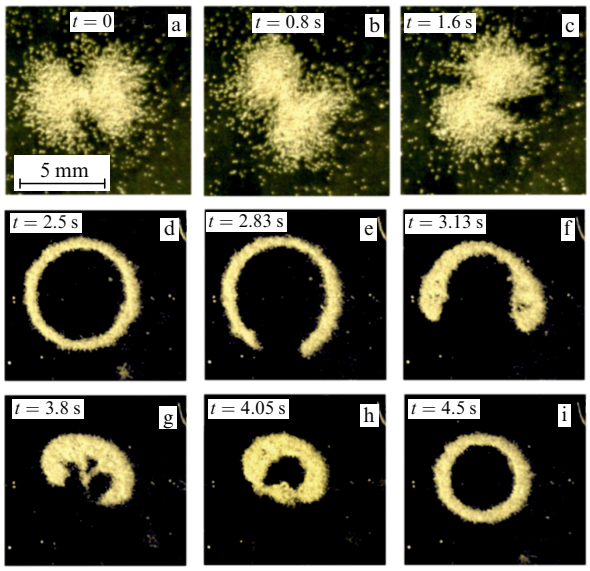


Figure 2. (Color and animation online.) Dynamically self-assembled structures formed by conducting microparticles (100-μm bronze spheres) in a weakly-conducting nonpolar liquid (toluene/alcohol mixture) in a dc electric field. Snapshots (a–c) show rotating binary star-like vortices. Each vortex excites toroidal electroosmotic flow in the suspending liquid (movie 2.1). Upon direction reversal of the electric field, the rotating 'binary stars' transform into pulsating rings (d–i) (movie 2.2). (From [48].)

mobility is sharply decreased. The authors argue that the particles in a lane can be regarded as being in a dynamically 'locked-in' state. For a review of various instabilities in oppositely charged colloidal systems, see [49].

Fascinating dynamically self-assembled structures emerge when conducting microspheres (100-μm bronze particles) are suspended in a thin layer of weakly conducting liquid (toluene/alcohol mixture) and energized by a strong dc electric field (up to 2–3 kV mm^{-1}) [48].[1] Some of these states are shown in Fig. 2. Depending on the amplitude of the dc electric field, its direction with respect to gravity, and alcohol to toluene ratio, a large variety of self-assembled structures is observed. For relatively low applied field values and low alcohol concentration, the patterns are mostly static: clusters, colloidal crystals, or honeycomb lattices. For higher applied field values and higher alcohol concentrations, the static patterns give way to dynamic states, like self-assembled rotating binary star-like vortices (Fig. 2a–c). Upon polarity reversal of the applied electric field, the rotating vortices transform into chaotically pulsating rings (Fig. 2d–i). The primary mechanisms responsible for the interactions between the particles are electrophoresis, i.e., the transport of charged particles by the electric field in the liquid, and electroosmosis, the transport of liquid by the electric field. While toluene is a nonpolar nonconducting fluid, the presence of alcohol makes the mixture a weak ionic conductor due to the small dissociation of the alcohol molecules. For typical experimental conditions (a few percent of alcohol in toluene), the corresponding Debye length is above 1–10 μm, i.e., many orders of magnitude higher than that in water. As a result, the static electric field in the mixture is practically unscreened, making it very different from the colloidal experiments in aqueous solutions [18, 49], where the Debye length is very short — on the order of 10 nm. The electric field causes motion of ions in the bulk of the liquid. Due to the presence of macroscopic conducting particles in the liquid, the flow of the ions is perturbed, resulting in the creation of toroidal electroosmotic flows in the proximity of the particles [48, 51]. The electrophoretic flows, together with the electrostatic interactions between the particles and gravity, are responsible for the creation of dynamic states in the experiment. A phenomenological theory of pattern formation in this system was developed in [52]. Experiments conducted with much smaller particles (2–3-μm gold spheres) revealed, in addition to vortices, dynamically self-assembled wires formed along the field direction and assembling/reassembling tree-like structures [53].

[1] A conducting particle in a contact with conducting plane acquires electric charge by direct electrification. It detaches from the plane and moves upwards if the electric force $F_e = \kappa a^2 E^2$ exceeds the force of gravity, $F_g = 4/3\pi\rho g a^3$, here a is the radius of the particle, E is applied dc vertical electric field, g is acceleration of gravity, ρ is the mass density of the particles, and κ is a constant dependent on the particle shape. For a conducting spherical particle one derives $\kappa = \zeta(3) + 1/3 \approx 1.36$, where ζ is Riemann ζ-function [50].

2.2 Colloidal systems energized by alternating magnetic fields

If colloidal particles possess a magnetic moment, either permanent (ferromagnetic) or induced (superparamagnetic), the self-assembly can be tuned and directed by an external magnetic field. A static field induces either chain-like clusters or bulk colloidal crystals [54]. In contrast, a variety of complex self-assembled structures can form in an alternating magnetic field, from quasi-static sheets and membranes [35] to dynamic self-assembled swimmers [33]. The applied magnetic field can be uniaxial [33], biaxial, or triaxial [35, 55].

Spectacular self-assembled structures emerge when a dispersion of ferromagnetic microparticles (100-μm nickel spheres) is suspended at the water–air interface and energized by an external ac magnetic field $H = H_0 \sin(2\pi f t)$, applied perpendicular to the interface (Fig. 3). Depending on the frequency f and amplitude H_0 of the external magnetic field, the particles self-assemble into linear snake-like objects

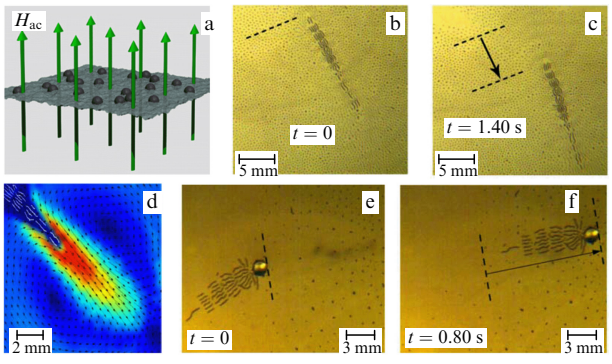


Figure 3. (Color and animation online.) Self-assembled surface swimmer (magnetic snake) formed by ferromagnetic microparticles (80–100-μm nickel spheres) at the water–air interface (a). Schematics of the experiment: a homogeneous ac magnetic field (60–80 Hz), applied perpendicular to the air/water interface, energizes suspended ferromagnetic particles. Images (b, c) illustrate the self-assembled swimmer (magnetic snake) formed spontaneously from a random dispersion of ferromagnetic particles (movie 3.1). (d) A large-scale streaming vortex flow generated by the tail of a magnetic snake; arrows show streamlines, and colors indicate the magnitude of the surface velocity (movie 3.2). (e, f) A swimmer formed by a snake attached to a 1-mm nonmagnetic glass bead (movie 3.3) [33].

(magnetic snakes) (Fig. 3b, c). The snakes are composed of several parallel ferromagnetically-aligned chains (segments). The segments, however, have an anti-ferromagnetic alignment. The distance between the neighboring segments λ is controlled by the frequency of the applied field, while the snake's width W is determined by the field's amplitude. A good approximation for the distance λ as a function of frequency of applied magnetic field f is given by the dispersion relation for gravity-capillary waves on the surface of a deep liquid:

$$f^2 \approx \frac{g}{2\pi\lambda} + \frac{2\pi\sigma}{\rho\lambda^3}\,, \tag{1}$$

where g is the acceleration of gravity, σ is the surface tension, and ρ is the density of the liquid. Thus, the primary mechanism of snake formation is an interplay between particles' magnetic dipolar interaction promoting chain formation and hydrodynamic surface flows excited by the oscillating chains at the air–water interface.

The snakes are immobile for low frequencies f of the field and become self-propelled objects for higher frequencies [33]. Due to the inertia of the fluid, the tails of a snake excite large-scale rectified streaming vortex flows [56], each tail serving as a powerful self-assembled pump (Fig. 3d) (compare to the Rayleigh or acoustic streaming phenomenon [57]). The magnitude of the rectified flow is proportional to H_0^2 and increases roughly linearly with the frequency f. While for relatively low frequencies of the applied field (below 60 Hz) the snake's large-scale flow is symmetric and the snake is immobile, for higher frequencies the symmetry of the flow becomes spontaneously broken, resulting in self-propulsion of the snake (Fig. 3b, c). Apparently, this mechanism of self-propulsion does not have a direct analog in nature. In addition, the symmetry of the snake's large-scale flows can be broken artificially, by placing a large nonmagnetic bead in the proximity of one of the snake's tails. The bead weakens the flow generated by the tail, resulting in an asymmetry of the streaming flow. In this case, the snake attaches to the bead and is propelled by a remaining free tail (Fig. 3e, f). Many aspects of dynamic self-assembly in this system are captured in the framework of a first-principle model based on the particle dynamics coupled to surface deformations in a thin liquid layer [58].

A rich variety of dynamic self-assembled patterns is observed when ferromagnetic microparticles are placed at the interface between two nonmiscible liquids, e.g., silicone oil and water. In this situation, the distance λ is determined by the density difference between two liquids [compare to Eqn (1)],

$$f^2 \approx \frac{g(\rho_1-\rho_2)}{2\pi\lambda(\rho_1+\rho_2)} + \frac{2\pi\sigma_{12}}{(\rho_1+\rho_2)\,\lambda^3}\,, \tag{2}$$

where $\rho_{1,2}$ are the densities of the bottom (1) and top (2) liquids, and σ_{12} is the interfacial tension.[2] Thus, by reducing the density contrast between the liquids $\rho_1-\rho_2$ or the interfacial tension σ_{12}, one can decrease the size of self-assembled snakes λ.

[2] To avoid the Rayleigh–Taylor instability of an interface between two liquids of different densities, the bottom liquid should be heavier than the top one [59].

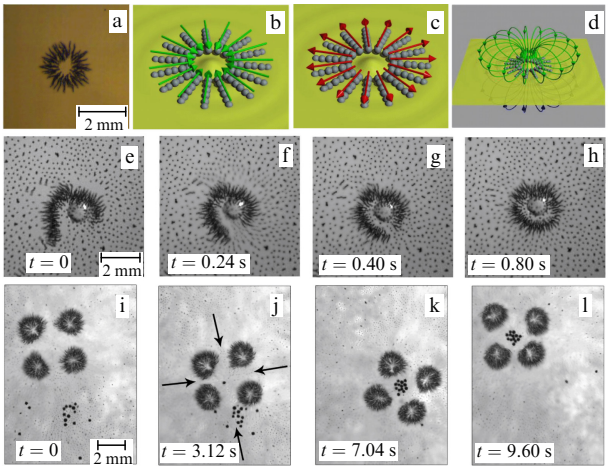


Figure 4. (Color and animation online.) Self-assembled colloidal structures at the interface between two nonmiscible liquids (silicone oil/water). (a) Isolated star-like object (aster). (b, c) Illustrations of aster and anti-aster magnetic orders. (d) Schematics of the toroidal vortex flows generated by aster in the top and bottom liquids [36]. (e–h) Action of a self-assembled aster-robot. The aster, remotely controlled by an external dc magnetic field applied parallel to the interface, performs simple robotic functions: capture, transport, and release of a large nonmagnetic particle (1-mm glass bead), (movie 4.1). (i–l) A cluster of four asters captures all nonmagnetic beads scattered at the oil/water interface; the particles are trapped in the interstitial space between the asters (movie 4.2); arrows show the direction of the surface flow created by the cluster [36].

For the rather large viscosity of the top liquid (the viscosity of silicone oil is about 20 times higher than the viscosity of water), in addition to magnetic snakes, a new type of dynamically self-assembled patterns is observed: axi-symmetric asters and clusters of asters. The asters are composed of chains with the magnetic moments pointing towards the center (asters) or out of the center (anti-asters) (Fig. 4a–c). While asters are immobile, they create a rectified toroidal vortex flow pointing down in the bottom liquid and up in the top liquid layer (Fig. 4d). Obviously, the magnetic order of asters is highly unfavorable under equilibrium conditions: asters rapidly disintegrate when the applied ac magnetic field is switched off. Application of a small dc magnetic field H_{dc} parallel to the interface between liquids results in controlled self-propulsion of the asters: they open up and swim in the direction of the field, with the asters and anti-asters swimming in opposite directions. It is important to emphasize that a homogeneous dc magnetic field exerts no net force on the aster: the propulsion is an outcome of a distortion of the aster's magnetic structure and consequent deflection of the rectified streaming flow from the vertical direction.

The in-plane component of the rectified flow propels the aster along the field direction. The swimming velocity V initially increases with the increase in in-plane field H_{dc}, reaches a maximum at some value, and finally vanishes for a relatively strong field (about 20 Oe for reported experimental conditions). A simple theoretical prediction based on approximation of an aster by a pair of magnetic spheres gives the following dependence of the aster's speed V on in-plane field H_{dc}:

$$V \sim H_{dc}(H_c^2 - H_{dc}^2)\,. \tag{3}$$

Thus, the shape of the asters and their swimming direction can be remotely controlled by a small in-plane magnetic field.

This functionality enables asters and clusters of asters to perform simple robotic functions: capture, transport, and positioning of target particles [36] (Fig. 4e–h). A cluster of four asters, shown in Fig. 4i–l, exhibits an additional functionality not present in a single aster: the possibility of capturing nonmagnetic particles in the interstitial space between the asters.

In addition to the colloidal assembly at the interface, the patterns can be formed in the bulk. It was shown in [35, 55, 60, 61] that in an alternating field, either electric or magnetic, the dipolar colloidal particles may assemble into planar sheets: a rotating planar field generates $1/r^3$ dipolar pair interaction with in-plane attraction (r being the distance between particles) and repulsion along the rotation axis. This anisotropic interaction forces the particles to assemble into flat sheets parallel to the field plane [60]. In the case of a triaxial field, the interactions are more complicated, and, as was shown in [55, 61], can become, under certain conditions, isotropic. The special case of a precessing triaxial magnetic field was studied in [35]: the field has a cone opening angle θ_m which is controlled by the ratio of the static component (z-component) to the rotating component (x, y components). The field-induced interaction between superparamagnetic particles decays as α/r^3 with the interparticle distance r. The constant α depends on the opening angle θ_m: for small θ_m the colloid behaves like a dipolar liquid. For the opening angle $\theta_m \approx 90^\circ$, the colloidal particles experience effective in-plane attraction [60, 62]. For the so-called 'magic angle',

$$\theta_m = \arctan \frac{1}{\sqrt{3}} \approx 54.7^\circ , \tag{4}$$

the time-averages interaction term vanishes irrespectively of the relative position of the particles. In this case, the interaction becomes isotropic and attractive, and decays as $1/r^6$, somewhat similar to the van der Waals interaction between molecules.

When a triaxial magnetic field with the opening angle close to the magic angle was applied to a suspension of micron-size superparamagnetic particles, a formation of ordered planar membranes consisting of closed-packed particles was observed [35]. The gradual formation of these membranes from a semi-dilute suspension of particles is illustrated in Fig. 5a–e. The particles initially assemble into dimers; the dimers form chains and Y-junctions. The Y-junctions eventually interconnect and coarsen, giving rise to patches of membranes of all orientations. The membranes show a remarkable self-healing capability: a membrane reconstitutes its structure after an artificial perforation is created. In addition to dense membranes, a rich variety of complex patterns was observed in a rotating (vortex) magnetic field: periodic lattices, honeycomb-like 'particle foams' (Fig. 5f–h) [61], sheets of spinning chains [63], various dynamically-assembled clusters, etc. [64]. When both electric and magnetic fields were applied, the reversible assembly/disassembly of magnetic Janus particles into novel staggered chain structures was observed [65].

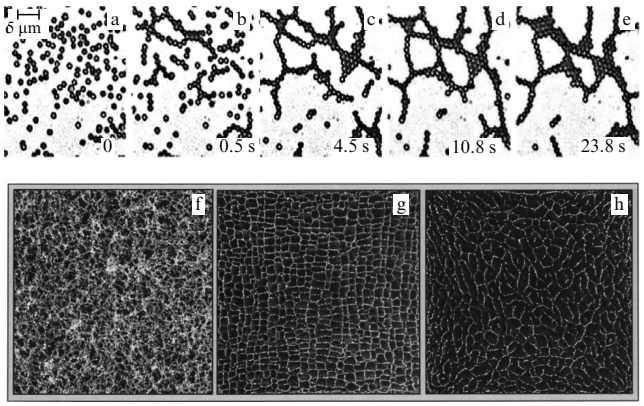


Figure 5. (Color online) (a–e) Formation of a self-healing membrane from a dilute suspension of superparamagnetic particles (0.15 particles per μm^3) after a rotating magnetic field field has been switched on. The appearance of short chains (0.5 s) is followed by their growth, branching, and the formation of a loose network (4.5 s). Most remaining unconnected clusters are rapidly captured, and the network is subsequently coarsened such that the membrane patches grow at the expense of the chainlike sections (10.8 s and 23.8 s). (Reproduced with the kind permission of the authors of [35] and the American Physical Society.) (f–h) Select structures made in triaxial magnetic fields. In a triaxial field, a disordered pattern is formed (particle gel) (f). In a two-dimensional heterodyning (with two different frequencies) field, a honeycomb structure forms (g). Three-dimensional heterodyning of the magnetic field leads to a particle foam (h). 50-μm Ni particles are used, images are 1 cm across. (Reproduced with the kind permission of the authors of [55] and the American Physical Society.)

3. Self-propelled colloidal systems

In this section, we consider colloidal suspensions of self-propelled particles. The interest in such systems is driven by a growing demand for nanotechnological applications based on autonomously moving devices capable of performing useful functions on a microscale. These applications include delivery of drug-laden nanoparticles to specific cellular targets (targeted drug delivery) or massive parallel assembly of microscopic machines by autonomously moving agents. A number of original design concepts of such microscopic swimmers have been developed in recent years, including chemically-propelled swimmers energized by a catalytic chemical reaction, and swimmers harvesting the energy of an electric or magnetic field or ultra-violet light radiation; for a recent review, see [41].

3.1 Practical realizations of self-propelled particles

Figure 6 illustrates several practical realizations of microscopic swimmers. A gold-platinum rod, designed by the Sen and Mallouk group [66, 70], swimming in a solution of hydrogen peroxide is shown in Fig. 6a. A typical swimming speed for the $2-3$-μm rod is on the order of 6–10 μm s^{-1}. The current consensus is that the rod is propelled by the self-induced electrophoretic flow powered by the catalytic decomposition of H_2O_2 at the gold/platinum contact. According to Ref. [70], the swimming speed V can be estimated by equating the propulsion force (due to electrochemical reaction) to the drag force:

$$V \sim \frac{SR^2\gamma}{\eta DL} , \tag{5}$$

where L and R are the length and radius of the rod, η is the dynamic viscosity of water, D is the diffusivity of oxygen, γ is the solution/solid interfacial tension, and S is the oxygen generation rate normalized by the surface area. A detailed theoretical consideration of the electrochemical locomotion of bi-metallic rods can be found in [71].

A polystyrene bead half coated by platinum (Janus particle), propelled by the gradient of osmotic pressure due

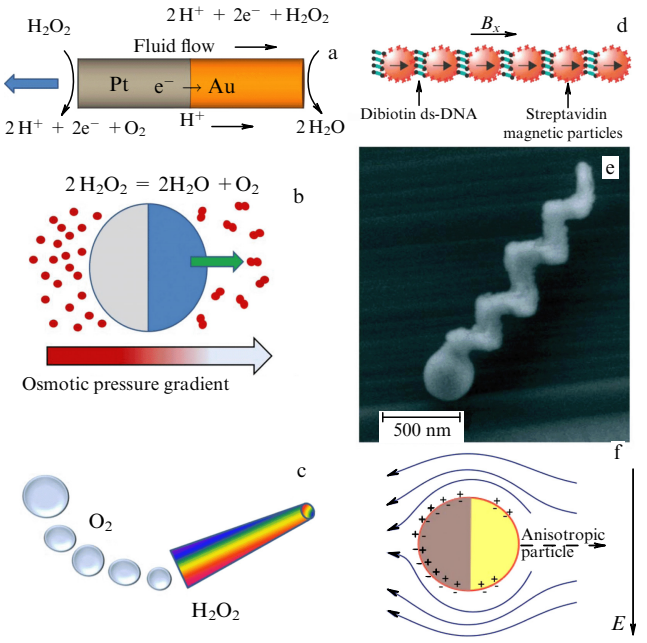


Figure 6. (Color online.) Several practical realizations of microscopic swimmers. (a) A gold-platinum rod propelled by self-electrophoresis in an aqueous solution of H_2O_2 (from [66]). The potential difference created at the Au/Pt contact in the course of catalytic decomposition of H_2O_2 causes electrophoretic flows propelling the rod. Blue arrow depicts swimming direction. (b) Two-sided (Janus) particle propelled by self-diffusiophoresis in an H_2O_2 solution, from [67]. Particles are 1.62-μm polystyrene spheres with one side coated with platinum. (c) High-speed bilayer polyaniline/platinum microtubes (rockets) grown in the conically shaped pores of a polycarbonate template membrane and propelled by O_2 microbubbles in an H_2O_2 solution [68]. (d) Artificial magnetic microswimmer. Superparamagnetic particles are coated with streptavidin (red cross symbols). Under an applied magnetic field B_x, the particles form chains along the field direction. Double-stranded DNA with biotin at each end binds the particles together (biotin and streptavidin are complementary proteins forming strong links). The chain is then attached to a red blood cell and energized by an ac magnetic field perpendicular to B_x (from [31]). (Reproduced with the kind permission of the Nature Publishing Group.) (e) Scanning electron microscopy image of an individual glass screw (nano-propeller) with nano-structured helicity. A 30-nm layer of a ferromagnetic material (cobalt) is deposited on one half of the helix. The propeller is then energized by a rotating triaxial magnetic field [69]. (Reproduced with the kind permission of the American Chemical Society.) (f) A metal–dielectric particle (a 4-μm polystyrene bead partly covered with gold) is propelled by ac electric field. The electric double layer on the gold side (black hemisphere) is more strongly polarized and thus drives a stronger induced charge electroosmosis slip (arrows) than the polystyrene side, resulting in induced charge electrophoresis motion in the direction of the dielectric side [43].

to self-diffusiophoresis, is shown in Fig. 6b. Platinum catalyzes the reduction of hydrogen peroxide (fuel) to oxygen and water, producing more molecules of reaction product than consumed fuel. The characteristic swimming speed is about 2–3 μm s^{-1} [67]. The propulsion velocity V can be estimated by using the lateral gradient of the excess solute concentration in the vicinity of the particle, resulting in

$$V \sim \frac{k_B T l^2 k}{\eta D}, \tag{6}$$

where k is the reaction rate, k_B is the Boltzmann constant, T is the temperature, and l is the range of of the interaction zone between the particles and the solute. While the details of the propulsion kinetics for bi-metallic rods and Janus particles are different (electrophoresis vs diffusionphoresis), the expressions for swimming speeds, Eqns (5) and (6), are similar.

A high speed microtube conical rocket developed by J Wang's group, UCSD, is shown in Fig. 6c. The rocket is propelled by O_2 bubbles in a hydrogen peroxide solution and swims up to 350 body lengths per second; typical rocket size is 5–10 μm [68]. Similar bubble-propelled particles (nanosubmarines) have also been designed by O Schmidt's group [72]. In contrast to Janus particles and gold-platinum rods, the decomposition of the hydrogen peroxide occurs on the inner Pt surface of the conical rocket, and thus does not have the ionic-strength limitation of catalytic bi-metallic rods.

In addition to chemical reactions, self-propelled colloidal particles can be energized by ac electric or magnetic fields. Some of these design concepts are inspired by Nature. Figure 6d shows artificial microswimmers (magnetic sperm) [31]. The swimmer is assembled from a dilute suspension of superparamagnetic nanoparticles linked by DNA molecules and attached to a red blood cell (cargo). The direction of swimming is prescribed by an applied dc magnetic field B_x, which orients the magnetic chain (flagellum). The swimmer is then energized by an ac magnetic field B_{ac} applied perpendicular to B_x. The typical size of the swimmer is about 5 μm, and the speed is about 20 μm s^{-1}, similar to that of motile bacteria. For this kind of artificial swimmer, the relevant dimensionless parameter characterizing the propulsion ability is the so-called sperm number S_p:

$$S_p = L\left(\frac{\xi_\perp \omega}{\kappa}\right)^{1/4}, \tag{7}$$

where L is the length of the filament, κ is its bending rigidity, $\omega = 2\pi f$, f is the frequency of the ac magnetic field B_{ac}, and $\xi_\perp$ is the perpendicular viscous friction coefficient.[3] The sperm number represents the relevance of viscous to elastic stresses on the filament. The experiments revealed that maximum dimensionless propulsion speed $V/L\omega$ occurs for $S_p \approx 3$ [31].

Another design of a microswimmer, inspired by a bacterium propelled by the rotation of helical flagella, is presented in Fig. 6e. A half of a nanoscale glass propeller is covered by ferromagnetic material (cobalt) and brought into motion by a rotating magnetic field [69]. These cork-screw particles can swim up to 40 μ s^{-1}. A magnetic swimmer based on a rotation of a magnetic doublet coupled to a boundary interaction was studied in Ref. [32]. In the broader context of magnetic swimmers, the self-assembled snake [33], discussed in the previous section, is different from all of these designs, since it utilizes self-propulsion due to a spontaneous symmetry breaking.

Metal-dielectric microparticles can be brought into motion by an ac electric field [43, 74]. In this case, the particles move perpendicular to the field direction by induced-charge electrophoresis, arising due to 'induced-charge electroosmosis', i.e., the action of an applied electric field on its own induced diffuse charge near a polarizable surface. For typical experimental conditions, the speed of particles perpendicular to the applied field was on the order of 30 μm s^{-1} (Fig. 6f). This phenomenon may find applications in various microfluidic devices.

[3] The hydrodynamic slender body theory gives for long thin filaments $\xi_\parallel/\xi_\perp \approx 2$, where $\xi_\parallel$ is the tangential friction coefficient [73].

3.2 Individual and collective motion of colloidal microswimmers

Figure 7 shows trajectories of individual self-propelled gold/platinum rods and platinum–silica Janus particles. The trajectories are not straight lines; the orientation of the particles changes randomly, likely due to thermal diffusion and interaction with gas microbubbles created earlier by other particles. Additional factors contributing to the diffusive-like behavior are imperfections of the rods, like roughness of the surface, curvature etc. The overall behavior is similar to the so-called run-and-tumble motion of motile bacteria [75]. In this context, one can discuss mean-square displacement (msd) of the colloidal particle, Δr^2. According to the Stokes–Einstein relation for a Brownian particle of radius R, the msd is linear in time, $\Delta r^2 \sim 6Dt$ in three dimensions and $\Delta r^2 \sim 4Dt$ in two dimensions, where $D = k_B T/(6\pi\eta R)$ is the particle's diffusion coefficient.[4] The particle will also perform rotational diffusion with the characteristic time scale τ_r. For a self-propelled particle moving with the velocity V, the direction of motion itself will be subject to rotational diffusion, leading to coupling between the translational and rotational motions. In this case, the self-propulsion results in a significant increase in the effective diffusion coefficient. According to [67], the two-dimensional projection of the msd is of the form[5]

$$\Delta r^2 = 4D\Delta t + \frac{V^2\tau_r^2}{2}\left[\frac{2\Delta t}{\tau_r} + \exp\left(-\frac{2\Delta t}{\tau_r}\right) - 1\right], \qquad (8)$$

where Δt is the observation time. For short times, $\Delta t \ll \tau_r$, the motion of the particles is roughly ballistic, $\Delta r^2 \approx 4D\Delta t + V^2\Delta t^2$. For long times, $\Delta t \gg \tau_r$, we obtain from Eqn (8)

$$\Delta r^2 \approx 4D\Delta t + V^2\tau_r\Delta t = 4D_{\text{eff}}\Delta t\,. \qquad (9)$$

Thus, due to the coupling between translation and rotation, the particle performs random walk with the effective diffusion coefficient $D_{\text{eff}} = D + V^2\tau_r/4$. Self-propulsion can, therefore, result in a significant increase in a particle's self-diffusion.

One of the fundamental issues in the context of artificial microswimmers is the possibility of collective motion for higher concentrations, similar to that exhibited by swimming bacteria, where large-scale patterns of collective locomotion, arising purely from collisions and hydrodynamic interactions, have been observed [76, 77] (see Fig. 10a in Section 4.2). Collective motion of microswimmers has many advantages compared to individual swimming, for example, from the perspective of collective cargo delivery or harvesting the mechanical energy of chaotic motion [78]. To date, no collective behavior similar to that of bacterial systems has been observed in the concentrated suspensions of artificial swimmers. The main reason is possibly due to a too high rotational diffusion of individual particles, which destroys the collective swimming state. In addition, gold/platinum particles show a tendency to form aggregates after a collision. Experiments [79] and simulations [80] with dense suspensions of self-propelled Janus colloids showed the merging and breaking of transient aggregates of particles; the average size of the aggregates grows linearly with the self-propelling velocity.

[4] Here we used the expression for viscous drag force F_d acting on a spherical particle: $F_d = 6\pi\eta RV$ [59].

[5] Systematic particle tracking for the majority of colloidal systems is possible only in two dimensions.

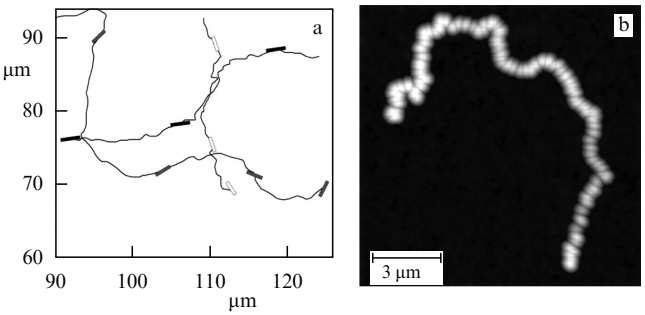


Figure 7. Trajectories of self-propelled particles. (a) Representative trajectories of several individual gold/platinum rods in H_2O_2 solutions, from [70]. (Reproduced with the kind permission of the American Chemical Society.) Intervals of ballistic motion are interrupted by sharp turns, similar to the run-and-tumble behavior of motile bacteria. (b) Projection of 120 frames of a 1-μm half-coated Pt-silica particle traveling in a 10% w/w H_2O_2 solution [81]. (Reproduced with the kind permission of the American Chemical Society.)

A variety of collective chemical responses of artificial swimmers to ultraviolet (UV) light irradiation was reported in Refs [82–84]. Upon ultraviolet light illumination, it was observed that 1 μm silver on silica Janus particles migrate away from the irradiated regions [83]. This behavior is reminiscent of a phototactic response (a tendency to navigate towards light) of biological systems, shown, for example, by some green algae. An interesting 'schooling' behavior was exhibited by micron-size silver chloride (AgCl) particles upon UV irradiation [82]. AgCl particles move autonomously in deionized water by self-diffusiophoresis due to asymmetric AgCl decomposition under the UV light. A moving AgCl particle releases ions, to which the other particles respond by drifting or 'schooling' into regions with higher particle concentrations. When photo-inactive silica particles are mixed with the AgCl, they respond to the ion release by swimming towards and surrounding individual AgCl particles (Fig. 8). Collective chemical oscillations were also observed in a suspension of AgCl particles in the presence of

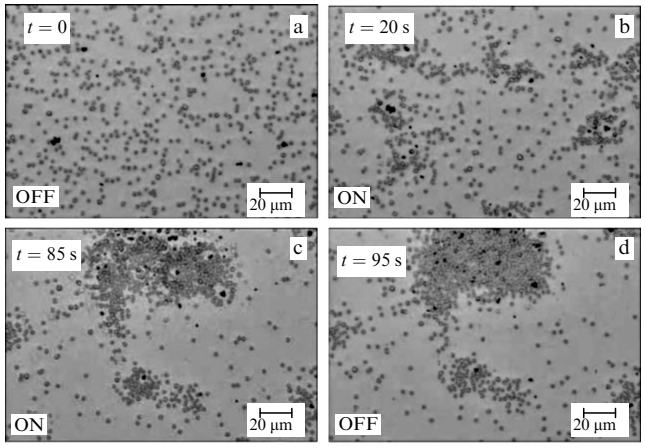


Figure 8. 'Predator–prey' behavior shown by 1-μm AgCl particles (darker objects) and 2.34-μm silica spheres mixed in deionized water. When irradiated with ultraviolet (UV) light (a–c), the silica spheres actively seek out the AgCl particles and surround them. A 2–3-μm exclusion zone is seen around the AgCl particles while the UV light is on; it disappears when the UV light is turned off (d) [82]. (Reproduced with the kind permission of John Wiley and Sons.)

UV light in a dilute hydrogen peroxide solution [84]. Both single-particle and collective, multi-particle responses were observed due to an oscillatory, reversible conversion of AgCl to silver metal on the particle surface. The collective motions of these light-powered microswimmers self-organize into clumped (clustered) chemical oscillators with significant spatiotemporal correlations between the particles.

4. Biocolloids: suspensions of swimming microorganisms

While there are obvious differences between the suspensions of living bacteria and inanimate colloidal systems, there are also many similarities, especially with the systems of artificial swimmers considered in Section 3. For example, the size of the common aerobic rod-shaped bacterium *Bacillus subtilis* is about 5 μm, and typical swimming speed is on the order of 20 μm s^{-1}, similar to that of many artificial swimmers. As a result, in addition to 'pure biological mechanisms', like chemotaxis, cell division etc., the bacteria experience the same spectrum of forces as colloidal self-propelled particles of a similar size: hydrodynamic entrainment, steric repulsion, buoyancy, and thermal fluctuations. Since the Reynolds number for an individual bacterium is exceedingly small, the effects of fluid inertia can be neglected.[6] In addition, some bacteria may possess a permanent magnetic moment (so-called magnetotactic bacteria like *Magnetospirillium magnetotacticum*), and, therefore, they can be manipulated and assembled by an external magnetic field [85, 86].

4.1 Collective behavior

Motile (swimming) microorganisms, like many bacteria, propel themselves in viscous fluid by a rotation of flagella and often develop large-scale patterns of collective locomotion with a characteristic length scale significantly exceeding the size of a single microorganism (see Section 4.2, Fig. 10a). The collective behavior of motile aerobic bacteria, especially at high concentrations, is governed by a subtle interplay among buoyancy, hydrodynamic interactions, oxygen consumption, and collisions. The recent consensus is that the collective motion and self-organization in the suspensions of swimming bacteria often arise due to pure 'physical mechanisms': hydrodynamic interactions between the organisms and short-range collisions (Fig. 9), whereas specific biological mechanisms, such as chemotaxis [75], play a relatively minor role. Large-scale self-organization is driven by the input of mechanical energy from the rotating bacterial flagella. Thus, in contrast to high Reynolds number hydrodynamic turbulence, the energy injection in bacterial suspensions occurs on a microscopic scale.

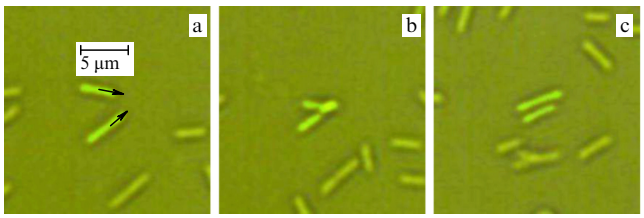


Figure 9. (Color online.) A sequence of images illustrating a collision event between two swimming bacteria *Bacillus subtilis* in a thin free-standing liquid film [77]. Black arrows indicate the swimming direction. The collision results in an alignment of bacteria.

[6] For a bacterium of length $L = 5$ μm swimming with the speed $V = 20$ μm s^{-1} in water, and kinematic viscosity of water $\nu = 0.01$ cm^2 s^{-1}, the corresponding Reynolds number is $\mathrm{Re} = LV/\nu \approx 10^{-4}$.

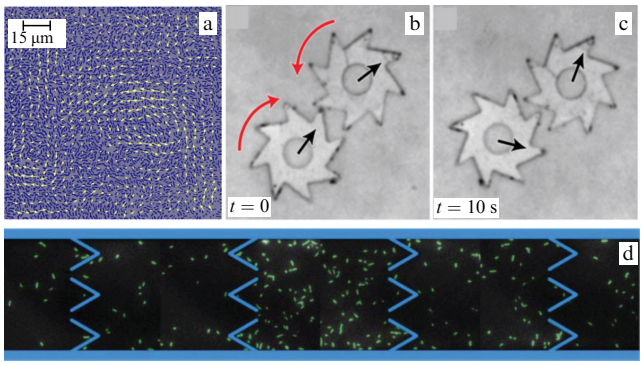


Figure 10. (Color and animation online.) (a) Collective motion in a concentrated suspension of motile bacteria *Bacillus subtilis*; bacteria are seen as short dark stripes, arrows indicate the direction of bacterial flow [77]. The collective swimming is manifested by a formation of swirls and jets with the characteristic scale exceeding the length of a bacterium by an order of magnitude (movie 10.1 courtesy of Andrey Sokolov). (b, c) Motile bacteria *Bacillus subtilis* power microscopic (0.5 mm) gears; each gear is about one million times heavier than an individual bacterium (movies 10.2 and 10.3) [78]. (d) Fluorescence micrograph of individual motile bacteria *Escherichia coli* (labeled green) inside a microchannel with funnel-shaped barriers. The chambers, where the rectification bias is reversed, have a higher concentration of bacteria [88].

Experimental studies of suspensions of *Bacillus subtilis* revealed that at a concentration above some critical one, a gradual transition from random swimming to collective locomotion occurs [76, 77]. The collective motion is characterized by a sixfold to eightfold increase in the velocity correlation length and up to a sixfold increase in the swimming speed of bacteria. Such a dramatic increase in the swimming speed in the collective motion mode is likely an outcome of two effects: hydrodynamic drag reduction in more dense bacterial 'packs' and more efficient energy injection due to synchronization of bacterial flagella by hydrodynamic coupling [87].

4.2 Rectification of random motion

Suspensions of swimming bacteria exhibit material properties very different from suspensions of passive particles. It was found recently that in concentrated suspensions of motile bacteria, such as *Bacillus subtilis*, *Escherichia coli*, the activity (i.e., swimming) results in a reduction of viscosity [44] and a dramatic increase in diffusivity [45, 46, 89]. Moreover, it was demonstrated that it is possible to harvest the mechanical energy of random bacterial motion [78]. While the laws of thermodynamics prohibit extraction of useful work from the Brownian motion of particles in equilibrium, these motions can be 'rectified' under nonequilibrium conditions, for example, in the presence of asymmetric geometrical obstacles. This is demonstrated by Fig. 10b, c, where aerobic bacteria *Bacillus subtilis* moving randomly in a suspended fluid film power submillimeter gears decorated with asymmetric teeth and an assembly of gears. The gears' angular velocities can be controlled by the amount of oxygen available to the bacteria. In contrast to passive particles, the bacteria slide along the slanted edges of the gear's teeth and become trapped in the corners at the tooth junctions for extended

periods of time before finally escaping these 'traps'. Since bacteria are self-propelled objects, they push against the gear wall during the trapping events and effectively transfer momentum to the gear.

Let us consider here how much power W_g a gear can extract from the chaotic motion of bacteria. This power can be estimated as $W_g \sim f_r \Omega^2$, where f_r is the rotational drag coefficient, and Ω is the gear's rotation rate. Approximating a gear by a thin disk of radius R, one obtains $f_r \approx 32\eta R^3/3$ [73]. For typical experimental conditions, $R \approx 200$ μm, and $\Omega \approx 1/10$ rad s^{-1} (1 rotation per minute), giving $W_g \sim 10^{-16}-10^{-15}$ W, i.e., on the order of a few femtowatts. A single bacterium delivers power on the order of $W_b \sim f_b V^2 \sim 10^{-18}-10^{-17}$ W, where f_b is the translational viscous drag coefficient for a single bacterium. Thus, only a few hundred bacteria are needed to rotate one microgear.[7] The ability to harness and control the power of collective motions appears to be an important requirement for further development of mechanical systems driven by microorganisms or synthetic swimmers.

Figure 10d illustrates rectification of bacteria concentration in microfluidic chambers with ratchet funnel walls [88]. In equilibrium, the averaged concentration of bacteria should be identical in all chambers. However, since the bacterial suspension is a nonequilibrium system, it is possible to trap most of the bacteria in one chamber due to a rectification bias imposed by the ratchet walls. The mechanism of rectification is similar to that described above.

4.3 Reduction of viscosity

The way in which swimming microorganisms alter the effective properties of the suspension depends on the type of propulsion. For example, most swimming bacteria are propelled from behind, by the rotation of helical flagella. In contrast, unicellular algae like *Chlamydomonas reinhardtii* swim by beating two cilia protruding forward from the body. As a result, the flow patterns created in the course of swimming by these microorganisms are very different. Since no net force is applied to a self-propelled particle, the microorganism imposes a *force dipole* on a suspending fluid — the propulsion force is compensated by viscous drag. Far from the microorganism, the hydrodynamic flow velocity $\mathbf{v}$ is therefore described by a point hydrodynamic dipole:

$$\mathbf{v} = -u_0 \mathbf{r}\left(\frac{3(\mathbf{rd})^3}{r^5} - \frac{1}{r^3}\right), \qquad (10)$$

where $\mathbf{d}$ is the unit vector of dipole orientation, $\mathbf{r}$ is the radius vector relative to the center of the dipole, and $u_0 = \alpha V_0 L^2$ is the hydrodynamic dipole strength [73] (V_0 is the magnitude of swimming speed, L is the length of a bacterium, and $\alpha \sim 0.1-0.3$ is a factor determined by the shape of a bacterium). The sign of the dipole u_0, however, depends on the details of propulsion, and is negative for bacteria-like swimmers (pushers) and positive for algae-like swimmers (pullers). The asymptotic flow field of a bacterium is shown in Fig. 11a. The corresponding flow fields of swimming bacteria [90] and unicellular algae [91] were measured directly.

To understand the effect of bacteria on suspension's rheological properties, consider the following simple arguments [92]. Since the bacterium has a rod-like cell body (aspect ratio is about 5), it will be oriented by a shear flow. For example, in a pure shear flow, the bacterium will align along the stable axis of the flow (Fig. 11b). Since the bacterium is a force dipole, it does not create an extra force on any external wall, and, therefore, does not change the measured shear stress. However, it will accelerate the liquid along the swimming direction, thus increasing the shear strain rate of the flow. As a result, the effective viscosity, being a ratio of shear stress to strain rate, will decrease in the presence of bacteria. In contrast, the algae-type swimmers (pullers) will increase the viscosity of the suspension.

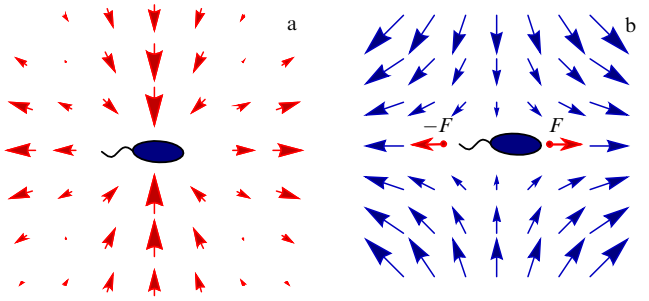


Figure 11. (Color online.) (a) Fluid velocity field created by a swimming bacterium; the swimming dipole creates a flow velocity field with a quadrupolar symmetry [73]. (b) Pure shear (elongational) flow orients a bacterium along the stable direction. The bacterium creates a force dipole with the magnitude F (indicated by red arrows) and accelerates the flow, resulting in a reduction of the effective viscosity.

Experimental studies on the suspensions of various microorganisms have indeed revealed a significant decrease in viscosity for the suspensions of bacteria [44] and an increase in viscosity for the suspensions of algae [93]. The dependence of viscosity on the concentration of bacteria n is shown in Fig. 12. A sevenfold decrease is observed for relatively high concentrations, in the collective motion mode. With a further

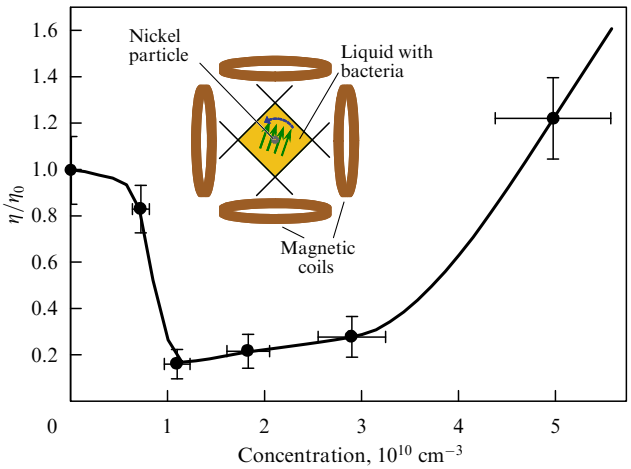


Figure 12. Illustration of viscosity reduction in suspensions of motile bacteria *Bacillus subtilis* [44]. With an increase in bacterial concentration, the effective viscosity of suspension η decreases up to sevenfold compared to the viscosity of suspending liquid η_0. Inset shows the schematics of the experimental setup. A magnetic (nickel) particle is immersed in a free standing liquid film containing bacteria *Bacillus subtilis*. The particle is spun by a rotating magnetic field created by a set of coils; the viscosity is extracted from the measurement of magnetic torque exerted on the particle.

[7] Here we used for the bacterial swimming speed $V \approx 20$ μm s^{-1}, effective bacterium size $a = 2-5$ μm, and translational viscous drag coefficient $f_b = 6\pi\eta a$.

increase in the concentration of bacteria, the viscosity increases again; the upturn is likely due to a jamming of the cells, slowdown of metabolism, an accumulation of waste products, etc.

While the experimental results qualitatively comply with the trend predicted by simple arguments [92], in reality the situation is far more complex. In particular, the experiments are performed in the geometry close to a planar shear flow (flow between two parallel moving walls) rather than pure shear (elongational) flow. In this case, the individual bacterium will be rotated by the flow, instead of aligning along the stable axis [73]. It was shown in theoretical studies that, in addition to the energy injection by swimming bacteria, in the case of planar shear flow one needs to take into account hydrodynamic interactions between the swimmers and their random reorientations (tumbling) in order to explain the reduction of viscosity [94, 95].

5. Theoretical approaches

Since active colloids are out of thermal equilibrium, the powerful machinery of thermodynamics and statistical physics developed for equilibrium systems, e.g., various techniques based on minimization of energy or maximization of entropy, can not be applied in the majority of cases. Thus, direct numerical modeling of colloidal suspensions and simplified phenomenological models have become the most popular research tools. Here, we present a very short (and far from complete) overview of some theoretical and computational approaches.

Different theoretical approaches are often applied to the same experimental system. Continuum coarse-grained models, derived for macroscopic degrees of freedom (concentrations, velocities, polarization, etc.), are more computationally effective than direct particle simulations, and, in general, provide better insights: they depend on a significantly smaller number of parameters, and allow studying bigger systems on longer times scales. On the flip side, systematic derivation of the coarse-grained equations in the out-of-equilibrium situation is a formidable challenge.

5.1 Discrete particle simulations

Figure 13 illustrates the comparison between the results of computer modeling and experiments in suspensions energized by electric [96] or magnetic fields [55] (see Figs 1 and 5). In most cases, there is good qualitative — and sometimes even quantitative — agreement between experiment and computer simulations. While the systems under study are rather different (charged particles in [18, 47] and magnetic particles in [55]), the simulation approaches, different in the details of implementation, have some similarities. In both cases the time evolution of the individual colloidal particle is governed by an overdamped Langevin equation with thermal noise (Brownian dynamics). Colloidal particles are typically treated as hard spheres or discs, with either a fixed dipole moment [55] or electric charge [96], while the hydrodynamic interactions between particles are neglected. Since the implementation of collisions between hard disks or spheres is associated with significant computationally challenges (see, e.g., [97]), 'softened' hard core potential is often used instead: the Lennard-Jones potential or the stiffer repulsive potential

$$V(r_{ij}) = \frac{V_0 \exp\left[-(r_{ij}-a)/\lambda\right]}{r_{ij}-a}\,,$$

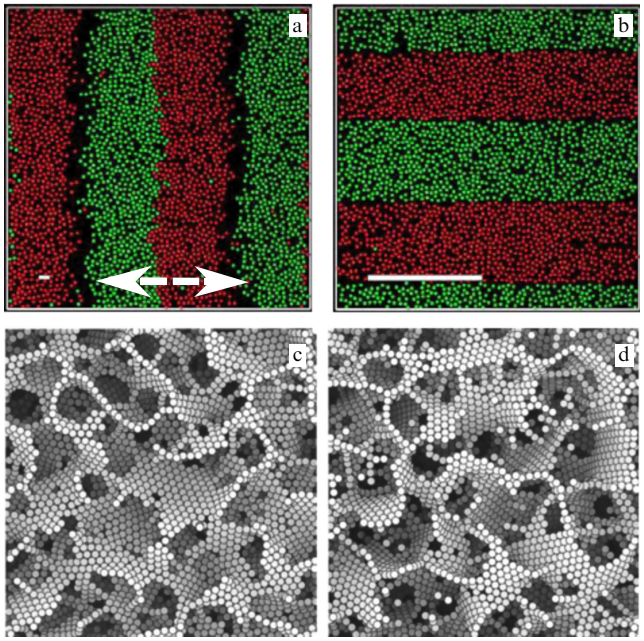


Figure 13. (Color online.) (a, b) Brownian dynamics simulations of a mixture of two oppositely charged colloids; direction of the driving ac electric field is indicated by the broken white arrow, the solid bar size is proportional to the driving force, i.e., field amplitude. In agreement with the experiment [18, 47] (see Fig. 1e, f), static bands are formed perpendicular to the field direction for small driving forces, and lanes are formed parallel to the field's direction for higher driving forces [96]. (c, d) Computer modeling of structures produced in a suspension of magnetic colloids by the ac tri-axial magnetic field. Image (c), produced by the three-dimensional heterodyning fields, is indistinguishable from the image (d), generated by simulated annealing (due to thermal fluctuations, no field; compare to Fig. 5h) [55]. (Reproduced with the kind permission of the American Physical Society.)

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| > a$ is the distance between particles, a is the particle radius, and V_0 and λ are the potential strength and decay length.

The Brownian dynamics can be illustrated for the system of two oppositely charged colloids energized by an electric field [96]. Particles of sorts A and B are subject to oscillatory force $\mathbf{f}^{\mathrm{A}}(t) = -\mathbf{f}^{\mathrm{B}}(t) = f_0 \sin(\omega t)\mathbf{e}_x$, where f_0 and ω are the magnitude and frequency of the driving field, respectively. Equations of motion have the form

$$\zeta^{\mathrm{A,B}}\dot{\mathbf{r}}_i^{\mathrm{A,B}} = \mathbf{f}^{\mathrm{A,B}} - \sum_{i\neq j} \frac{\mathrm{d}V(r_{ij})}{\mathrm{d}r_j} + \xi_i(t)\,, \tag{11}$$

where $\zeta^{\mathrm{A,B}}$ are the particles' mobilities, $\mathbf{r}_i^{\mathrm{A,B}}$ are the positions, and ξ is a random force (thermal noise). Despite these drastic simplifications, the simulations often reproduce key experimental observations: the formation of lanes [96] and 'colloidal foams' [55]. It was shown in [98] that the hydrodynamic interactions between the particles, while critical in some cases, appear to be unimportant for strongly driven oppositely charged colloids, i.e., for lanes.

5.2 Continuum models

The theoretical description becomes significantly more challenging when the flows generated by colloidal particles are crucial, i.e., in the case of pulsating rings and rotating vortices studied in [48] and self-assembled magnetic swimmers [33]. The phenomenological continuum model based

on coupled equations for the evolution of particle concentration and self-induced electro-osmotic flows was proposed in Ref. [52]. The model operates in terms of two-dimensional concentrations (averaged over the vertical coordinate z) of two types of particles: the concentration of gas, i.e., rapidly moving particles ρ_g, and the concentration of immobile particles, or precipitate ρ_p. While the separation into gas and precipitate is somewhat artificial, it allows us to reduce the computational complexity of the problem. Since the total number of particles N is conserved, $N = \int(\rho_p + \rho_g)\,dx\,dy = \text{const}$, the dynamics is described by the corresponding conservation laws

$$\partial_t \rho_g = -\nabla \mathbf{J}_g + \tilde{f}, \qquad \partial_t \rho_p = \nabla \mathbf{J}_p - \tilde{f}, \tag{12}$$

where $\mathbf{J}_{g,p}$ are the corresponding mass fluxes, and function $\tilde{f}$ describes the gas/precipitate conversion which depends on $\rho_{g,p}$, the applied vertical electric field E, and conductivity of the electrolyte (proportional to the concentration of alcohol c). The fluxes can be written in the following generic form (a combination of diffusive and advective fluxes)

$$\mathbf{J}_{g,p} = -D_{g,p}\nabla\rho_{g,p} - \alpha_{g,p}(E)\,\rho_{g,p}\mathbf{v}_\perp\left(1 - \beta_{g,p}(E)\,\rho_{g,p}\right). \tag{13}$$

Here, $D_{g,p}$ are the diffusivities, and $\alpha_{g,p}$, $\beta_{g,p}$ are linear/quadratic advection coefficients for each fraction. The term $1 - \beta_{g,p}\rho_{g,p}$ describes flux saturation for large particle concentrations. $\mathbf{v}_\perp$ is the horizontal (in-plane) velocity of the liquid. The horizontal velocity $\mathbf{v}_\perp$ can be obtained from the incompressibility condition $\nabla\mathbf{v}_\perp + \partial_z v_z = 0$. Assuming that the vertical vorticity $\omega_z = \partial_y v_x - \partial_x v_y$ is small compared to the in-plane vorticity, for $\omega_z = 0$ the horizontal velocity can be obtained from the quasipotential ϕ: $\mathbf{v}_\perp = -\nabla\phi$. The incompressibility condition then yields

$$\nabla^2\phi = \partial_z v_z\,. \tag{14}$$

In turn, the depth-averaged vertical (z) component of velocity $V = h^{-1}\int_0^h v_z\,dz$, where h is the thickness of the cell, can be obtained from the corresponding Navier–Stokes equation (see for detail [52]):

$$\partial_t V = \nu\nabla^2 V - \zeta V + cE\int K(|\mathbf{r} - \mathbf{r}'|)(\rho_g + \rho_p)\,d\mathbf{r}'\,, \tag{15}$$

where ν is the kinematic viscosity, ζ accounts for the dissipation due to friction with the walls, and the last term describes the depth-averaged force acting on charged weakly conducting fluid containing charged microparticles. $K(|\mathbf{r} - \mathbf{r}'|)$ is a phenomenological localized kernel which determines the scale of emergent patterns. With the appropriate choice of the model parameters and functional form of the conversion rate $\tilde{f}$, the model captured, on a qualitative level, the entire complexity of the experimental phase diagram, as well as the primary patterns observed in the experiment: static crystals, honeycomb lattices, and even dynamic pulsating rings and rotating vortices (Fig. 14a–f).

A hybrid model based on discrete particle dynamics and a continuum hydrodynamic model for the liquid was used to describe the formation of self-assembled magnetic snakes and swimmers [58] (Fig. 14g–i). In this approach, the Newton equations for positions and orientations of colloidal particles at the interface between air and water were coupled to the Navier–Stokes equation for the fluid flow. The full nonlinear Navier–Stokes equation (rather than the linear Stokes equation) is necessary for describing the rectified large-scales flows generated by the snakes. To reduce the computational load associated with the numerical solution of the three-dimensional nonlinear partial differential equation for the fluid flow, the Navier–Stokes equation was solved in a much simpler shallow water approximation [59] for the depth-averaged horizontal velocity $\mathbf{v}$, valid when the thickness of water layer h becomes small compared to the size of a magnetic snake. These equations are of the form

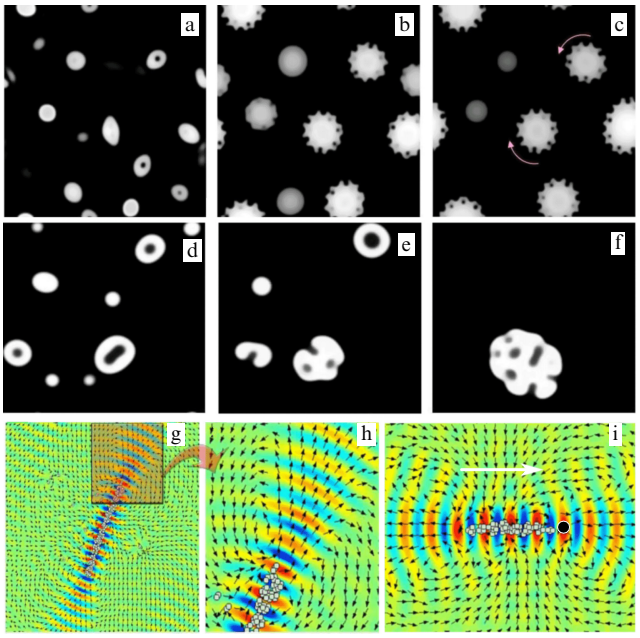


Figure 14. (Color and animation online.) (a–c) Formation of rotating vortex 'stars' in the model of electrostatically driven nonaqueous suspension of conducting microparticles, (movie 14.1) [52]. (d–f) Rotating vortices transform into randomly pulsating rings upon reversal of the direction of an applied electric field (movie 14.2). The structures obtained by computer modeling qualitatively resemble those observed experimentally; compare Fig. 2a–c, rotating stars and (d–i), pulsating rings. (g) Magnetic snake formed from 225 initially randomly distributed particles at the water–air interface; colors indicate elevation of the water surface, and arrows denote hydrodynamic surface flow generated by the snake (movie 14.3). (From [58].) (h) Flow pattern in the vicinity of the snake's tail. (i) Self-assembled swimmer formed by a snake attached to a nonmagnetic bead (large black circle); white arrow indicates the direction of swimming (movie 14.4); compare to images shown in Fig. 3b–f.

$$\partial_t h + \nabla h\mathbf{v} = 0\,, \tag{16}$$

$$\partial_t\mathbf{v} + \mathbf{v}\nabla\mathbf{v} = \nu(\nabla^2\mathbf{v} - \zeta\mathbf{v}) - \nabla h + \sigma\nabla\nabla^2 h + H_0\sin(\omega t)\sum_j s(\mathbf{r} - \mathbf{r}_j)\mathbf{p}_j\,. \tag{17}$$

Here, σ is the surface tension, and ζ describes friction with the bottom. The last term describes the flow generated by the vertical ac magnetic field with the magnitude H_0 and frequency ω acting on magnetic dipoles with the orientation $\mathbf{p}_j = (\cos(\phi_j), \sin(\phi_j))$, and function $s(r)$ describes the particle's shape. The positions $\mathbf{r}_j$ and orientations ϕ_j of the particles are governed by the following Newton equations of motion:

$$m_p\ddot{\mathbf{r}}_j + \mu_t\dot{\mathbf{r}}_j = \mathbf{F}_j + \mu_t\mathbf{v} - \beta\nabla h\,, \tag{18}$$

$$I_p\ddot{\phi}_j + \mu_r\dot{\phi}_j = T_j + \Sigma_j + \kappa H_0\sin(\omega t)\nabla h\times\mathbf{p}_j\,, \tag{19}$$

where m_p, I_p, μ_t, and μ_r are the particle's mass, moment of inertia, and translational and rotational viscous drag coefficients, and $\mathbf{F}_j$ and T_j are forces and torques due to magnetic dipole-dipole interaction and steric repulsion between the particles (torques have only one nonzero component). Σ_j is the torque exerted by shear hydrodynamic flow on a solid particle.[8] Equations (17)-(19) have the following meaning: in addition to magnetic and hydrodynamic forces and torques $\mathbf{F}_j$, T_j, and Σ_j, the particles are subject to advection by the hydrodynamic flow ($\sim \mathbf{v}$), sliding down the gradient of the surface due to gravity ($\beta\nabla h$). The last term in Eqn (19) deserves special attention: it describes the magnetic alignment of particles along the direction of projection of the external ac magnetic field parallel to the surface of the fluid. The in-plane component of the field, oscillating with the same frequency ω, appears due to the deformation of the surface, and is proportional to the slope ∇h. This in-plane field facilitates antiferromagnetic ordering of the neighboring segments and is crucial for the formation of snakes and asters. The model Eqns (16)-(19) successfully reproduced the assembly of a snake from random dispersion of the particles, as well as formation of the self-propelled snake-bead hybrid (Fig. 14g–i).

5.3 Kinetic approach

Significant progress has been achieved in the understanding of collective motion in suspensions of self-propelled swimmers via analysis of simplified probabilistic models, while the detailed role of the specific physical mechanism is a matter of active discussion. The most recent theoretical approaches are based on the mean-field Fokker–Plank–Smoluchowski-type kinetic equations for the time-dependent probability density $P(\mathbf{r}, \mathbf{d}, t)$ of finding a particle in a certain position $\mathbf{r}$ with an orientation $\mathbf{d}$. In the majority of these approaches, the hydrodynamic interactions are described in the dilute limit, when the flow field generated by many bacteria is close to a superposition of individual fields [99, 100]. The bacterial collisions are often incorporated via binary collision integral [101]. In the most general form, the kinetic equation can be written as follows:

$$\partial_t P = -\nabla_r\big((V_0\mathbf{d} + \mathbf{u})P\big) - \nabla_n(\omega P) + I_c + D_t\nabla_r^2 P + D_r\nabla_d^2 P\,, \tag{20}$$

where, V_0 is the speed of an individual bacterium with respect to the liquid in the direction of its unit orientation vector $\mathbf{d}$, $\mathbf{u}$ is the hydrodynamic velocity induced by all swimming particles, $\omega = (\mathbf{I} - \mathbf{dd})(\gamma\mathbf{E} + \mathbf{W})\mathbf{d}$ is the rotational velocity induced by bacteria, $\mathbf{E}$ and $\mathbf{W}$ are the rate of strain and vorticity tensors, γ is the shape parameter which is close to one for slender bodies like bacteria, D_t and D_r are translational and rotational diffusions (e.g., due to tumbling), and ∇_r and ∇_d are vector differential operators for positions $\mathbf{r}$ and orientations $\mathbf{d}$ [73]. I_c is a binary collision integral describing short-range alignment interaction between the bacteria: direct steric collisions between the bacteria make their orientations after collision more aligned, and, consequently, more correlated [101] (see Fig. 9). In addition to collisions, bacteria interact via long-range hydrodynamic forces described by the translational and rotational velocities $\mathbf{v}$, ω. Various versions of Eqn (20) are solved directly (mostly in 2D), or reduced to a much simpler system of equations for coarse-grained quantities, such as concentration $\rho = \int P\,\mathrm{d}\mathbf{d}$ or average orientation $\tau = \int \mathbf{d}P\,\mathrm{d}\mathbf{d}$. While these kinetic models are able to reproduce some key experimental observations, like the onset of motion at some critical concentration, the agreement with experiment remains mostly qualitative.

[8] For typical experimental conditions, this torque appears to be small compared to other terms in Eqn (19).

6. Conclusions

Active colloids is a rapidly expanding and developing area, and many scientific and technological breakthroughs may occur in the near feature. One of the intriguing research avenues is the collective behavior of shape-anisotropic colloids, like rods, platelets, or even more complex shapes, e.g., chiral. Similar to liquid crystals [6], where a variety of nontrivial phases arises due to the shape anisotropy of the molecules, we can anticipate a rich spectrum of nontrivial dynamic states and structures when the suspension of anisotropic particles is energized by electric or magnetic fields. These states, showing unique optical and mechanical properties, will play a bigger role in reconfigurable smart materials for emerging technologies based on the self-assembly of anisotropic colloids, from photonics [102, 103] to microfluidic machines [62] and robotics [36, 104, 105].

Scientific breakthroughs are expected in the field of artificial swimmers. Technical progress will likely happen in the design of swimmers, optimization of their swimming abilities, and functionalization for specific applications. Current swimmer designs are based mostly on hydrogen peroxide propulsion, which makes them unsuitable for the majority of medical applications related to targeted drug delivery. The use of biologically common fuels, such as glucose, is highly desirable and promising in this respect. Another intriguing application is the design and optimization of microswimmers for specific collective tasks, like targeted cargo delivery, parallel assembly, scavenging of contaminants, and other tasks generally associated with 'swarm intelligence', an approach to the coordination of multirobot systems consisting of large numbers of simple microrobots [106].

Thanks to the revolutionary increase in computing power, a significant fidelity increase in the numerical algorithms for active colloidal suspensions is widely anticipated. The majority of current theoretical approaches are based on direct discrete particle simulations with stochastic forces due to thermal fluctuations, whereas hydrodynamic interactions are included on a highly simplified level, mostly via viscous friction acting on the particle or in the mean-field approximation (see, e.g., [58, 61, 98, 107]). A more adequate treatment of the hydrodynamic forces between colloidal particles beyond pair-wise interactions is highly desirable, but still computationally prohibitive. In addition to discrete particle simulations, a number of continuum phenomenological approaches have been proposed to describe dynamic states, like rotating binary clusters and self-propelled magnetic snakes [33, 52]. Thus, further progress is expected in the derivation of coarse-grained models for dynamic states in colloidal suspension. The models derived from the first principles, for example, in the framework of kinetic theories, should incorporate long-range forces (hydrodynamic, magnetic), as well as short range interactions (collisions, lubrications), etc.

There has been a rapidly growing number of computational and theoretical studies on the generic dynamic and statistical properties of collective behavior exhibited by self-

propelled particles with simplified interactions, from point-like particles to rigid rods and swimming flagella (see, e.g., [108–113]). A variety of dynamic phases have been predicted, including moving clusters, bands, and swarming states. However, to date, the connection between these simulations and experimentally the observed dynamics of self-propelled colloids has remained not fully satisfactorily explained. Refined computational models will likely incorporate, on a more realistic level, the specific properties of self-propelled colloids and a variety of inter-particles forces, whether electric, magnetic, steric, or hydrodynamic.

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Science and Engineering, under Contract DEAC02-06CH11357.

References

1. Hunter R J, White L R, Chan D Y C *Foundations of Colloid Science* Vol. 1 (Oxford: Clarendon Press, 1987)
2. Norde W *Colloids and Interfaces in Life Sciences* (Boca Raton, Fla.: CRC, 2003)
3. Anscombe N, Craig F, Harris S *Eng. Technol.* **7** (2) 68 (2012)
4. Sagis L M C *Rev. Mod. Phys.* **83** 1367 (2011)
5. Frenkel D *Physica A* **313** 1 (2002)
6. Kléman M, Lavrentovich O D *Soft Matter Physics*: *an Introduction* (Berlin: Springer-Verlag, 2003)
7. Witten T A *Rev. Mod. Phys.* **71** S367 (1999)
8. Aranson I S, Tsimring L S *Granular Patterns* (Oxford: Oxford Univ. Press, 2009)
9. Israelachvili J N *Intermolecular and Surface Forces* (Burlington, MA: Academic Press, 2011)
10. Ebert F, Keim P, Maret G *Eur. Phys. J. E* **26** 161 (2008)
11. Royall C P et al. *J. Phys. Condens. Matter* **20** 404225 (2008)
12. Hiltner P A, Krieger I M *J. Phys. Chem.* **73** 2386 (1969)
13. Subramanian G et al. *Adv. Mater.* **11** 1261 (1999)
14. Hynninen A-P et al. *Nature Mater.* **6** 202 (2007)
15. Cai W, Shalaev V *Optical Metamaterials*: *Fundamentals and Applications* (Berlin: Springer-Verlag, 2009)
16. Lapine M et al. *Nature Mater.* **11** 30 (2011)
17. Miszta K et al. *Nature Mater.* **10** 872 (2011)
18. Vissers T et al. *Soft Matter* **7** 2352 (2011)
19. Vissers T et al. *J. Colloid Interface Sci.* **361** 443 (2011)
20. Leunissen M E et al. *Nature* **437** 235 (2005)
21. Leunissen M E, Vutukuri H R, van Blaaderen A *Adv. Mater.* **21** 3116 (2009)
22. Martinez-Veracoechea F J et al. *Phys. Rev. Lett.* **107** 045902 (2011)
23. Löwen H et al. *J. Phys. Condens. Matter* **17** S3379 (2005)
24. Frenkel D, Wales D J *Nature Mater.* **10** 410 (2011)
25. Löwen H *J. Phys. Condens. Matter* **20** 404201 (2008)
26. Sacanna S et al. *Nature* **464** 575 (2010)
27. Chen Q, Bae S C, Granick S *Nature* **469** 381 (2011)
28. Erb R M et al. *Nature* **457** 999 (2009)
29. Nykypanchuk D et al. *Nature* **451** 549 (2008)
30. Kraft D J et al. *Proc. Natl. Acad. Sci. USA* **109** 10787 (2012)
31. Dreyfus R et al. *Nature* **437** 862 (2005)
32. Tierno P et al. *J. Phys. Chem. B* **112** 16525 (2008)
33. Snezhko A, Belkin M, Aranson I S, Kwok W-K *Phys. Rev. Lett.* **102** 118103 (2009)
34. Kumacheva E et al. *Phys. Rev. Lett.* **91** 128301 (2003)
35. Osterman N et al. *Phys. Rev. Lett.* **103** 228301 (2009)
36. Snezhko A, Aranson I S *Nature Mater.* **10** 698 (2011)
37. Sanchez C, Arribart H, Guille M M G *Nature Mater.* **4** 277 (2005)
38. Grzybowski B A et al. *Soft Matter* **5** 1110 (2009)
39. Li F, Josephson D P, Stein A *Angew. Chem. Int. Ed.* **50** 360 (2011)
40. Snezhko A *J. Phys. Condens. Matter* **23** 153101 (2011)
41. Ebbens S J, Howse J R *Soft Matter* **6** 726 (2010)
42. Lauga E, Powers T R *Rep. Prog. Phys.* **72** 096601 (2009)
43. Gangwal S et al. *Phys. Rev. Lett.* **100** 58302 (2008)
44. Sokolov A, Aranson I S *Phys. Rev. Lett.* **103** 148101 (2009)
45. Wu X-L, Libchaber A *Phys. Rev. Lett.* **84** 3017 (2000)
46. Sokolov A, Goldstein R E, Feldchtein F I, Aranson I S *Phys. Rev. E* **80** 031903 (2009)
47. Vissers T, van Blaaderen A, Imhof A *Phys. Rev. Lett.* **106** 228303 (2011)
48. Sapozhnikov M V, Tolmachev Y V, Aranson I S, Kwok W-K *Phys. Rev. Lett.* **90** 114301 (2003)
49. Löwen H *Soft Matter* **6** 3133 (2010)
50. Aranson I S et al. *Phys. Rev. Lett.* **84** 3306 (2000)
51. Yeh S-R, Seul M, Shraiman B I *Nature* **386** 57 (1997)
52. Aranson I S, Sapozhnikov M V *Phys. Rev. Lett.* **92** 234301 (2004)
53. Sapozhnikov M V, Aranson I S, Kwok W-K, Tolmachev Y V *Phys. Rev. Lett.* **93** 84502 (2004)
54. Yethiraj A, Van Blaaderen A *Nature* **421** 513 (2003)
55. Martin J E et al. *Phys. Rev. E* **69** 021508 (2004)
56. Belkin M, Snezhko A, Aranson I S, Kwok W-K *Phys. Rev. Lett.* **99** 158301 (2007)
57. Riley N *Theor. Comput. Fluid Dyn.* **10** 349 (1998)
58. Belkin M, Glatz A, Snezhko A, Aranson I S *Phys. Rev. E* **82** 015301(R) (2010)
59. Landau L D, Lifshitz E M *Fluid Mechanics* (Oxford: Pergamon Press, 1987) [Translated from Russian: *Gidrodinamika* (Moscow: Nauka, 1986)]
60. Elsner N et al. *J. Chem. Phys.* **130** 154901 (2009)
61. Martin J E, Anderson R A, Williamson R L *J. Chem. Phys.* **118** 1557 (2003)
62. Grzybowski B A, Stone H A, Whitesides G M *Proc. Natl. Acad. Sci. USA* **99** 4147 (2002)
63. Solis K J, Bell R C, Martin J E *J. Appl. Phys.* **107** 114911 (2010)
64. Nagaoka Y, Morimoto H, Maekawa T *Langmuir* **27** 9160 (2011)
65. Smoukov S K et al. *Soft Matter* **5** 1285 (2009)
66. Paxton W F et al. *J. Am. Chem. Soc.* **128** 14881 (2006)
67. Howse J R et al. *Phys. Rev. Lett.* **99** 048102 (2007)
68. Gao W et al. *J. Am. Chem. Soc.* **133** 11862 (2011)
69. Ghosh A, Fischer P *Nano Lett.* **9** 2243 (2009)
70. Paxton W F et al. *J. Am. Chem. Soc.* **126** 13424 (2004)
71. Moran J L, Posner J D *J. Fluid Mech.* **680** 31 (2011)
72. Sanchez S et al. *J. Am. Chem. Soc.* **133** 701 (2011)
73. Kim S, Karrila S J *Microhydrodynamics. Principles and Selected Applications* (Boston: Butterworth-Heinemann, 1991)
74. Bazant M Z, Squires T M *Phys. Rev. Lett.* **92** 66101 (2004)
75. Berg H C *E. coli in Motion* (New York: Springer, 2004)
76. Dombrowski C et al. *Phys. Rev. Lett.* **93** 98103 (2004)
77. Sokolov A, Aranson I S, Kessler J O, Goldstein R E *Phys. Rev. Lett.* **98** 158102 (2007)
78. Sokolov A, Apodaca M M, Grzybowski B A, Aranson I S *Proc. Natl. Acad. Sci. USA* **107** 969 (2010)
79. Theurkauff I et al. *Phys. Rev. Lett.* **108** 268303 (2012)
80. Thakur S, Kapral R *Phys. Rev. E* **85** 026121 (2012)
81. Ke H et al. *J. Phys. Chem. A* **114** 5462 (2010)
82. Ibele M, Mallouk T E, Sen A *Angew. Chem. Int. Ed.* **48** 3308 (2009)
83. Sen A et al. *Faraday Discuss.* **143** 15 (2009)
84. Ibele M E et al. *ACS Nano* **4** 4845 (2010)
85. Cēbers A, Ozols M *Phys. Rev. E* **73** 021505 (2006)
86. Godoy M et al. *J. Appl. Phys.* **111** 044905 (2012)
87. Elfring G J, Lauga E *Phys. Rev. Lett.* **103** 088101 (2009)
88. Lambert G, Liao D, Austin R H *Phys. Rev. Lett.* **104** 168102 (2010)
89. Miño G et al. *Phys. Rev. Lett.* **106** 048102 (2011)
90. Drescher K et al. *Proc. Natl. Acad. Sci. USA* **108** 10940 (2011)
91. Drescher K et al. *Phys. Rev. Lett.* **105** 168101 (2010)
92. Hatwalne Y et al. *Phys. Rev. Lett.* **92** 118101 (2004)
93. Rafaï S, Jibuti L, Peyla P *Phys. Rev. Lett.* **104** 098102 (2010)
94. Haines B M et al. *Phys. Rev. E* **80** 041922 (2009)
95. Ryan S D et al. *Phys. Rev. E* **83** 050904(R) (2011)
96. Wysocki A, Löwen H *Phys. Rev. E* **79** 041408 (2009)
97. Aranson I S, Tsimring L S *Rev. Mod. Phys.* **78** 641 (2006)
98. Rex M, Löwen H *Eur. Phys. J. E* **26** 143 (2008)
99. Koch D L, Subramanian G *Annu. Rev. Fluid Mech.* **43** 637 (2011)
100. Saintillan D, Shelley M J *Phys. Rev. Lett.* **100** 178103 (2008)
101. Aranson I S et al. *Phys. Rev. E* **75** 040901(R) (2007)
102. Tang S K Y et al. *Adv. Mater.* **23** 2413 (2011)
103. Golovin A B, Lavrentovich O D *Appl. Phys. Lett.* **95** 254104 (2009)
104. Ilievski F et al. *Angew. Chem.* **123** 1930 (2011)

105. Donald B R et al., in *Algorithmic Foundation of Robotics VIII. Selected Contributions of the Eighth Intern. Workshop on the Algorithmic Foundations of Robotics* (Springer Tracts in Advanced Robotics, Vol. 57, Eds G S Chirikjian et al.) (Berlin: Springer-Verlag, 2010) p. 69
106. Beni G, in *Swarm Robotics. SAB 2004 Intern. Workshop, Santa Monica, CA, USA, July 17, 2004. Revised Selected Papers* (Eds E Sahin, W M Spears) (Berlin: Springer, 2005) p. 1
107. Martin J E *Phys. Rev. E* **79** 011503 (2009)
108. Grégoire G, Chaté H *Phys. Rev. Lett.* **92** 025702 (2004)
109. Ginelli F et al. *Phys. Rev. Lett.* **104** 184502 (2010)
110. Ripoll M et al. *Phys. Rev. Lett.* **101** 168302 (2008)
111. Bialké J, Speck T, Löwen H *Phys. Rev. Lett.* **108** 168301 (2012)
112. Saintillan D, Shelley M J *Phys. Rev. Lett.* **99** 58102 (2007)
113. Yang Y, Marceau V, Gompper G *Phys. Rev. E* **82** 031904 (2010)

Physics – Uspekhi **56** (1) 93 – 102 (2013)

METHODOLOGICAL NOTES PACS numbers: **02.70. – c**, **05.45. – a**, **64.60. – i**

Critical phenomena far from equilibrium

E N Rumanov

DOI: 10.3367/UFNe.0183.201301f.0103

Contents

Abstract. Stationary regimes of active systems — those in which dissipation is compensated by pumping — are considered. Approaching the bifurcation point of such a regime leads to an increase in susceptibility, with soft modes making the dominant contribution. Weak noise, which is inherent to any real system, increases. Sufficiently close to bifurcation, the amplitude of random pulsations is comparable to the average value of the fluctuating quantity, as in the case of developed turbulence. The spectrum of critical pulsations is independent of the original noise. Numerical simulation of the neighborhood of a bifurcation point is considered unreliable because of the poor reproducibility of results. Due to the high susceptibility, calculation roundings result in 'chaotic' jumps of the solution in response to a smooth change in the parameters. It is therefore necessary in the simulation process to introduce a small random function of time, white noise. The solutions of the Langevin equations obtained in this way should be processed statistically. Their properties (except for the intensity of pulsations) are independent of the noise induced. Examples of the statistical description of bifurcations are given.

1. Introduction

By an active system, we mean one that, while submerged in a heat bath, is kept far from equilibrium by a certain external action (for example, an applied voltage produces a current in a conductor). In the absence of an external action, the system relaxes to equilibrium with the heat bath. The time this takes is called the relaxation time, and, importantly, the final state is independent of what the initial state was. Switching on an interaction results in the dependence on the initial conditions and on the time scale of relaxation disappearing, leading to a steady-state regime that is the same for the entire set of initial conditions. This set forms a basin of attraction in the space of states, containing an attractor to which all the trajectories of the basin converge.

For a time-independent interaction, steady states can occur among the regimes that set in. Equilibrium can be considered a limit case of a steady-state regime in which the intensity of the external influence tends to zero. One of the characteristics of equilibrium is susceptibility, which can be defined as follows [1]. Let a system be subject to a small sinusoidal perturbation. The state of the system then changes at the same frequency, and the susceptibility is the ratio of the amplitude of this change to that of the perturbation. A similar definition applies to a steady-state far-from-equilibrium regime, in which case a small sinusoidal addition to a constant perturbation should be considered.

As system and/or perturbation parameters change, the steady-state regime can lose its stability and transform into a different regime. For point-like systems, the space of states is finite dimensional, and the point corresponding to a steady state in this space is a limit one. The proximity of bifurcation often suggests a contraction, at least in one direction, of the distance of this point from the boundary of its attraction basin. If the state is somewhat away from the limit point in this direction, the 'force' that returns the system to its original regime is small. For equilibrium systems, there is a visual way to explain why this is the case. Corresponding to a stable point is a minimum of a (thermodynamic) potential, and to an unstable point, a maximum (in the given direction). Bringing the two points together decreases the potential gradient (i.e., force), and thus creates a situation close to indifferent equilibrium (for both point-like and distributed systems). Even a small change in the perturbation causes a significant deviation from the steady-state regime under critical conditions (when the system parameters are close to the bifurcation

E N Rumanov Institute of Structural Macrokinetics and Materials Science, Russian Academy of Sciences,
ul. Akad. Osipyana 8, 142432 Chernogolovka, Moscow region, Russian Federation. E-mail: ed@ism.ac.ru

Received 24 June 2011, revised 22 December 2011
Uspekhi Fizicheskikh Nauk **183** (1) 103 – 112 (2013)
DOI: 10.3367/UFNr.0183.201301f.0103
Translated by E G Strel'chenko; edited by A M Semikhatov

point), and the deviation relaxation time increases; in other words, the susceptibility increases (primarily in its low-frequency part, because at low frequencies the system has enough time to deviate considerably before the external influence changes its direction).

Critical phenomena are due to high susceptibility and, as far as equilibrium is concerned, were first studied in connection with the liquid–vapor transition. The important results obtained then were the increase in susceptibility (in this case, compressibility) and (primarily soft-mode) density fluctuations [2] (specifically, thermal fluctuations, whose variance can be found by thermodynamic calculations). Far from the equilibrium, thermal noise often occurs (random input voltage oscillations being an example). Weak noise has no noticeable effect on the behavior of the system as long as the system parameters are far from the bifurcation point, but approaching this point causes the noise to increase due to the high susceptibility. Indeed, sufficiently deep in the critical region, the random amplitude of pulsations is comparable to the average value of the fluctuating quantity, similarly to the developed turbulence case. The steady-state regime turns out to be chaotic. The pulsation spectrum is independent of the 'core' noise and is dominated by low frequencies. The closer to the bifurcation point, the higher and sharper the low-frequency peak is. Unlike flicker noise [3], this peak disappears as the parameter values move farther and farther from their critical values.

Numerical simulation of the bifurcation region is considered unreliable due to poorly reproducible results. The number of digits is always constant, and unavoidable roundings result in smooth variations of the parameters, leading to jumps in solutions at high susceptibility. The way out, as we see it, is to bring the situation closer to experiment by adding a small random function of time (white noise) to the constant source. The solutions of the Langevin equation [4] obtained in this way are also random functions and should be processed statistically. For given values of the parameters, it is possible to find the variance of the pulsations, their frequency spectra, critical exponents, n-point correlators, etc., thus obtaining a reliable description of critical phenomena. The bifurcation theory must have a bearing on statistics.

In a real system, noise is not white in general, and its statistics are typically unknown. As the bifurcation point is approached, however, the pulsation correlation time increases and definitely exceeds the noise correlation time, with the result that any noise looks like a δ-correlated (white) one. For a complex system, especially for a distributed one, the susceptibility can be difficult to calculate; sufficiently deep in the critical region, nonlinear susceptibility is important. On the other hand, these quantities should not necessarily be used. As noted above, solutions of the Langevin equations are chaotic time series similar to what is obtained in real life experiments. If the series are sufficiently long (which means a long computation time), high accuracy can be achieved in determining bifurcation statistics.

Systems with a relatively small ($n \geqslant 3$ [5]) number of dependent variables allow chaotic solutions without noise addition (dynamical chaos). But the phase trajectories corresponding to such solutions are locally smooth and should be monitored for a long time (for example, by constructing a Poincaré map [6]) in order to detect chaos. The trajectories of the Langevin equations are Brownian. In typical cases, extending the spectral density of critical pulsations to all frequencies (up to the inverse atomic collision time) would give a temperature $\sim 10^5$ K. In fact, the spectrum falls off sharply with increasing the frequency, only soft modes are excited, and the substance remains cold. On a scale that is larger than atomic (but, normally, smaller than the system size), intense mixing occurs.

In bifurcation theory (see, e.g., Refs [7, 8]), following the spirit of the geometry of smooth maps [9], much attention is given to state space singularities, in which context exotic objects such as the 'saddle–node', 'saddle–focus', and, in the case of many dimensions, even more complex objects are introduced, which, elegant as they are, are still not observable in real life. We see below that (1) in the region of criticality, the neighborhood of a limit point becomes a chaotic attractor consisting of Brownian trajectories and (2) singularities are smeared out.

There is a large amount of literature on dynamic chaos with locally smooth trajectories (see, e.g., Ref. [10] for a review). Because noise cannot be fully eliminated in reality, the question arises: are such attractors of interest for physics? Far from bifurcation, where the susceptibility is not high, the Brownian nature of trajectories is not apparent. Our objective is to draw attention to an unusual type of *critical attractor* that becomes relevant if the parameters involved are close to the bifurcation point of a steady-state regime. A critical attractor consists of Brownian trajectories. A similar remark should be made concerning Hamiltonian systems with a saddle point in the phase space. The appearance of a 'stochastic layer' next to the separatrix [11], in fact, implies that noise cannot be neglected in this situation.

The above suggests that it is in principle possible to create devices that, given a complex natural or artificial system, can warn of the nearing bifurcation of its stationary regime (i.e., a catastrophe) based on the enhancement of soft modes in the system noise spectrum. Of course, the soft mode range is specific for each system. For example, for a continuously stirred tank reactor (CSTR), frequencies smaller than the inverse mixture residence time should be considered. In designing such a device, at least a basic understanding is needed of the processes in the system the device is intended for. Macrokinetic equations, including those for mass, momentum, and energy balance, are indispensable for this purpose.

2. Critical susceptibility

We first consider the case of a point-like system, i.e.,

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = f_i(X_1, X_2, \ldots) + y_i(t)\,, \qquad i = 1, 2, \ldots, n\,. \tag{1}$$

Here, X_i are coordinates (dependent variables), t is time, the vector field f_i depends on the coordinates and the constant force applied to the system, and y_i are small time-dependent force additions. If these additions are neglected, system (1) can have constant solutions X_i^{s} satisfying the equations

$$f_i(X_1^{\mathrm{s}}, X_2^{\mathrm{s}}, \ldots) = 0\,. \tag{2}$$

Such solutions correspond to steady-state regimes of an active system specified by the field f_i. If deviations from the steady state, $x_i = X_i - X_i^{\mathrm{s}}$, are assumed to be small, we can write the linear equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -\lambda_{ik}x_k + y_i\,, \tag{3}$$

where repeated indices imply summation and $-\lambda_{ik} = \partial f_i/\partial X_k$ is the Jacobi matrix calculated at the limit point $X_i = X_i^s$ of the state space.

Close to bifurcation, the limit point is approached by the boundary of its attraction basis along at least one direction. We let X denote the coordinate corresponding to this direction; the relaxation time for deviations $x = X - X^s$ is large compared to that for the other degrees of freedom, due to the small 'restoring' force for this coordinate (bifurcation variable). Deviations for the other, fast-relaxing, coordinates are averaged, and hence, relative to x, they can be considered to be equal to the steady-state values $x_i = 0$. This leaves us with only one of Eqns (3),

$$\frac{dx}{dt} = -\lambda x + y\,. \tag{4}$$

Here, the coefficient λ plays the role of a bifurcation parameter, its value at the bifuraction point being $\lambda = 0$. The susceptibility $\sigma(\omega)$ at a frequency ω is defined as the Fourier component ratio x_ω/y_ω. According to Eqn (4),

$$\mathrm{Re}\,\sigma = \frac{\lambda}{\lambda^2+\omega^2}\,, \qquad \mathrm{Im}\,\sigma = \frac{\omega}{\lambda^2+\omega^2}\,. \tag{5}$$

As $\lambda \to 0$, both the real and imaginary parts of the susceptibility increase primarily at low frequencies, $\omega < \lambda$. The closer to the bifurcation point, the higher and sharper the low-frequency peak is.

Weak noise, an inherent feature of any real influence (as exemplified by random oscillations in an input voltage), becomes higher as the susceptibility increases. For sufficiently small λ, the amplitude of random pulsations is comparable to the average value of the fluctuating quantity (i.e., X^s), as in the case of developed turbulence. To put it another way, approaching bifurcation causes the regime to change from steady state to chaotic. As is clear from the above, for an arbitrary initiating noise, the pulsations must have a low-frequency peak in their spectrum. But in contrast to the flicker noise [3], this peak disappears as the parameters are moved away from the bifurcation point.

In the mathematical theory of bifurcations (see, e.g., Refs [7, 8]), much attention is focused on rearrangements in the state space, which are treated using the geometry of smooth maps [9]. We suppose, for example, that there are two singular points, a node and a saddle, which approach one another as the parameters are varied. According to Refs [7, 8], a saddle–node pair forms at the critical values of the parameters, after which the singularity disappears. As we have seen, including noise makes saddle–node and saddle–focus pairs (and, in higher dimensions, more complex objects) unobservable. The trajectories randomly intersect one another and, because of the increased noise, cannot even approximately be considered autonomous. A further point to emphasize is that critical chaos differs from the chaotic solutions that autonomous systems with a relatively small number of equations provide in a certain parameter range [5]. In this last case, the trajectories are locally smooth, and to see the presence of chaos, it is necessary to monitor them for some finite period of time (for example, by constructing the Poincaré map [6]). The critical trajectory turns out to be Brownian and densely covers the neighborhood of the limit point. As already noted above, however, here, in contrast to the standard Brownian motion (thermal fluctuations), only soft modes are excited. The chemically interesting fact of an increased mixing intensity (critical diffusion) should also be noted [13].

If the bifurcation is such that n variables relax anomalously slowly, we can introduce a susceptibility matrix $\sigma_{ik}(\omega)$ as the resolvent of the Jacobi matrix in Eqn (3),

$$\sigma_{ik}(\omega) = (\lambda_{ik} - \mathrm{i}\omega\delta_{ik})^{-1}\,. \tag{6}$$

Here, the exponent -1 denotes the inverse matrix. In the case of equilibrium, the variance of thermal fluctuations can be calculated by thermodynamic formulas. The coefficients λ_{ik} are expressed in terms of the derivatives of the entropy S,

$$\lambda_{ik} = -\frac{1}{2}\frac{\partial^2 S}{\partial x_i\,\partial x_k}\,, \tag{7}$$

taken at the limit point, where the entropy has a maximum. The most detailed studies have been performed on equilibrium bifurcations corresponding to continuous phase transitions. In this case, the bifurcation variable is referred to as an order parameter. Near the critical liquid–vapor transition point in the pressure–temperature plane, the bifurcation curves of a spinodal decomposition converge to the phase equilibrium line (boiling curve), forming a cusp, the end of the bistable (overheating and overcooling) region. In the neighborhood of this point, the compressibility (the susceptibility with respect to the transition order parameter, the density) is large. The theory of continuous transitions and critical phenomena under equilibrium conditions has been worked out in detail (see, e.g., Ref. [12]). We do not discuss these questions here.

3. Langevin equations

Numerical simulation cannot be considered a reliable way of studying the neighborhood of bifurcations because (1) the computed results show a lack of reasonable reproducibility; and (2) a smooth change in parameter values leads not to smooth changes but to random jumps in the solution. The reason for this is obvious: as a result of high susceptibility, roundings involved in the numerical integration of finite-difference equations cause significant changes in these solutions. The way out is to bring the problem formulation closer to the experimental situation. Random pulsations in the critical region not only are observed in numerical simulations but also are seen in experiment. Along with thermal fluctuations, nonthermal noise can be a source of such pulsations, which, in equilibrium, are due to inhomogeneities, impurities, and other factors that act to 'smear out' a continuous phase transition. In active systems, an influence is inevitably accompanied by weak noise at least, necessitating that the governing equations be augmented by a controlled noise term that makes solutions more chaotic compared to what the roundings produce in an uncontrollable way. The resulting solutions — which are random functions of time — should then be subjected to statistical processing.

Equations containing a weak random source are referred to as Langevin equations [4]. In the case of Eqn (4) alone, we take the function $y(t)$ to be white noise:

$$\langle y(t)\,y(t')\rangle = \mathrm{const}\,\delta(t-t')\,. \tag{8}$$

Random pumping gives rise to random motion near the limit point. According to Eqn (4), the correlator is

$$\langle x(t_1)\,x(t_2)\rangle = \exp\left[-\lambda(t_1+t_2)\right] \times \int_{-\infty}^{t_1} \mathrm{d}t \int_{-\infty}^{t_2} \mathrm{d}t'\,\langle y(t)\,y(t')\rangle \exp\left[\lambda(t+t')\right] \tag{9}$$

or

$$\langle x(t_1)\,x(t_2)\rangle = \frac{\text{const}}{2\lambda}\exp\left(-\lambda|t_1-t_2|\right). \tag{10}$$

The constant in Eqns (8) and (10) is $(y^2)_\omega$, where the white noise Fourier component is independent of frequency. For the Fourier components of the correlator,

$$(x^2)_\omega = \int_{-\infty}^{\infty} \langle x(t)\,x(t-t')\rangle \exp(\mathrm{i}\omega t')\,\mathrm{d}t', \tag{11}$$

we have

$$(x^2)_\omega = \frac{2\lambda}{\lambda^2+\omega^2}\,\langle x^2\rangle = \frac{(y^2)_\omega}{\lambda^2+\omega^2}. \tag{12}$$

Formulas (9)–(12) are identical to similar expressions for equilibrium [1]. In Ref. [1], however, the variance $\langle x^2\rangle$ is a known quantity calculated from thermodynamic formulas. For example, for the temperature, $\langle \Delta T^2\rangle = k_\mathrm{B}T^2/c_V$, where k_B is the Boltzmann constant and c_V is the specific heat at constant volume. Therefore, the constant in Eqn (8) is not arbitrary and should be chosen such that equalities (10) and (12) hold at a given variance $\langle x^2\rangle$. In the case of an active system far from equilibrium, the intensity of a random source $y(t)$ should be regarded as a specified noise component of the influence. Pulsations due to this noise are usually stronger than thermal fluctuations.

Using the Langevin equations, high-susceptibility systems can be numerically simulated in a way closely corresponding to the experimental situation. The order of actions is as follows. One or more of the equations is augmented by a weak white noise term and the resulting system is then integrated numerically over a chosen time interval, which should be large compared to the time scales of the original problem (relaxation time, etc.). This procedure is repeated for different values of the bifurcation parameter in the critical region and for a number of values of the noise intensity (spectral density). The data array (or rather a set of time series) so obtained is processed statistically. In Section 4, we consider a simple example of the numerical simulation of critical phenomena.

4. Critical attractor

Viewed from the perspective of nonlinear physics, exothermic reactions are interesting because of the feedback between temperature and the reaction rate in the sense that a reaction causes heating, which in turn accelerates the reaction. A mathematical model for processes involving such reactions should include equations with exponential sources (the reaction rate is usually activation-temperature dependent). Strong nonlinearity makes the physical picture more contrasting and new effects easier to discover. An exothermal transformation in a CSTR is described by the equations [14, 15]

$$\frac{\mathrm{d}\eta}{\mathrm{d}t} = \Phi(\eta, T) - \eta\tau^{-1}, \tag{13}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{Q}{c}\,\Phi - \alpha(T-T_\mathrm{a}) \tag{14}$$

for the reaction product concentration (transformation depth) η and temperature T. The reaction rate in the simplest case is given by

$$\Phi = (1-\eta)k\exp\left(-\frac{E}{T}\right), \tag{15}$$

where k is a constant of dimension frequency and E the activation energy expressed in the same units as temperature. Usually, $E \gg T$. Other notations are: τ is the reactor residence time of the mixture, Q is the reaction heat, c is the specific heat, $\alpha = \tau^{-1} + \tau_\mathrm{f}^{-1}$, τ_f is the reactor cooling time under no-flow conditions, i.e., for $\tau \to \infty$, and T_a is the temperature of the heat bath and the supplied mixture.

It was noted as early as Ref. [14] that system (13), (14) exhibits bistability in the sense that depending on the initial conditions, either a cold (T_a) steady-state regime or a hot ($T - T_\mathrm{a} \approx Q/c\alpha\tau$) regime is established. It is assumed that the mixing time is small compared to τ. Instead of Eqns (13) and (14), a single second-order equation can be considered. For example, expressing η in terms of T and $\mathrm{d}T/\mathrm{d}t$ using Eqn (14), we obtain [16]

$$\ddot{T} = -\frac{\mathrm{d}V}{\mathrm{d}T} - \gamma(T,\dot{T})\,\dot{T} \tag{16}$$

(with the dot denoting time differentiation),

$$-\frac{\mathrm{d}V}{\mathrm{d}T} = -\frac{Q}{c\tau}\,k\exp\left(-\frac{E}{T}\right) + \left[k\exp\left(-\frac{E}{T}\right) + \frac{1}{\tau}\right]\alpha(T-T_\mathrm{a}), \tag{17}$$

$$\gamma = \alpha + \frac{1}{\tau} + k\exp\left(-\frac{E}{T}\right) - \frac{E}{T^2}\left[\alpha(T-T_\mathrm{a}) + \dot{T}\right]. \tag{18}$$

Mathematically, Eqn (16) is equivalent to the equation of motion in one dimension T for a particle of unit mass subject to potential force (17) and the friction force $-\gamma\dot{T}$. If η and T are specified at the initial instant, the initial 'velocity' $\dot{T}$ is determined by Eqn (14). The 'friction coefficient' in Eqn (18), depending on the parameters, can change sign and become negative — a feature that makes Eqn (16) similar to the Van der Pol equation [17]. Whether the inequality $\gamma < 0$ holds depends on the pumping, i.e., the material flow through the reactor.

The potential $V(T)$ has one minimum outside and three extrema (two minima and one maximum in between) inside the bistability region. The extrema $T_\mathrm{d} < T_\mathrm{m} < T_\mathrm{u}$ clearly have positions coincident with the temperatures of steady-state solutions of system (13), (14). A closer look at $V(T)$ allows us to see which of the two stable states corresponds to a deeper minimum. Let the problem parameters be taken as the Damköhler and Semenov numbers,

$$D = \tau k\exp\left(-\frac{E}{T_\mathrm{c}}\right), \qquad S = D(\alpha\tau)^{-1}, \tag{19}$$

and the scale temperature T_c be the temperature at the cusp where the bistable region in the (D, S) plane terminates (and

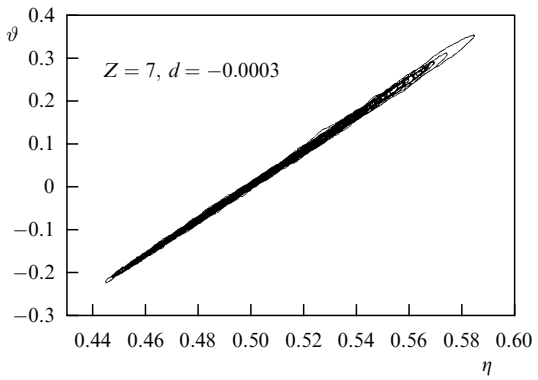


Figure 1. Overall view of the critical attractor.

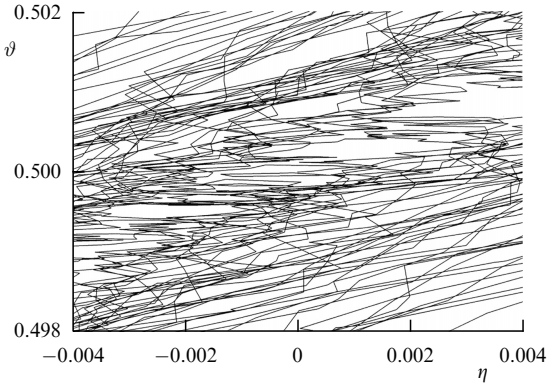


Figure 2. A magnified view (100×) of a small portion of the attractor shown in Fig. 1. The Brownian nature of the trajectory can be seen.

where all three extrema merge). It is readily seen [18] that the cusp is similar to a critical point in the pressure–temperature plane for the liquid–vapor system. The curve $V_d = V_u$ is similar to the boiling curve (phase equilibrium), and the boundaries of the bistable region are similar to spinodals. For the purposes of numerical simulation, it is useful to introduce the dimensionless temperature

$$\vartheta = \frac{E}{T_c^2}(T - T_c)\,, \qquad \exp\left(-\frac{E}{T}\right) \approx \exp\left(-\frac{E}{T_c}\right)\exp\vartheta\,. \quad (20)$$

At the cusp, $D = 1$ and $S = 4/Z$, where $Z = EQ/(cT_c^2)$ is the Zeldovich number, $\vartheta = 0$, and the heat bath temperature is $\vartheta_a = -2$. The convenient parameters to use in the neighborhood of this point are d and s such that $D = 1 + d$ and $S = (4/Z)(1 + s)$, in terms of which the boundary of the bistability region takes the form

$$s = \frac{d}{2} \pm \frac{2\sqrt{2}}{3}(-d)^{3/2}\,. \quad (21)$$

The straight line $s = d/2$ corresponds to phase equilibrium for $d < 0$ and is similar to a critical isochore for $d > 0$. On this line, as before, $\vartheta \approx 0$, and in the bistability region, $\vartheta \approx \pm\sqrt{-6d}$. On the spinodals, $\vartheta \approx \mp\sqrt{-2d}$.

As trajectories approach the limit point in the state space, they start oscillating in a random manner, covering quite a large region, a critical attractor. We examine the properties of such an attractor with the example of the system

$$\dot\eta = (1 - \eta)\exp\vartheta - \eta D^{-1}\,, \quad (22)$$

$$\dot\vartheta = Z(1 - \eta)\exp\vartheta - \frac{2 + \vartheta}{S} - \frac{2}{S}\,y(t)\,, \quad (23)$$

obtained from Eqns (13) and (14) using the transformations discussed above. The source of weak noise in Eqn (23) is turbulence in the reactor cooling system. Figure 1 shows a trajectory [19] in the (η, ϑ) plane for $D_c - 1 = 3 \times 10^{-4}$ and $S_c - 1 = 1.5 \times 10^{-4}$, demonstrating the shape and size of the critical attractor. In Fig. 2, a small portion of this plot is shown on an enlarged scale, demonstrating the Brownian nature of this trajectory.

5. Van der Waals and fluctuation regions

We linearize Eqn (16) in the vicinity of the minimum ϑ_0 to obtain

$$\ddot\vartheta_1 + \gamma_0\dot\vartheta_1 + \omega_0^2\vartheta_1 = 0\,, \quad (24)$$

where the time scale is $(1/k)\exp(E/T_c)$, $\vartheta_1 = \vartheta(t) - \vartheta_0$, $\gamma_0 = \gamma(\vartheta = \vartheta_0, \dot\vartheta = 0)$, and

$$\omega_0 = \left[\exp\vartheta_0\left(\frac{Z}{D} - \frac{3 + \vartheta_0}{S}\right) - \frac{1}{DS}\right]^{1/2}. \quad (25)$$

The frequency ω_0 vanishes at the cusp and is small near it (which implies a small value of the restoring force that occurs when the reactor deviates from the steady-state operation). A small excitation $\varepsilon\exp(-i\omega t)$ in the right-hand side of Eqn (24) produces a response $A\exp(-i\omega t)$. Defining the susceptibility $\sigma(\omega)$ as the ratio A/ε, we obtain

$$\mathrm{Re}\,\sigma = \frac{\omega_0^2 - \omega^2}{R}\,, \qquad \mathrm{Im}\,\sigma = \frac{\gamma_0\omega}{R}\,, \qquad R = (\omega_0^2 - \omega^2) + \gamma_0^2\omega^2\,. \quad (26)$$

With the right-hand side of Eqn (24) taken to be weak white noise, the Fourier components of the correlation function $\langle\vartheta_1(t)\,\vartheta_1(t')\rangle$ are found as

$$(\vartheta^2)_\omega = \rho\gamma_0^{-1}\,\frac{\mathrm{Im}\,\sigma}{\omega}\,, \quad (27)$$

where ρ is the frequency-independent spectral noise density (an analog of the fluctuation–dissipation theorem, with $\rho/2\gamma_0$ playing the role of temperature). The integral over spectrum (27) is the variance, and hence in the limit $\vartheta_0 \to 0$, all modes increase, but the soft one especially so. The closer the critical point is, the higher and sharper the low-frequency peak.

Equations (24)–(27) hold for what can be called the van der Waals region. In a sufficiently small neighborhood of the cusp, the nonlinearity is important. By analogy with Ginzburg's criterion [20], we can compare the variance of critical fluctuations $\langle\vartheta_1^2\rangle$ and the characteristic value $\vartheta_0^2 \sim |d|$ to determine the boundaries of this neighborhood. Nonlinear effects have been investigated in [21] by integrating system (22), (23) numerically. The weak noise in Eqn (23) may be due to turbulence in the reactor cooling system.

Figure 3 shows the variation of the variance $\langle\vartheta_1^2\rangle$ with $|d|$ on a log–log scale. For $d > 0$, i.e., outside the bistability region, we see that the initial power-law increase is followed by saturation. The critical exponent is, according to Eqn (27), -1, to be compared with the least-square value -0.99.

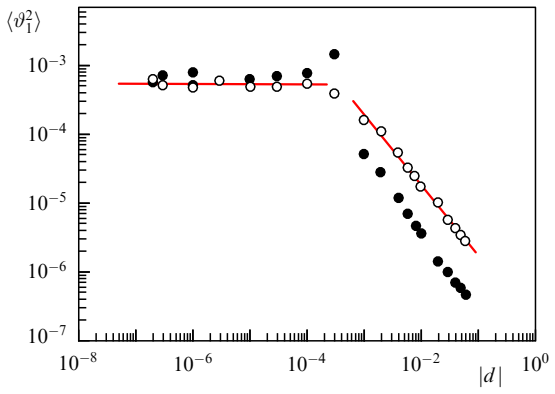


Figure 3. Temperature fluctuation variance as a function of the distance to the cusp: (○) $d > 0$, (●) $d < 0$. Straight lines are obtained by the least-square processing of the corresponding portion of the plot.

6. Kramers transitions

For $d < 0$, there is a variance peak between the van der Waals and fluctuation regions in the plot in Fig. 3. In the bistable region, the potential has two minima, with the barrier between them lowering with decreasing $|d|$. This allows spontaneous transitions between the minima [22] due to the noise-induced diffusion in the field $\mathrm{d}V/\mathrm{d}\vartheta$. The transition probability W obeys the Arrhenius law, which, in this particular case, has the form

$$W \propto \exp\left[-\frac{2\gamma_0}{\rho}(V_\mathrm{m} - V_{\mathrm{d,u}})\right]. \qquad (28)$$

Figure 4 shows a plot of $\vartheta(t)$ for two values of d of the same magnitude but opposite sign. For $d < 0$, along with critical pulsations, large temperature jumps are observed: instead of oscillations about one of the limit points in the (η, ϑ) plane, oscillations about another limit point in the same plane occur. A kind of intermittency is observed in the plot. Transitions between hot and cold states lead to a significant increase in the variance $\langle(\vartheta - \vartheta_0)^2\rangle$. As $|d| \to 0$, the difference $V_\mathrm{m} - V_{\mathrm{d,u}}$ decreases and the contribution from the jumps vanishes. As $|d|$ increases, probability (28) falls off exponentially, and no jumps occur during the computation run. Nonlinear effects in the bistability region are more pronounced due to the quadratic terms in the expansion of $\mathrm{d}V/\mathrm{d}\vartheta$ in powers of ϑ around the potential minimum (for $d > 0$, no such term is present). As can be seen from Fig. 3, this region also contains the interval of a power-law behavior $\langle\vartheta_1^2\rangle \propto |d|^{-n}$, but with an exponent noticeably larger than unity ($n \approx 1.2$).

Similar jumps in a bistable system are described in Ref. [23], which is concerned with stimulated oscillations of a micromechanical vibrator, with bistability due to an S-like nonlinear resonance curve. We also note the numerical simulation of fluctuations at the transition from one potential well to two [24] (where only one — single-minimum — phase was studied).

7. Limit cycle developing out of critical chaos

Unlike the critical point for transitions between phases of the same symmetry, system (22), (23) can have periodic solutions [25–27]. According to Eqn (16), the oscillatory instability boundary is determined by the conditions $\gamma(T = T_0, \dot{T} = 0) = 0$, where T_0 is the temperature of the steady-state regime under consideration. If the parameters vary along the straight line $s = d/2$ ($d > 0$), γ vanishes at $Z = Z_v \approx 8 + 2d^2$. Close to this threshold, for $Z = 8$, the pulse spectrum does not look qualitatively different from that

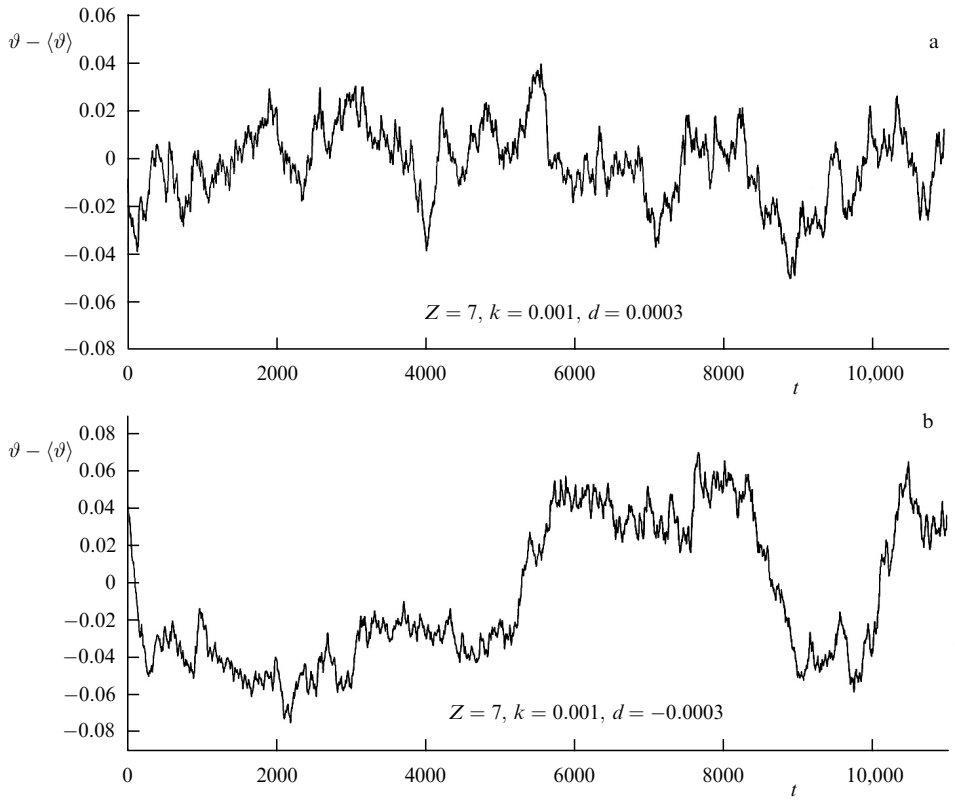


Figure 4. Example plots of the time dependence of temperature, $|d| = 3 \times 10^{-4}$: (a) $d > 0$, (b) $d < 0$.

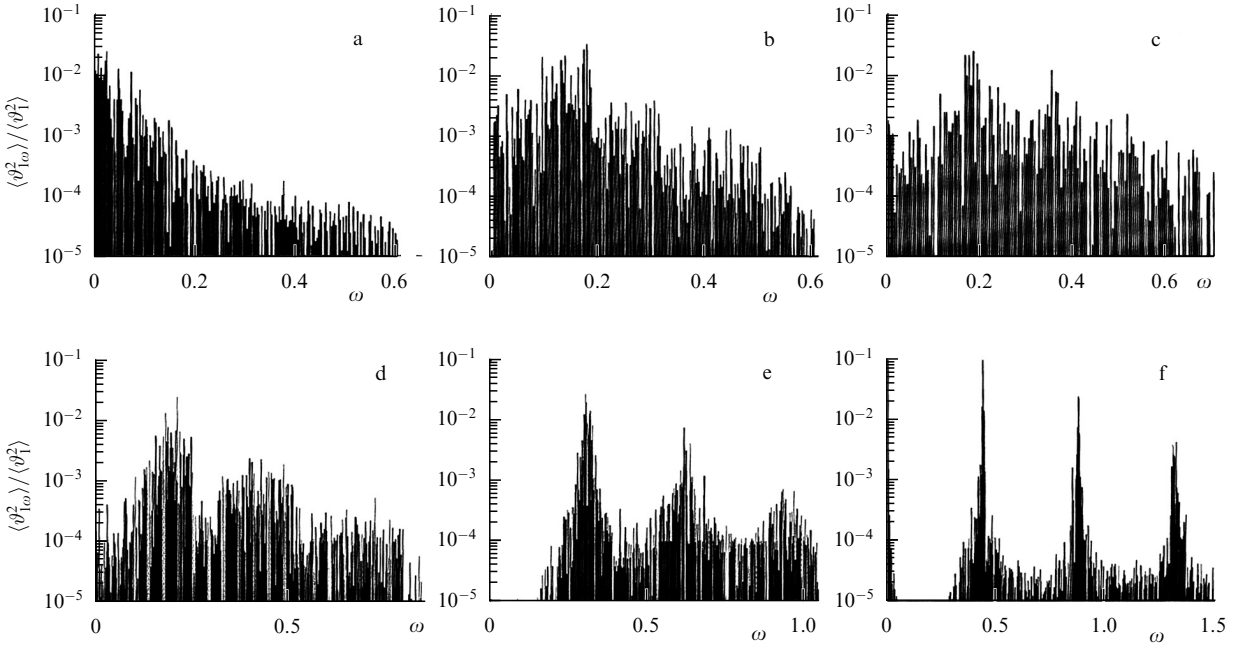


Figure 5. Frequency spectra for various values of Z: (a) 7, (b) 8, (c) 8.1, (d) 8.2, (e) 8.5, and (f) 9. Abscissa: $\omega = 2\pi n$, n are integers, t_0 is the computation time.

for $Z = 7$. Further evolution of the spectrum is illustrated in Fig. 5, which shows that the spectrum broadens at $Z = 8$, a small 'friction' causes the excitation of a large number of modes, but the spectrum remains continuous. The expansion continues at $Z = 8.1$, i.e, where the regime of regular oscillations would set in in the absence of noise. It is only at large values of Z, deep in the instability region, that peaks at the fundamental frequency and higher harmonics appear. The fact that the power of the continuous spectrum condenses into lines can be interpreted as a synchronization effect [28] or a classical analog of Bose–Einstein condensation. To see that this is indeed the case, we note that a state of a time-periodic system is characterized by a phase. Upon condensation, the number of particles (or quanta) in a given state is large, and the uncertainty relation for the number of particles and the phase imposes no significant restriction on how accurately the phase can be determined. A similar discussion on the macroscopic filling of long-wavelength modes at the expense of short-wavelength ones can be found in Ref. [29].

8. Traveling fronts and pulses near the propagation threshold

As an example of a distributed-parameter system, we consider a set of small CSTR systems interacting via diffusion and heat conduction. Instead of Eqns (22), (23), we now write

$$\dot{\eta} = L\Delta\eta + (1-\eta)\exp\vartheta - \eta D^{-1}\,, \tag{29}$$

$$\dot{\vartheta} = \Delta\vartheta + Z(1-\eta)\exp\vartheta - \frac{2+\vartheta}{S} - \frac{2}{S}\,y(t)\,, \tag{30}$$

where $\sqrt{(\chi/k)\exp(E/T_*)}$ is the length scale, χ is the heat diffusivity, L is the Lewis number (the ratio of diffusivity to heat diffusivity), and the scale temperature is chosen to be $T_* = T_a + Q/c$. The reactors can be arranged in the form of a chain or a membrane, and Δ is the one- or two-dimensional Laplace operator. For solitary waves along the chain, the boundary conditions for Eqns (29) and (30) can be written as

$$x = \pm\infty\,, \qquad \frac{\partial\eta}{\partial x} = \frac{\partial\vartheta}{\partial x} = 0\,. \tag{31}$$

We use a no-flow, $\tau \to \infty$ scenario to study a traveling front. With the noise $y(t)$ turned off in Eqn (30), such a front consists of three zones: heating, reaction, and cooling [30]. The reaction zone concentrates near the temperature maximum and is narrow compared to the other two. In the heating zone, the reaction can be neglected due to the insufficiently high temperature, and in the cooling zone, the original material is already used up. In the wave-comoving frame of reference, Eqns (29) and (30) are steady state, with $\partial/\partial t \to u\,\partial/\partial x$, where u is the wave velocity. We approximate the chemical source in Eqn (30) with the function $\delta(x)$, where x is the reaction zone coordinate, and replace Eqn (29) with the condition

$$u = u_0 \exp\vartheta_m\,, \tag{32}$$

where u_0 is the front velocity under no-loss conditions for $S \to \infty$ and ϑ_m is the maximum temperature. It is easily seen that problem (29)–(32) has two solutions: two branches of $u(S)$, the upper increasing and the lower decreasing with increasing S. This situation is clearly unstable. The threshold is located where the branches merge.

For the problem with a δ source, it is easy to write the nonsteady-state wave solution:

$$\frac{1}{2\sqrt{\pi L}}\int_0^t \frac{\varphi(t')}{\sqrt{t-t'}}\exp\left[-\frac{l^2(t',t)}{4L(t-t')}\right]\mathrm{d}t' = 1\,, \tag{33}$$

$$\frac{1}{2\sqrt{\pi}}\int_0^t \frac{\varphi(t')}{\sqrt{t-t'}}\exp\left[-\frac{l^2(t',t)}{4(t-t')} - \frac{t-t'}{S}\right]\mathrm{d}t' = 1 + Z\vartheta_m\,, \tag{34}$$

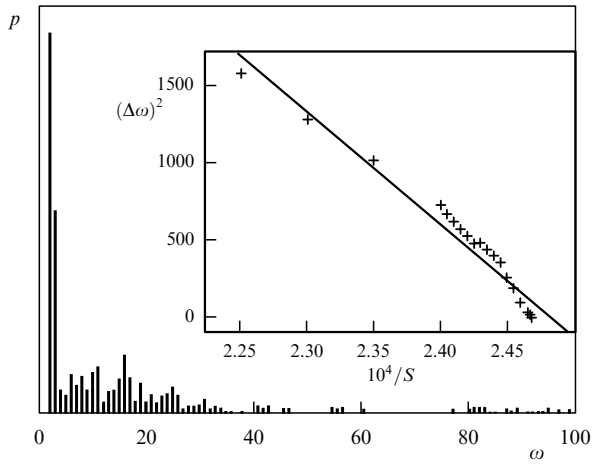


Figure 6. Typical pulsation spectrum $u(t) - \langle u \rangle$. Abscissa: the frequency $\omega = 2\pi n/t_0$, with n being an integer and $t_0 = 60{,}000$ the computation time. Inset: spectral maximum width versus distance to the threshold.

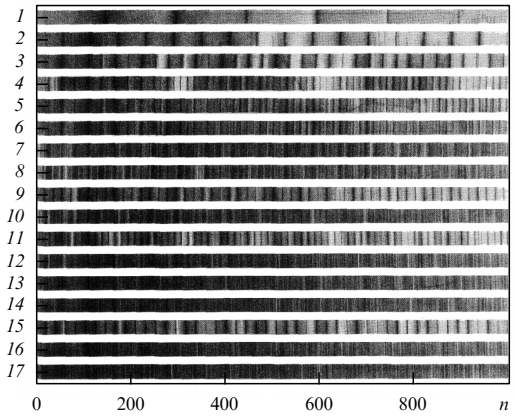


Figure 7. (See in color online.) Frequency spectra of the velocity pulsations as a function of $\omega = 2\pi n/t_0$, computation time $t_0 = 10^5$. The values of S: 350 (*1*), 328 (*2*), 326.914 (*3*), 326.640 (*4*), 326.620 (*5*), 326.611 (*6*), 326.551 (*7*), 326.500 (*8*), 326.466 (*9*), 326.400 (*10*), 326.375 (*11*), 326.350 (*12*), 326.315 (*13*), 326.250 (*14*), 326.230 (*15*), 326.219 (*16*), 326.213 (*17*). Discrete spectrum changes to continuous via intermittency.

where $\varphi^2 = u_0^2 \exp\vartheta_m$ and $l(t', t) = \int_{t'}^{t} u(t'')\,dt''$. In general, however, Eqns (33) and (34) are nonlinear integro-differential equations for u and ϑ_m, by no means easier to solve than the original problem. Near the threshold, the velocity varies slowly, allowing an expansion in the small parameter, which leads to the quasi-steady-state equation [31]

$$\frac{dv}{dt} = \frac{s - 2v^2}{t_*}, \qquad t_* = \frac{Z\sqrt{e}}{u_0^2}\left[L - 1 + \frac{2}{Z}(L+2)\right], \qquad (35)$$

where $v = u/u_{th} - 1$ and $s = S/S_{th} - 1$, with the index th denoting threshold values. Above the threshold, there are two steady-state points, the smaller of which is unstable; for $s < 0$, the velocity decreases and the wave decays. Because a realistic medium always contains at least small-size inhomogeneities and because the near-threshold susceptibility is high, the wave motion becomes chaotic. A point to note is the dominance of soft modes. The velocity fluctuation spectrum behaves as shown in Fig. 6.

The quasi-steady-state equation is in fact valid only for L close to unity. For $L > 1$, the planar front is unstable and becomes curved, whereas for $L < 1$, an oscillatory instability develops. Near the threshold, this instability produces random velocity pulsations in the absence of external noise (dynamical chaos). The period doubles as the parameter values shift to the threshold. The intervals between the doublings decrease as a geometric progression in accordance with Ref. [32]. The chaotic regimes are bounded by $S_* = 326.611$. The time dependence of the velocity shows an intermittent behavior between S_* and $S_{th} = 326.2$, as does the evolution of the frequency spectra with increasing S. Example spectra are shown in Fig. 7 [33].

If the flow transverse to the chain has a nonzero velocity, the solutions can be either switching waves between the hot and cold regions or traveling pulses. In the pulse after the reaction zone, the initial state restores itself due to the flow. Numerical simulations were performed for $L = 1$ [34]. It was found (Fig. 8) that as the parameters approach their threshold values, the time it takes for a pulse to attain the uniform motion regime increases, and oscillations occur. The oscillation period is large compared to the characteristic time $(1/k)\exp[E/(T_a + Qc^{-1})]$ and increases as $S \to S_{th}$. As

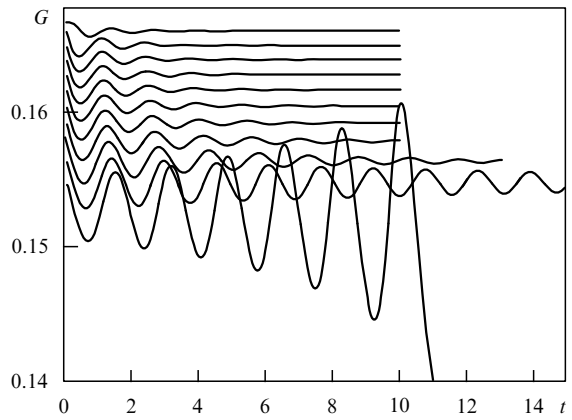


Figure 8. Time dependence of the pulse power (integral over the source) for S values close to the threshold. The top curve is constructed for $S = 770$, and each subsequent one, with S decreased by unity. For $S \leqslant 761$, the pulse decays irreversibly.

before, the dominance of soft modes near the bifurcation is seen.

9. Conclusions

The growth of soft modes near a bifurcation is a universal phenomenon. To study it, in addition to numerical simulation, real experiments were conducted [35] using an electrical circuit with a dinistor as a working element. This device, which exhibits an S-shaped current–voltage characteristic, was connected in series with a constant (current-independent) resistor R. The direct current regimes are determined by the intersection of the $I(V)$ characteristic and the loading straight line $I = (V_0 - V)/R$ (I is the current, V_0 is the applied voltage, and V is the voltage across the dinistor). Depending on the values of R and V_0, either one or three intersections are possible (with the middle one unstable in the latter case). For each R, V_0 pair, the variation of the pulsations of voltage V with time was recorded, with a recording time of 100 s. In total, 256 experimental runs were made. Example spectra are

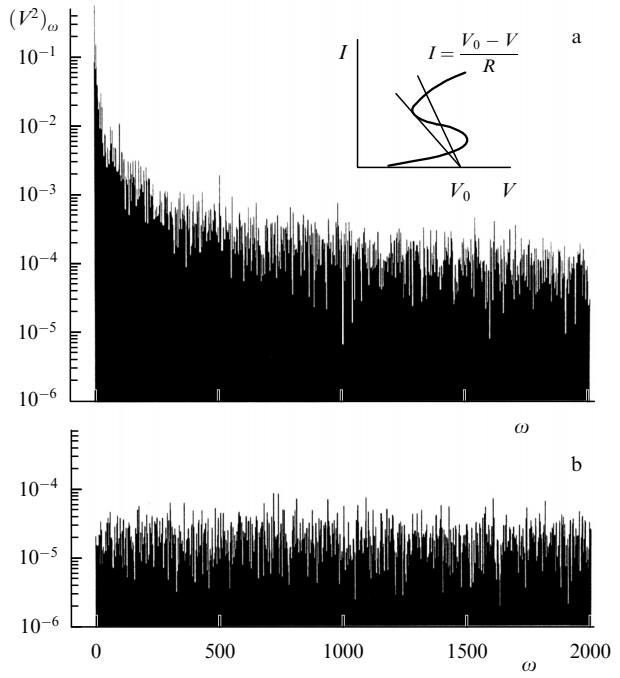


Figure 9. (See in color online.) Frequency spectra of pulsations close to and at a distance from the bifurcation point. (a) $R = 75$ kΩ, $V_0 = 30$ V, (b) $R = 60$ kΩ, $V_0 = 32$ V. Frequency is in units of $2\pi n/t_0$, t_0 is the duration of the experiment.

shown in Fig. 9. Far from the bifurcation, usual white noise is observed, but close to it, all the modes increase (the spectrum is normalized to the variance), but the soft modes especially so. At zero frequency, a maximum, or more precisely a resonance, forms.

The approach of a complex system (artificial or natural) to a bifurcation of its steady-state regime (i.e., to a catastrophe) can be predicted beforehand from the enhancement of soft modes in its noise spectrum, obviously motivating the development of spectrum-monitoring devices. The report by the Federal Service for Ecological, Technological and Atomic Surveillance (Rostechnadzor) on the Sayano-Shushenskaya hydroelectric station accident (for the online version, visit http://www.scientific.ru/trv/archive/act.doc) contains only one plot of the vibration amplitude for the turbine bearing cover. What happened in the accident was that the cover broke off and water rushed in to flood the station engine hall. During the observation period, the average value of the amplitude increased threefold, and its maximum value increased fifteenfold. As the report states, no resonances were detected, implying that the difference between the average and maximum values is due to the compression of the spectrum (zero frequency resonance, as discussed above). The term 'catastrophe precursors' is used in [36] in referring to soft modes. In an alternative 'compressive sensing' approach [37] for predicting catastrophes, unknown equations are restored from experimental data (time series). The right-hand side of each equation is written as an expansion in powers of dependent variables, with the expansion coefficients calculated by a computer program (most of them, incidentally, turn out to be negligibly small). But as the description in Ref. [37] implies, this procedure is realistic only for a small number of degrees of freedom (or the number of equations), not to mention distributed systems. By contrast, the method we propose here is simple and has no such restrictions, and we see it as our duty that it be made known as widely as possible throughout the science and technology community.

Acknowledgments

This paper was supported by the Russian Foundation for Basic Research Grant 11-03-00058-a and by the Russian Academy of Sciences Presidium program P-02, subprogram 1, code 2114.

References

1. Landau L D, Lifshitz E M *Statistical Physics* Vol. 1 (Oxford: Pergamon Press, 1980) [Translated from Russian: *Statisticheskaya Fizika* Vol. 1 (Moscow: Fizmatlit, 1995)]
2. Ornstein L S, Zernike F *Proc. Acad. Sci. Amsterdam* **12** 793 (1914)
3. Weissman M B *Rev. Mod. Phys.* **60** 537 (1988)
4. Langevin P *C.R. Acad. Sci.* **146** 530 (1908) [Translated into Russian: *Izbrannye Trudy* (Selected Works) (Moscow: Izd. AN SSSR, 1960) p. 338]
5. Lorenz E N *J. Atmos. Sci.* **20** 130 (1963)
6. Poincaré H *Sur les propriétés des fonctions définies par les équations aux différences partielles* (Paris: Gauthier-Villars, 1879) [Translated into Russian: *Izbrannye Trudy* (Selected Works) Vol. 1 *Novye Metody Nebesnoi Mekhaniki* (New Methods of Celestial Mechanics) (Moscow: Nauka, 1971)]
7. Arnol'd V I *Sov. Phys. Usp.* **26** 1025 (1983) [*Usp. Fiz. Nauk* **141** 569 (1983)]
8. Arnold V I *Catastrophe Theory* 3rd ed. (Berlin: Springer-Verlag, 1992) [Translated from Russian: *Teoriya Katastrof* 5th ed. (Moscow: URSS, 2007)]
9. Whitney H *Ann. Math.* **62** 374 (1955)
10. Loskutov A *Phys. Usp.* **53** 1257 (2010) [*Usp. Fiz. Nauk* **180** 1305 (2010)]
11. Sagdeev R Z, Usikov D A, Zaslavsky G M *Nonlinear Physics. From the Pendulum to Turbulence and Chaos* (Chur: Harwood Acad. Publ., 1988); Zaslavsky G M, Sagdeev R Z *Vvedenie v Nelineinuyu Fiziku. Ot Mayatnika do Turbulentnosti i Khaosa* (Introduction to Nonlinear Physics. From the Pendulum to Turbulence and Chaos) (Moscow: Nauka, 1988)
12. Binney J J et al. *The Theory of Critical Phenomena. An Introduction to the Renormalization Group* (Oxford: Clarendon Press, 1998)
13. Ivleva T P et al. *Chem. Eng. J.* **168** 1 (2011)
14. Frank-Kamenetskii D A *Acta Physicochim. USSR* **10** 365 (1939); *Zh. Fiz. Khim.* **13** 738 (1939)
15. Zel'dovich Ya B *Zh. Tekh. Fiz.* **11** 493 (1941)
16. Rumanov É N *Dokl. Phys.* **51** 238 (2006) [*Dokl. Ross. Akad. Nauk* **408** 325 (2006)]
17. Van der Pol B *Radio Rev.* **1** 701 (1920)
18. Merzhanov A G, Rumanov É N *Sov. Phys. Usp.* **30** 293 (1987) [*Usp. Fiz. Nauk* **151** 553 (1987)]
19. Vaganova N I, Rumanov E N *Dokl. Phys.* **56** 507 (2011) [*Dokl. Ross. Akad. Nauk* **440** 463 (2011)]
20. Ginzburg V L *Sov. Phys. Solid State* **2** 1824 (1961) [*Fiz. Tverd. Tela* **2** 2031 (1960)]
21. Vaganova N I, Rumanov É N *JETP* **108** 349 (2009) [*Zh. Eksp. Teor. Fiz.* **135** 395 (2009)]
22. Kramers H A *Physica* **7** 284 (1940)
23. Chan H B, Stambaugh C *Phys. Rev. Lett.* **99** 060601 (2007)
24. Kravtsov Yu A, Surovyatkina E D *Phys. Lett. A* **319** 348 (2003)
25. Frank-Kamenetskii D A , Sal'nikov I E *Zh. Fiz. Khim.* **17** 79 (1943)
26. Aris R, Amundson N R *Chem. Eng. Sci.* **7** 121 (1958)
27. Vaganov D A, Samoilenko N G, Abramov V G *Chem. Eng. Sci.* **33** 1131 (1978)
28. Pikovsky A, Rosenblum M, Kurths J *Synchronization. A Universal Concept in Nonlinear Sciences* (Cambridge: Cambridge Univ. Press, 2001) [Translated into Russian (Moscow: Tekhnosfera, 2003)]
29. Keldysh L V, Tikhodeev S G *Sov. Phys. JETP* **64** 45 (1986) [*Zh. Eksp. Teor. Fiz.* **91** 78 (1986)]
30. Zel'dovich Ya B *Zh. Eksp. Teor. Fiz.* **11** 159 (1941)
31. Dovzhenko A Yu, Maklakov S V, Rumanov I E, Rumanov E N *JETP* **95** 973 (2002) [*Zh. Eksp. Teor. Fiz.* **122** 1125 (2002)]

32. Feigenbaum M J *J. Stat. Phys.* **19** 25 (1978)
33. Dovzhenko A Yu, Rumanov É N *JETP* **98** 359 (2004) [*Zh. Eksp. Teor. Fiz.* **125** 406 (2004)]
34. Dovzhenko A Yu, Rumanov É N *JETP* **104** 508 (2007) [*Zh. Eksp. Teor. Fiz.* **131** 567 (2007)]
35. Dovzhenko A Yu, Dovzhenko M, Mashkinov L B, Rumanov É N *Dokl. Phys.* **48** 13 (2003) [*Dokl. Ross. Akad. Nauk* **388** 181 (2003)]
36. Rumanov E N, Vaganova N I *Priroda* (10) 23 (2011)
37. Wang W-X et al. *Phys. Rev. Lett.* **106** 154101 (2011)

PACS number: **01.90.+g**

Physics news on the Internet (based on electronic preprints)

DOI: 10.3367/UFNe.0183.201301d.0086

1. Direct observation of *T*-invariance violation in a system of B mesons

The direct measurements of the violation of *T*-invariance (invariance of processes under time reversal) for K mesons were previously performed at CERN and at the Fermi National Accelerator Laboratory, but the results obtained were not free of considerable uncertainty. Violation of *T*-invariance in a system of B mesons was earlier established only indirectly, by examining the violation of *CP*-invariance. Now, it was proved possible to measure the effect of *T*-invariance violation for B mesons directly, without resorting to the *CPT*-theorem, by using a new method of data analysis in the BaBar experiment conducted at the National Accelerator Laboratory (SLAC). In the BaBar experiment, decays of $\Upsilon(4S)$ resonances created pairs of $B^0\bar{B}^0$ in quantum-entangled states. Entanglement made it possible to compare the rates of the processes corresponding to different ordering of B^0 and $\bar{B}^0$ decays in time, and also under the permutation of the final states (of decay products). As a result, the violation of *T*-invariance has been established with high statistical significance — 14σ. The measured parameters that characterize the violation of *T*-invariance correspond to magnitudes previously derived from the effect of *CP*-invariance violation.

Source: *Phys. Rev. Lett.* **109** 211801 (2012)
http://dx.doi.org/10.1103/PhysRevLett.109.211801

2. Superconductivity in $La_{2-x}Sr_xCuO_4$

A team led by Ivan Bozovic of Brookhaven National Laboratory has continued to work on the experiments described earlier [see, e.g., *Phys. Usp.* **51** 170 (2008)] and discovered that, under certain conditions, a drop in temperature, instead of resulting in the transition to superconducting state, suppresses superconductivity in the compound $La_{2-x}Sr_xCuO_4$. A $La_{2-x}Sr_xCuO_4$ layer was grown on a substrate by an improved molecular-beam epitaxial technique which allows controling the doping level x. Near the superconducting transition temperature, Bozovic et al. observed superconducting fluctuations which normally precede the emergence of superconductivity. Unexpectedly, superconducting fluctuations were suppressed in specimens with $x = 0.055 - 0.06$, plus they were completely absent in sufficiently high magnetic fields. As temperature was further reduced, suppression was enhanced, and superconductivity would not emerge. One cause of this could supposedly lie in structural defects which at low temperatures hamper electron flow (the electron localization effect). As the doping level increased, the suppression effect disappeared. For example, a specimen with $x = 0.07$ became superconducting when cooled to $T_c = (9 \pm 1)$ K.

Source: *Nature Materials* **12** 47 (2013)
http://dx.doi.org/10.1038/nmat3487

3. Acoustic analogue of the dynamic Casimir effect

C I Westbrook and his colleagues at the C Fabry Laboratory (Université Paris-Sud, France) have implemented an acoustic analogue of the dynamic Casimir effect, which was first observed in 2011. In the dynamic Casimir effect, virtual particles transform into real ones due to rapid nonadiabatic changes in boundary conditions. In the experiment by C I Westbrook et al., researchers varied the potential of the optical trap holding a Bose–Einstein condensate of helium atoms, which led to a change in the speed of sound and in the spectral composition of oscillations in the condensate. The potential was varied by changing the intensity of the laser beams that formed the trap. In version I of the experiment, the potential was changed sharply once, while in version II it was subjected to 10% sinusoidal modulation for 25 ms until the trap was turned off and the condensate cloud broke apart. As a result of these changes, the thermal fluctuations in the condensate transformed into pairs of elementary excitations — quasiparticles moving in opposite directions with momenta of identical magnitude and frequency equal to one half of the modulation frequency. Excitations corresponded to lateral components in the velocity distribution of gas particles in the expanding cloud. These excitations satisfied the Bogoliubov–de Gennes dispersion relation both in the phonon mode, when excitation consisted of several correlated atoms, and in the single atoms mode. The researchers are looking forward to fabricating an acoustic analogue of Hawking radiation, similarly to how in 2009 an experiment by J Steinhauer et al. generated an acoustic analogue of the black hole horizon.

Source: *Phys. Rev. Lett.* **109** 220401 (2012)
http://dx.doi.org/10.1103/PhysRevLett.109.220401

4. Effect of light on the conductivity of insulators

F Krausz (Institute for Quantum Optics, Max Planck Society, Germany) and colleagues have been able to demonstrate a method of ultrafast control of dielectric conductivity using high-power femtosecond pulses of near-infrared radiation (NIR), comprising a mere several oscillations of the light wave. The conductivity of amorphous silicon dioxide (SiO_2) exposed to these pulses increased over ≈ 1 fs by about 18 orders of magnitude and dropped back over the same time. A wave field with an intensity of several volts per angström substantially altered the electronic structure but, nevertheless, this transition occurred reversibly, without destroying the atomic structure of the specimen. Conductivity measurements were made by spectroscopic methods and

Uspekhi Fizicheskikh Nauk **183** (1) 86 (2013)
DOI: 10.3367/UFNr.0183.201301d.0086
Translated by V I Kisin

by recording the current flowing across the electrodes. The observed properties are well explained by the theoretical model developed by V Apalkov and M Stockman. Even though the conductivity of semiconductors is much simpler to control than that of insulators, the changes caused in conductivity are much slower. In principle, the new effect offers the possibility of ultrafast control of electrical signals in promising devices operating in the terahertz, and even pentahertz, ranges (10^{15} Hz).

Sources: *Nature* **493** 70 (2013), *Nature* **493** 75 (2013)
http://dx.doi.org/10.1038/nature11567
http://dx.doi.org/10.1038/nature11720

5. Gamma-ray bursts caused by lightning

Sometimes lightning discharges generate gamma-ray flashes several thousandths of a second long, known as terrestrial gamma-ray flashes. A GBM detector aboard the NASA's Fermi Gamma-ray Space Telescope is currently recording approximately two gamma-ray flashes due to lightning per week with a time resolution of about 2 μs. It was assumed in the past that powerful radio bursts which are also generated by lightning discharges are not directly traceable to the generation of gamma-rays. However, as follows from the new data collected with the GBM detector, gamma-ray flashes and some broad peaks in radio bursts in fact occur simultaneously and have similar pulse shapes. Consequently, these gamma and radio signals appear to be of the same origin, being generated in one and the same area of the electrical discharge. It is highly probable that 'runaway electrons', whose theory was developed by A V Gurevich and his colleagues (Lebedev Physical Institute, RAS), are responsible for high-energy phenomena in the lightning.

Source: http://www.nasa.gov/mission_pages/GLAST/
news/vision-impro ve.html

Prepared by *Yu N Eroshenko*
(e-mail: erosh@ufn.ru)

ISSN 1063-7869 (Print)
ISSN 1468-4780 (Online)

CODEN: PHUSEY

PHYSICS–USPEKHI

Uspekhi Fizicheskikh Nauk (Успехи Физических Наук, **www.ufn.ru**) publishes reviews of the current state of the most topical problems in physics and in associated fields under the general headings: reviews of topical problems, physics of our days, instruments and methods of investigation, methodological notes, from the history of physics, conferences and symposia, personalia, physics news on the Internet, and bibliography. The journal was founded in 1918 and is published monthly.

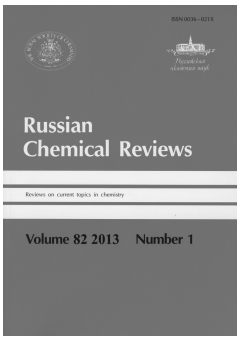
Physics–Uspekhi (Advances in Physical Sciences) is the English edition of the Russian monthly journal *Uspekhi Fizicheskikh Nauk*. Translation into English started with Russian volume 66. From 1958 until 1992 the journal was published by American Institute of Physics under the title *Soviet Physics – Uspekhi* and in 1993 under its current title *Physics – Uspekhi*. Since 1994 *Physics – Uspekhi* has been published jointly by Uspekhi Fizicheskikh Nauk and Turpion Ltd. From the beginning of 1996 *Physics – Uspekhi* is being translated, typeset and edited in Moscow by Uspekhi Fizicheskikh Nauk (UFN).
From 2009 published by Uspekhi Fizicheskikh Nauk, Moscow
Printed by Page Bros, Norwich, UK.

Institutional subscription information (volume 56, 2013, monthly)
For all countries, except the United States, Canada, Central and South America, the subscription rates are: Print + Online (1958 – 2013) £ 1717; Online only (1958 – 2013) £ 1545; Single issue £ 172. Delivery is by air-speeded mail from the UK.

Orders, back issues, change of address to*: *Physics-Uspekhi, Journals Subscription Fulfilment, IOP Publishing, Temple Circus, Temple Way, Bristol BS1 6HG, United Kingdom

For the United States, Canada, Central and South America, the subscription rates are:
Print + Online (1958 – 2013) US $3091; Online only (1958 – 2013) US $ 2782; Single issue US $ 309. Delivery is by transatlantic airfreight and onward mailing. ***United States Postal Identification Statement*** *Physics – Uspekhi*, volume 56, published monthly. Periodicals Postage Paid at Continental Station, PA and additional mailing offices. POSTMASTER: Send orders, address changes to *Physics – Uspekhi*, IOP Publishing, PO Box 320, Congers, NY 10920-0320, USA.
Online services: Since 2008 electronic access to the journal content is hosted by IOP Publishing. All subscriptions to the journal include electronic access dating from the first English translation volume. The electronic version of the journal is available at http://iopscience.org/phu. All questions regarding online access should be sent to customer services at custserv@iop.org or custserv@turpion.ru

Editorial Office: P N Lebedev Physical Institute, RAS, Leninskiĭ prospekt 53, 119991 Moscow, Russian Federation
Tel. (7-499) 132 62 65, (7-499) 132 63 48, (7-499) 190 34 52
Tel./Fax (7-499) 190 42 44. E-mail: ufn@ufn.ru



Russian Chemical Reviews

iopscience.org/rcr
www.turpion.org/journal/rc

Russian Chemical Reviews (RCR) is the English translation of the monthly review journal *Uspekhi Khimii*, one of the leading Russian journals in chemistry, founded in 1932. The journal achieved the highest Impact Factor among all Russian translation journals in all subject areas.

With work written by authorities in their individual fields, the journal gives researchers around the world access to the advances and achievements of chemists from Russia and other countries of the former Soviet Union in most aspects of modern chemistry:

chemical physics; physical chemistry, including catalysis; the structure of molecules and quantum chemistry; coordination chemistry; analytical chemistry; organic and organometallic chemistry; chemistry of macromolecules; biochemistry and bioorganic chemistry; materials chemistry.

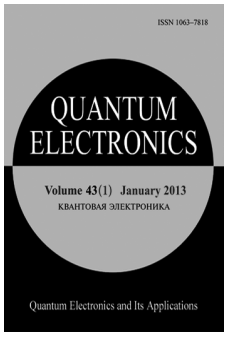
The journal's historic archive provides access to the golden age of Russian science in chemistry and associated fields, including research by Nobel Laureates and other leading and pivotal characters in the history and development of Russian science.

Its combination of expertise and interdisciplinary approach means that articles published in RCR appeal to scientists at all levels, including postgraduate students, teachers, and researchers, in chemistry and related scientific disciplines such as physical chemistry, chemical physics, materials science, nanochemistry, nanostructures, and nanotechnologies.

Year	2013
Volume	82
Frequency	12 issues per year
Print ISSN	0036-021X
Online ISSN	1468-4837
CODEN	RCRVAB

Online archive
1960–2012 available free with journal subscription.
1960–2002 available in Turpion's Historic Archive: Turpion offers the option to acquire perpetual rights of Turpion journals content for a one-time purchase. Since 2008, electronic access to the content back to the first English translation volume has been hosted by IOP Publishing at **iopscience.org/rcr**.

Subscription price
- Print and Online £ 1714
- Online only £ 1543



Quantum Electronics

iopscience.org/qe
www.turpion.org/journal/qe

Established alongside the first publication of the Russian journal *Kvantovaya Elektronika* in 1971, the English translation *Quantum Electronics* (QE) is produced just a few weeks after each original edition, providing efficient access to unique research from more than 300 world-class Russian institutions and specialists from 25 countries.

QE is the only journal that provides comprehensive results in topics such as quantum electronic devices, laser physics and optics, interaction of laser radiation with matter, and the transmission and processing of information at basic and applied research levels. It is a valuable resource for those working with all aspects of laser research or with the practical application of laser technologies in the metrological, biological, and medical fields, or in the electronics, engineering, defense, and materials industries. The journal's historic archive provides access to pioneering research in these areas, including research by Nobel Laureates and other leading and pivotal characters in the history and development of Russian science.

With an Editorial Board and council consisting of more than 40 world-class experts, the journal also covers laser plasmas, nonlinear optical phenomena, nanotechnologies, fiber and integrated optics, and active media, and continues to build on the strong foundation established by Nobel Prize Laureate Nikolay G Basov.

Year	2013
Volume	43
Frequency	12 issues per year
Print ISSN	1063-7818
Online ISSN	1468-4799
CODEN	QUELEZ

Online archive
1971–2012 available free with journal subscription.
1971–2002 available in Turpion's Historic Archive: Turpion offers the option to acquire perpetual rights of Turpion journals content for a one-time purchase. Since 2008, electronic access to the content back to the first English translation volume has been hosted by IOP Publishing at **iopscience.org/qe**.

Subscription price
- Print and Online £2500
- Online only £2250