Physics – Uspekhi **55** (8) 808 – 837 (2012)

CONFERENCES AND SYMPOSIA

PACS numbers: 01.10.Fv, **04.20. – q**, **04.70. – s**, 47.27.Gs, 47.35.Pq, 68.03.Kn, 72.25.Hg, 72.25.Pn, 72.25.Rb, **71.35. – y**, 73.63.Hs, **78.47. – p**, **78.67. – n**, **96.50. – e**, **96.60. – j**, **95.75. – z**, **98.80. – k**

Modern problems in the physical sciences
(Scientific session of the Physical Sciences Division of the Russian Academy of Sciences, 30 November 2011)

DOI: 10.3367/UFNe.0182.201208f.0869

On 30 November 2011, a scientific session of the Physical Sciences Division of the Russian Academy of Sciences (RAS) was held in the conference hall of the Lebedev Physical Institute, RAS.

The agenda of the session announced on the RAS Physical Sciences Division website www.gpad.ac.ru included the following reports:

(1) **Ivchenko E L** (Ioffe Physical Technical Institute, RAS, St. Petersburg) "Spin physics in semiconductor nanosystems";

(2) **Golub L E** (A F Ioffe Physical Technical Institute, RAS, St. Petersburg) "Spin transport in heterostructures";

(3) **Levchenko A A** (Institute of Solid State Physics, RAS, Chernogolovka, Moscow region) "Capillary turbulence on the surface of quantum liquids";

(4) **Babin S A** (Institute of Automation and Electrometry, Siberian Branch of the RAS) "New generation modes in fiber lasers";

(5) **Kurt V G** (Astro-Space Center of the Lebedev Physical Institute, RAS, Moscow) "Motion of the Sun through the interstellar medium";

(6) **Lukash V N** (Astro-Space Center of the Lebedev Physical Institute, RAS, Moscow) "Cosmological flow generation in general relativity".

Papers written on the basis of oral reports 1–3, 5, and 6 are presented below.

PACS numbers: **71.35. – y**, **78.47. – p**, **78.67. – n**
DOI: 10.3367/UFNe.0182.201208g.0869

Spin physics in semiconductor nanosystems

E L Ivchenko

1. Introduction

The rapid development of nanotechnologies in the past two decades has aroused sustained interest in semiconductor spin physics (see Refs [1–3] for a review). The driving factor behind this interest is the paradigm of using spin, an additional degree of freedom of an electron, as a tool for the quantum processing of information. In classical computing, the basic unit of information is a bit, which, regardless of its particular physical realization, takes one of the two mutually exclusive values, either 0 or 1 (yes/no, yin or yang, etc.). A qubit (q-bit, or quantum bit), similar to a bit, admits two eigenstates, $|0\rangle$ and $|1\rangle$. The difference between the bit and the qubit is fundamentally that, whereas a bit can only be in one of the available states, a qubit has the possibility of being not only in one of the two basis states but also in any normalized superposition of these:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \qquad |\alpha|^2 + |\beta|^2 = 1\,,$$

with the complex coefficients α and β. One of the qubit implementation scenarios (of which there are many) uses the spin $\pm 1/2$ states of a charge carrier (an electron or a hole). For pure quantum-mechanical states, the coefficients α, β and the average projections of the electron spin onto the x-, y-, z-axes of the Cartesian coordinate system are related by the well-known expressions

$$s_x = \mathrm{Re}\,\{\alpha^*\beta\}\,, \quad s_y = \mathrm{Im}\,\{\alpha^*\beta\}\,, \quad s_z = \frac{1}{2}\left(|\alpha|^2 - |\beta|^2\right). \quad (1)$$

The states of a qubit can be described by a vector **s** on the Bloch sphere. A qubit can store much larger amounts of information compared to a bit, the exact amount depending on how accurately the position of the vector **s** is known. Spin-dependent optical and transport phenomena and their practical applications advancing the prospective technologies based on electron spin devices and apparatus are the subject of the field generally known as spintronics. Semiconductor spintronics has as its tasks to study the orientation (injection), accumulation, and detection of spins and their ability to be optically and electrically controlled. While spintronics is currently still in its infancy and in need of new conceptual ideas to realize effective electron spin-based devices, the attractive and exciting world of spin-dependent phenomena is definitely worth exploring and, as is often the case in other fields of physics, basic research will be sooner or later translated into practical applications. This brief report starts with a bird's eye view of spin-dependent phenomena that are explored in the physics of semiconductor nanostructures, followed by a more detailed review of two groups of

E L Ivchenko Ioffe Physical Technical Institute, Russian Academy of Sciences, St. Petersburg, Russian Federation
E-mail: ivchenko@coherent.ioffe.ru

Uspekhi Fizicheskikh Nauk **182** (8) 869 – 900 (2012)
DOI: 10.3367/UFNr.0182.201208f.0869
Translated by E G Strel'chenko, S D Danilov, K A Postnov; edited by A Radzig, A M Semikhatov

Uspekhi Fizicheskikh Nauk **182** (8) 869 – 876 (2012)
DOI: 10.3367/UFNr.0182.201208g.0869
Translated by E G Strel'chenko; edited by A Radzig

phenomena related to the optical control of electron spins in planar arrays of quantum dots and to the magneto-optics of single quantum dots.

2. Problems in spintronics

In this section we consider one by one the problems (illustrated schematically in Fig. 1) whose solution will potentially benefit the practical application of spintronics. Sections 3 to 5 will provide some illustrative examples of how electron spins can currently be controlled.

Effective g factor. This parameter, also know as the Landé factor, is key to describing the interaction of the electron spin with the magnetic field (Zeeman effect). In a typical semiconductor, the effective g factor is highly sensitive to the band gap width and to the valence band spin–orbit splitting. The bulk semiconductor values of the g factor vary from $+2$ to -50. In nanostructures, superlattices, quantum wells, quantum wires, and quantum dots, the g factor is strongly dependent on the shape and geometrical sizes of the nanoobject. In a magnetic field $\mathbf{B}$, the electron spin rotates about the vector $\mathbf{B}$ with an angular velocity $\mathbf{\Omega_B} = g\mu_B\mathbf{B}/\hbar$, where μ_B is the Bohr magneton (Larmor precession).

The theory of the Zeeman effect in heterostructures developed in the 1990s [4] yields the following formula for the transverse electron g factor ($g_{xx} = g_{yy}$) in the lower conduction subband of a quantum well (QW), quantum wire (QWR), and quantum dot (QD) [5]:

$$g = g_0 + \left(g_A(E_{e1}) - g_0\right) w_A + \left(g_B(E_{e1}) - g_0\right) w_B + \left(g_B(E_{e1}) - g_A(E_{e1})\right) V_{3-d}(R) f^2(R)\,. \quad (2)$$

The notation here is as follows: g_0 is the free electron g factor ($g_0 \approx 2$); d is the nanostructure dimensionality ($d = 0, 1, 2$ for a QD, QWR, and QW, respectively); V_n is the generalized volume ($4\pi R_{QD}^3/3$ for a spherical quantum dot of radius R_{QD}, πR_{QWR}^2 for a cylindrical quantum wire of cross-section radius R_{QWR}, and $2R_{QW}$ for a quantum dot of width $a \equiv 2R_{QW}$); A(B) stands for the composite material the well (barrier) is made of; E_{e1} is the dimensional quantization energy of the ground state electron (e1) as counted from the conduction band bottom of material A; w_A (w_B) is the probability of finding the electron in material A (B); $g_A(E) = g(E)$ and $g_B(E) = g(E - \Delta E_c)$:

$$g(E) = g_0 - \frac{4}{3}\frac{|p_{cv}|^2}{m_0}\frac{\Delta}{E_g(E_g + \Delta)} + \Delta g\,; \quad (3)$$

E_g is the band gap width; Δ is the valence band spin–orbit splitting; ΔE_c is the conduction band discontinuity at the A/B interface; $p_{cv} = \langle S|\hat{p}_x|X\rangle$ is the interband matrix element of the momentum operator calculated between the orbital Bloch functions Γ_1 and Γ_{15}; Δg is the remote band contribution to the g factor treated as a small fitting parameter, and, finally, $f(R)$ is the boundary value of the electron envelope wave function.

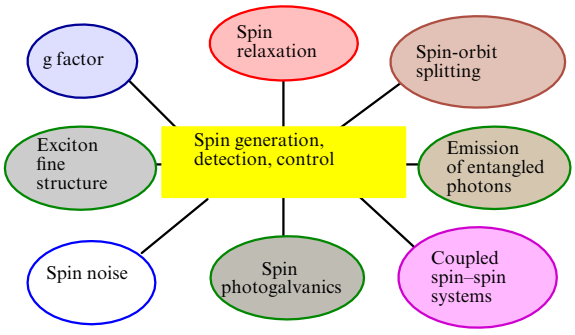

Figure 1. Purposes (center) and problems of spintronics.

Figure 2 reproduces a comparison taken from Ref. [6] of the experimental data on the electron g factor in quantum dot structures with the calculated results based on the theory developed in Refs [4, 5]. The theory of the longitudinal g factor (g_{zz}) of light holes in quantum wells has recently been presented in Ref. [7].

Spin splitting of dimensional quantization subbands. In a d-dimensional medium ($d = 1, 2, 3$) with no inversion center, free carrier states are spin-split. In particular, the effective Hamiltonian for quantum dot electrons in the lower dimensional quantization subband e1 contains, in addition to the standard parabolic dispersion $\hbar^2(k_x^2 + k_y^2)/2m^*$ (m^* is the effective mass), spin-dependent terms linear in the wave vector [8, 9]:

$$\mathcal{H}_{\mathbf{k}}^{(1)} = \beta_{lm}\sigma_l k_m \quad (4)$$

(where σ_l are Pauli matrices), which have the effect that, even in the absence of a magnetic field, the spin of an electron in the state with wave vector $\mathbf{k}$ precesses at an angular frequency $\mathbf{\Omega_k}$ with components $\Omega_{\mathbf{k}l} = 2\beta_{lm}k_m/\hbar$. The fact that the components β_{lm} of the pseudotensor are electric field-dependent allows the spin to be controlled by both a magnetic and an electric field. The theory of the spin splitting of electronic subbands is described in book [10], which also includes a basic bibliography on the topic. In centrosymmetric bulk semiconductors, for example, in Si crystals, symmetry prevents such splitting. However, there is no inversion center in an Si quantum dot with an odd number of atomic planes, so that the splitting is nonzero; the components of the corresponding

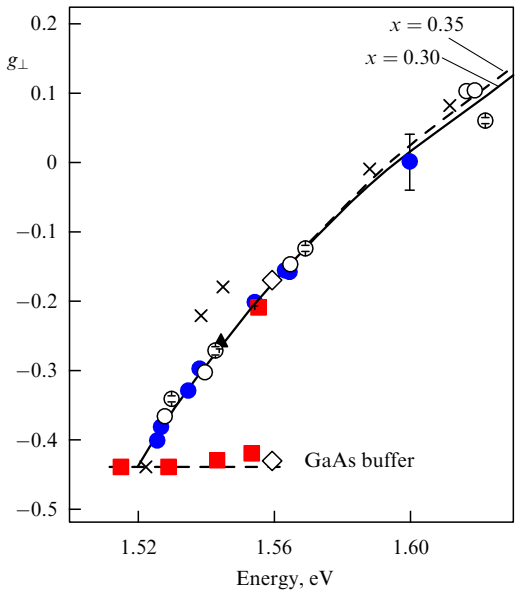

Figure 2. Transverse electron g factor ($g_\perp$) in a GaAs/$Al_xGa_{1-x}As$ quantum dot structure as a function of optical resonance energy. Different symbols refer to experimental values obtained from different samples. Solid and dashed lines were calculated for structure compositions with $x = 0.30$ and $x = 0.35$. Horizontal dashed line points to the electron g factor in a bulk GaAs crystal [6].

tensor β were calculated using the microscopic strong coupling method [11].

Spin relaxation. Spin relaxation time is another key parameter in spintronics. In semiconductor quantum wells, spin relaxation is predominantly via the Dyakonov–Perel mechanism, due to spin precession with angular frequency $\mathbf{\Omega}_{\mathbf{k}}$ (see above). An electron scattering from a state with wave vector $\mathbf{k}$ to that with $\mathbf{k}'$ changes the direction of its axis of rotation, resulting, as the electron undergoes multiple successive collisions, in its spin vector exhibiting random diffusive motion over the Bloch sphere. But this is exactly spin relaxation, because diffusive random walk reduces the average spin value exponentially. Prior to Ref. [12], the corresponding relaxation time was believed to be determined by the transport relaxation time for defect or phonon scattering, i.e. by the mobility-determining time. In Refs [12–14] it was shown that, while electron–electron collisions have no effect on the mobility, they have an effect on the spin relaxation time. Indeed, for an electron to change the direction of its wave vector — and hence the direction of a spin precession axis — it does not matter whether the scattering is by a defect, a phonon, or another electron. As seen from Fig. 3, which compares theoretical and experimental results, the Dyakonov–Perel relaxation is due to electron–electron collisions in perfectly doped quantum well structures at temperatures between 10 K and 100 K.

Fine structure of exciton energy spectrum. What is commonly referred to as a mechanical exciton is a bound electron–hole state calculated by allowing only for the direct pairwise Coulomb interaction in the semiconductor. Because electron (hole) states in a quantum dot are (doubly) spin-degenerate, the ground state of a mechanical exciton is fourfold degenerate. Including the exchange electron–hole interaction removes the degeneracy of the exciton level, at least partially. The theory of the fine structure of exciton states in semiconducting structures can be found in books [10, 15] and review paper [16] (see also original paper [17]). In Section 5, the nanostructure potential shape and the crystallographic orientation of the quantum dot will be discussed in terms of how they affect the splitting behavior of the exciton sublevels, and new experimental and theoretical results will be presented on the spin properties of unstrained GaAs/AlGaAs (111) quantum dots [18].

Emission of entangled photon pairs. In quantum dots possessing D_{2d} (D_{3h}, C_{3v}) symmetry, there is a twofold degeneracy for the exciton sublevels that are optically active in the directions normal to the growth axis. The photoluminescence of biexcitons in such dots allows entangled photon pairs to be generated that are described by the wave function

$$\frac{1}{\sqrt{2}}\left(|\sigma_+\rangle_{\text{biexc}}\,|\sigma_-\rangle_{\text{exc}} + |\sigma_-\rangle_{\text{biexc}}\,|\sigma_+\rangle_{\text{exc}}\right), \qquad (5)$$

where $|\sigma_\pm\rangle_{\text{biexc}}$ is a photon with polarization σ_+ or σ_-, emitted due to biexciton recombination with the production of a photon and an exciton, and $|\sigma_\pm\rangle_{\text{exc}}$ is the second photon emitted on the recombination of the remaining exciton, with

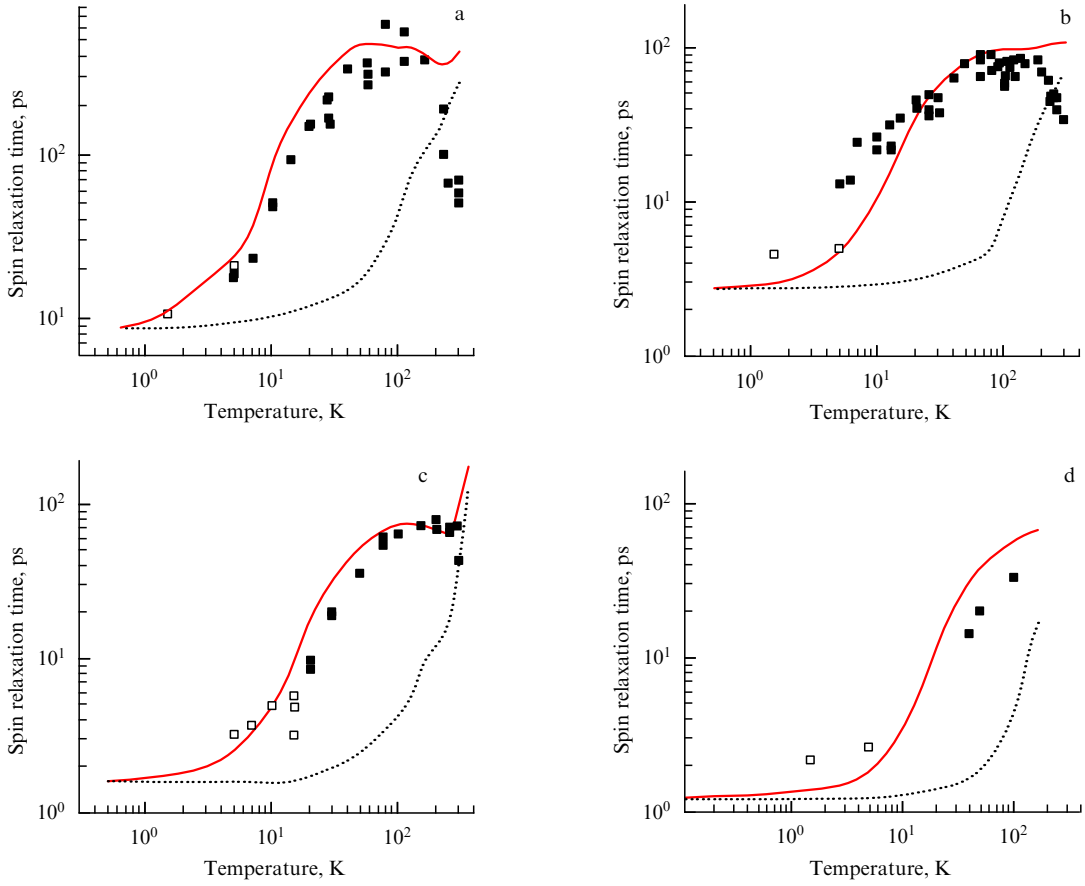

Figure 3. Temperature dependence of an electron spin relaxation time for four GaAs/AlGaAs quantum dot structures with electron concentrations: 1.75×10^{11} cm^{-2} (a), 2.3×10^{11} cm^{-2} (b), 3.1×10^{11} cm^{-2} (c), and 3.3×10^{11} cm^{-2} (d). Experimental results are marked by squares. Solid and dashed lines are calculated, respectively, with and without account for electron–electron collisions. In either case, account was taken of electron scattering by defects and phonons, which is described by the momentum relaxation time τ_p determined independently from transport measurements [14].

polarization being opposite to that of the first photon. Due to the interparticle interaction, the two phonons differ in energy by 1–2 meV and are distinguishable spectroscopically. In quantum dots of lower symmetry, the radiation doublet splits into two linearly polarized dipoles, resulting, if the splitting exceeds the uncertainty $\hbar/\tau$ (τ is the exciton lifetime), in the disappearance of entanglement, i.e. of coherence between the two bracketed states in expression (5). References [19, 20] consider two independent mechanisms, one paramagnetic and one diamagnetic, in which a magnetic field suppresses the original splitting of the radiation doublet.

Coupled spin–spin systems. What in particular distinguishes quantum dots from quantum wells (or wires) is the absence of free two-dimensional (or one-dimensional) motion in them and the fact that the Dyakonov–Perel spin relaxation mechanism does not work. However, with as many as $10^5 - 10^6$ host lattice nuclei in a typical quantum dot, the random nuclear spin exerts influence on the electron spin in this case [21]. Given the optical orientation of the electrons, the nuclear spins can also be polarized, thus giving rise to a coupled spin–spin system, a subject of active research in recent years (see, for example, Refs [22, 23]).

Spin noise. Noise determines the minimum signal amplitudes which can be processed with the means of electronics (in particular, spintronics). Moreover, the study of spin fluctuations provides independent information about the properties of the spin system in equilibrium and nonequilibrium conditions. While fluctuations of the free carrier spin polarization in semiconductors have been under theoretical study since the 1970s [24], it is only relatively recently that spin noise was detected in experiments on semiconductors (see Ref. [25] for a review) — many years, incidentally, after a similar observation in atomic physics back in 1981 [26]. Under equilibrium conditions, the spectral fluctuation density has the Lorentzian form

$$\langle \delta s_i^2 \rangle_\omega = \frac{1}{2} \frac{n\tau_s}{1 + (\omega\tau_s)^2}, \qquad (6)$$

where n is the electron concentration, τ_s is the electron spin relaxation time, and $i = x, y, z$. In a magnetic field $\mathbf{B}$, the spin polarization has its transverse component $\delta\mathbf{s} \perp \mathbf{B}$ fluctuating according to formula (6), with the frequency ω replaced by the difference $\omega - \Omega_\mathbf{B}$. As a result, the spectral peak shifts from point $\omega = 0$ to the Larmor precession frequency $\Omega_\mathbf{B}$, which is exactly what is observed in a 2D electron gas [25]. A similar phenomenon has been observed for an assembly of quantum dots [27]. Some theoretical aspects of spin fluctuations in quantum wires were treated recently in Ref. [28].

Spin photogalvanics. Similar to wheel or propeller rotation converting to translation, gyrotropic media allow the conversion, direct or inverse, of the angular momentum into translational motion. In the electron Hamiltonian, it is the spin-dependent terms linear in $\mathbf{k}$ [see formula (4)] which incorporate this possibility and, hence, account for the circular photogalvanic effect theoretically predicted in Ref. [29] in 1978 (more details are in paper [30] and books [10, 15, 31]).

3. Faraday and Kerr spin effects

Currently, the most popular optical spectroscopy approach to the study of electron spin orientation in nanostructures is to adopt the two-beam pump–probe method. Theoretical paper [32] was apparently the first publication to suggest applying this method to bulk semiconductors. The paper considered two monochromatic beams: one highly intensive and circularly polarized, which is responsible for the optical orientation of electron spins, and the other polarized linearly, which acts as a probe. The aim was to study how the probe beam rotates its polarization plane in the transmission geometry (the so-called Faraday spin effect) or in the reflection geometry (Kerr spin effect) (the magnitude of rotation being proportional to the induced electron spin). The current pump–probe setup uses short, picosecond pulses and measures the angle of rotation θ of the probe pulse polarization plane as a function of the interpulse time delay t (Fig. 4a).

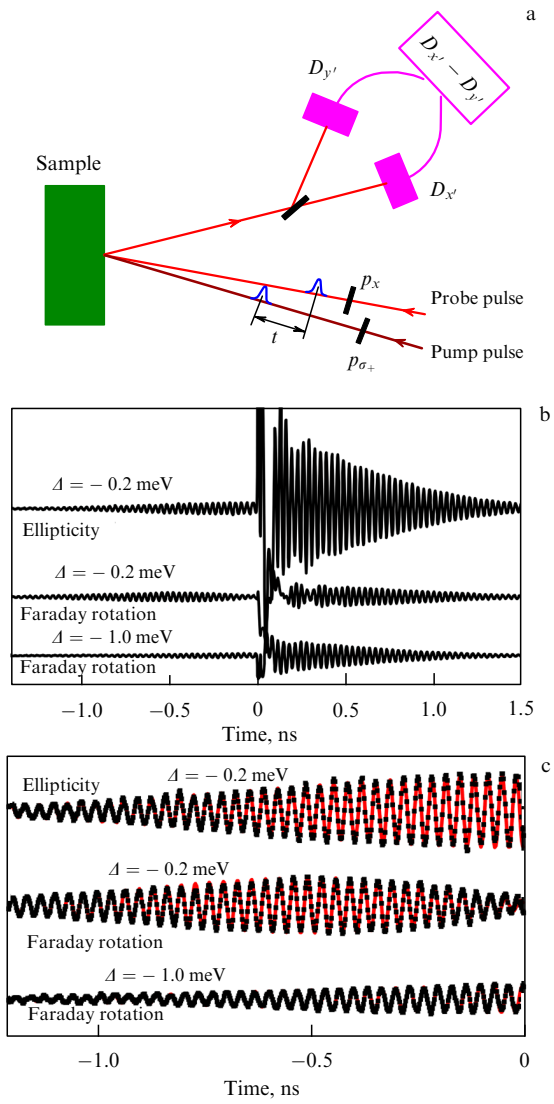

Figure 4. (a) Schematic setup for measuring the rotation angle of the probe beam polarization plane in the spin Faraday effect induced by a circularly polarized pump beam. (b) Ellipticity and Faraday rotation temporal signals measured at near-equal pumping and probing frequencies with mismatch $\Delta \equiv \hbar(\omega_P - \omega_{pr}) = -0.2$ meV and $\Delta = -1$ meV. (c) Comparison of theoretical (solid lines) and experimental (dots) results in the negative time delay range. Note that the dots nearly fit to the curves [33].

4. Spin synchronization in an assembly of charged quantum dots

Figure 4b shows the beats of the spin Faraday rotation and induced ellipticity signals measured at a temperature of 4 K as a function of time delay t on an array of charged InGaAs/GaAs quantum dots. The sample under study,

subjected to a transverse magnetic field ($B = 4$ T, and Larmor precession period ≈ 33 ps), was periodically excited by short circularly polarized pump pulses at a repetition period $T_R = 13.2$ ns, with the periodic probe pulse train shifted by time t (see Ref. [34]). In the performance of an experiment it was possible to vary the mismatch between the carrying frequencies of the pump pulses (ω_P) and the probe pulses (ω_{pr}). The pump pulse can cause a so-called trion to form in the quantum dot, namely a system of two opposite-spin electrons and a heavy hole. Because the trion recombination time equals ≈ 0.4 ns, most of the time each quantum dot contains one electron, whose optical orientation is precisely what produces the spin Faraday effect. Increasing the delay t decreases the amplitude of the spin beats. The reason for this loss of spin coherence does not relate to the time of the actual spin relaxation, which greatly exceeds the repetition period T_R, but rather is due to the spread in the electron Larmor precession frequencies in the excited quantum dots, whose resonance frequency ω_0 differs from the carrying frequency ω_P by no more than the inverse pulse duration τ_p^{-1}. Surprisingly, at first sight, the occurrence of a signal at negative values of t (Fig. 4b). To describe this unusual effect, a microscopic theory was developed [35] to account for the excitation and measurement of long-lived spin coherence and how to control it in singly charged quantum dots by using short-duration optical pulses in the pump–probe mode. As seen from Fig. 4b, calculations for negative values of t using this theory agree well with the experiment.

The key stages of the theoretical treatment are briefly as follows. The starting point is to find out the effect of a short-duration pulse on a single, charged quantum dot. The resonance approximation yields the following equations:

$$i\hbar\dot{\psi}_{3/2} = \hbar\omega_0\psi_{3/2} + V_+(t)\,\psi_{1/2}\,, \qquad i\hbar\dot{\psi}_{1/2} = V_+^*(t)\,\psi_{3/2}\,,$$

$$i\hbar\dot{\psi}_{-3/2} = \hbar\omega_0\psi_{-3/2} + V_-(t)\,\psi_{-1/2}\,, \quad i\hbar\dot{\psi}_{-1/2} = V_-^*(t)\,\psi_{-3/2}\,.$$

Here, $\psi_{\pm 1/2}$ and $\psi_{\pm 3/2}$ are the respective detection probability amplitudes of a single spin $\pm 1/2$ electron or a hole spin $\pm 3/2$ trion in the quantum dot, $V_\pm(t) = -\int d(\mathbf{r})\,E_{\sigma^\pm}(\mathbf{r},t)\,d^3r$ are the interaction matrix elements with the electric field $E_{\sigma^\pm}(\mathbf{r},t)$ of the right and left circularly polarized light wave, and $\dot{\psi} \equiv \partial\psi/\partial t$. The optical transition dipole moment, a quantity characterizing the efficiency of interaction, is given by

$$d(\mathbf{r}) = -i\,\frac{ep_{cv}}{\omega_0 m_0}\,F(\mathbf{r},\mathbf{r})\,, \tag{7}$$

where e is the electron charge, and m_0 is the free electron mass. The two-particle envelope F is defined as

$$F(\mathbf{r},\mathbf{r}) = \varphi_h(\mathbf{r})\,\varphi_e^{(tr)}(\mathbf{r})\int d^3r'\varphi_e(\mathbf{r}')\,\varphi_e^{(tr)}(\mathbf{r}')\,, \tag{8}$$

where $\varphi_e^{(tr)}$ and φ_h are the electron and hole one-particle enveloping functions in the trion, respectively, and φ_e is the envelope of the single (resident) electron in the quantum dot.

The spin $\mathbf{S}^+$ of an electron after being acted by the pump pulse is related linearly to the spin $\mathbf{S}^-$ at the moment when the pulse with polarization σ_+ arrives:

$$S_\alpha^+ = \mathcal{L}_{\alpha\beta}S_\beta^- + \frac{Q^2-1}{4}\,\delta_{\alpha z}\,,$$

$$\hat{\mathcal{L}} = \begin{bmatrix} Q\cos\Phi & Q\sin\Phi & 0 \\ Q\cos\Phi & -Q\sin\Phi & 0 \\ 0 & 0 & \dfrac{Q^2+1}{2} \end{bmatrix}. \tag{9}$$

For rectangular pulses with $f(t) \equiv V_\pm(t)\exp(i\omega_P t)/\hbar = f_0$ for $|t| < \tau_p/2$, and $f(t) = 0$ outside this interval, we obtain

$$Q = \sqrt{1 - \frac{\Theta^2}{x^2}\sin^2\frac{x}{2}}\,, \qquad \Phi = \pi y - \phi\,, \tag{10}$$

where $\Theta = 2f_0\tau_p$, $x = \sqrt{(2\pi y)^2 + \Theta^2}$ is the effective Rabi frequency, $y = (\omega_P - \omega_0)\,\tau_p/2\pi$ is the dimensionless frequency mismatch, and $\sin\phi = (y/Qx)\sin(x/2)$. In a magnetic field $\mathbf{B} \parallel x$, the transverse spin components oscillate: $S_\beta(\Delta t) = \mathcal{M}_{\beta\gamma}(\Delta t)\,S_\gamma^+$, where the time Δt is measured from the moment at which one of the periodic pump pulses arrives, and where the nonzero components of the matrix $\mathcal{M}$ have the form $\mathcal{M}_{yy} = \mathcal{M}_{zz} = \cos(\Omega_B\Delta t)$, $\mathcal{M}_{zy} = -\mathcal{M}_{yz} = \sin(\Omega_B\Delta t)$, and $\mathcal{M}_{xx} = 1$. As a result, the following closed linear equation can be utilized to find the stationary value of the vector $\mathbf{S}^+$:

$$S_\alpha^+ = \mathcal{L}_{\alpha\beta}\,\mathcal{M}_{\beta\gamma}(T_R)\,S_\gamma^+ + \delta_{\alpha z}\,\frac{Q^2-1}{4}\,.$$

A similar computation yields the probability amplitude corrections $\delta\psi_{\pm 1/2}$, $\delta\psi_{\pm 3/2}$ linear in the probe pulse electric field, thus allowing the angle of rotation θ to be calculated by first finding the spin-dependent amplitude corrections for the transmitted (or reflected) probe pulse and then summing them over the quantum dots.

Quantum dots with a Larmor precession period that is a multiple of T_R, i.e. $\Omega_B T_R = 2\pi N$ (where N is an integer), exhibit a resonant accumulation of spin and contribute dominantly to the spin Faraday rotation. Due to the spread in the frequency Ω_B, the commensurability condition is satisfied by quantum dots with N's differing by ± 1, ± 2, ... Thus, the attenuation of the signal upon increasing $t > 0$ is due to the fact that in quantum dots with different values of N, electron spins rotate with different angular velocities. However, by the time of the arrival of the next pump pulse, the spins steadily align themselves along the z-axis. This phenomenon is similar to the synchronization of laser modes [36].

5. Spectroscopy of single quantum dots

In a quantum dot grown along the [001]-axis and characterized by the point symmetry C_{2v}, the electron and hole ground states transform according to the equivalent spinor representations Γ_5 (or, in an alternative notation, E'). In the absence of exchange interaction, the ground state of an exciton e1 – hh1 is fourfold degenerate [37]. The exchange interaction between an electron and a hole removes the degeneracy completely and leads to the splitting of the exciton level to the sublevels Γ_1, Γ_2, Γ_3, Γ_4 (or, accordingly, A_1, B_1, A_2, B_2 in other notations of irreducible representations). In Cartesian coordinates ($x_1 \parallel [1\bar{1}0], y_1 \parallel [110], z \parallel [001]$), the states Γ_2 and Γ_4 are optically active for $\mathbf{e} \parallel x_1$ and $\mathbf{e} \parallel y_1$ polarizations, whereas transitions to the other two states, Γ_1 and Γ_3, are forbidden for $\mathbf{e} \perp z$. In a longitudinal magnetic field $\mathbf{B} \parallel z$, the pair of states Γ_1, Γ_3 undergoes mixing, as does the pair Γ_2, Γ_4 — but there is no mixing between the pairs of sublevels. Therefore, the absorption or emission spectra exhibit two lines linearly polarized in the absence of a magnetic field, and circularly polarized in a strong longitudinal magnetic field. For a similar reason, only two of each four processes $(X^-, j) \to (e1, s)$ and $(X^+, s) \to (hh1, j)$ are permitted in

the radiative recombination of a trion X^- (two singlet-state electrons and a heavy hole) or a trion X^+ (two singlet-state holes and an electron) in an external longitudinal magnetic field, so that the emission spectrum exhibits doublets, not quartets. Here, s, j are the indices of the split Zeeman sublevels of a trion or a single carrier in a quantum dot.

GaAs/AlGaAs quantum dots grown along (111) direction obey totally different selection rules. Figure 5 displays the photoluminescence spectrum of an individual quantum dot for stationary over-barrier optical excitation. In this case, time-integrated spectra contain both emission lines of neutral excitons X^0 and those of trions X^- and X^+. As seen from the figure, in a longitudinal magnetic field $\mathbf{B}\|[111]$, instead of doublets comprising circularly polarized σ_+ and σ_- lines, one observes a quartet, two lines of which are right-hand circularly polarized, σ_+, the other two being left-hand circularly polarized, σ_-. The fact that quantum dots grown along [001]- and [111]-axes differ in the structure of their photoluminescence spectra is naturally explained by the difference in their symmetry point groups C_{2v} and C_{3v}. It should be remembered that in the C_{2v} group the heavy hole states $|\pm 3/2\rangle$ transform according to the two-dimensional irreducible representation Γ_5, and the direct product $\Gamma_5 \times \Gamma_5^* = \Gamma_1 + \Gamma_2 + \Gamma_3 + \Gamma_4$ contains only one representation Γ_3, according to which the magnetic field component B_z transforms. In C_{3v}, the states $|\pm 3/2\rangle$ form the basis of the reducible representation $\mathcal{D} = \Gamma_5 + \Gamma_6$, and the direct product $\mathcal{D} \times \mathcal{D}^* = 2\Gamma_1 + 2\Gamma_2$ includes two representations Γ_2, according to which the component $B_{z'}$ ($z' \| [111]$) transforms in this group. As a result, the Zeeman Hamiltonian in the field $\mathbf{B}\|[111]$ is described by two linearly independent parameters g_{h1}, g_{h2}, and the basis $|\pm 3/2\rangle$ has the form of a 2×2 matrix:

$$\mathcal{H}_\mathbf{B} = \frac{1}{2}\,\mu_B\, B_{z'} \begin{bmatrix} g_{h1} & g_{h2} \\ g_{h2} & -g_{h1} \end{bmatrix}. \tag{11}$$

The eigenvalues of the matrix (11) are given by

$$E_\pm = \pm g_h\, \mu_B\, B_{z'}, \quad g_h \equiv \sqrt{g_{h1}^2 + g_{h2}^2}, \tag{12}$$

and the corresponding eingenfunctions can be reduced to the form

$$|h, +\rangle = C_1 \left|\frac{3}{2}\right\rangle + C_2 \left|-\frac{3}{2}\right\rangle,$$

$$|h, -\rangle = -C_2 \left|\frac{3}{2}\right\rangle + C_1 \left|-\frac{3}{2}\right\rangle$$

with the coefficients

$$C_1 = \sqrt{\frac{1}{2}\left(1 + \frac{g_{h1}}{\sqrt{g_{h1}^2 + g_{h2}^2}}\right)},$$

$$C_2 = \mathrm{sign}\,(g_{h2}) \sqrt{\frac{1}{2}\left(1 - \frac{g_{h1}}{\sqrt{g_{h1}^2 + g_{h2}^2}}\right)}.$$

It should be noted that the coefficients C_1, C_2 are independent of the magnetic field. For $C_2 \neq 0$, a spin 1/2 (or $-1/2$) electron can recombine both with a hole $|h, +\rangle$ and with a hole $|h, -\rangle$ to emit a σ_- (or σ_+) photon. Thus, all

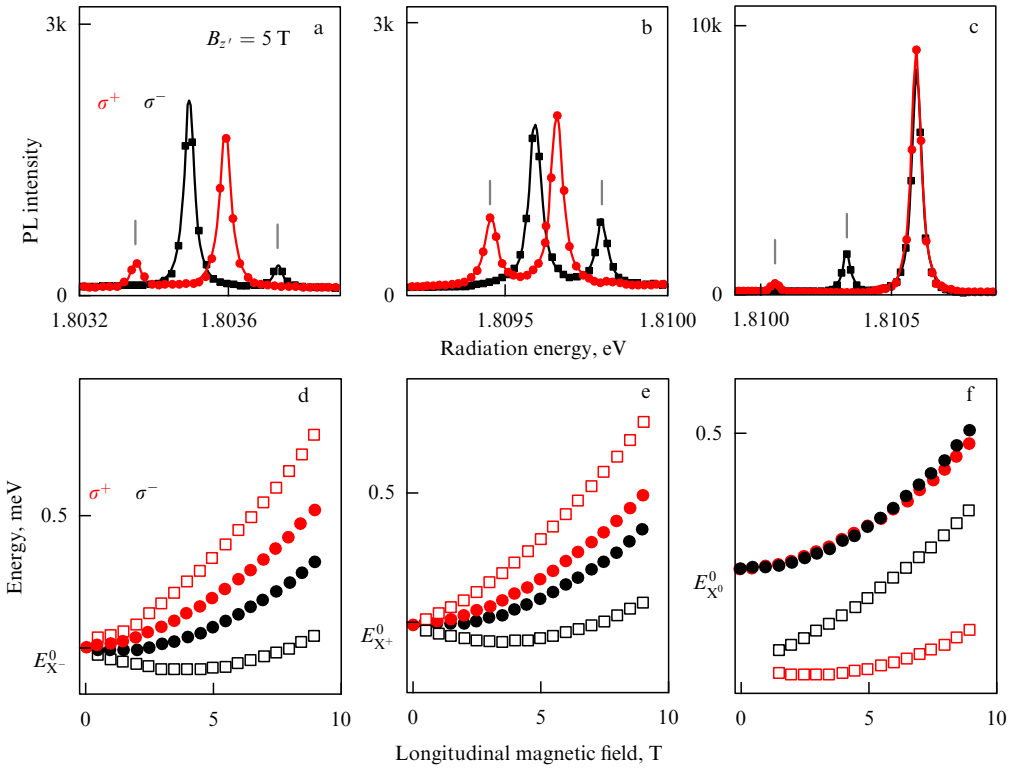

Figure 5. Photoluminescence (PL) spectrum of a single GaAs/AlGaAs (111) quantum dot measured in a magnetic field $B_{z'} = 5$ T in the emission range of trion X^+ (a), trion X^- (b), and neutral exciton X^0 (c); k is taken as a conditional intensity measurement unit. Circles and squares show the right and left circularly polarized emission lines, respectively. (d–f) Change in the position of the corresponding emission line with increasing magnetic field [18].

four transitions turn out to be optically active, whereas in (001) quantum dots the parameter g_{h2} is zero, implying that $C_2 \equiv 0$ and that only two recombination processes are allowed. It should be emphasized that the mixing effect of $\pm 3/2$ states in a longitudinal magnetic field can show its worth in trigonal systems with arbitrary dimensionality $d = 0-3$, including an exciton in Ge crystals formed by an L-valley electron and a Γ_8^+ hole and bound on a neutral donor [38].

A nonzero value g_{h2} of the g factor can be obtained by noting that in bulk zinc blende lattice semiconductors the Zeeman interaction of Γ_8 holes with a magnetic field is described by the Hamiltonian

$$\mathcal{H}_{\mathbf{B}}^{(\Gamma_8)} = -2\mu_{\mathrm{B}}\left[\kappa \mathbf{J}\,\mathbf{B} + q(J_x^3 B_x + J_y^3 B_y + J_z^3 B_z)\right], \quad (13)$$

which contains two dimensionless coefficients, κ and q. Here, x, y and z are the crystallographic axes [100], [010] and [001], and J_x, J_y, and J_z are the angle momentum matrices in the Γ_8 basis. Let us go over in Hamiltonian (13) to the coordinates $x' \| [11\bar{2}]$, $y' \| [\bar{1}10]$, $z' \| [111]$ and introduce the basis functions $|3/2\rangle'$, $|-3/2\rangle'$, which transform according to the reducible representation $\mathcal{D} = \Gamma_5 + \Gamma_6$ of the C_{3v} group. Then, the Zeeman splitting in the field $\mathbf{B} \| [111]$ will be described by a 2×2 matrix with $g_{h1} = -6\kappa$, and $g_{h2} = 2\sqrt{2}q$.

Acknowledgments. This work was supported by an RAS Presidium Program and by RFBR grants. The author is grateful to M M Glasov for helpful discussions of the manuscript.

References

1. Dyakonov M I (Ed.) *Spin Physics in Semiconductors* (Berlin: Springer, 2008)
2. Kusrayev Yu, Landwehr G (Guest Eds) *Semicond. Sci. Technol.* **23** (11) (2008), Special issue on optical orientation
3. *Fiz. Tekh. Poluprovodn.* **42** (8) (2008) [*Semicond.* **42** (8) (2008)], Vladimir Idelevich Perel' (on the 80th anniversary of his birthday), Special issue
4. Ivchenko E L, Kiselev A A *Fiz. Tekh. Poluprovodn.* **26** 1471 (1992) [*Sov. Phys. Semicond.* **26** 827 (1992)]
5. Kiselev A A, Ivchenko E L, Rössler U *Phys. Rev. B* **58** 16353 (1998)
6. Yugova I A et al. *Phys. Rev. B* **75** 245302 (2007)
7. Durnev M V, Glazov M M, Ivchenko E L *Physica E* **44** 797 (2012); arXiv:1111.6837
8. Rashba E I *Fiz. Tverd. Tela* **2** 1224 (1960) [*Sov. Phys. Solid State* **2** 1109 (1960)]
9. Bychkov Yu A, Rashba E I *Pis'ma Zh. Eksp. Teor. Fiz.* **39** 66 (1984) [*JETP Lett.* **39** 78 (1984)]
10. Ivchenko E L *Optical Spectroscopy of Semiconductor Nanostructures* (Harrow, UK: Alpha Sci. Intern. Ltd, 2005)
11. Nestoklon M O, Golub L E, Ivchenko E L *Phys. Rev. B* **73** 235334 (2006)
12. Glazov M M, Ivchenko E L *Pis'ma Zh. Eksp. Teor. Fiz.* **75** 476 (2002) [*JETP Lett.* **75** 403 (2002)]
13. Glazov M M, Ivchenko E L *Zh. Eksp. Teor. Fiz.* **126** 1465 (2004) [*JETP* **99** 1279 (2004)]
14. Leyland W J H et al. *Phys. Rev. B* **75** 165309 (2007)
15. Ivchenko E L, Pikus G E *Superlattices and other Heterostructures: Symmetry and Optical Phenomena* (Berlin: Springer-Verlag, 1995); *Superlattices and other Heterostructures: Symmetry and Optical Phenomena* 2nd ed. (Berlin: Springer, 1997)
16. Ivchenko E L *Phys. Status Solidi A* **164** 487 (1997)
17. Gupalov S V, Ivchenko E L, Kavokin A V *Zh. Eksp. Teor. Fiz.* **113** 703 (1998) [*JETP* **86** 388 (1998)]
18. Sallen G et al. *Phys. Rev. Lett.* **107** 166604 (2011)
19. Stevenson R M et al. *Phys. Rev. B* **73** 033306 (2006)
20. Glazov M M et al. *Phys. Rev. B* **76** 193313 (2007)
21. Kusrayev Yu G *Usp. Fiz. Nauk* **180** 759 (2010) [*Phys. Usp.* **53** 725 (2010)]
22. Glazov M M, Yugova I A, Efros Al L *Phys. Rev. B* **85** 041303(R) (2012); arXiv:1103.3249
23. Glazov M M *Fiz. Tverd. Tela* **54** 3 (2012) [*Phys. Solid State* **54** 1 (2012)]
24. Ivchenko E L *Fiz. Tekh. Poluprovodn.* **7** 1489 (1973) [*Sov. Phys. Semicond.* **7** 998 (1973)]
25. Müller G M et al. *Physica E* **43** 569 (2010)
26. Aleksandrov E B, Zapasskii V S *Zh. Eksp. Teor. Fiz.* **81** 132 (1981) [*Sov. Phys. JETP* **54** 64 (1981)]
27. Crooker S A et al. *Phys. Rev. Lett.* **104** 036601 (2010)
28. Glazov M M, Sherman E Ya *Phys. Rev. Lett.* **107** 156602 (2011)
29. Ivchenko E L, Pikus G E *Pis'ma Zh. Eksp. Teor. Fiz.* **27** 640 (1978) [*JETP Lett.* **27** 604 (1978)]
30. Ivchenko E L *Usp. Fiz. Nauk* **172** 1461 (2002) [*Phys. Usp.* **45** 1299 (2002)]
31. Ivchenko E L, Ganichev S D, in *Spin Physics in Semiconductors* (Ed. M I Dyakonov) (Berlin: Springer, 2008) p. 245
32. Aronov A G, Ivchenko E L *Fiz. Tverd. Tela* **15** 231 (1973) [*Sov. Phys. Solid State* **15** 160 (1973)]
33. Glazov M M et al. *Phys. Rev. B* **82** 155325 (2010)
34. Greilich A et al. *Science* **313** 341 (2006)
35. Yugova I A et al. *Phys. Rev. B* **80** 104436 (2009)
36. Haken H *Light* Vol. 2 *Laser Light Dynamics* (Amsterdam: North-Holland, 1985) [Translated into Russian (Moscow: Mir, 1988)]
37. Dzhioev R I et al. *Pis'ma Zh. Eksp. Teor. Fiz.* **65** 766 (1997) [*JETP Lett.* **65** 804 (1997)]
38. Averkiev N S et al. *Fiz. Tverd. Tela* **23** 3117 (1981) [*Sov. Phys. Solid State* **23** 1815 (1981)]

PACS numbers: 72.25.Hg, 72.25.Pn, 72.25.Rb, 73.63.Hs
DOI: 10.3367/UFNe.0182.201208h.0876

Spin transport in heterostructures

L E Golub

1. Introduction. Spin splittings

In the absence of an external magnetic field, electronic states can be spin-split if the system has no space inversion center. The reason for these spin splittings is the spin–orbit interaction. The simplest example of a noncentrosymmetric medium is a surface. The Hamiltonian of the spin-orbit interaction in a half-infinite medium assumes the following (Rashba [1, 2]) form:

$$H_{\mathrm{so}} = \alpha(\boldsymbol{\sigma} \times \mathbf{k})\,\mathbf{n}\,. \quad (1)$$

Here, the vector $\boldsymbol{\sigma}$ is composed of Pauli matrices, $\mathbf{k}$ is the electron wave vector, α is a certain number, and $\mathbf{n}$ is a unit normal vector to the surface. This form of such spin–orbit interaction occurs in various noncentrosymmetric semiconductors, metals, and superconductors.

L E Golub Ioffe Physical Technical Institute,
Russian Academy of Sciences, St. Petersburg, Russian Federation
E-mail:golub@coherent.ioffe.ru

Uspekhi Fizicheskikh Nauk **182** (8) 876 – 879 (2012)
DOI: 10.3367/UFNr.0182.201208h.0876
Translated by E G Strel'chenko; edited by A Radzig

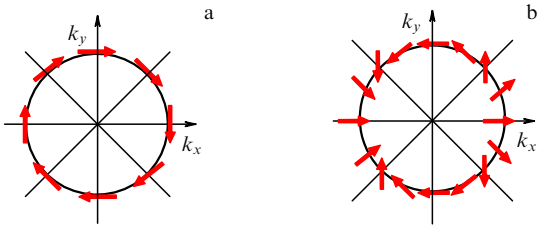

Figure 1. Effective magnetic field $\mathbf{\Omega}(\mathbf{k})$ for electrons (a) and polaritons (b).

In addition to expression (1), an interaction of a different form — Dresselhaus — occurs in heterostructures grown from noncentrosymmetric semiconductors (for example, GaAs). The Hamiltonian of any type of such interaction has the form $H_{so} \propto \sigma_i k_j$, $\sigma_i k_j k_l k_m$ and can be conveniently written down in the Zeeman form

$$H_{so} = \frac{\hbar}{2}\,\boldsymbol{\sigma}\,\mathbf{\Omega}(\mathbf{k})\,. \qquad (2)$$

Here, the effective Larmor frequency $\mathbf{\Omega}(\mathbf{k})$ is a pseudovector odd in $\mathbf{k}$. The spin splitting energy is $\Delta_{so}(\mathbf{k}) = \hbar|\mathbf{\Omega}(\mathbf{k})|$. The directions of $\mathbf{\Omega}(\mathbf{k})$ in $\mathbf{k}$ space for the Rashba interaction are shown by arrows in Fig. 1a. In real heterostructures, both types of splitting exist. Their interference results in suppressing electron spin relaxation for one of the spin orientations in the plane of a structure grown along the [001] direction if the Rashba and Dresselhaus fields are equal in magnitude [3–5]. Based on these ideas, a model of a spin transistor capable of realizing diffusive spin transport was proposed [6].

2. Weak localization of electrons and exciton polaritons

A transport phenomenon which clearly displays spin splittings is weak localization — an effect consisting in enhanced backward scattering due to the interference of the incident and scattered waves. There are a variety of pathways for a wave propagating through a system of scatterers, some of them forming a special class of self-intersecting or loop-containing trajectories. Trajectories with loops following clockwise and counterclockwise are accounted for independently, and the phase incursions of the wave passing along these trajectories are equal. As a result, the waves reflected after passing along two such trajectories interfere, and the passage through a system of scatterers becomes less probable than in a classical calculation. Such a reduction in passage is a localization in character, even though it occurs for freely propagating particles, which is the reason why this phenomenon — enhanced backward scattering — came to be known as the *weak localization* effect.

Because of their wave properties, electrons also exhibit this effect, if to a small degree. In electronic systems, weak localization shows its worth as a decrease in conductivity, as opposed to the classical Drude value σ_{cl}. The correction to the conductivity due to the weak localization is equal to $\Delta\sigma_{WL} \sim (\lambda/l)\,\sigma_{cl}$, where λ is the de Broglie wavelength, and l is the electron mean free path. As a result, we have $\Delta\sigma_{WL} \sim e^2/\hbar$ [7].

What is remarkable about the correction $\Delta\sigma_{WL}$ is its sensitivity to a classically weak magnetic field which does not produce a considerable Lorentz force. The field dependence is due to the fact that electron waves that have passed a self-intersecting trajectory in two opposite directions differ in phase by an amount equal to the magnetic field flux through the loop. As a result, the field destroys the wave interference, and the conductivity returns to its classical value, i.e. increases. There are a variety of semiconductors, metals, and heterostructures that exhibit positive magnetoconductivity or negative magnetoresistance. This magnetoresistance is referred to as *anomalous* due to its strong dependence on the magnetic field and temperature.

The presence of spin splittings we touched upon in the Introduction changes the weak localization picture markedly. From Fig. 1a it is seen that, if the spin follows the change in the effective field $\mathbf{\Omega}$, a clockwise or counterclockwise path tracing imparts to the electron a (Berry) phase equal to $\pm\pi$. Considering the existence of an electron spin, this difference in sign is significant, because the wave function of a backward scattered electron differs by a factor $\pm$i from the incident wave function. As a result, the constructive interference becomes destructive, the backward scattering is suppressed rather than enhanced, the correction to the conductivity is positive, and the magnetic field destroying this correction leads to positive magnetoresistance. Because all this is opposite to the spinless case, this mode of electron behavior is called *weak antilocalization*. In the case of moderately strong spin–orbit interaction, when Δ_{so} is much less than the Fermi energy, the time between collisions is too short for an electron to fully align along the field $\mathbf{\Omega}$, so that magnetoresistance is sign-variable and is represented by a curve with a maximum.

The mid-1990s theory [8] of weak localization for heterostructures with Rashba and Dresselhaus (see Eqn (2)) spin splittings was successful in describing magnetoresistance in low-mobility heterostructures available at the time [9]. Anomalous magnetoresistance was typically observed in much weaker magnetic fields than the 'transport' field in which the magnetic length becomes equal to the electron mean free path, $B_{tr} = \hbar/(2el^2)$.

However, starting from the 2000s, experimental studies began to appear that demonstrated anomalous magnetoresistance in fields $B \lesssim B_{tr}$. The application of the theory developed in Ref. [8] sometimes led to meaningless fitting parameters. Especially challenging for theorists was the finding [10] that the formulas of Ref. [8] produce different fitting parameters for the decreasing and increasing portions of the magetoconductivity curve. It became clear that high-mobility heterostructures require for their description a new expression for anomalous magnetoresistance, which is suitable for both the diffusion and ballistic modes of weak antilocalization.

It was Ref. [11] which came up with the new theory. In the presence of the spin–orbit interaction (2), an electron which resided in one of the spin states $|\alpha\rangle$ before tracing the loop is, generally, in a different state, $|\beta\rangle$, after tracing it due to the spin having rotated in the effective field $\mathbf{\Omega}$. In accordance with the electron spin having two possible projections onto the growth axis before and after the loop tracing ($\alpha, \beta = \pm 1/2$), there are four interferential contributions to the conductivity. It is convenient to introduce two variables, the total electron moment $\mathbf{S}$ before and after tracing the loops, and its projection m onto the heterostructure growth axis. It turns out that the interference is constructive or destructive if, respectively, the electron spin states form a singlet ($S = 0$, $m = 0$) or a triplet ($S = 1$, $m = 0, \pm 1$) before and after the loop tracing. The correction to the conductivity is calculated

by the Green function method. In a magnetic field, the probability of backward scattering is calculated as a function of Landau level indices N. Whereas in the singlet channel all Landau levels contribute independently, in the triplet channel contributions from different levels are engaged. For example, independent contributions in the case of the Rashba interaction come from the states (N, m) with equal $N + m$: $(N, 1)$, $(N + 1, 0)$, and $(N + 2, -1)$, whereas in the case of the Dresselhaus interaction they come from those with equal $N - m$. In either case, the triplet contribution to the conductivity is described by the trace of a third-rank matrix [11]. According to the new theory, the magnetoconductivity may have its minimum equally well in fields larger and smaller than $B_{\rm tr}$. For $B \gg B_{\rm tr}$, the magnetoconductivity comes to the asymptotic value

$$\Delta\sigma_{\rm WL} = -0.25\sqrt{\frac{B_{\rm tr}}{B}}\,\frac{e^2}{\hbar}\,,$$

whatever the magnitude of the spin–orbit interaction.

The theory proved successful in describing experimental data on anomalous magnetoresistance in various heterostructures [12, 13]. Experimental results described by the theory for various temperatures and various concentrations of two-dimensional electrons are reviewed in Ref. [14].

Reference [15] considers anisotropic spin splitting for the case with both the Rashba and Dresselhaus contributions present. In this case, the contributions to $\Delta\sigma_{\rm WL}$ from different Landau levels do not break up in groups, and the triplet contribution is described by an infinite-rank matrix. In the case of a spin splitting cubic in the wave vector, which occurs for a high electron concentration, the triplet contribution is made by trios of states with equal values of $N + 3m$ [15]. The expressions obtained are also valid for hole type quantum wells, for which spin splitting is also cubic in the wave vector. The anomalous magnetoresistance in hole heterostructures has recently become a subject of study for several experimental groups simultaneously.

Weak localization comprises a wave interference phenomenon and, as such, also occurs for light. A suitable heterostructure system to study light interference is the microcavity. Experiments on the Rayleigh scattering of exciton polaritons demonstrate the enhancement of backward scattering [16]. In polarization-resolved measurements, light of different polarizations excites different polariton states. These are conveniently described by introducing the pseudospin vector $\mathbf{S}$ composed of the polarization Stokes parameters. The longitudinal–transverse splitting of polaritons is described by the same Hamiltonian (2) as the electron spin splitting, with the important difference that the effective Larmor frequency of polaritons is quadratic in their momenta in the microcavity plane, $\mathbf{k}$. Its directions in $\mathbf{k}$ space are shown in Fig. 1b. Due to the evenness of $\boldsymbol{\Omega}(\mathbf{k})$, the Berry phase — a phase incursion of a polariton as it traces a loop — equals 2π (if the spin follows the direction of $\boldsymbol{\Omega}$) rather than π, as for electrons. Therefore, polariton interference is constructive, backward scattering is enhanced, and localization takes place. As a result, backward scattering increases in intensity compared to scattering through other angles [17].

In polarization-resolved experiments, the scattering of polaritons with a fixed pseudospin projection can be studied separately. This makes a polariton system advantageously different from its electronic counterpart, in that in the latter the only quantity which is measured is the conductivity equal to a sum of the singlet and all the triplet interferential contributions. It turns out that for a pseudospin oriented in the microcavity plane (light is polarized linearly), the backward scattering intensity enhances, whereas for the pseudospin directed along the growth axis (circular polarization of light) it shows a dip [18]. If the cavity is nonsymmetric, the longitudinal–transverse splitting becomes anisotropic with respect to various directions of $\mathbf{k}$ in the plane of the structure, resulting in the difference in backscattering for two linear polarizations. Further still, interferential effects act to convert linear polarization to circular [18].

The presence of a longitudinal–transverse splitting also affects the classical dynamics of polarization in microcavities in which the multiple elastic light scattering mode is realized [19]. This requires that the polariton elastic scattering time τ_1 be much shorter than the lifetime τ_0. The kinetics of a polariton pseudospin is described by the equation [17]

$$\frac{\mathbf{S}_\mathbf{k}}{\tau_0} + \mathbf{S}_\mathbf{k} \times \boldsymbol{\Omega}(\mathbf{k}) + \frac{\mathbf{S}_\mathbf{k} - \langle \mathbf{S}_\mathbf{k} \rangle}{\tau_1} = \mathbf{g}_\mathbf{k}\,. \tag{3}$$

Here, $\mathbf{S}_\mathbf{k}$ is the pseudospin of a polariton with a wave vector $\mathbf{k}$, the angular brackets stand for averaging over the directions of $\mathbf{k}$, and $\mathbf{g}_\mathbf{k}$ is the generation rate to the $\mathbf{k}$ state. Pseudospin rotations in the field $\boldsymbol{\Omega}(\mathbf{k})$ lead to the emission of circularly polarized light when the exciting polarization is linear. Instead of the labor-consuming computer calculations that were done in Ref. [19], the kinetic theory of Ref. [18] can be used to describe the experimental results.

3. Spin orientation by electric current

The conversion of linearly polarized radiation to circularly polarized one is a special case of the more general transformation of translational motion to rotational. In electronic systems, an example of such a phenomenon is the electric current-assisted orientation of spin.

The appearance of spin $\mathbf{s}$ due to the flow of electric current with density $\mathbf{j}$ is described phenomenologically by the equality

$$s_i = Q_{im} j_m\,, \tag{4}$$

where $\hat{Q}$ is a second-rank pseudotensor. This relation applies only to gyrotropic media. Although bulk GaAs type semiconductors are not gyrotropic, the fact is that, virtually regardless of the material used to grow a heterostructure, its $\hat{Q}$ has nonzero components. Symmetry analysis shows that if gyrotropy is due to the structural asymmetry of the system (similar to the Rashba splitting), the spin aligns along the plane of the structure perpendicular to the current (Fig. 2a). The absence of a spatial inversion center in a bulk material (for example, in GaAs) leads, in addition to Dresselhaus

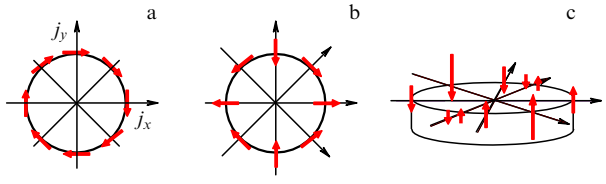

Figure 2. Spin orientation by current in (z-axis grown) heterostructures with strong structural asymmetry (a), and in A_3B_5 quantum dots with $z \parallel [001]$ (b) and $z \parallel [110]$ (c).

splitting, to the possibility of a spin being oriented by an electric current. In this case, the effect is crystallographically sensitive. In A_3B_5 heterostructures grown along the cubic [001]-axis, the current orients the spin to lie in the plane, as shown in Fig. 2b: parallel or perpendicular to the current depending, respectively, on whether the current flows along the axes [100], [010] or along [110], [1$\bar{1}$0]. In [110]-grown heterostructures, a current flowing along the [1$\bar{1}$0]-axis causes the spin to align along the growth axis (Fig. 2c).

Experimental studies of current-assisted spin orientation were carried out on heterostructures grown from GaAs, InAs, ZnSe, and GaN in different crystallographic directions [20–22].

The microscopic mechanism of spin orientation by current is related to the spin–orbit splitting (2). In an external electric field **E**, the electronic system acquires a drift quasimomentum $\mathbf{k}_{dr} = e\mathbf{E}\tau_{tr}/\hbar$, where τ_{tr} is the mobility-controlling transport relaxation time. The presence of a finite quasimomentum formally implies a nonzero spin splitting $\mathbf{\Delta}_{so}(\mathbf{k}_{dr}) = \hbar\mathbf{\Omega}(\mathbf{k}_{dr})$ due to the effective magnetic field. As a result, spin polarization emerges, which, by analogy with the Zeeman effect, can be represented in the form

$$s = -c\,\frac{\mathbf{\Delta}_{so}(\mathbf{k}_{dr})}{\bar{E}}\,, \qquad (5)$$

where $\bar{E}$ is the average electron energy equal to the Fermi energy or the temperature in the degenerate and Boltzmann cases, respectively.

For a system in a real magnetic field, the constant c would be 1/4. For spin orientation by an electric field, this coefficient is also on the order of unity. However, a difference shows up in this case between the thermodynamic effect of spin orientation by a magnetic field and the kinetic effect of electric current-assisted spin orientation. In the presence of an electric field, the electron distribution in **k** space undergoes a shift in any of the spin branches and, taking into account the spin–orbit splittings of the energy spectrum, has the form shown in Fig. 3a. However, until there is no coupling between the spin subsystems, the spin up and spin down electrons are equal in number, thus preventing the appearance of an average spin. It is spin relaxation which allows the spin subbands to 'communicate'. The arrows in Fig. 3a indicate spin flip processes. It is seen that, if the probability of a spin flip scattering event depends on the momentum transfer, spin generation will occur: transitions from the spin-up subband to the spin-down subband and those in the opposite direction will occur at different rates. However, because spin relaxation is controlled by the same processes, the spin flip probability does not affect the stationary value of spin. This leaves us with formula (5) as the correct one — with the caveat, however, that theoretical studies (of which there are many) yield different values for c (for a review, see Ref. [23] and references cited therein). The spread is due to the fact that the degree of spin orientation depends on the ratio of the spin to energy relaxation rates. If energy relaxation processes are faster than those of spin relaxation, they rapidly mix spin between electrons of different energies. In this case, the spin distribution over energy is equilibrium, albeit $c = 1/2$, i.e. twice the value in a real magnetic field. In the opposite limit of slow energy relaxation, spin establishes itself for each energy independently. Although this results in $c = 1/4$, the distribution of spins over the energy is nonequilibrium [23]. Experimentally, energy relaxation processes are slow at temperatures from 4–10 K, but at higher temperatures energy relaxation becomes faster than the spin relaxation. For comparable values of energy and spin relaxation times, the coefficient c ranges between 1/4 and 1/2.

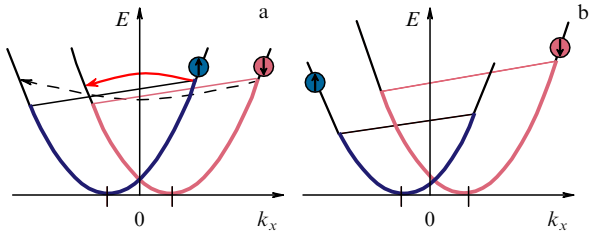

Figure 3. Microscopic picture of electric spin orientation. (a) Spin subband populations in the presence of an electric field. Spin–orbit splitting is taken into account, and spin relaxation processes are neglected. Solid and dashed arrows denote spin flip processes running with different probabilities. (b) Distribution in spin subbands due to spin relaxation processes.

Of interest is the case of symmetrically doped (110) quantum wells. In such systems, symmetry allows the normal spin component to be oriented by a current flowing along the [1$\bar{1}$0]-axis, but the Rashba spin splitting is zero, and the Dyakonov–Perel spin relaxation is absent. Allowing fluctuations in the position of a doping impurity makes a quantum well locally nonsymmetric, resulting in the fluctuation of the spin–orbit interaction and giving rise to spin relaxation [24]. This also enables current-assisted spin orientation but, remarkably, the coefficient c in formula (5) becomes a function of the correlation length of the spin–orbit field fluctuations [23]. Also of interest are the kinetics of current-assisted spin orientation in such a system, in particular, the fact of their becoming nonmonoexponential for slow energy relaxation [23].

Two-dimensional topological insulators realizable, for example, on the surface of three-dimensional Bi_2Se_3 compounds, are yet another class of systems allowing for spin orientation by current. The electronic spectrum of such systems is linear in the two-dimensional wave vector and is described by the effective Hamiltonian (1). Importantly, because of the absence of terms quadratic in k, contribution (1) is not a small correction in this case. Spin in topological insulators is oriented perpendicular to the current (see Fig. 2a), and its magnitude is $s = k_{dr}/(2k_F)$, with k_F being the Fermi wave vector [23]

Acknowledgments. This work was supported by RAS and RFBR programs, and by a grant from the President's Fund for Young Scientists.

References

1. Rashba E I *Fiz. Tverd. Tela* **2** 1224 (1960) [*Sov. Phys. Solid State* **2** 1109 (1960)]
2. Rashba E I *Usp. Fiz. Nauk* **84** 557 (1964) [*Sov. Phys. Usp.* **7** 823 (1965)]
3. Averkiev N S, Golub L E *Phys. Rev. B* **60** 15582 (1999)
4. Averkiev N S, Golub L E *Semicond. Sci. Technol.* **23** 114002 (2008)
5. Averkiev N S *Usp. Fiz. Nauk* **180** 777 (2010) [*Phys. Usp.* **53** 742 (2010)]
6. Schliemann J, Egues J C, Loss D *Phys. Rev. Lett.* **90** 146801 (2003)
7. Altshuler B L, Aronov A G, in *Electron-Electron Interactions in Disordered Systems* (Eds A L Efros, M Pollak) (Amsterdam: Elsevier, 1985) p. 1

8. Iordanskii S V, Lyanda-Geller Yu B, Pikus G E *Pis'ma Zh. Eksp. Teor. Fiz.* **60** 199 (1994) [*JETP Lett.* **60** 206 (1994)]
9. Knap W et al. *Phys. Rev. B* **53** 3912 (1996)
10. Studenikin S A et al. *Phys. Rev. B* **68** 035317 (2003)
11. Golub L E *Phys. Rev. B* **71** 235310 (2005)
12. Guzenko V A et al. *Phys. Status Solidi C* **3** 4227 (2006)
13. Yu G et al. *Phys. Rev. B* **78** 035304 (2008)
14. Glazov M M, Golub L E *Semicond. Sci. Technol.* **24** 064007 (2009)
15. Glazov M M, Golub L E *Fiz. Tekh. Poluprovodn.* **40** 1241 (2006) [*Semiconductors* **40** 1209 (2006)]
16. Gurioli M et al. *Phys. Rev. Lett.* **94** 183901 (2005)
17. Glazov M M, Golub L E *Phys. Rev. B* **77** 165341 (2008)
18. Glazov M M, Golub L E *Phys. Rev. B* **82** 085315 (2010)
19. Amo A et al. *Phys. Rev. B* **80** 165325 (2009)
20. Ganichev S D et al. *J. Magn. Magn. Mater.* **300** 127 (2006); cond-mat/0403641
21. Silov A Yu et al. *Appl. Phys. Lett.* **85** 5929 (2004)
22. Sih V et al. *Nature Phys.* **1** 31 (2005)
23. Golub L E, Ivchenko E L *Phys. Rev. B* **84** 115303 (2011)
24. Glazov M M, Sherman E Ya, Dugaev V K *Physica E* **42** 2157 (2010)

PACS numbers: 47.27.Gs, 47.35.Pq, 68.03.Kn
DOI: 10.3367/UFNe.0182.201208i.0879

Kinetic and discrete turbulence on the surface of quantum liquids

L V Abdurakhimov, M Yu Brazhnikov, A A Levchenko, I A Remizov, S V Filatov

1. Introduction

Wave turbulence is a nonequilibrium state in a system of interacting nonlinear waves in which the energy pumping and dissipation ranges are well separated in the wave number space. A turbulent state is characterized by a directed energy flux P in the k-space. Wave turbulence states can be realized in many nonlinear systems, for example, in plasmas [1], magnetic systems in solids [2], and on the surface of seas and oceans [3]. In our experiments, we explore the turbulence in a system of capillary waves, where surface tension plays the main role. Waves on a water surface are conventionally referred to as gravity waves if their wavelength exceeds 17 mm, and as capillary waves otherwise.

The frequency ω of capillary waves on the surface of a liquid is defined by the modulus of the wave vector k together with the surface tension coefficient σ and fluid density ρ:

$$\omega = \left(\frac{\sigma}{\rho}\right)^{1/2} k^{3/2} . \qquad (1)$$

Dispersion law (1) for capillary waves is of a decaying type, i.e., it permits the three-wave processes of wave decay into two waves or coalescence of two waves into a single wave while such that the energy and momentum are conserved,

$$\omega_1 \pm \omega_2 = \omega_3 , \qquad k_1 \pm k_2 = k_3 . \qquad (2)$$

When the surface of a liquid is excited by an external force, a turbulent state can develop in the system of capillary waves, in which the energy flux P in the k-space is directed from the pumping range toward large wave numbers (high frequencies), forming a direct cascade. Under the assumption that wave interactions are weak and hence the main contribution to energy transfer comes from three-wave processes, the theory of (weak) wave turbulence [4] predicts a power law for the energy distribution over frequencies, $E(\omega) \sim \omega^{-3/2}$.

However, exploring the energy frequency spectrum in experiments with capillary waves is a rather difficult task. From the standpoint of an experimentalist, it is most convenient to explore not the energy distribution E_ω but the pair correlation function $I(\tau) = \langle \eta(r, t+\tau)\, \eta(r, t) \rangle$ for the deviation of the surface elevation from equilibrium at a point r, because the deviation $\eta(r, t)$ from a planar surface is directly measurable.

The wave turbulence theory [4] for a system of capillary waves on the surface of a liquid predicts the formation of a turbulent cascade in the inertial range bounded by the pumping at low frequencies and the dissipation range at high frequencies. Within the inertial range, the pair correlation function $I(\tau)$ in the Fourier representation is described by a power-law function of the frequency (turbulent cascade):

$$I_\omega \sim \omega^{-m} , \qquad (3)$$

and $E(\omega) \sim \omega^{4/3} I_\omega$. The exponent m depends on the spectral characteristics of the driving force. Under the excitation of the surface of the liquid by a low-frequency noise in a broad band $\Delta\omega$ (with a bandwidth exceeding the characteristic pumping frequency ω_p, $\Delta\omega \geqslant \omega_\mathrm{p}$), the turbulent cascade I_ω is described by the function ω^{-m} with the exponent $m = 17/6$. Numerical simulations [5] provide an estimate of m that is close to the theoretical prediction. The results of numerical modeling in Ref. [6] indicate that as the bandwidth of noise pumping $\Delta\omega$ is reduced, a series of equidistant peaks emerges in the turbulent cascade, with their widths behaving as a linear function of the frequency. For a narrow-band pumping, $\Delta\omega < \omega_\mathrm{p}$, the decrease in the height of these peaks as the frequency increases is described by a power-law function of frequency with an exponent that exceeds the value for broadband noise pumping by one, i.e., $m = 23/6$.

Our experimental studies on the surface of liquid hydrogen have shown that the spectral characteristic of the applied force determines the value of the power-law exponent [7]. When the surface is perturbed by a low-frequency harmonic force, the correlation function I_ω exhibits a set of narrow peaks whose frequencies are multiples of the pumping frequency ω_p. The peak maxima are well described by a power law ω^{-m} with $m = 3.7 \pm 0.3$. When, in addition to pumping at a single resonance frequency, a harmonic force at another resonance frequency is applied, the exponent decreases to $m = 2.8 \pm 0.2$. The exponent was also close to $m = 3 \pm 0.3$ when the surface was excited by a broadband low-frequency noise. In these experiments, we have qualitatively shown that in passing from the surface excitation with broadband noise to pumping by a harmonic force at the single-cell resonance frequency, the exponent m increases. Detailed results characterizing the evolution of a turbulent

L V Abdurakhimov, M Yu Brazhnikov, A A Levchenko, I A Remizov, S V Filatov Institute of Solid State Physics, Russian Academy of Sciences, Chernogolovka, Moscow region, Russian Federation
E-mail: levch@issp.ac.ru

Uspekhi Fizicheskikh Nauk **182** (8) 879 – 887 (2012)
DOI: 10.3367/UFNr.0182.201208i.0879
Translated by S D Danilov; edited by A M Semikhatov

cascade under variations in the spectral characteristic of the excitation force (in passing from broadband pumping to a narrowband one) are obtained in Ref. [8].

In a stationary turbulent spectrum in a system of capillary waves, the energy is transferred to the region of high frequencies, where it is converted to heat through viscous losses, and the turbulent cascade decays. To keep the turbulent cascade in a stationary state, the energy has to be permanently supplied at low frequencies. The high-frequency boundary of the inertial range can be estimated by assuming that at the bounding frequency ω_b, the time scale of nonlinear wave interaction τ_{nl} becomes comparable in the order of magnitude to the viscous dissipation time τ_v [3],

$$\omega_b \sim \left(\frac{P^{1/2}}{\nu}\right)^{6/5} \sim \left(\frac{\eta_0^2\,\omega_0^{17/6}}{\nu}\right)^{6/5}, \quad (4)$$

where η_0^2 is the wave amplitude squared at the pumping frequency ω_p and ν is the kinematic viscosity of the liquid.

At high frequencies, the spectrum behavior is determined by the details of energy dissipation and nonlinear wave interaction. When waves in the dissipation range interact mainly with the nearest neighbors but not with waves from the inertial range, the wave distribution at high frequencies becomes close to exponential [9]. A detailed analysis [10] gives a quasi-Planck spectrum for the correlation function in the dissipation range,

$$P_\omega \sim \omega^s \exp\left(-\frac{\omega}{\omega_d}\right), \quad (5)$$

where ω_d is the characteristic frequency of the distribution. Results of numerical modeling for capillary waves [10] corroborated the exponential dependence of the wave distribution in the dissipative range. In our research, we therefore concentrated on a detailed study of the evolution of a turbulent cascade as the bandwidth $\Delta\omega$ was varied: from broadband noise pumping (kinetic regime) to narrow-band noise excitation, and further to monochromatic pumping (discrete mode).

In bounded geometries, the continuous spectrum of capillary waves becomes discrete, whereas the separation between the resonance modes increases with the frequency. In the case of monochromatic forcing of the liquid surface, the turbulent cascade consists of harmonics whose frequencies are multiples of the pumping frequency. Simple reasoning leads to the conclusion that for such surface excitation, system of equations (2) does not have any solutions [11, 12]. However, as shown in Ref. [13], this limitation is lifted if nonlinear broadening of resonance peaks is taken into account. In that case, the conservation laws have to be written as

$$\left||k_1|^{3/2} \pm |k_2|^{3/2} - |k_3|^{3/2}\right| < \delta, \qquad k_1 \pm k_2 - k_3 = 0, \quad (6)$$

where δ is the characteristic nonlinear broadening of the resonance peak. Additionally, it is necessary to take into account that for a classical liquid at high frequencies, the discrete spectrum becomes quasicontinuous because of viscous broadening of the resonance peaks. But in liquid hydrogen and, especially, helium, the kinematic viscosity coefficient is less than in water, by a factor of 10 for hydrogen and 100 for helium. Therefore, the quantization can play an important role in the energy transfer in the cascade [14] under monochromatic excitation. Our work [15] presents results obtained on the surface of superfluid helium under harmonic forcing when the discrete character of the system is essential (discrete turbulence). At the same time, for broadband pumping applied to surfaces of liquid hydrogen and helium, the case of kinetic turbulence is realized, which is closest to the model system worked out theoretically in Ref. [4].

2. Experimental procedure

In recent years, owing to advances in experimental technology and computer methods for processing rapidly varying signals, substantial progress has been achieved in exploring capillary turbulence on the surface of water [16–19], ethanol [20], silicon oil [21], and even mercury [22]. Our previous experiments [23] have shown that using liquid helium and hydrogen in studies of turbulence offers a number of advantages over the use of traditional media owing to the small density and low kinematic viscosity of helium and hydrogen.

In our research, we used a technique [24] based on measuring the power of a laser beam reflected from the oscillating surface of a liquid. Measurements were conducted in optical cells located in the vacuum cavity of a helium cryostat. A plane horizontal capacitor was installed inside the cells. Gaseous hydrogen or helium was condensed in a cylindrical copper cup. Its diameter was 60 mm in experiments with hydrogen and 30 mm in experiments with helium. The cup height was varied in the range 4–6 mm. Above the cup, the upper horizontal metallic plate of a capacitor was mounted, leaving a gap of 3.5 mm. The liquid was accumulated until it reached the edge of the cup. The temperature during measurement was $T = 15.5$ K in experiments with hydrogen and 1.7 K in experiments with helium.

The lower plate of the capacitor is fitted with a radioactive source emitting β-electrons with the mean energy 5 keV. Under the action of radiation, an ionized layer of liquid is formed in the vicinity of the source. A voltage of 1000 V applied to the capacitor plates drives positive ions out of the ionized layer toward the surface of the liquid. In this manner, the charged liquid surface and the upper metallic plate form a plane capacitor.

Waves on the charged fluid surface are excited by an alternating electric field generated by AC voltage with an amplitude of 1–100 V applied to the metallic cup, in addition to DC voltage. Pumping either is harmonic at frequencies close to the resonant frequencies of the cylindrical cell or represents broadband noise. The noise pumping signal was synthesized through the inverse Fourier transformation given the power spectrum and applying random phases. Using an electric field to excite the surface of the liquid offers a number of advantages. Indeed, it allows applying the force only to the surface and controlling the symmetry of forcing together with its spectral characteristics.

Changes in the power of the reflected laser beam were measured by a Hamamatsu s3590-08 semiconductor receiver. The AC output of the photoreceiver $P(t)$, proportional to the power of the reflected ray, was stored in computer memory at a sampling rate up to 100 kHz with the help of a fast 24-bit analog-to-digital converter (ADC). The recording time of the signal $P(t)$ was varied from 3 s to 100 s. The dependences $P(t)$ were processed with a fast Fourier transform (FFT) routine. As a result, we obtained the distribution of the squared amplitude of the harmonics over frequency P_ω^2, which for a

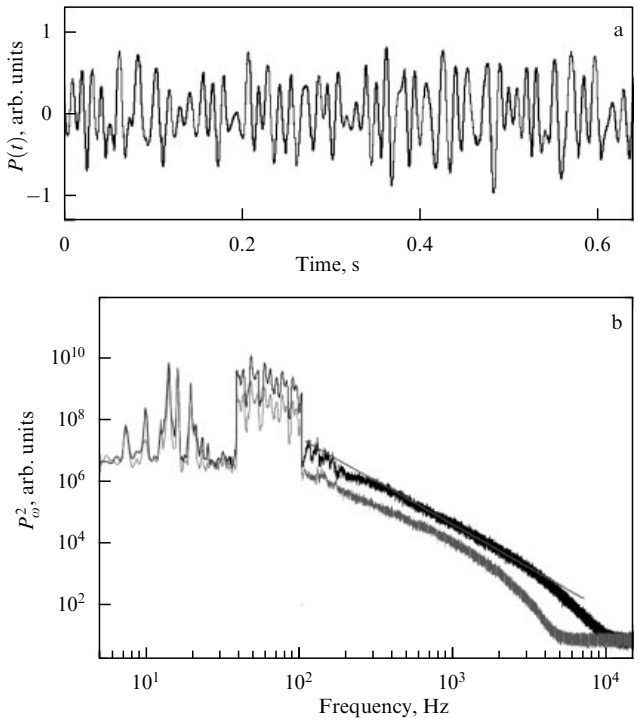

Figure 1. (a) Part of the experimental record of $P(t)$. Pumping is carried out with a random force in the frequency range 39–103 Hz. (b) Turbulent cascade on the surface of liquid hydrogen excited by a random force in the frequency range 39–103 Hz for two pumping levels. The straight line is the dependence $\omega^{-17/6}$.

broad beam, as shown in Ref. [24], is proportional to the pair correlation function for the surface deviation from equilibrium, $I_\omega \sim P_\omega^2$.

3. Modification of a turbulent cascade with a reduction in the forcing bandwidth

Figure 1a shows part of the signal $P(t)$ recorded from the surface of liquid hydrogen excited by noise. The bandwidth of the electric signal $V(t)$ applied to the guard ring was 64 Hz (from 39 Hz to 103 Hz), i.e., the surface was excited by broadband noise. The maximum amplitude of the noise forcing signal was 10 V, while maximum steepness of waves (angular amplitude) did not exceed 0.03 in the pumping range.

Figure 1b shows the distribution P_ω^2 (the dark curve) that corresponds to the signal in Fig. 1a. In the frequency range from 200 Hz to 8 kHz, a turbulent cascade formed whose frequency dependence was described by a power law with the exponent $m = 2.8 \pm 0.1$. For comparison, the straight line shows a function proportional to $\omega^{-17/6}$. The deviation from the power law at high frequencies (4–8 kHz) is a manifestation of the impact of viscous losses in the liquid on the turbulent distribution [4]. The dissipation range becomes well pronounced as the excitation force amplitude decreases. The light curve in Fig. 1b corresponds to the spectrum P_ω^2 with pumping in the same frequency range, but with the amplitude two and half times smaller. The high-frequency boundary of the inertial range decreased to 2.5 kHz. In the frequency range above 2.5 kHz, a sharp reduction in the oscillation amplitude is observed, which is characteristic of the spectrum in the dissipation range.

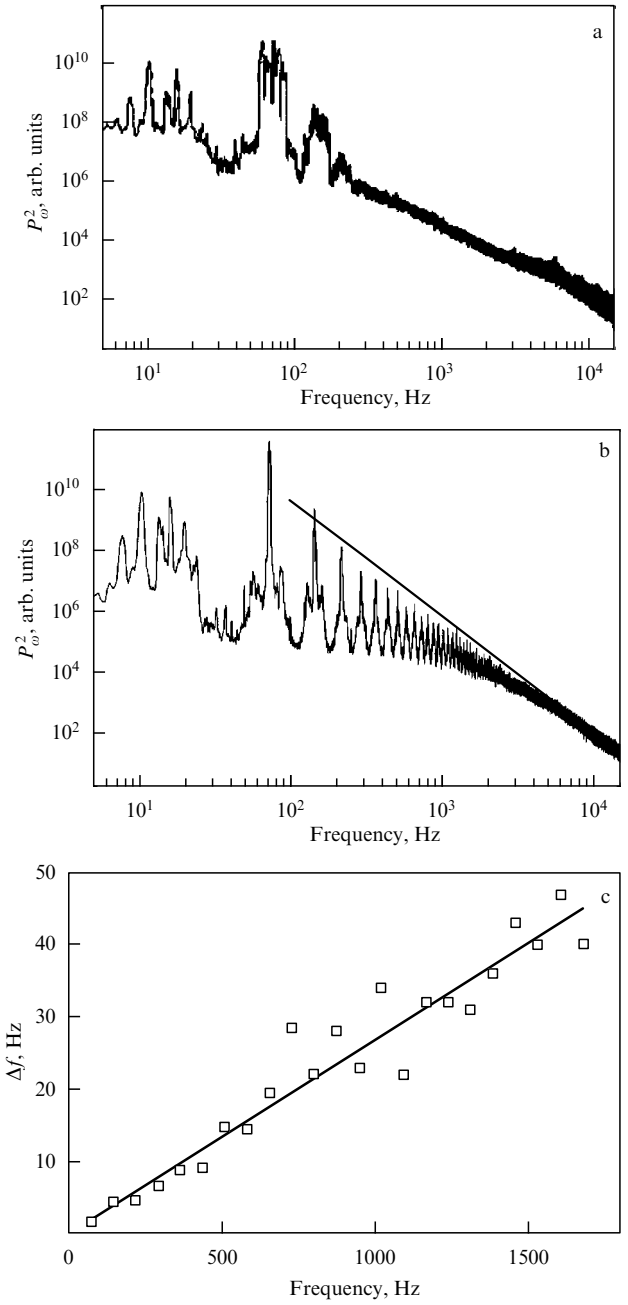

Figure 2. (a) The spectrum P_ω^2 of surface oscillations forced by noise in the frequency range 57–89 Hz. (b) The same as (a), but in the frequency range 71–75 Hz. The straight line corresponds to the dependence $\omega^{-23/6}$. (c) The dependence of the peak width Δf in the spectrum of panel (a) on frequency. The straight line is the linear law 0.027ω.

When the bandwidth of noise pumping was reduced relative to the mean pumping frequency such that $\Delta\omega \approx \omega/2$, several peaks emerged on the turbulent cascade. The distribution P_ω^2 for noise pumping in the frequency range 57–89 Hz is shown in Fig. 2a. The angular wave amplitude was 0.03, the same as in the previous case. The first peak lies in the pumping range. The second and third peaks correspond to waves appearing as a result of a nonlinear interaction. The separation between the centers of the peaks is approximately equal to the pumping frequency $\omega_p = 73$ Hz. It is clearly seen that a well-developed turbulent cascade formed in the frequency range 250 Hz–6 kHz. At frequencies above the

upper boundary of the inertial range $\omega_b = 6$ kHz, the decay of the cascade is observed, caused by viscous losses.

As the pumping bandwidth $\Delta\omega$ is reduced further, the peaks in the turbulent cascade become more expressed, and the minima go deeper. Figure 2b shows the distribution P_ω^2 when the surface is excited by noise in the frequency band of 4 Hz (from 71 Hz to 75 Hz). About 30 peaks are distinctly visible in the turbulent cascade. The separations between the peaks stays the same (73 Hz). The inertial range extends from 200 Hz to 15 kHz. The frequency dependence of peak maxima within the inertial range is close to the power law $\omega^{-3.8\pm0.1}$. We note that the difference in exponents in cases of narrow- and broadband pumping is 1.0 ± 0.2.

The peak widths Δf increase with frequency. The experimental dependence of peak widths on frequency is presented in Fig. 2c for noise pumping in the frequency range 71–75 Hz. Obviously, the increase in Δf with frequency is described by a linear function, and satisfactory agreement is observed between experimental data and the linear dependence predicted by the theory. The solid line in Fig. 2c corresponds to the linear function $\Delta f = 0.027\omega$. This means that the effective width of the pumping range in the capillary wave system is about 2 Hz, while the noise bandwidth in the electric signal applied to the guard ring is 4 Hz. This discrepancy comes from the discreteness of the capillary wave spectrum in the experimental cell and the finite width of resonance modes. The separation between two neighboring resonances at frequencies about 100 Hz amounts to ≈ 10 Hz for the peak width about 1 Hz. Thus, for a surface forced with noise in the frequency band of 4 Hz, there is such a position of the pumping interval relative to the resonance frequencies of the cell that only a single resonance harmonic can be excited. Arguably, this happens in the case displayed in Fig. 2b.

The linear dependence of the peak width on frequency can readily be explained [6]. Indeed, if nonlinear waves are excited in the range $\omega_p \pm \Delta\omega$, then, by virtue of the nonlinear interaction between them, waves appear in the frequency range $2\omega_p \pm 2\Delta\omega$, and so on. Hence, the linear dependence of the peak width on frequency must pass through the coordinate origin. Precisely these considerations have been used to draw the solid line in Fig. 2c.

The experimental results presented above showed that the change in the spectral characteristics of noise forcing (the bandwidth) leads to a qualitative modification of the turbulent cascade in the system of capillary waves on the surface of liquid hydrogen. For broadband pumping, the turbulent distribution is described well by a monotonically decaying power-law function with the exponent close to $m = 2.8 \pm 0.1$. By contrast, for narrowband forcing, a set of peaks appears in the turbulent cascade, their maxima following the power law with the exponent $m = 3.8 \pm 0.1$. Our results turn out to be in very good agreement with theory.

4. The decay of the turbulent cascade in the dissipation range

In the same fashion as in the experiments described in Section 3, capillary waves on the surface of liquid hydrogen were excited by a random force in the range 39–103 Hz. The mean amplitude of AC voltage (pumping) V_p, averaged over the frequency range, was varied from zero (in the absence of pumping) to the maximum amplitude $V_p = 30$ V, which was limited by the maximum angular wave amplitude allowed by the geometry of our low-temperature optical system.

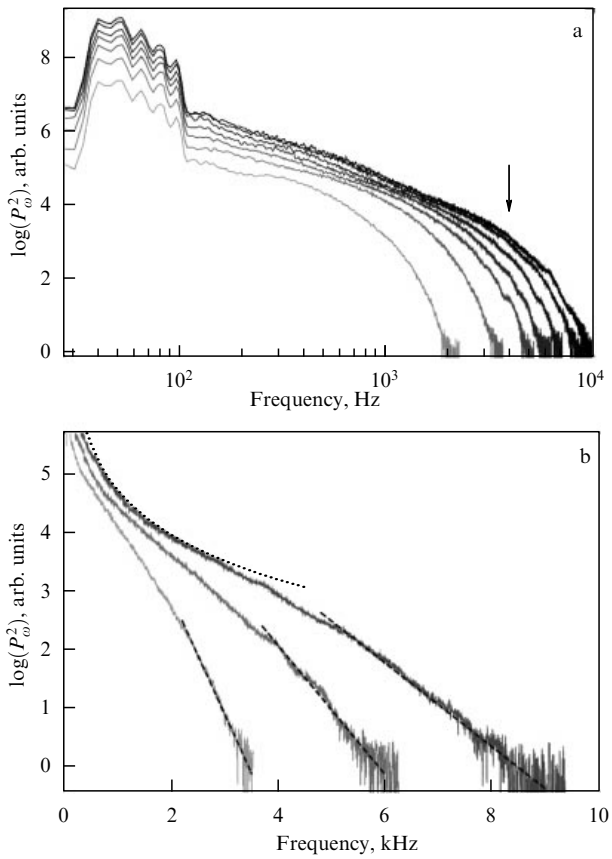

Figure 3. (a) The spectrum of surface oscillations P_ω^2 excited by a random force in the frequency range 39–103 Hz with different pumping amplitudes. The pumping amplitude is varied from 4 V (light curve) to 30 V (dark curve). The arrow marks the high-frequency boundary of the inertial range for maximum pumping. (b) The spectra P_ω^2 obtained for pumping with amplitudes $V_p = 8$, 16, and 26 V. The dotted line corresponds to the power law $\omega^{-2.8}$. The dashed lines represent the approximation by the function $\sim \exp(-\omega/\omega_d)$ with $\omega_d \approx 0.2$, 0.4, and 0.6 kHz for $V_p = 8$, 16, and 26 V, respectively.

Figure 3a shows the Fourier spectrum of the power of a reflected laser beam for various excitation amplitudes. The pumping range can well be discerned on the low-frequency side. It is followed by the inertial range within which the spectrum P_ω^2 can be described by power law (3). The width of the inertial range, as is clearly seen, depends on the pumping amplitude. When the surface is excited with a force that corresponds to the amplitude $V_p = 4$ V, the dissipative range begins immediately after the pumping range, while the inertial range is absent. The increase in the amplitude of the driving force expands the inertial range, and its high-frequency boundary ω_b shifts toward higher frequencies. The widest inertial range, from 0.3 to 4 kHz, is observed for the maximal pumping amplitude $V_p = 30$ V. At frequencies above the inertial range boundary, surface perturbations decay because of viscous losses, and the cascade smoothly fades out, disappearing in instrumental noise. The turbulent cascades recast in linear coordinates (Fig. 3b) show that the decay in the wave amplitude in the dissipative range can be approximated rather well by the exponential dependence $P_\omega^2 \sim \exp(-\omega/\omega_d)$. In the fitting, it is assumed that $\omega \gg \omega_d$, i.e., that the characteristic frequency (viscous boundary) ω_d is much lower than the frequency of waves from the dissipation range. For example, the spectrum obtained for pumping with the amplitude $V_p = 26$ V is approximated by the exponential

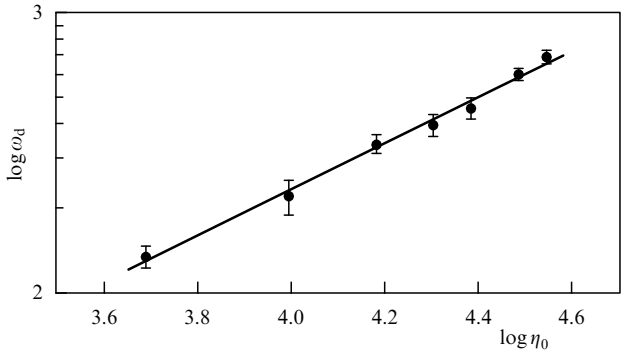

Figure 4. The dependence of the frequency ω_d on the amplitude of low-frequency pumping η_0; the straight line is the fit $\eta_0^{0.85}$.

with $\omega_d = 0.6$ kHz in the frequency range 5–9 kHz. Unfortunately, the fitting interval turns out to be insufficiently broad for reliably estimating the power of the pre-exponential factor in quasi-Planck distribution (5). The values of ω_d thus found are several times smaller than the visible boundaries separating the inertial range and dissipation range (Fig. 3b). This discrepancy can be explained by some arbitrariness in the definition of the high-frequency inertial range boundary, and hence the value of ω_b is only known by an order of magnitude. The characteristic frequency ω_d obtained by fitting an exponential to the spectra grows with the pumping amplitude in the dissipation range. To correctly plot the dependence of ω_d on the pumping amplitude, we used the surface response $\eta_0 = P$ at the frequency 53 Hz as a measure of the excitation level. The quantity η_0 is directly proportional to the mean wave amplitude at this frequency. Figure 4 shows the dependence of ω_d on the amplitude η_0. The experimental dependence is described by the power law $\omega_d(\eta_0) \sim \eta_0^n$ with the exponent $n = 0.85 \pm 0.05$.

It is noteworthy that the approximation of experimental spectra with quasi-Planck distribution with a small value of the exponent s (no higher than 2) does not significantly affect the value of ω_d (by less than 20%), nor does it influence the power-law exponent n in the amplitude dependence of the characteristic frequency $\omega_d(\eta_0)$.

The obtained value of the exponent $n = 0.85$ differs substantially from the value 12/5 expected from Eqn (4), which is surprising because for a turbulent cascade formed by a harmonic force, the measured amplitude dependence proves to be in good agreement with the theoretical estimate $\omega_d(\eta) \sim \eta^{1.3}$.

We note especially that the turbulent cascade in the dissipation range decays noticeably faster for monochromatic than for broadband pumping. Figure 5a shows the turbulent cascade at the surface of superfluid helium forced monochromatically at a frequency of 79.7 Hz. The turbulent cascade very closely resembles distributions observed previously in experiments on the surface of liquid hydrogen [7] and in our first studies on the superfluid helium surface [25]. The spectrum consist of equidistant harmonics: the first harmonic corresponds to pumping, while the others are generated as a result of the nonlinear interaction of waves with frequencies that are multiples of the pumping frequency. At frequencies greater than 4 kHz, the cascade decays extremely rapidly because of viscous losses and disappears, being buried in instrumental noise. The value of 4 kHz can be treated as the high-frequency boundary ω_b of the inertial range. The decay is described by an exponential dependence with the characteristic frequency $\omega_d = 170$ Hz close to the

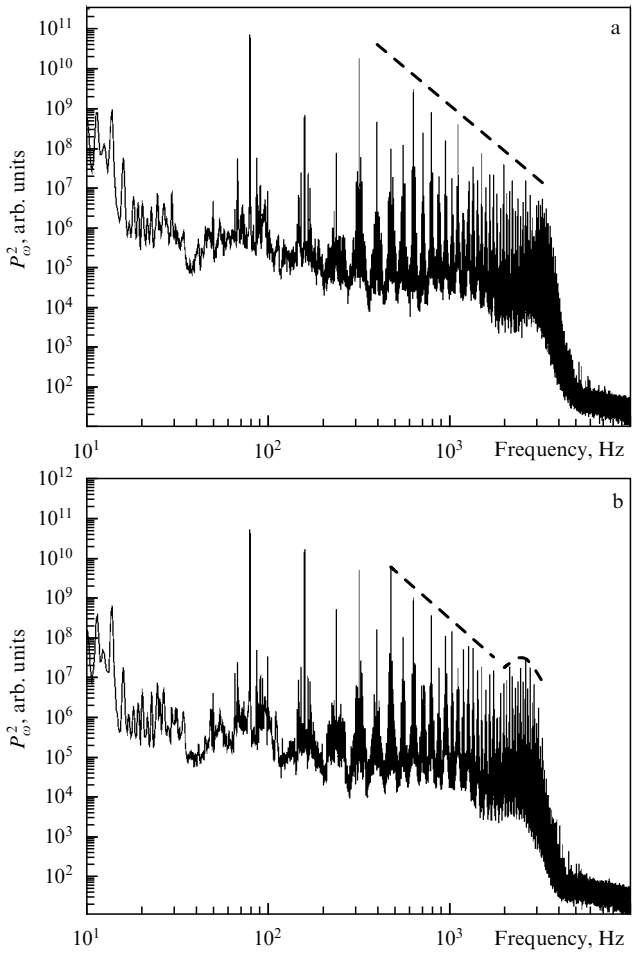

Figure 5. (a) The distribution P_ω^2 on the surface of superfluid helium at the temperature 1.7 K. The surface was excited by a sinusoidal force at a frequency of 79.7 Hz. The pumping amplitude was 11 V. The dashed line corresponds to the power law $P_\omega^2 \sim \omega^{-3.7}$ theoretically predicted for random narrowband pumping. (b) Turbulent spectrum as the amplitude of forcing is reduced to 10 V. A local maximum is observed near the frequency 2.5 kHz (marked by the dashed line).

pumping frequency ω_p [26]. We mention that for broadband forcing of the superfluid helium surface, the turbulent distribution in the dissipation range smoothly decays following an exponential law with the characteristic frequency ω_b that is close to the frequency of the high-frequency edge of the inertial range (Fig. 6b).

We can assume that the condition of locality for waves from the dissipation range is violated under harmonic excitation. For these waves, the interaction with waves from the inertial range proves to be dominating. Indeed, the nonlinear interaction time for three-wave processes with strongly different wave vectors ($k_1 \ll k_2$, $k_2 \approx k_3$) satisfies the relationship $\tau_{non} \sim k^{-1/2}$ [10], whereas for the local interaction ($k_1 \approx k_2 \approx k_3$) the time $\tau_{loc} \sim k^{1/4}$ [27]. Thus, the waves from the dissipation interval most efficiently interact with low-frequency waves from the inertial range, concentrated in the vicinity of the pumping range.

5. Discrete turbulence

In experimental studies of turbulent distributions on the surface of superfluid helium, we used two types of driving forces to generate surface waves. In experiments of the first type, the surface was excited by a sinusoidal force at one of the

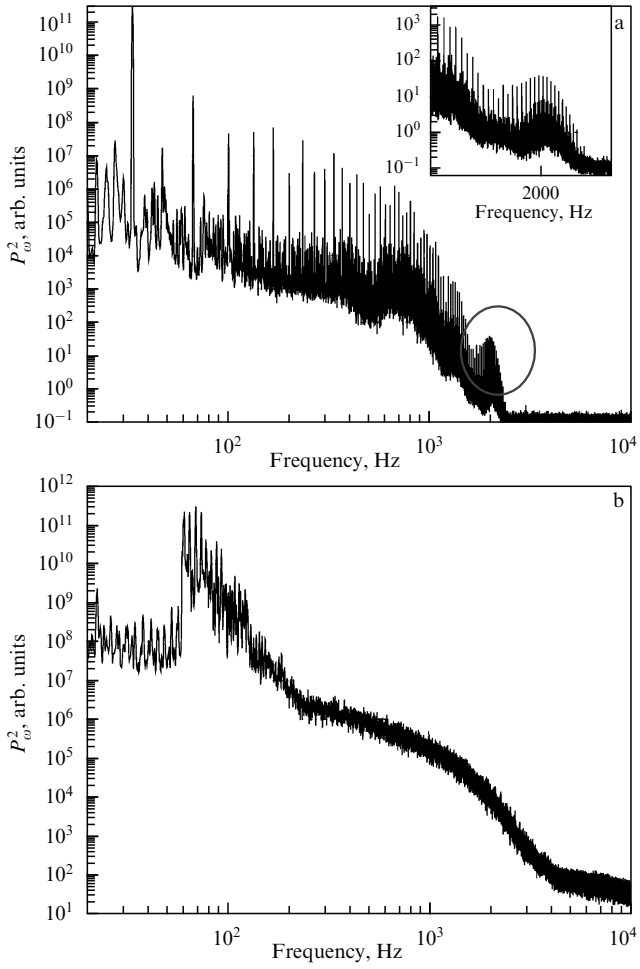

Figure 6. (a) Turbulent spectrum for pumping at the frequency $\omega/2\pi = 34$ Hz. The local maximum in the dissipation range is zoomed in in the inset. (b) The turbulent cascade of capillary waves pumped with noise in the frequency range 60–130 Hz.

cell resonance frequencies. In experiments of the second type, the surface was excited by noise in a bounded frequency range.

Figure 5a displays the spectrum P_ω^2 of capillary waves obtained in an experiment in which the superfluid helium surface was forced by an AC voltage at the frequency 79.7 Hz and amplitude 11 V. In the inertial range, the amplitudes of the harmonics decay with frequency according to the power law $P_\omega^2 \sim \omega^{-m}$ with $m \approx 3.7$, in agreement with the theoretical prediction for narrowband forcing [6].

If the forcing amplitude is slightly reduced, to 10 V, the shape of the turbulent cascade experiences qualitative changes. The high-frequency boundary of the inertial range shifts toward low frequencies, in agreement with the results of our previous experiments [7]. However, a new phenomenon is observed: the formation of a distinct local maximum close to the high-frequency boundary ω_b of the inertial range (marked by the dashed line in Fig. 5b). As the pumping amplitude is reduced further, the local maximum shifts to low frequencies. The spectrum consist of several harmonics for minimum pumping, and the local maximum is not observed.

We summarize common features of the results obtained in experiments. First, a local maximum is formed at high frequencies close to the end of the inertial range. Second, the shape and position of the maximum depend on the pumping frequency ω_p and the pumping wave amplitude. For example, Fig. 6a shows the spectrum P_ω^2 obtained for the surface of helium forced sinusoidally at a frequency of 34 Hz, with a well-pronounced maximum in the dissipation range, and not in the inertial range as in Fig. 5b.

When the surface is forced by noise, the local maximum is not observed. As an illustration, Fig. 6b presents a turbulent distribution for a helium surface forced with noise in the frequency range 60–130 Hz. The pumping amplitude was selected such that the end of the inertial range was located at the same frequencies as in Fig. 5b. It can be seen that this spectrum is qualitatively different from spectra obtained with monochromatic pumping: it is smooth and continuous.

The formation of a local maximum can be interpreted as energy accumulation in a narrow frequency interval near the end of the inertial range where the transition from nonlinear energy transfer to viscous decay occurs. A possible reason for this accumulation can be a bottleneck that impedes energy transfer to the dissipation range. It was shown in [10] that an insufficient rate of energy dissipation through viscous losses can in principle essentially modify the cascade shape at high frequencies in the inertial range. However, judging by the frequency and amplitude dependences of the maximum position, the formation of the cascade is not related to the trivial influence of viscosity. On the other hand, we are dealing with a weakly interacting nonlinear discrete wave system, and, as shown in Refs [11, 12], we can expect the discreteness of the system to affect wave interactions. In the later study [13], a model of frozen turbulence was proposed, and it was shown that the discreteness can result in an oscillating turbulent spectrum for surface waves in a square geometry, when the wave number space is two-dimensional. In our experiment, the geometry is circular, surface oscillations are described by the Bessel functions, and the wave number space is one-dimensional. Because the cell shape and size define the density of resonance modes, we made some estimates in order to understand the influence of the discrete character of spectra on the turbulent cascade in our experiments. We suppose that the main reason for the bottleneck and, consequently, the maximum formation is a detuning between frequencies of two discrete spectra, namely the spectrum of surface oscillations in a finite-size cell and the spectrum of turbulent cascade harmonics. If the surface is excited by a harmonic force, the frequency of the first peak in the turbulent cascade coincides with the frequency of resonance harmonics ω_p, which satisfies dispersion relation (1).

For surface waves in a cylindrical cell of diameter D, the resonance values of wave vectors satisfy the equation $J_1(kD/2) = 0$, where $J_1(x)$ is the Bessel function of the first order. For large magnitudes of wave vector k, the resonances becomes equidistant with the step $\Delta k \approx 2\pi/D$. Consequently, the distance between two nearest resonances in the frequency space increases with frequency:

$$\Delta\omega = \frac{\delta\omega}{\delta k}\,\Delta k \approx 2\pi\,\frac{\delta\omega}{\delta k}\,\frac{1}{D} = \frac{3\pi}{D}\left(\frac{\sigma}{\rho}\right)^{1/3}\omega^{1/3}. \qquad (7)$$

In other words, resonances in the case of capillary waves are not equidistant, in contrast to frequencies of harmonics in the turbulent cascade, which are multiples of the pumping frequency ω_p.

Obviously, frequency detuning can be essential only when the resonance broadening $\delta\omega$ is small compared to the

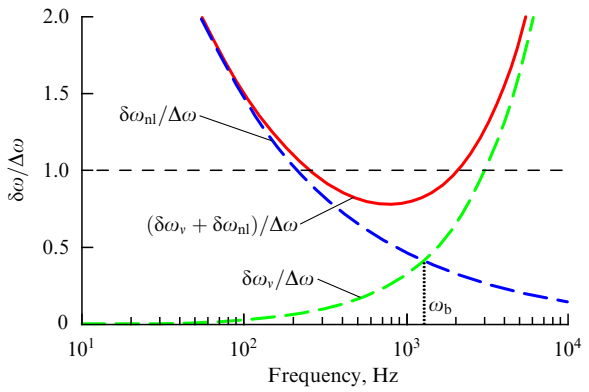

Figure 7. The estimate of the relative broadening $\delta\omega/\Delta\omega$ for harmonic pumping of moderate amplitude. The increasing dashed curve represents the relative viscous broadening $\delta\omega_v/\Delta\omega$, the decreasing dashed curve depicts the relative nonlinear broadening $\delta\omega_{\rm nl}/\Delta\omega$, and the solid curve with a minimum corresponds to the total relative broadening $\delta\omega/\Delta\omega$.

separation between resonances $\Delta\omega$ (Fig. 7),

$$\frac{\delta\omega}{\Delta\omega} < 1\,. \tag{8}$$

The resonance broadening $\delta\omega$ can be represented as a sum of viscous $\delta\omega_v$ and nonlinear $\delta\omega_{\rm nl}$ broadening:

$$\delta\omega = \delta\omega_v + \delta\omega_{\rm nl}\,. \tag{9}$$

Broadening of the resonance peak caused by viscous losses increases with the frequency,

$$\delta\omega_v = 4\nu\kappa_\omega^2 = 4\nu\left(\frac{\rho}{\sigma}\right)^{2/3}\omega^{4/3}\,, \tag{10}$$

while the characteristic time of viscous damping decreases, $\delta\omega_v = \tau_v^{-1}$.

Broadening due to the nonlinear energy transfer in the turbulent cascade can be estimated from the characteristic time of nonlinear interaction $\tau_{\rm nl}$ as $\delta\omega_{\rm nl} = \tau_{\rm nl}^{-1}$. In the case of harmonic pumping, we assume that $\tau_{\rm nl} \sim \omega^{1/6}$ [7], whence $\delta\omega_{\rm nl} \sim \omega^{-1/6}$. As the driving force amplitude A increases, the nonlinearity of waves increases, if the frequency is kept fixed. We can therefore write

$$\delta\omega_{\rm nl} \approx \varepsilon(A)\,\omega^{-1/6}, \tag{11}$$

where $\varepsilon(A)$ is an increasing function of the wave amplitude A at the pumping frequency. Substituting Eqns (7) and (9)–(11) in Eqn (8), we obtain the condition for the detuning to be important:

$$\frac{4\nu(\rho/\sigma)^{2/3}\omega^{4/3} + \varepsilon(A)\,\omega^{-1/6}}{(3\pi/D)(\sigma/\rho)^{1/3}\omega^{1/3}} < 1\,.$$

On the high-frequency boundary $\omega_{\rm b}$ of the inertial range, the nonlinear energy transfer along the spectrum gives way to viscous dissipation. Therefore, as noted above, it is assumed that at the frequency $\omega_{\rm b}$, the characteristic time of viscous damping is close to that of nonlinear interaction, $\tau_v(\omega_{\rm b}) \approx \tau_{\rm nl}(\omega_{\rm b})$. Although the precise form of the function $\varepsilon(A)$ is unknown, we can conclude that the resonance peak broadenings induced by the nonlinear interaction and by viscous losses are also close to each other at the frequency $\omega_{\rm b}$, $\delta\omega_v(\omega_{\rm b}) \approx \delta\omega_{\rm nl}(\omega_{\rm b})$.

It follows from estimates that in our experiments, the total relative broadening of the resonance peak satisfies condition (8) in a finite frequency interval located near the high-frequency boundary $\omega_{\rm b}$ (see Fig. 7). In this interval, the frequency detuning between harmonics in the turbulent cascade and resonance peaks becomes essential, and the discrete regime of capillary turbulence is realized. Following the logic of Ref. [10], we suppose that the energy flux bottleneck forms in that region, which determines the specific shape of the distribution P_ω^2. At high pumping amplitudes (or at high frequencies), the relative broadening exceeds unity, the system becomes quasicontinuous, and the kinetic regime of turbulence is realized. However, based on this simple assumption, we cannot compute the exact position of the local maximum or its form. Rigorous theoretical analysis and numerical simulations are needed.

The proposed model presumes that the following conditions, needed for the energy to accumulate in a system of capillary waves, are satisfied: insignificant viscous broadening of the resonance peak, not very strong nonlinear broadening (moderate pumping amplitudes), and large separation between neighboring resonance frequencies (a relatively small cell size). We stress that it is the use of superfluid helium with extremely low viscosity [28] that enabled us to observe the energy condensation in the turbulent cascade.

6. Conclusions

Passing from broadband to narrowband and further to harmonic pumping in experiments on the formation of a turbulent state in a system of capillary waves leads to a qualitative change in the turbulent distribution: a set of peaks evolves in the cascade modifying the frequency dependence of the correlation function. A quasi-Planck distribution of waves over frequency is formed in the dissipation range with the characteristic frequency defined by spectral characteristics of pumping.

Using superfluid helium with an extremely low value of kinematic viscosity enabled observing the discrete regime of capillary turbulence. The influence of the discreteness of the surface oscillation spectrum on the turbulent distribution is manifested in the formation of a local maximum near the end of the inertial range — the energy condensation in a narrow frequency range. The fundamental cause of this phenomenon is the frequency detuning between the harmonics in the turbulent cascade and resonance modes of the cylindrical resonator, and also in the formation of a bottleneck for the energy flux toward higher frequencies.

The authors are grateful to L P Mezhov-Deglin, E A Kuznetsov, and G V Kolmakov for the useful discussions. The research was supported in part by the RFBR grant 11-02-12147.

References

1. Musher S L, Rubenchik A M, Zakharov V E *Phys. Rep.* **252** 177 (1995)
2. L'vov V S *Wave Turbulence under Parametric Excitation: Applications to Magnets* (Berlin: Springer-Verlag, 1994)
3. Zakharov V E, Filonenko N N *Zh. Priklad. Mekh. Tekh. Fiz.* **8** 37 (1967) [*J. Appl. Mech. Tech. Phys.* **8** 62 (1967)]
4. Zakharov V E, L'vov V S, Falkovich G *Kolmogorov Spectra of Turbulence* Vol. 1 (Berlin: Springer-Verlag, 1992)
5. Pushkarev A N, Zakharov V E *Phys. Rev. Lett.* **76** 3320 (1996)

6. Fal'kovich G E, Shafarenko A B *Zh. Eksp. Teor. Fiz.* **94** 172 (1988) [*Sov. Phys. JETP* **67** 1393 (1988)]
7. Brazhnikov M Yu, Kolmakov G V, Levchenko A A *Zh. Eksp. Teor. Fiz.* **122** 521 (2002) [*JETP* **95** 447 (2002)]
8. Abdurakhimov L V, Brazhnikov M Yu, Levchenko A A *Pis'ma Zh. Eksp. Teor. Fiz.* **89** 139 (2009) [*JETP Lett.* **89** 120 (2009)]
9. Malkin V M *Zh. Eksp. Teor. Fiz.* **86** 1263 (1984) [*Sov. Phys. JETP* **59** 737 (1984)]
10. Ryzhenkova I V, Fal'kovich G E *Zh. Eksp. Teor. Fiz.* **98** 1931 (1990) [*Sov. Phys. JETP* **71** 1085 (1990)]
11. Kartashova E A *Physica D* **46** 43 (1990)
12. Kartashova E A *Physica D* **54** 125 (1991)
13. Pushkarev A N, Zakharov V E *Physica D* **135** 98 (2000)
14. Zakharov V E et al. *Pis'ma Zh. Eksp. Teor. Fiz.* **82** 544 (2005) [*JETP Lett.* **82** 487 (2005)]
15. Abdurakhimov L V, Brazhnikov M Yu, Remizov I A, Levchenko A A *Pis'ma Zh. Eksp. Teor. Fiz.* **91** 291 (2010) [*JETP Lett.* **9**1 271 (2010)]
16. Wright W B, Budakian R, Putterman S J *Phys. Rev. Lett.* **76** 4528 (1996)
17. Henry E, Alstrøm P, Levinsen M T *Europhys. Lett.* **52** 27 (2000)
18. Punzmann H, Shats M G, Xia H *Phys. Rev. Lett.* **103** 064502 (2009)
19. Brazhnikov M Yu, Kolmakov G V, Levchenko A A, Mezhov-Deglin L P *Europhys. Lett.* **58** 510 (2002)
20. Falcón C et al. *Europhys. Lett.* **86** 14002 (2009)
21. Westra M-T *Patterns and Weak Turbulence in Surface Waves* (Eindhoven: Technische Univ. Eindhoven, 2001)
22. Falcon É, Laroche C, Fauve S *Phys. Rev. Lett.* **98** 094503 (2007)
23. Kolmakov G V et al. *Prog. Low Temp. Phys.* **16** 305 (2009)
24. Brazhnikov M Yu, Levchenko A A, Mezhov-Deglin L P *Prib. Tekh. Eksp.* (6) 31 (2002) [*Instrum. Exp. Tech.* **45** 758 (2002)]
25. Abdurakhimov L V, Brazhnikov M Yu, Levchenko A A *Fiz. Nizk. Temp.* **35** 127 (2009) [*Low Temp. Phys.* **35** 95 (2009)]
26. Abdurakhimov L V, Brazhnikov M Yu, Levchenko A A *J. Phys. Conf. Ser.* (2012), accepted
27. Kolmakov G V *Pis'ma Zh. Eksp. Teor. Fiz.* **83** 64 (2006) [*JETP Lett.* **83** 58 (2006)]
28. Donnelly R J, Barenghi C F *J. Phys. Chem. Ref. Data* **27** 1217 (1998)

PACS numbers: **95.75.–z**, **96.50.–e**, **96.60.–j**
DOI: 10.3367/UFNe.0182.201208j.0887

Motion of the Sun through the interstellar medium

V G Kurt, E N Mironova

1. Introduction

The motion of the Sun includes many components with different velocities, directions, and reference systems. For example, the Sun moves in a complicated open trajectory around the solar system barycenter. The maximum amplitude of this motion sometimes exceeds the radius of the Sun itself. This excursion is due to the motion of the most massive planets in the Solar System, Jupiter and Saturn, with the respective orbital periods 11.859 and 29.428 years. The Sun also moves relative to the 100 nearest stars in the direction toward the Hercules constellation with a velocity of 19.2 km s^{-1}. This motion was discovered by W Herschel (1738–1822) at the end of the 18th century based on the analysis of proper motions of the brightest (and correspondingly closest) stars. Naturally, Herschel could not express the value of this velocity in units such as km s^{-1} because he did not know the distance to these stars. Distances to stars were measured only in the early 1830s almost simultaneously by V Ya Struve (1793–1864) (Russia), F Bessel (1784–1846) (Prussia), and T Henderson (1798–1844) (England) using annual parallaxes of stars, which amount to only fractions of an arc second, even for the nearest stars. The direction to the apex of this motion is $\alpha = 270^\circ$ and $\delta = 30^\circ$.

The Sun also participates in an almost circular orbital motion around the galactic center with a velocity of 220 km s^{-1} in the direction perpendicular to the direction to the galactic center. With the distance 7.9 kpc to the galactic center, the orbital period of this motion is about 200 mln years, and during its life (5 billion years), the Sun has already made about 25 revolutions around the galactic center.

The Sun, together with the Galaxy, also has a peculiar velocity relative to nearby galaxies of the Local Group of galaxies. For example, the Galaxy approaches the center of the Andromeda nebula (M31) with a velocity of 290 km s^{-1} relative to the Solar System barycenter.

Finally, the Sun, together with the Milky Way and the Local Group, moves relative to the isotropic 3 K cosmic microwave background with the velocity (667 ± 22) km s^{-1} in the direction $l = 276^\circ \pm 3^\circ$ and $b = 30^\circ \pm 3^\circ$ (galactic coordinates). In a certain sense, this reference frame is a peculiar, singular coordinate system. Just this motion is responsible for the presence of the dipole component in the decomposition of cosmic microwave background in spherical functions. The amplitude of the dipole component is 6.706 mK.

This paper is focused on the study of the motion of the Sun relative to the local interstellar medium (LISM) on scales smaller than one or several parsecs but larger than 1000 astronomical units (a.u.).

2. Brief history of the discovery of the motion of the Sun relative to the local interstellar medium

In 1959, a group of astronomers from the Naval Research Laboratory (NRL) in the USA headed by G Fridman discovered a bright UV glowing of the sky from the rocket Aeroby-χ, which was capable of reaching only a 140 km altitude. The glow was measured in the atomic hydrogen line L_α ($\lambda = 1215.7$ Å) with the intensity reaching 20 kR (1 Rayleigh (R) = 10^6 photons cm^{-2} s^{-1} $(4\pi$ sr$)^{-1}$). The minimum of this glow was found to come from the anti-solar direction, and its intensity at a distance exceeding 90 km from Earth slowly decreased with increasing height [1]. In the same year, using a cell filled with molecular hydrogen supplied with a filament for its dissociation, which provided a sufficient amount of neutral hydrogen atoms for L_α line absorption, Morton determined that 7% of the discovered UV emission has a temperature exceeding 7000 K [2]. This could be explained by the presence of both a hot atomic hydrogen component in the upper atmosphere of Earth and an extra-atmosphere 'hot' emission component.

At almost the same time, starting in 1961, a research program of the Moon, Venus, and Mars explorations using automatic interplanetary stations (AISs) started in this country.

V G Kurt, E N Mironova Astro-Space Center, Lebedev Physics Institute, Russian Academy of Sciences, Moscow, Russian Federation
E-mail: vkurt@asc.rssi.ru

Uspekhi Fizicheskikh Nauk **182** (8) 887–894 (2012)
DOI: 10.3367/UFNr.0182.201208j.0887
Translated by K A Postnov; edited by A M Semikhatov

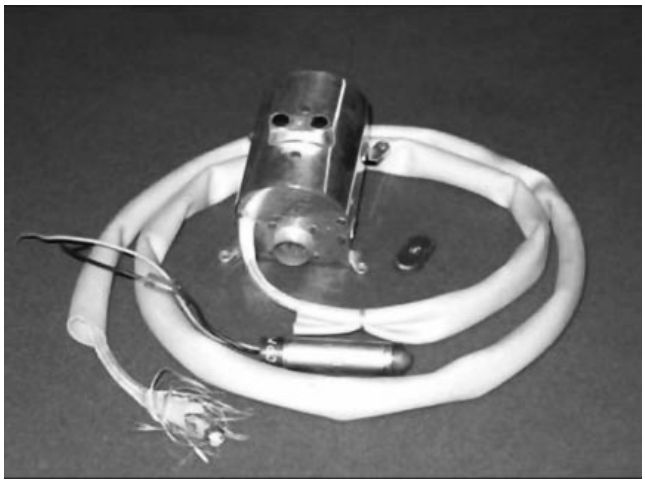

Figure 1. LA-2 two-channel ultraviolet (UV) photometer: the detector unit and Geiger counter.

In 1961, our group from the Sternberg Astronomical Institute of Moscow State University carried out similar studies of the upper atmosphere glow at altitudes up to 500 km using the R5 geophysical rocket. For this purpose, we constructed the two-channel ultraviolet photometer, LA-2 (Fig. 1), for observations in both the hydrogen L_α line and atomic oxygen triplet $\lambda = 1302$, 1304, and 1305 Å. A specially designed Geiger counter of photons filled with NO and an LiF or CaF_2 window with a diameter of 2 mm was used. From the shortwave side, the sensitivity of the detector was limited by the window transparency at $\lambda = 1050$ Å and $\lambda = 1225$ Å for LiF and CaF_2 respectively, while from the longwave side, the sensitivity of these detectors was limited by the ionization potential 9.3 eV ($\lambda = 1340$ Å) of nitrogen oxide. The first detector measured the total intensity in the L_α line and in the atomic oxygen triplet lines, while the detector with the CaF_2 window measured only the intensity of the atomic oxygen triplet lines.

Using the opportunity to install our detectors aboard the AIS launched to Venus and Mars, we were able to measure the extended hydrogen corona of Earth up to distances of 125,000 km, i.e., 20 Earth radii [3, 4]. To our surprise, at large distances (up to 100 mln km from Earth), the intensity did not vanish, but stayed almost constant at the level of 300–500 R, which, with the sensitivity of our detectors, corresponded to about 1000 counts per second. The counter with the calcium fluoride window allowed measuring the cosmic ray background, which was about 20–30 counts per second. Using the charged particle background suppression system enabled us to decrease this background rate to 2–3 counts per second. Theoretical estimates for the dependence of the interplanetary atomic hydrogen density on the solar wind particle flux and atomic hydrogen dissipation from the upper atmosphere of large planets could not explain such a high intensity of the neutral hydrogen atomic density in the interplanetary space. Nor could a detailed calculation of this emission intensity from even the nearest stars explain the observed effect, because the absorption by the interstellar dust and atoms with a low ionization potential (alkali metals) totally excluded the L_α line emission with such a high intensity. Scattering of solar L_α emission by the interstellar neutral hydrogen near the Sun was the only explanation for the effect observed. However, it was clear that an H II zone, similar to Stroemgren zones (a cavity filled with fully ionized hydrogen) found around hot early type O, B, and A giants, must exist around the Sun.

Calculations [5] showed that the observed intensity can be provided by the formation of an empty spherical region (cavern) around the Sun with a radius of 100 a.u. in an infinite medium (the Milne problem) filled with atomic hydrogen with a density of 0.01 cm^{-3}.

Indeed, exactly such zones of fully ionized hydrogen exist around hot blue early type (O, B, A) giants. The size of such a stationary Stroemgren zone is determined by the photon flux emitted by the star with $\lambda < 912$ Å and by the electron number density n_e of the medium. Equating the number of photons emitted by the star per second to the number of recombinations at all levels except the first one in the total volume of the Stroemgren sphere (the H II zone), we can calculate the Stroemgrem sphere radius

$$R_S = n_e^{-2/3}\left[\frac{4\pi F_*(L_c)(3/4)\,\pi}{\alpha_\Sigma - \alpha_1}\right]^{1/3},$$

where $4\pi F_*(L_c)$ is the total flux of quanta with the energy $h\nu > 13.56$ eV emitted by the star (L_c is the notation for the Lyman continuum), α_Σ is the recombination coefficient at all levels, and α_1 is the recombination coefficient at the first level.

The mass of gas confined inside the Stroemgren sphere exceeds that of the star itself by 100 or even 1000 times, and the radius of the Stroemgren sphere reaches several parsecs.

The boundary between the H II zone and the neutral-hydrogen zone H I is very thin (~ 0.05 pc), its width is $\Delta R_S = 1/n_H\sigma_i$, where n_H is the neutral hydrogen number density in the interstellar medium, of the order of 1 cm^{-3} and σ_i is the ionization cross section of atomic hydrogen near the ionization threshold, which is about 8×10^{-18} cm^2.

However, it is clear that the Stromgren theory is totally inapplicable to cold solar-type dwarf stars. Indeed, the time of ionization of a hydrogen atom at the distance 1 a.u. from the Sun is about 3×10^6 s, while the recombination time is 100–1000 times as long. Clearly, even with the velocity 10–30 km s^{-1} relative to the interstellar medium, the Sun 'flies' a distance of about 10^4 a.u. over the recombination time, forming a recombination tail 0.1 pc or even longer in size. This fact was first noted by theorists Blum and Far [6] from Bonn University.

Ionization of neutral hydrogen atoms passing by the Sun is in turn determined by two mechanisms: photoionization by hard emission from the solar corona with $h\nu > 13.56$ eV and charge exchange reactions with solar-wind protons. The effective cross section of the resonance charge exchange process of neutral hydrogen atoms with solar-wind protons is very high: at the maximum (close to 15 eV), it reaches 1.5×10^{-15} cm^2, and this process is 2–3 times more effective than photoionization.

Interstellar neutral hydrogen atoms, obviously, move in a hyperbolic trajectory, which is determined by the impact parameter P relative to the velocity at infinity V_∞. In polar coordinates, the trajectory is written as

$$\frac{1}{r} = \frac{GM_{\rm eff}(1+\cos\theta)}{V_\infty^2 P^2} + \frac{\sin\theta}{P},$$

where $M_{\rm eff} = M_\odot(1-\mu)$ and G is the gravitational constant. The factor $1-\mu$ takes the radiation pressure acting on a

hydrogen atom into account, which, as gravitation, is inversely proportional to the square of the distance from the Sun. This allows introducing the effective mass $M_{\rm eff}$ instead of the mass of the Sun, where μ is the ratio of the radiation pressure in the L_α line to the gravitational attraction force. At $\mu = 1$, hydrogen atoms move in straight trajectories. In this case, only ionization is effective.

The critical intensity in the L_α line center at which $\mu = 1$ is equal to 3.32×10^{11} photons per cm^2 per s. Such a high value of the intensity is hardly possible, even at the solar maximum activity. Nevertheless, the linear trajectory model is very convenient for approximate analytic calculations of the intensity of the observed scattered L_α radiation.

Next, to calculate the volume emissivity of the interstellar medium in the L_α line, we should include the Doppler shift from the projection of the hydrogen atom velocity in a hyperbolic orbit on the radius vector (centered in the Sun) of a point in the interplanetary space. Here, it is also necessary to take the complex profile of the L_α solar emission line with exponential wings into account. The total width of the L_α emission is about 1 Å, while the Doppler shift at the velocity about 30 km s^{-1} can reach 0.1 Å at a distance of the order of 1 a.u., which is quite substantial.

The probability that an atom flies without being ionized up to a point with polar coordinates r, θ is

$$p = \exp\left(-\frac{r_0^2\theta}{V_\infty \tau P}\right),$$

where $r_0 = 1$ a.u., V_∞ is the atom velocity 'at infinity', and τ is the lifetime, $1/\tau = 1/\tau_1 + 1/\tau_2$, where τ_1 and τ_2 are respectively the atom lifetimes at the 1 a.u. distance due to photoionization and charge exchange reactions with solar-wind protons.

Clearly, at each point of space, there are two types of atoms with different impact parameters P with the densities

$$n_1 = n_\infty \frac{(\sqrt{A}+1)^2}{4\sqrt{A}},$$

$$n_2 = n_\infty \frac{(\sqrt{A}-1)^2}{4\sqrt{A}},$$

where $A = 1 + 4M_{\rm eff}\,G/[V_\infty\, r\,(1+\cos\theta)\,]$.

It is clear that for $\mu = 1$, i.e., when gravitational attraction is fully compensated by radiation pressure acting on the hydrogen atom, $n_2 = 0$, which corresponds to the 'linear trajectory' model.

The volume emissivity $j(r,\lambda)$ [cm^{-3} 4π sr] can be written as

$$j(r,\lambda) = n(t,|\dot r|)\,\pi\Phi_{\rm S}(\lambda)\,p(\theta)\left(\frac{r_0}{r}\right)^2\sigma,$$

where $n(r,|\dot r|)$ is the atomic hydrogen density as a function of coordinates and the radial velocity, $\pi\Phi_{\rm S}(\lambda)$ is the solar radiation flux as a function of wavelength, and

$$p(\theta) = \frac{1}{16\pi}\left(\frac{11}{3} + \cos^2\theta\right)$$

is the scattering diagram.

To obtain the observed intensity, the volume emissivity should be integrated along the line of sight and over wavelengths within the solar L_α line width.

Such a primitive model assumes, clearly, that velocities of all atoms 'at infinity' are equal to V_∞ and are parallel to each other, which corresponds to zero temperature of the local interstellar medium. This model, called the 'cold model', can be applied with sufficient accuracy only in the region ahead of the solar motion direction in the LISM within the angle $\theta_{\rm cr} = \arctan(V_{\rm T}/V_\infty)$, which approximately corresponds to $\theta = 27°$, where $V_{\rm T}$ is the thermal velocity of atoms, which is equal to $\sqrt{2kT/m_{\rm p}}$. For larger angles, the Maxwellian distribution of hydrogen atoms over both velocity modulus and direction must be taken into account. For this, we should integrate 'cold models' with a weight proportional to the Maxwellian velocity distribution and homogeneous distribution over angles.

Therefore, to tune the 'cold model' to the observational data, the following parameters of the Sun motion in the LISM should be determined:

V_∞, the absolute value of the Sun's velocity;

λ and β, the two angles characterizing the direction of the Sun's motion in the ecliptic coordinates, which can then be easily transformed into the galactic coordinates;

n_∞, the hydrogen atom number density at infinity;

μ, the ratio of the radiation pressure force to the gravitational attraction force;

$\pi\Phi(0)$, the intensity in the solar emission L_α line center;

T_∞, the temperature of hydrogen atoms at infinity.

Hence, the minimum number of sought parameters is 7. However, the agreement between theory and observations can be improved by introducing additional parameters, for example, the solar proton flux anisotropy as a function of the heliographic altitude, because the charge exchange process dominates in the ionization of hydrogen atoms.

Figure 2 shows the distribution of the L_α volume emissivity inside the Solar System.

In [7], using the 'hot model', an analytic expression for the helium atom density along the Sun's direction of motion was also obtained, which allows 'sewing' the data on the helium atom number density ahead of the Sun's motion to that in other directions and thus obtaining the complete picture of

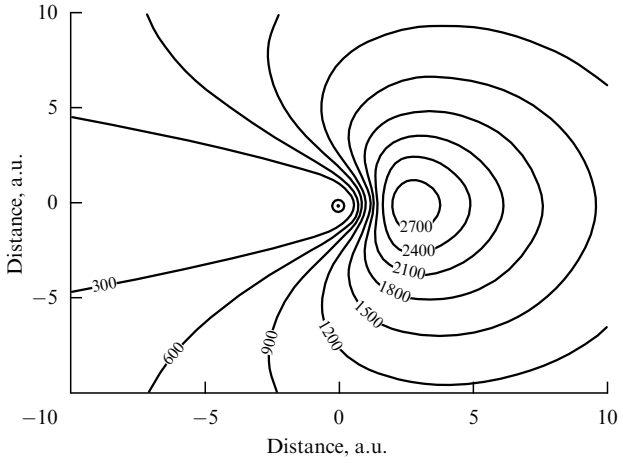

Figure 2. The volume L_α-line luminosity distribution [cm^{-3} s^{-1}] for the model parameters $V_\infty = 20$ km s^{-1}, $\tau = 1.2 \times 10^6$ s, and $\mu = 1$. The plane of the figure represents one of the cuts passing through the direction of the wind and the Sun. The upwind direction is to the right. The region with the maximum volume luminosity is located between 2 a.u. and 4 a.u. in the upwind direction.

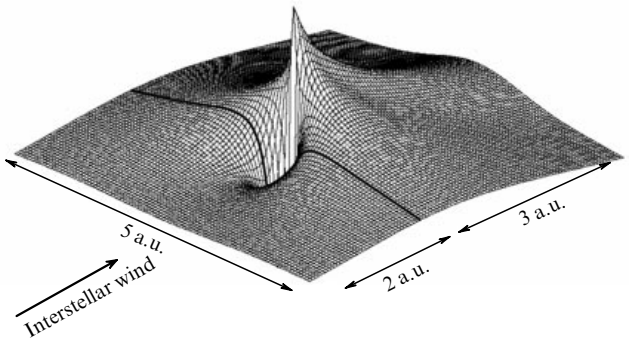

Figure 3. The volume luminosity in the He I $\lambda = 584$ Å line in the plane passing through the Sun and the wind axis (for the model parameters $V_\infty = 25.5$ km s^{-1}, $T = 7000$ K, and $\tau = 1.0 \times 10^7$ s, the Doppler width of the solar $\lambda = 584$ Å line is $W_D = 35.5$ km s^{-1}). Solar parameters are assumed to be isotropic.

the atom number density inside the Solar System. Figure 3 shows the atomic helium density at distances up to 10 a.u. [8]. The sharp minimum (about 0.3 a.u.) is due to photoionization of helium atoms near the Sun.

In addition, it is necessary, although it is unclear how, to account for the interaction of hydrogen atoms with the high-temperature plasma between two shocks (probably supersonic) that were formed by solar wind stopping at a distance of about 100 a.u. and by interstellar medium stopping at a distance of 200 a.u. Clearly, this effect is not dominant, because the effective conventional 'optical depth' for number densities about 10^{-3} or 10^{-4} cm^{-3}, the cross section less than 10^{-15} cm^2, and the length of the intermediate zone between the LISM and the heliosphere about 10^{15} cm does not exceed 10^{-3} or 10^{-2}. Nevertheless, this question remains open (see [9, 10]) and its solution can possibly slightly change the LISM parameters discussed below.

An important question is the determination of the profile of scattered radiation in the L_α line, which would allow an independent measurement of the incident interstellar matter flow. However, to solve this problem, an optical dispersion or interferometric spectrometer is needed, which is not an easy task for the radiation intensity of only 500 R at the maximum and the line width less than 0.1 Å. To solve this problem, very much as R Wood did in his experiments with sodium vapors, we used an absorbing cell filled with atomic hydrogen, similar to what was used by Morton in the first rocket observations of the L_α-emission.

Clearly, the observed bolometric intensity after passing through the cell filled with atomic hydrogen can be written as

$$I(\tau_0, T_E, T_A) = I(0)\,\frac{1}{\pi}\int_0^\infty \exp\left\{-x^2 - \tau_0 \exp\left[-\frac{T_E}{T_A}\left(x - \frac{V_r}{c}\right)\right]^2\right\} dx,$$

where τ_0 is the optical depth of the cell in the L_α line center, T_E and T_A are the respective radiation temperatures of the LIMS and atomic hydrogen in the cell, V_r is the radial velocity component of a hydrogen atom, and c is the speed of light in the vacuum. As the first approximation, we used the Doppler profile of the LISM emission line. The cell temperature (close to 300 K) was measured by a temperature probe on the cell surface. For the ratio $T_E/T_A \gg 1$ and $\tau_0 = 10$, the approximate expression:

$$RF = \frac{I}{I_0} = 1 - \frac{32.2}{\sqrt{T_E}} \exp\left(-\frac{60.6 V_r^2}{T_E}\right),$$

is valid for $T_E = 7000$ K with an accuracy better than 5%.

A more precise expression, taking both the L_α- absorption and scattering inside the cell into account, was obtained in [11].

3. Observations and data processing

Observations of the intensity of scattered L_α-radiation by a diffraction spectrometer and photometers both with and without an absorbing cell filled with hydrogen were carried out by us from the Venera and Mars IASs. However, the best observations were apparently obtained in the joint Soviet–French experiments aboard the Prognoz-5 and Prognoz-6 satellites (Fig. 4). Both satellites moved in highly eccentric orbits with an apogee of about 200,000 km, a perigee of about 1000 km, and an orbital period of 4 days. The principal inertia moment axes of the satellites were pointed toward the Sun, and the satellites rotated around them with a spin period of 2 min, corresponding to the angular velocity 3 deg s^{-1}. During the axial rotation, special optical trackers periodically registered Earth, the Moon, and the Sun passing across their field of view, which allowed calculating the Euler angles and extrapolating the orientation of the satellites for the subsequent 12 h interval of time. Then, new rotation parameters were determined. The reorientation of the main inertia axis to the Sun was repeated every 4–10 days. The correctness of the orientation was checked by matching the results at the beginning of the next 12 h interval. The system of orientation of the apparatus axes was controlled by observing hot bright stars with a UV photometer. Using this method, the orientation of the optical axis of the photometer was determined with an accuracy of 1–2 deg, and sometimes better. The photometers had two optical axes; the main one

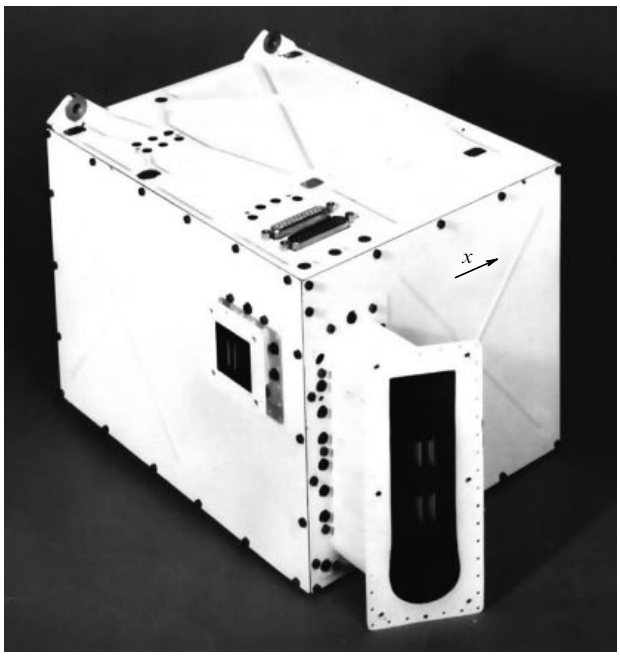

Figure 4. General view of the four-channel UV photometer installed aboard the Prognoz-5 and Prognoz-6 satellites.

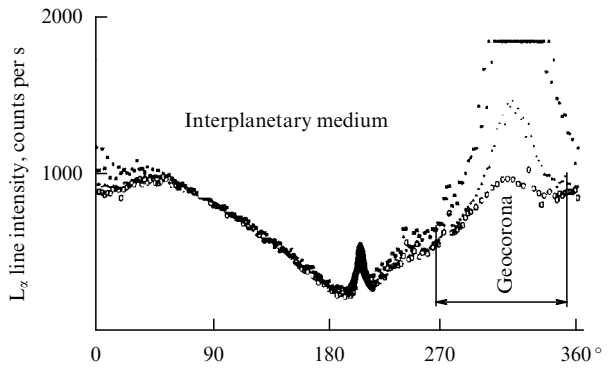

Figure 5. The L_α-line intensity distribution for three scan circles corresponding to three positions of the satellite in orbit. The narrow peak is a hot star, wide maxima near 310 deg are due to the geocorona. The signal between 50 and 150 deg is due to the interplanetary L_α-line emission.

had three detectors centered on the H I L_α line and the He I ($\lambda = 584$ Å) and He II ($\lambda = 304$ Å) lines supplied with the absorbing cell and a narrow-band interferometric filter in the L_α-channel. For the He I and He II channels, broadband filters made of thin metal films with a bandwidth of 100 Å were used. The optical axes of these channels were directed normally to the spin axis of the satellite, i.e., normally to the direction to the Sun. An additional He I channel was oriented at an angle of 10 deg to the anti-solar direction. The Prognoz-5 and Prognoz-6 satellites actively operated for four and six months, respectively. The telemetry was received once every 10.7 s, but because the spin period of the satellite (120 s) and the time the telemetry was taken were incommensurable, the intensity and reduction factor (the ratio of switched-on to switched-off cells of the H I L_α-detector) depended on the spin angle of the satellite continuously with a high spatial resolution (about 1–3 deg). Clearly, all scan trajectories passed through both ecliptic poles, which provided constant control of the detector sensitivity stability. Figures 5 and 6 show examples of scans in the H I L_α line and FR. The passage of the field of view through Earth's hydrogen corona is clearly seen.

In the data processing, all points with the impact parameter of the optical axis passed at distances smaller than 50000 km for the L_α line and 70000 for the He I ($\lambda = 584$ Å) line were rejected. This excluded the influence of extended hydrogen and helium atmospheres of Earth. The He II ($\lambda = 304$ Å) channel was used by us as the 'background' channel registering cosmic rays and induced radiation from the spacecraft. Its counts were subtracted from those of all other detectors. We note that in the H I L_α channel, the count rate was 100 times higher than in the He I ($\lambda = 584$ Å) channel. Figure 7 clearly demonstrates how the optical axis of the detector crosses the focusing cone of helium atoms 'behind the Sun'.

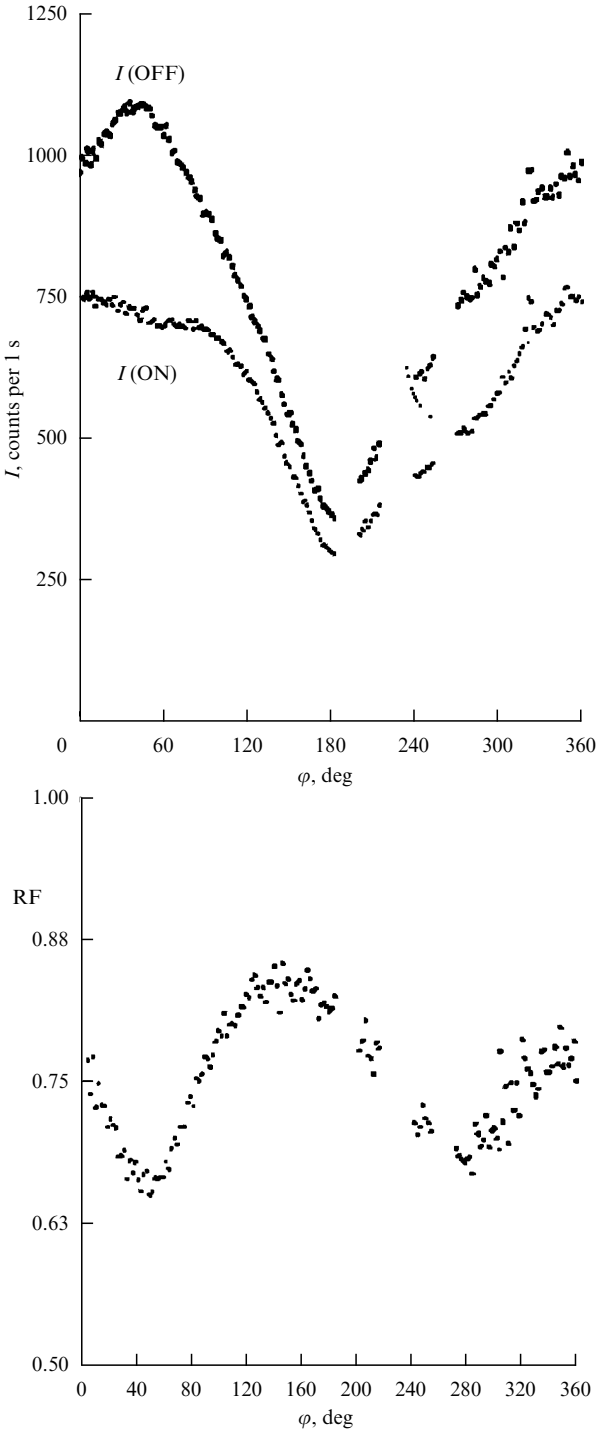

Figure 6. Observed intensity of the L_α line along the scan circle with a turned-on, I(ON), and turned-off, I(OFF), cell. The reduction factor is RF = I(ON)/I(OFF).

4. Measurement of LISM parameters

When interpreting H I observations, all data were fitted to obtain all LISM parameters simultaneously [12]. In exactly the same way, the French group processed the He I results [13]. Our group divided all parameters into two parts. The solar data (the He I line width and the lifetime of a helium atom at a distance of 1 a.u.) were taken from the literature. The direction of the Sun's motion in the LISM (two angles in the ecliptic coordinate system) was determined from the easily observed maximum of the glow on the He I ($\lambda = 584$ Å) line, i.e., when the optical axis of the detector crossed the focusing cone axis. Two parameters were fitted: the LISM temperature and the absolute value of the Sun's velocity [14–16]. In our opinion, the fitting with only two parameters gives more reliable results than fitting with a larger number of parameters. We note that the determination error of the Sun's velocity relative to the LISM does not exceed 2 km s^{-1}, while the temperature is determined with a much worse accuracy.

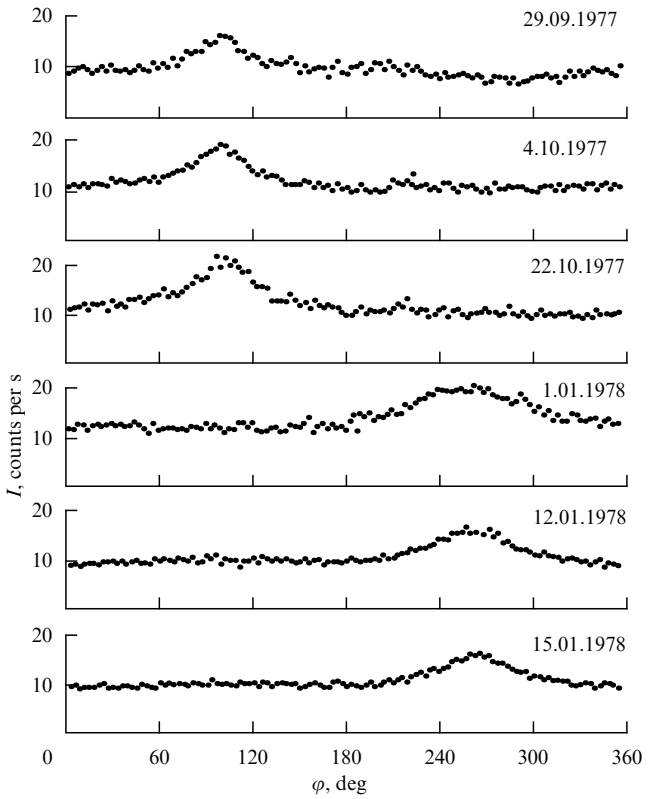

Figure 7. The He I ($\lambda = 584$ Å) line emission intensity along six scan circles during the passage through the maximum helium atom density cone.

Tables 1 and 2 list the results obtained by us and from measurements by the EUVE (Extreme Ultraviolet Explorer) and SOHO/SWAN (Solar and Heliospheric Observatory/ Solar Wind Anisotropy) satellites. The coordinates of the downwind direction measured by the Prognoz-5 and Prognoz-6 satellites are related to the epoch 1950,0, while those measured by the EUVE and SOHO satellites, to the epoch 2000,0.

Both hydrogen and helium densities are factors in the expression for the observed intensity. Clearly, their determination accuracy depends only on the absolute calibration of the detectors, i.e., on the recalculation of the observed count rate into physical units erg cm^{-2} s^{-1} sr^{-1} or Rayleighs. For calibration, we used two sources whose intensity can be calculated based on several parameters; for a blackbody (BB) source, only one parameter (its temperature) is needed. However, for our observations in the vacuum UV range, a source with a temperature of 10,000 K are even 20,000 K was required. Such sources were manufactured by us jointly with the Institute of High Temperatures, RAS using high-pressure lamps filled with gas (for example, xenon). However, determination of the temperature with the required accuracy (about 25 K) is a very complicated problem. Obviously, the black-body approximation can be used only in lines, because in the continuum, the hot gas emission is optically thin and the black-body formula cannot be applied. The radiation is optically thick and corresponds to the black-body intensity only in the emission lines. We used the method of the Balmer line width determination for hydrogen, which was added to the noble gas in small amounts (1–3%). This allowed estimating the electron density in the lamp, which, in turn, is determined by the Saha formula, and thus the temperature was ultimately calculated. Unfortunately, the absolute calibration accuracy obtained by this method was rather low (more than 50%) due to errors in the temperature determination. In addition, it proved to be very difficult to take a huge radial temperature gradient in the gas-discharge tube into account, with the temperature changing from 25,000 K on the tube axis to 300 K on the fresh-water-cooled walls.

Synchrotron radiation is the second source with the known absolute intensity. To calculate its intensity, the curvature radius of the relativistic electron trajectory, magnetic field strength, energy of the electron, and the number of electrons in the accelerator channel should be known. We used the VEPP-2M electron–positron accelerator of the Nuclear Physics Institute, Siberian Branch of RAS, where our group was invited by A M Budker, who was the director of the institute at that time. Using this method allowed increasing the absolute calibration accuracy to 15–20%. Finally, during the satellite flight, we could test or improve the calibration by observing bright blue and hot stars accidentally occurring in the field of view of the detectors.

5. Direct (nonoptical) methods of measurements of the local interstellar medium parameters

As far as we know, the first successful measurements of the atomic helium density were carried out by a group of researchers from the Max-Planck Institute for Aeronomy (Lindau, Germany) using detectors aboard the Ulysses extra-ecliptic station of the European Space Agency, launched on 6 October 1990 and operated until 1 July 2008, i.e., for 17 years [17]. Ulysses crossed the ecliptic

Table 1. LISM parameters from Prognoz 5, 6 observations of the He I ($\lambda = 584$ Å) line.

Satellite	$\lvert V\rvert$, km s^{-1}	λ, deg	β, deg	T, K	n (He I), cm^{-3}
Prognoz-6 (our results [12])	25.3 ± 1.9	77.5 ± 2.6	-6.2 ± 2.1	$13{,}500 \pm 1700$	0.018 ± 0.002
Prognoz-5 (results of the French group [13])	27 ± 3	74.5 ± 3	-6 ± 3	11,000 – 24000	0.015 ± 0.0023
EUVE (USA)	24.5	74.7 ± 0.5	-5 ± 0.1	7000	0.0135 ± 0.08

Table 2. LISM parameters from Prognoz 5, 6 observations of the L_α line.

Satellite	$\lvert V\rvert$, km s^{-1}	λ, deg	β, deg	T, K	n (H), cm^{-3}
Prognoz-5 and Prognoz-6	20 ± 1	71 ± 2	-7.5 ± 3	8000 ± 1000	0.065 ± 0.015
SOHO (SWAN)	21.5 ± 0.5	72.3 ± 0.7	-8.7 ± 0.9	$11{,}500 \pm 500$	?

Table 3. LISM parameters from the Ulysses measurements of the helium atom flux.

$\lvert V\rvert$, km s^{-1}	λ, deg	β, deg	T, K
26 ± 1	72 ± 2	-2.5 ± 2.7	6700 ± 1500

plane four times and could observe both solar poles. In particular, the latitude asymmetry of the solar wind was measured, which directly confirmed our results obtained from the Prognoz satellites [18].

To register helium atoms that entered the Solar System from the interstellar medium, a group of researchers headed by H Rosenbauer at the Max-Planck Institute for Aeronomy designed a detector based on the interaction of helium atoms having an energy of about 15 eV (which corresponds to the velocity of 25 km s^{-1}) with a thin gold foil. Electrons kicked out from the foil were detected by a channel photomultiplier, which, with the known velocity of the electrons, allowed estimating their local density (in contrast, in our integral method, the radiation on the $\lambda = 584$ Å line was measured along the entire path from the measurement point to infinity). The detector scanned almost the entire sky, enabling the determination of the direction of helium atom trajectories. The measurements were carried out from different points in the Solar System, owing to which a high accuracy was achieved (Table 3). Data obtained in a completely different way are in good agreement with our results obtained by 'optical' means.

6. Conclusion

From the brief consideration of the problem of the motion of the Sun relative to the local interstellar medium given above, we conclude that the basic parameters of the LISM are known quite well. These include the density of atomic hydrogen and helium in the close vicinity of the Solar System (at distances exceeding 20 a.u.) and the direction of the motion of the Sun relative to LISM. The LISM temperature is known with less accuracy. Undoubtedly, the temperature values derived by us from measurements on L_α lines of hydrogen ($\lambda = 1215.7$ Å) and helium ($\lambda = 584$ Å) are different, and this fact calls for explanation. Second, the temperature measurement error on these two lines highly exceeds the relative measurement errors of the solar motion direction relative to the LISM and of the hydrogen and helium atomic density.

It is very plausible that the temperature difference for hydrogen and helium atoms is due to the interstellar atoms crossing the transient zone between the heliosphere and 'pure' interstellar space. Hopefully, new results obtained by the IBEX (Interstellar Boundary Explorer) satellite can be used to improve the parameters of the local interstellar medium.

References

1. Kupperian J E (Jr.) et al. *Planet. Space Sci.* **1** 3 (1959)
2. Morton D C, Purcell J D *Planet. Space Sci.* **9** 455 (1962)
3. Bertaux J L et al. *Kosmich. Issled.* **16** 269 (1978) [*Cosmic Res.* **16** 214 (1978)]
4. Bertaux J L et al. *Astron. Asrophys.* **46** 19 (1976)
5. Kurt V G, Germogenova T A *Astron. Zh.* **44** 352 (1967) [*Sov. Astron.* **11** 278 (1967)]
6. Blum P W, Fahr H J *Astron. Astrophys.* **4** 280 (1970)
7. Blum P W, Pfleiderer J, Wulf-Mathies C *Planet. Space Sci.* **23** 93 (1975)
8. Lallement R et al. *Astron. Astrophys.* **426** 875 (2004)
9. Baranov V B *Space Sci. Rev.* **143** 449 (2009)
10. Katushkina O A, Izmodenov V V *Pis'ma Astron. Zh.* **36** 310 (2010) [*Astron. Lett.* **36** 297 (2010)]
11. Bertaux J L, Lallement R *Astron. Astrophys.* **140** 230 (1984)
12. Bertaux J L, Lallement R, Kurt V G, Mironova E N *Astron. Astrophys.* **150** 1 (1985)
13. Dalaudier F, Bertaux J L, Kurt V G, Mironova E N *Astron. Astrophys.* **134** 171 (1984)
14. Burgin M S et al. *Kosmich. Issled.* **21** 83 (1983) [*Cosmic Res.* **21** 72 (1983)]
15. Kurt V G, Mironova E N, Bertaux I L, Dalaudier F *Kosmich. Issled.* **22** 97 (1984) [*Cosmic Res.* **22** 86 (1984)]
16. Kurt V G, Mironova E N, Bertaux J L, Dalaudier F *Kosmich. Issled.* **22** 225 (1984)
17. Witte M et al. *Adv. Space Res.* **13** (6) 121 (1993)
18. Lallement R, Bertaux J L, Kurt V G *J. Geophys. Res.* **90** 1413 (1985)

PACS numbers: **04.20.–q**, **04.70.–s**, **98.80.–k**
DOI: 10.3367/UFNe.0182.201208k.0894

Generation of cosmological flows in general relativity

V N Lukash, E V Mikheeva, V N Strokov

1. Introduction

The Copernicus principle is known to cast doubt on the uniqueness of our Universe. Therefore, it seems likely that there is some physical mechanism of gravitational reproduction of cosmological flows of matter expanding from super-large to small curvatures and densities. We relate the solution of the cosmogenesis problem (the origin of universes) to black holes, in which regions with high space–time curvature are formed in a natural evolutionary manner during gravitational collapse. We only need to continue the singular states that appeared in this way in time and to see what geometrical structures are found beyond them in the future.

An analytic continuation of general relativity (GR) solutions through singular hypersurfaces $r = 0$ is realized in the model class of 'black-white' holes with integrable singularities [1, 2]. In these models, the space–time of a black hole can be connected with a white hole endowed with the metric of a homogeneous cosmological model, allowing an explicit realization of the geometrical concept of a multisheet universe (hyperverse). These topics are discussed in this paper.

2. How the Schwarzschild metric can be continued

A black hole with a positive external mass $M > 0$ without rotation or charge is described in GR by the Schwarzschild metric in the vacuum:

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{r}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1 - 2GM/r} - r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right), \quad (1)$$

V N Lukash, E V Mikheeva, V N Strokov Astro-Space Center, Lebedev Physics Institute, Russian Academy of Sciences, Moscow, Russian Federation
E-mail: lukash@asc.rssi.ru, helen@asc.rssi.ru, strokov@asc.rssi.ru

Uspekhi Fizicheskikh Nauk **182** (8) 894 – 900 (2012)
DOI: 10.3367/UFNr.0182.201208k.0894
Translated by K A Postnov; edited by A M Semikhatov

where the variable $r > 0$ is defined as the internal curvature radius of a closed homogeneous isotropic 2-space $\mathrm{d}Y^2 = \gamma_{ij}\,\mathrm{d}y^i\,\mathrm{d}y^j$, which does not depend on y^i $(i = 1, 2)$, and G is the gravitational constant. The Y-space is invariant under the group of motions G_3 (two translations that allow moving to an arbitrary point Y, and rotation around a point), and it can be reduced to the form $\mathrm{d}Y^2 = r^2\,\mathrm{d}\Omega$, where $\mathrm{d}\Omega \equiv \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$ decribes the unit-radius 2-sphere $\mathbb{S}^2$, $y^i = (\theta, \varphi)$. Any of the $\mathbb{S}^2$ points can be taken as the pole $\theta = 0$, and turning around the pole is described by the angle φ.

The space X orthogonal to Y is given in the Eulerian gauge, in which one of the coordinates x^I coincides with r. Additionally, metric (1) is independent of the coordinate $t \in \mathbb{R}^1$. Therefore, the black hole in the vacuum has the group of motions G_4 acting on the hypersurface $\mathbb{C}^3 = \mathbb{R}^1 \times \mathbb{S}^2$ (three translations in $\mathbb{C}^3$ and homogeneous rotations in $\mathbb{S}^2$).

Topologically, geometry (1) represents a 4-cylinder with a homogeneous 3-surface $\mathbb{C}^3$ and the radial coordinate r, which in the region $r > 0$ determines the Schwarzschild sector of a black (or white) hole. Attempts to extend this solution inevitably lead to regions occupied by matter. Therefore, the problem of the analytic continuation of metric (1) must be solved using more general GR metrics with matter, which we restrict by requiring the spherical symmetry G_3 in $\mathbb{S}^2$.

We consider a class of such metrics that in the orthogonal gauge ($g_{Ii} = 0$) have the form

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}X^2 - \mathrm{d}Y^2\\ &= -\mathcal{N}\mathcal{K}\,\mathrm{d}t^2 + \frac{\mathrm{d}\rho^2}{4\rho\mathcal{K}} - \rho(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2)\,,\end{aligned} \tag{2}$$

where the real variable ρ^{-1} is not restricted by the sign and is defined as the Y-space internal curvature scalar that is independent of y^i:

$$R^{(Y)}_{ij} = \rho^{-1}\gamma_{ij}\,. \tag{3}$$

The Ricci 2-tensor is constructed from the metric γ_{ij}, which by the spherical symmetry can always be reduced to the form $\mathrm{d}Y^2 = \rho\,\mathrm{d}\Omega$. For general reference frames, ρ, $\mathcal{K}$, and $\mathcal{N}$ are 4-scalars depending on the coordinates of the X-space $\mathrm{d}X^2 = n_{IJ}\,\mathrm{d}x^I\,\mathrm{d}x^J$; here, $\mathcal{K}$ is the kinetic term of ρ and $\mathcal{N}$ is restricted to have positive values:

$$\mathcal{K} = \frac{\rho_{,\mu}\,\rho^{,\mu}}{4\rho} \equiv -1 - 2\Phi\,, \qquad \mathcal{N} \equiv N^2 > 0\,, \tag{4}$$

with Φ being the metric gravitational potential. The energy–momentum tensor has the form $T^{\nu}_{\mu} = \mathrm{diag}\,(T^J_I, -p_\perp, -p_\perp)$, where $T^J_I = (\epsilon + p)\,u_I\,u^J - p\delta^J_I$. For $\rho > 0$, the functions ϵ, p, $p_\perp$, and $u^\mu = (u^I, 0, 0)$ respectively describe the energy density, longitudinal and transversal tension, and 4-velocities of matter ($u_I u^I = 1$). Metric (2) transforms into solution (1) in the part of the domain $\sqrt{\rho} = r > 0$ that is free of matter. Therefore, by definition, we call the domain $\mathcal{K} > 0$ the T-region of space (2). There, in particular, the component T^t_t describes the pressure[1] and the meaning of the other components depends on the sign of ρ.

The GR equations relate the metric and matter scalars:

$$\Phi' = 2\pi G P\,, \qquad \dot{\Phi} = 2\pi G T^{\rho}_{t}\,, \tag{5}$$

[1] In particular, under the comoving condition $T^t_t = -p$ in (2).

$$\frac{N'}{N} = \frac{2\pi G(E+P)}{\mathcal{K}}\,, \tag{6}$$

$$\frac{(\rho N E)'}{N} + p_\perp - \frac{P}{2} = -\frac{(N^3 T^{t\rho})^{\cdot}}{4N^3\mathcal{K}}\,, \qquad \dot{P} = (T^{\rho}_{t})'\,, \tag{7}$$

where the prime and dot respectively denote partial derivatives with respect to ρ and t, and

$$P \equiv -T^t_t - \frac{\Phi}{4\pi G\rho}\,, \qquad E \equiv T^{\rho}_{\rho} + \frac{\Phi}{4\pi G\rho}\,. \tag{8}$$

By modeling the state of the effective matter, we can determine P and E from Bianchi identities (7), and integrating (5) and (6) over $\mathrm{d}x^I$ from external solution (1) into the future, we can reconstruct the metric potentials. Similarly, Eqns (5) are transformed for the mass function $m = m(x^I)$:

$$\Phi \equiv -\frac{Gm}{\sqrt{\rho}}\,, \qquad m_{,I} = 4\pi\rho\, e_{IK}\, T^K_J\, \frac{\partial x^J}{N\partial t}\,, \tag{9}$$

where $(\ldots)_{,I} \equiv \partial/\partial x^I$ and the totally antisymmetric tensor in X is given by

$$\begin{aligned} e_{IJ} &= \left|\det\,(n_{IJ})\right|^{1/2}\begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix},\\ e^{IJ} &= \left|\det\,(n_{IJ})\right|^{-1/2}\begin{pmatrix} 0 & 1\\ -1 & 0\end{pmatrix}.\end{aligned} \tag{10}$$

3. Space–time near an integrable singularity

We are interested in solutions (2) generated by metric (1) that are geodesically complete. The sufficient condition for the metric to pass through $\rho = 0$ is the finiteness of the potentials Φ and N and their derivatives. We call such models black-white (or black/white) holes with integrable singularity [1]. They are described by the mean (average over vacuum state (1)) 4-dimensional metric space without punctures, which provides a continuous extension of the affine parameters of world lines of test particles through the singular hypersurface $\rho = 0$. Using reference frames constructed on these particles, we investigate the geometry of black-white holes outside the Schwarzschild sector.

The effective matter generated by a strong gravitational field near the singularity changes the space–time near $\rho = 0$ and is the physical reason for the existence of an integrable singularity. The formation of matter in extremely strong gravitational fields corresponds to the Le Chatelier principle: this is how Nature reacts to a sharp increase in the metric potential amplitudes as $\rho \to 0$ (which would continue to increase in the absence of matter) and thus precludes their divergence. Specific quantum gravitational mechanisms of mutual transformations of matter and gravitational degrees of freedom under extreme conditions need to be considered separately. Here, developing papers [1, 2], we postulate the continuity of gravitational potentials (2) in the presence of effective matter.

In the R-region of space–time (2) with the signature $(+,-,-,-)$ specified by generating metric (1), we have $\mathcal{K} < 0$ and $\rho = r^2 > 0$. Here, the hypersurface $r = 0$ is degenerate and represents a time-like worldline of the center of a spherically symmetric distribution of matter (for example, the center of a star). If there is no matter in the

R-region, it follows from (1) that $r > 2GM$, i.e., the hypersurface $r = 0$ does not lie in the R-region.

In the T-region, we have $\mathcal{K} > 0$, and the variable ρ can have any sign because the physics in this region is determined by nonlinear quantum effects, and *a priori* there are no grounds to believe that the signature remains indefinite [3]. Here, the singular hypersurface $\rho = 0$ splits the 4-space into domains with different signatures $(-,+,-,-)$ for $\rho > 0$ and $(-,-,+,+)$ for $\rho < 0$. The complete geometry depends on the distribution and properties of the effective matter near $\rho = 0$.

We assume that the effective matter distribution maintains the symmetry of the generating field. For example, the region that is evolutionary adjacent to that with the Schwarzschild metric (for example, $T_{\mu\nu} \neq 0$ for $r \leqslant r_0 = \text{const} < 2GM$ and $T_{\mu\nu} = 0$ for $r > r_0$) preserves the Killing t-vector and depends only on ρ. The region inside the star keeps the spherical symmetry and the field homogeneity of the star. We use these constraints in constructing the models in Sections 4–6.

Several properties of geometries (2) should be noted.

• The continuity of the potential Φ implies the integrability of $P(t,\rho)$ along the lines $t = \text{const}$ [see (5)]:

$$\Phi(t,\rho) = \Phi_0 + 2\pi G \int_0 P\,\mathrm{d}\rho\,, \qquad \Phi_0 = \Phi(t,0)\,. \tag{11}$$

• The continuity of Φ and N implies the integrability of $E(t,\rho)$ along the lines $t = \text{const}$ [see (6)]:

$$N(t,\rho) = N_0 \exp\left(2\pi G \int_0 \frac{E+P}{\mathcal{K}}\,\mathrm{d}\rho\right), \quad N_0 = N(t,0)\,. \tag{12}$$

• For $\mathcal{K}_0 \equiv -1 - 2\Phi_0 = K_0^2 = \text{const} > 0$, the space–time at $\rho \to 0$ has the structure

$$\mathrm{d}s^2 = -\,\mathrm{d}\tilde{t}^2 + \mathrm{d}U\,\mathrm{d}V - UVK_0^2 \sin^2\theta\,\mathrm{d}\varphi^2\,, \tag{13}$$

$$\tilde{t} = K_0 \int N_0\,\mathrm{d}t\,, \qquad \rho = K_0^2\,UV\,, \qquad \theta = \frac{1}{2K_0}\ln\left|\frac{U}{V}\right|.$$

The direction θ can be taken along any meridian in $\mathbb{S}^2$ and counted from an arbitrarily chosen pole $\theta = 0$.

It follows that the integrable singularity of a black-white hole includes horizon hypersurfaces $U = 0$ and $V = 0$ lying in the T-region and intersecting along the space-like bifurcation line $U = V = 0$. They separate the T-region in cone sectors of the black $(U < 0, V < 0)$ and white $(U > 0, V > 0)$ holes and the static zone $(UV < 0)$. The horizons have the cylindrical symmetry $\mathbb{R}^1 \times \mathbb{S}^2$ with the space-like longitudinal axis $t \in \mathbb{R}^1$ and null geodesics in $\mathbb{S}^2$ ($t = \text{const}$).[2] Photons propagating in these directions are confined gravitationally at $\rho = 0$, where they perform infinite oscillations in Y within a finite interval of the affine parameter. Trajectories of other particles, including photons with a projection in $\mathbb{R}^1$, intersect the singular hypersurface $\rho = 0$ and go away into other metric domains.

4. Geometric maps of an oscillating hole

The structure of a 4-space can be described by 2- and 3-dimensional cross sections and cuts, which can be covered with a net of trajectories of test particles on them. It is convenient to choose coordinates such that the light trajectories have a slope of 45° as in the flat world [4]. By the spherical symmetry, all geodesics in (2) can be represented by three groups of cross sections:

- X-planes: longitudinal cross sections $(\theta,\varphi) = \text{const}$;
- UV-planes: transversal cross sections $(t,\varphi) = \text{const}$;
- tUV-hypersurfaces: longitudinal–transversal cuts $\varphi = \text{const}$.

The X-planes are filled with radial or longitudinal geodesics, the tUV-planes contain particles propagating in both longitudinal and transversal directions, and the UV-planes orthogonal to the bifurcation line are filled with spiral geodesics that do not escape beyond the T-region (see the Appendix) (Fig. 1).

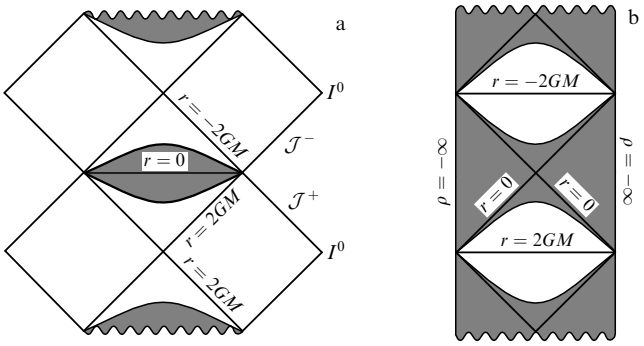

Figure 1. (a) Longitudinal and (b) transversal cross sections of an oscillating black-white hole. $\mathcal{J}^+$ is the light-like infinity of the future for observers in the region $r > 0$, $\mathcal{J}^-$ is the light-like infinity of the past in the region $r < 0$, and I^0 are spatial infinities of the R-regions.

As an example, we consider metric (2), which is independent of t, with matter in the T-region $r \leqslant r_0 = \text{const} < 2GM$. The matter is distributed in the UV-planes along bundles of straight lines $(t,\theta,\varphi) = \text{const}$ intersecting at the points $U = V = 0$. Integrating (9) along the time lines $r = \sqrt{\rho}$ from black-hole region (1) to the future, we obtain continuous potentials

$$\Phi = -\frac{Gm}{r}\,, \quad m = M - 4\pi\int_{r_0} p(r)\,r^2\,\mathrm{d}r = -4\pi\int_0 p\rho\,\mathrm{d}r\,, \tag{14}$$

which directly connect both holes. The black-hole metric is in the sector $r > 0$, while a white hole is obtained from the extension of solution (14) into the region $r < 0$ under the finiteness condition $p\rho = \Phi_0/4\pi G = \text{const}$ at $\rho = 0$. The complete geometry can be restored by integrating from the bifurcation line points along all bundles of lines, including regions $\rho < 0$.

For illustration, we consider the model of an oscillating black-white hole with a triggered matter distribution depending on $\rho \leqslant \rho_0 \equiv r_0^2$ (see [1, 2], case B). In other words, the transversal pressure increases jump-wise: $p_\perp = \lambda_0\theta(\rho_0 - \rho)$. The longitudinal pressure, which is chosen to be vacuum-like ($p = -\epsilon$), is calculated using Bianchi identities (7):

$$\rho \leqslant \rho_0: \quad P = -E = \frac{2}{3}\,\lambda_0\,, \quad \Phi = \frac{1}{2}\,H_1^2(\rho - 3\rho_0)\,, \tag{15}$$

$$\rho > \rho_0: \quad P = -E = \frac{M}{4\pi|r|^3}\,, \quad \Phi = -\frac{GM}{|r|}\,, \tag{16}$$

[2] A comparison is suggested with horizons $r = 2GM$ of an eternal black (white) hole for the Kruskal metric, which are null in the radial direction and space-like in $\mathbb{S}^2$ and intersect over a space-like bifurcation 2-sphere.

where $H_1^2 \equiv 8\pi G\lambda_0/3 \equiv GM/r_0^3 = \text{const}$. Substituting (15) and (16) in (2), we obtain the metric with matter ($\rho \leqslant \rho_0$):

$$\begin{aligned} \mathrm{d}s^2 &= -\mathcal{K}\,\mathrm{d}t^2 + \frac{\mathrm{d}\rho^2}{4\rho\mathcal{K}} - \rho\,\mathrm{d}\Omega \\ &= \frac{-(1-\tilde\rho)^2\,\mathrm{d}\tilde t^2 + H_1^{-2}\tilde\rho^{-1}\,\mathrm{d}\tilde\rho^2 - 4r_1^2\tilde\rho\,\mathrm{d}\Omega}{(1+\tilde\rho)^2} \end{aligned} \tag{17}$$

and the metric in the vacuum ($\rho = r^2 > \rho_0$)

$$\mathrm{d}s^2 = \left(1 - \frac{2GM}{|r|}\right)\mathrm{d}t^2 - \frac{\mathrm{d}r^2}{1-2GM/|r|} - r^2\,\mathrm{d}\Omega\,, \tag{18}$$

where $\tilde t = K_0 t$, $\rho = 4r_1^2\tilde\rho(1+\tilde\rho)^{-2}$, $\mathcal{K} = K_0^2 - H_1^2\rho$, $r_1 = K_0/H_1 = r_0(3 - r_0/GM)^{1/2}$, $K_0^2 = -1 - 2\Phi_0$, and $\Phi_0 = \Phi(0) = -3GM/2r_0$. The value of the potential of the integrable singularity is one and a half times lower than that at the matter boundary, and this ratio is independent of the parameter r_0.

We represent (17) using the proper interval on the ρ axis:

$$\begin{aligned} \rho = r^2 \in [0,\rho_0]:\ \ \mathrm{d}s^2 \\ = \mathrm{d}\tau^2 - \cos^2(H_1\tau)\,\mathrm{d}\tilde t^2 - r_1^2\sin^2(H_1\tau)\,\mathrm{d}\Omega\,, \end{aligned} \tag{19}$$

$$\begin{aligned} \rho \leqslant 0:\ \ \mathrm{d}s^2 = -\mathrm{d}x^2 - \cosh^2(H_1x)\,\mathrm{d}\tilde t^2 \\ + r_1^2\sinh^2(H_1x)\,\mathrm{d}\Omega\,, \end{aligned} \tag{20}$$

where $r = -r_1\sin(H_1\tau) \in [-r_0, r_0]$ and $\rho = -r_1^2\sinh^2(H_1x) \leqslant 0$. The space–time of static domains is asymptotically anti-de Sitter:

$$\rho \ll -r_0^2:\quad T_\nu^\mu = -\lambda_0\delta_\nu^\mu\,,\quad R_\nu^\mu = 3H_1^2\delta_\nu^\mu\,. \tag{21}$$

For this reason, the regions $\rho < 0$ are also referred to as anti-de Sitter (AdS) zones.

5. Source of an eternal black (white) hole

Eternal holes that are almost everywhere free of matter can be obtained in the limit $r_0 \to 0$ in Eqns (15)–(18):

$$r \in \mathbb{R}^1:\qquad \epsilon = -p = 2p_\perp = M\,\frac{\delta(r)}{2\pi r^2}\,, \tag{22}$$

$$\rho < 0:\quad p = -\epsilon = p_\perp = \lambda_0 \equiv \frac{3M}{8\pi r_0^3}\,. \tag{23}$$

The AdS zones are strongly curved by the dense vacuum λ_0 [see (20)]:

$$\begin{aligned} \rho \ll -r_0^2:\ \ \mathrm{d}s^2 = -\mathrm{d}x^2 + \frac{3}{4}\,r_0^2\,\exp(2H_1x) \\ \times(-H_1^2\,\mathrm{d}t^2 + \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2), \end{aligned} \tag{24}$$

which produces a δ-like material source with the geometry of eternal black hole (18) in the region $r \in \mathbb{R}^1$. This source, localized at $|r| \leqslant r_0$, has the density $\sim\lambda_0$ and the total mass M. The value of r_0 can be calculated in quantum field theory. In our classical treatment, r_0 is a free parameter of the problem. We also note that the relation between the longitudinal and transversal tensions in (22) is not universal: it depends on the chosen condition of the model $p = -\epsilon$.

Solutions (22)–(24) show that the polarized vacuum in static AdS zones is the source of eternal black holes. The gravitational mass of the effective matter of each of the AdS zones, via its light hypersurface of the future $\rho = 0$, generates the causally connected Schwarzschild metric of a black-white hole in the vacuum, which extends in time to the next integrable singularity $\rho = 0$, where the process repeats (see Fig. 1). The full geometry is invariant under the reversal $r \to -r$ of the spherical coordinate system relative to the bifurcation lines. In this sense, phase transitions between gravitational and material degrees of freedom in this model are reversible.

In Section 6, we construct one more example of a reversible geometry, in which the parent star, collapsed from the R-region with the formation of a singularity in the T-region, is the source of a black hole. During the evolution, this model passes the stage of the effective matter and transforms into a white hole with the metric of a homogeneous cosmological model.

6. Astrophysical black-white hole

We consider the model of a black-white hole with an integrable singularity, assuming that the black hole was formed during the collapse of a star from the parent universe. The star is modeled as a homogeneous sphere with the radius $r(T, R=1) = a(T)$, which was at rest in a flat space–time and had a mass M ($a \gg 2GM$ at $T \to -\infty$), and then started collapsing due to self-gravitation, with the initial pressure being negligible. Here, T and R are the Lagrange coordinates comoving with the star material, and the proper time $T \in \mathbb{R}^1$ and radial markers of spherical shells $R \geqslant 0$ are normalized such that $R = 1$ on the star surface.

The symmetry of the field is determined by the initial and boundary conditions of the problem. Inside the star ($R \leqslant 1$), the symmetry of $\mathbb{R}^1 \times \mathbb{E}^3$ is the 6-parameter group G_6 on $\mathbb{E}^3$ and the potentials a, H, ϵ, p are functions of T. Outside the star ($R > 1$), the group of motions G_4 on $\mathbb{C}^3$ and all potentials depend on r. To avoid crossing matter flows, we assume that the longitudinal tension, which is produced outside the star at large curvatures, is vacuum-like. Then, at $R > 1$, $N = 1$ and the energy–momentum tensor is invariant under motions in X: $T_\nu^\mu = -\mathrm{diag}\,(p, p, p_\perp, p_\perp)$. Inside $R \leqslant 1$, the matter is Pascalian, $T_\nu^\mu = (\epsilon + p)\,u_\nu u^\mu - p\delta_\nu^\mu$ (with the 4-velocity $u_\nu = T_{,\nu}$), and its state is calculated from the boundary conditions. The homogeneity of the star suggests the continuity of Φ and p at the boundary $R = 1$ ($r = a$), whereas the density and transversal pressure can change discontinuously.

Inside the star, the metric has the form [5]

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}T^2 - a^2(\mathrm{d}R^2 + R^2\,\mathrm{d}\Omega) \\ &= \frac{a^4H^2\,\mathrm{d}t^2 - \mathrm{d}r^2}{1+2\Phi} - r^2\,\mathrm{d}\Omega\,, \end{aligned} \tag{25}$$

where the Euler and Lagrange coordinates are related as

$$R \leqslant 1:\quad r = aR\,,\quad t = -\int\frac{\mathrm{d}T}{a^2H} - \frac{R^2}{2}\,, \tag{26}$$

and the Hubble function $H = \mathrm{d}a/a\,\mathrm{d}T$ can be found from the Friedman equations:

$$H^2 = \frac{8\pi G}{3}\,\epsilon = \frac{2Gm}{r^3} = -\frac{2\Phi}{r^2}\,,\quad \frac{\mathrm{d}\epsilon}{\mathrm{d}T} + 3H(\epsilon + p) = 0\,. \tag{27}$$

Initially, the pressure is absent and the star collapses freely ($H < 0$, $a^3H^2 = 2GM$). The Newtonian potential in this limit

is $\Phi_N = GM(R^2-3)/2a$. The internal tension in the star appears at $a \leqslant r_0 < 2GM$, and it can be calculated from the matching conditions with the effective matter at the boundary $R = 1$.

Outside the star, the metric has the form

$$\begin{aligned} ds^2 &= (1+2\Phi)\,dt^2 - \frac{dr^2}{1+2\Phi} - r^2\,d\Omega \\ &= dT^2 + 2\Phi\,dR^2 - r^2\,d\Omega\,, \end{aligned} \tag{28}$$

where the Lagrange reference frame comoves with the shells of free dust particles following the collapse of the star surface:

$$R > 1: \quad R - T = \int \frac{dr}{\sqrt{-2\Phi}}\,, \quad t = T - \int \frac{\sqrt{-2\Phi}\,dr}{1+2\Phi}\,, \tag{29}$$

$$\Phi = -\frac{Gm}{r} \equiv -\frac{1}{2}\,H^2 r^2\,, \quad m = M - 4\pi \int_{r_0} p(r)\,r^2\,dr\,. \tag{30}$$

The source of the Schwarzschild metric is the mass M of the star with the continuous potential $\Phi = -GM/r$ at the star boundary. The discontinuity of the functions N, t, and g_{RR} is due to the density jump at $R = 1$. The effective matter outside the star emerges at $r \leqslant r_0$ and preserves the symmetry of the parent metric. The continuity of Φ suggests that $m(0) = 0$ and leads to the formulas in Section 4.

By extending the reference frame of free particles to the entire cone $\rho = r^2 \geqslant 0$, we continue the metric from the black-hole region ($r > 0$) to the white-hole region ($r < 0$), bypassing the body of the star itself ($R > 1$) [see (14)–(18)]. Closing the external metric, we then restore the complete solution on the entire manifold $R \geqslant 0$ by matching p at the star boundary (Fig. 2). As a result, everywhere in the effective matter zone, we obtain $P = 2\lambda_0/3$ and

$$|r| \leqslant r_0\tilde{R}: \quad r = -\sqrt{3}r_0\tilde{R}\sin(H_1\tilde{T})\,,$$

$$\Phi = \frac{3}{2}\,\Phi_0\tilde{R}^2\cos^2(H_1\tilde{T})\,, \tag{31}$$

where $\tilde{R} \equiv \min(R, 1)$ and $\tilde{T} \equiv T + (1-R)\,\theta(R-1)$. The state of effective matter inside the star is described by the equation $\epsilon + 3p = 2\lambda_0$.

Freely falling shells of the star intersect simultaneously at one point $r = 0$ separating the time-like and space-like parts of the axis $r = 0$ (see Fig. 2). At this bifurcation point, the equation of state of the effective matter is $p = -\epsilon/3$ and the motion is rectilinear: $a \propto T$. Gravity is 'switched off' due to the mass ar $r = 0$, which provides the finite amplitude of tidal forces and the continuation of matter world lines into the future.

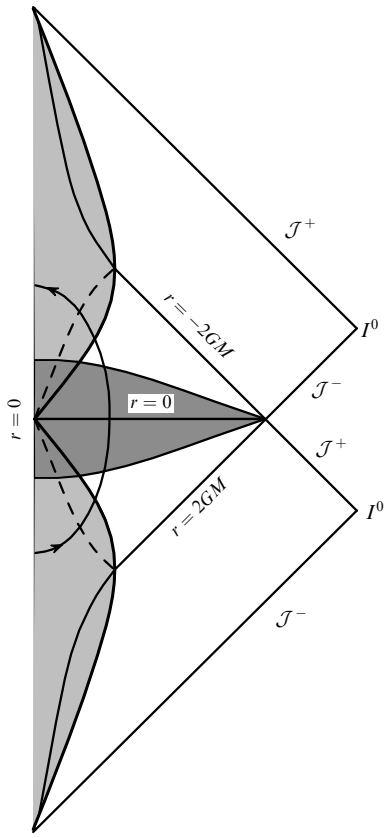

Figure 2. Penrose diagram of an astrophysical black-white hole. The light gray region is the body of the star, the dark gray region is the effective matter, the dashed line separates the R- and T-regions inside the star, and the line with arrows is a contour $t = \text{const}$.

7. Physical nature of integrable singularities

The concept of integrable singularities allowed us to radically advance in solving the cosmogenesis problem using the new class of black-white-hole models in GR. Are these solutions only research tools or do they really exist? If they do, then why and how do they form? These questions require separate studies. Here, we discuss some ideas on the physical nature of integrable singularities.

Reasons for the formation of singular structures with a finite gravitational potential are related to the redistribution of matter and space–time degrees of freedom in strong gravitational fields of the singularities themselves. (At high intensities and small scales, quantum effects apparently play the key role.) That is why we talk about the effective matter, which includes both material and gravitational degrees of freedom that remain after averaging the metric over the quantum state. However, we keep notions of energy–momentum and laws of motion in the form of Bianchi identities. This allows describing the back reaction using the Einstein equations in which the left-hand side is defined using the mean metric $g_{\mu\nu}$ and the right-hand side contains the energy–momentum tensor of the effective matter T^{μ}_{ν}, including all polarization, gravitationally modified, and other terms of quantum theory that need to be calculated.

We illustrate the method by considering the development of one degree of freedom in the given physical symmetry of a collapsing star. Let the physical variable be described by a massless field φ minimally coupled to metric (25). Secondary quantization leads to the following equation for the amplitudes of Fourier harmonics $\varphi_k = v_k(\eta)/a$ [5]:

$$v_k'' + (k^2 - U)\,v = 0\,, \tag{32}$$

$$U = U(\eta) = \frac{a''}{a} = a^2H^2 + (aH)' = -2\Phi + \frac{\Phi'}{\sqrt{-2\Phi}}\,, \tag{33}$$

where the prime denotes the derivative with respect to the conformal time $\eta = \int dT/a$, k is the wave number, and Φ is the gravitational potential on the stellar surface.

As long as the pressure in the star is low, the tidal potential increases with decreasing the radius: $U = GM/a = a^2H^2/2$.

At large a, the state of the field is vacuum-like: it oscillates near an equilibrium point in the adiabatic zone $U < k^2$.

During the collapse, the field enters the parametric zone, and its amplitude unlimitedly increases with increasing $U > k^2$:

$$\varphi_k = \frac{\exp(-\mathrm{i}k\eta)}{a\sqrt{2k}} \to \frac{H}{k^{3/2}}\,, \quad \varphi_{,\mu}\varphi^{,\mu} \simeq H^4 \ln U\,. \tag{34}$$

At this stage, the pressure cannot be neglected any more; the vacuum is polarized and in turn affects the metric, which changes the gravitational potential and the rate of collapse near $a = 0$ (for the solar mass and the number of degrees of freedom ~ 100, this occurs at $H \simeq 0.1\, M_{\mathrm{P}} (\ln U_0)^{-1/2} \sim 10^{17}$ GeV).

We can suppose that the back reaction described by the Friedman equations restructures the solution such that tidal potential (33) stops increasing infinitely and saturates: $U(r \to 0) = U_0 = -2\Phi_0 = \mathrm{const}$. This precludes an ultraviolet catastrophe and saves oscillators from being destroyed, because high-frequency modes with $k^2 > U_0$ remain in the adiabatic zone and are not polarized. Thus, the notion of the Heisenberg state vector stays valid, which corresponds to the Minkowski vacuum in the R-region, and the main requirement of the metric theory on the finiteness of the gravitational potential on a singular hypersurface is secured. We call such a structure an integrable singularity.

This example illustrates the difference between our theory and models with bounces (see, e.g., [6]), which have no singularity, $a > 0$, and $H = \Phi = 0$ at the bounce. The last requirement, in our opinion, is redundant and is not a necessary condition. In our models, the potential Φ reaches an extremum at the singularity $r = 0$, which weakens the singularity but does not fully eliminate it. For $a \propto T$, the situation resembles the explosive model of nongravitating particles by Milne. At the moment of crossing, particles move by inertia and do not feel the central mass attraction. Gravitation is switched off at this moment ($m(0) = 0$); therefore, the space does not curve, although the energy density diverges.

8. The arrow of time and natural selection

Gravitational tidal models of black-white holes containing expanding flows of matter elucidate many fundamental unanswered questions in modern physics. One of them is the causality principle, which puts the causal relationships in correspondence with the arrow of time. In the context of our work, we can discuss the origin of the cosmological arrow of time, which we regard as the orientation of the future light cone in the direction of the volume expansion of a large-scale flow of matter.

As is well known, dynamical equations describing microscopic processes are invariant under the change in the sign of time. However, the local dynamics have to be completed with an external arrow of time, because the invariance under time reversal is lost in the limit transition to the global geometry.

Geodesically full geometries with integrable singularities suggest the origin of the arrow of time. They include different space–time domains separated by event horizons $r = 0$ and $r = 2GM$: nonstationary regions of black and white holes (lying between the Schwarzschild and singular horizons), alternating static R-zones (connecting the Schwarzschild horizons of white and black holes), and AdS zones (connecting singular horizons of black and white holes). Here, all possibilities are realized. Each collapsing and anticollapsing (cosmological) region has its own time. The static regions are time-independent, but are not spatially homogeneous.

We are dealing with a unique geometry that is split into sectors with time and space parity. When crossing any of the horizons, the meaning of the coordinate r on which the metric depends changes [7]. In some domains, r is the time coordinate (and we then obtain black holes and/or cosmological models), while in other domains, it is a spatial coordinate (static zones). The complete geometry can then remain invariant under the change $r \to -r$.

Therefore, the origin of the cosmological arrow of time is related to the initial conditions. We (observers) belong to a cosmological flow of matter and live in its proper time, which started 14 bln years ago at the moment $r = 0$. The time coordinate can be extrapolated into the past, to the pre-cosmological epoch, but there it described the time in the T-zone of the parent black hole. In the even more remote past, this coordinate represented the radial coordinate of an asymptotically flat space of the parent universe in which the star had lived before it collapsed into a black hole. The integrable singularity $r = 0$ that emerged during the collapse 'kindled' our Universe, and after several billion years, the nonlinear large-scale structure began developing to initiate the process of star formation. As a result, new black holes appeared; they can be entrances to new universes.

This process of evolution of a multisheet space–time resembles the growth of a tree (a genealogical tree, so to speak). Such a tree can flourish, or can wither if no new black holes form in the daughter universes. The critical situation appears when the development conditions do not provide the production of seed density fluctuations to form gravitationally bound matter clumps and their collapse into black holes. But another scenario can be realized: one collapse under favorable conditions, which gave rise to inflationary parameters and phase transitions in a white hole, can lead to the flourishing of a whole tree with nondecaying chains of new universes. Because of these processes, cosmological natural selection works [8]: only universes where black holes can be formed survive and develop, and this is possible for a certain set of parameters, world constants, etc.

This concept of a multisheet universe is based on the gravitational instability processes, which resemble oscillating tides. Anticollapsing space–time regions (white holes) stem from collapsing black holes, and, conversely, an expanding quasihomogeneous flow of matter of a white hole disintegrates into clumps collapsing into black holes. The former process is related to the $r = 0$ horizon and the latter is related to the $r = 2GM$ horizon. At both horizons, the gravitational potential is relativistic, and quantum gravitational processes of vacuum polarization and pair creation should be taken into account [9]. However, while these effects are suppressed by the mass parameter at the Schwarzschild horizon (the Hawking evaporation), they dominate at $r = 0$ and form structures of integrable singularities.

9. Conclusions

In the framework of our concept of integrable singularities, a new class of black-white holes in GR is obtained, which can be the key to solving the cosmogenesis problem. An integrable singularity at $r = 0$ can be compared with a classical cusp, in which the energy density or the longitudinal

tension of matter diverges, but the mass is zero,[3] $m(0) = 0$, and the gravitational potential $r = 0$ is bounded. Because of this property, the tidal forces are finite and any geodesics freely extend from a black hole into a white hole, where the geometry is equivalent to that of an expanding cosmological model.

The mass of matter in a multisheet universe can be arbitrarily large, because it is compensated by the negative gravitational binding energy. Hence, the total energy of holes, measured in static zones, is constant in time. The integrable singularities are reminiscent of machines for reprocessing gravitational degrees of freedom into material ones; however, the quantitative characteristics of this process can be determined only in the self-consistent quantum theory.

The work was supported by the Ministry of Education and Science of Russia grant 16.740.11.0460 of 13.05.2011 and the RFBR grant ofi-11-02-00857. VNS thanks FAPEMIG for support.

10. Appendix. Motion of test particles in a black-white hole

We consider the motion of test particles in metric (2), which is independent of the time t. Let $k^\mu = \mathrm{d}x^\mu/\mathrm{d}\lambda$ and λ be the tangent vector, and the affine parameter along the trajectory be $x^\mu = x^\mu(\lambda)$. From the geodesic equation $k^\mu k_{\nu;\mu} = 0$, we obtain

$$k \equiv k_t = \mathrm{const}\,, \quad k_\perp \equiv k_\theta + \frac{k_\varphi}{\sin^2\theta} = \mathrm{const}\,, \tag{35}$$

where $k_\varphi = \mathrm{const}$ is the azimuthal angular momentum. Unlike invariants of motion for the longitudinal and transversal momenta, the value of the azimuthal number depends on the orientation of the polar coordinate system. By adjusting the pole $\theta = 0$ with one of the points on the trajectory, we have $\varphi = \mathrm{const}$ and $k_\varphi = 0$ everywhere on the world line of a particle. In the projection on the 2-sphere, the particle moves along a meridian with the angle θ monotonically increasing within the interval 2π by the number of turns in $\mathbb{S}^2$:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = -\frac{k_\perp}{\rho}\,, \quad \left(\frac{\mathrm{d}\rho}{2\,\mathrm{d}\lambda}\right)^2 - (k_\perp^2 + n\rho)\,\mathcal{K} = \frac{k^2\rho}{N^2}\,. \tag{36}$$

The equation for $\lambda(\rho)$ follows from the normalization integral $k_\mu k^\mu \equiv n = \mathrm{const}$, where $n = 0$ for the null and $n = 1$ for the time-like geodesics. Inside the cone $\rho = r^2 \geqslant 0$ of a black-white hole, Eqn (36) has the form

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + \phi = \frac{k^2}{N^2} - n\,, \quad \phi = \frac{k_\perp^2}{r^2}\,(1 + 2\Phi) + 2n\Phi\,, \tag{37}$$

where the potential $\phi = \phi(r)$ tends to zero as $r \to \infty$ (Fig. 3).

The longitudinal light geodesics ($n = k_\perp = 0$) propagate directly from the black hole into the white hole with the continuous affine time $\lambda = -k^{-1}\int N\,\mathrm{d}r$. The spiral light geodesics on the cross sections ($n = k = 0$) lie in the T-region of the black or white hole, $\mathcal{K} = K^2 \geqslant 0$, coming from AdS zones and going back to AdS zones: $\theta = -k_\perp \int \mathrm{d}\lambda/\rho = \pm\int \mathrm{d}\rho/(2\rho K)$. Of special interest are photons with $n = k = k_\perp = 0$ living on the horizons $\rho = 0$:

$$\theta = \pm\frac{\ln|\lambda - \lambda_0|}{2K_0}\,, \quad \rho = \mp 2K_0(\lambda - \lambda_0)\,k_\perp \to 0\,, \tag{38}$$

where λ_0 is the value of the affine parameter on the bifurcation line.

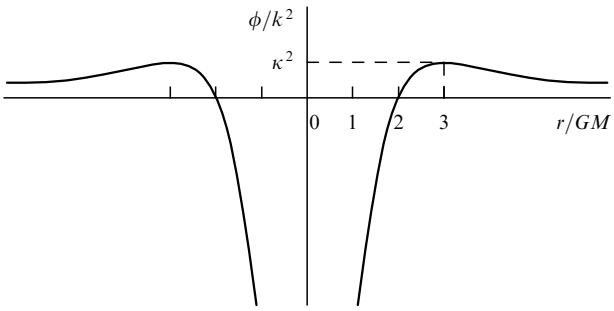

Figure 3. The potential $\phi(r)$ of light geodesics in a black-white hole.

Trajectories of photons with the impact parameter $\kappa \equiv k_\perp(3\sqrt{3}GMk)^{-1} < 1$ connect the R- and T-regions. Photons with $\kappa = 1$ return to R-regions at the radius $r = 3GM$. For $\kappa > 1$, there are photons of two kinds: in the R-region with $r > 3GM$, and in the zone $r < 3GM$ that unites the T-region and the inner part of the R-region adjacent to it.

[3] The time-like axis of the zero mass (the line of the star center) lies in the R-region, and the mass function can be found by integrating the density over radial shells [see (9)]: $m = 4\pi\int_0 \epsilon(r)\,r^2\,\mathrm{d}r$. The space-like axis of the zero mass (bifurcation lines [see (13)]) lies in the T-region, and the mass function of a black-white hole can be obtained by integrating the longitudinal tension over time: $m = -4\pi\int_0 p(r)\,r^2\,\mathrm{d}r$.

References

1. Lukash V N, Strokov V N, arXiv:1109.2796
2. Lukash V N, Mikheeva E V, Strokov V N *Usp. Fiz. Nauk* **182** 216 (2012) [*Phys. Usp.* **55** 204 (2012)]
3. Sakharov A D *Zh. Eksp. Teor. Fiz.* **87** 375 (1984) [*Sov. Phys. JETP* **60** 214 (1984)]
4. Hawking S, Penrose R *The Nature of Space and Time* (Princeton, N.J.: Princeton Univ. Press, 1996)
5. Lukash V N, Mikheeva E V *Fizicheskaya Kosmologiya* (Physical Cosmology) (Moscow: Fizmatlit, 2010)
6. Novikov I D *Pis'ma Zh. Eksp. Teor. Fiz.* **3** 223 (1966) [*JETP Lett.* **3** 142 (1966)]
7. Novikov I D *Astron. Zh.* **38** 564 (1961) [*Sov. Astron.* **5** 423 (1961)]
8. Smolin L *The Life of the Cosmos* (Oxford: Oxford Univ. Press, 1999)
9. Frolov V P, Shapiro I L *Phys. Rev. D* **80** 044034 (2009)