Physics – Uspekhi **53** (4) 397 – 424 (2010)

CONFERENCES AND SYMPOSIA

PACS numbers: 01.10.Fv, **26.30.–k**, **26.35.+c**, **95.35.+d**, **95.36.+x**, 95.55.Ka, 95.85.Pw, 98.38.Dq, 98.62.Hr, **98.65.–r**, 98.70.Rz, 98.80.Cq

Extragalactic astronomy
(Scientific session of the Physical Sciences Division of the Russian Academy of Sciences, 28 October 2009)

DOI: 10.3367/UFNe.0180.201004e.0415

A scientific session of the Physical Sciences Division of the Russian Academy of Sciences (RAS), "Extragalactic astronomy," was held in the Conference Hall of the Kapitza Institute for Physical Problems, RAS, on 28 October 2009.

The following reports were put on the session agenda posted on the web site www.gpad.ac.ru of the Physical Sciences Division, RAS:

(1) **Varshalovich D A, Ivanchik A V, Balashev S A** (Ioffe Physical Technical Institute, RAS) "Big Bang nucleosynthesis of deuterium and HD/H_2 molecular abundances in interstellar clouds of 12 Gyr ago";

(2) **Aptekar R L, Golenetskii S V, Mazets E P** (Ioffe Physical Technical Institute, RAS) "Studies of cosmic gamma-ray bursts and gamma repeaters with the Ioffe Institute Konus experiments";

(3) **Beskin G M, Karpov S V** (Special Astrophysical Observatory, RAS), **Bondar S V** (Scientific Research Institute of Precision Instrument Making) "Discovery of the fast optical variability of the GRB 080319B gamma burst and the prospects for wide-angle high time resolution optical monitoring";

(4) **Starobinskii A A** (Landau Institute for Theoretical Physics, RAS) "Experimental and theoretical investigation of dark matter in the Universe";

(5) **Zasov A V, Sil'chenko O K** (Shternberg State Astronomical Institute, Lomonosov Moscow State University) "Galactic disks and their evolution";

(6) **Burdyuzha V V** (Astro-Space Center of the Lebedev Physics Institute) "Dark components of the Universe."

Papers based on reports 1–3, 5, and 6 are published below. A A Starobinskii's extended report will be presented in the form of a review, which is planned for publication in one of the forthcoming issues of *Physics–Uspekhi*.

Uspekhi Fizicheskikh Nauk **180** (4) 415 – 444 (2010)
DOI: 10.3367/UFNr.0180.201004e.0415
Translated by R L Aptekar, K A Postnov, E N Ragozin; edited by A M Semikhatov

PACS numbers: **26.30.–k**, **26.35.+c**, 98.38.Dq
DOI: 10.3367/UFNe.0180.201004f.0415

Big Bang nucleosynthesis of deuterium and HD/H_2 molecular abundances in interstellar clouds of 12 Gyr ago

D A Varshalovich, A V Ivanchik,
S A Balashev, P Petitjean

1. Big Bang nucleosynthesis

It has been widely discussed recently that previously unknown forms of matter and energy, cold dark matter (CDM) and dark energy (DE), dominate in the Universe, while ordinary baryonic matter (consisting of protons and neutrons) contributes only 4–5% to the total energy density. This statement is based on the detailed quantitative analysis of the Big Bang nucleosynthesis (BBN) and careful measurement of the deuterium abundance in the interstellar matter of the most distant galaxies and protogalaxies [1, 2].

As is well known, there was a short stage in the early Universe when the temperature decreased such that the synthesis of nuclides from protons and neutrons became effective. At that time, the Universe was a unique thermonuclear reactor filled with almost homogeneous fully ionized plasma with the temperature $T \sim 10^9 - 10^8$ K. But that stage was so short that only the lightest nuclides ^{2}D, ^{3}T, ^{3}He, ^{4}He, ^{6}Li, ^{7}Li, and ^{7}Be could be formed in appreciable amounts. The relative abundance of relic nuclides formed during BBN can be reliably calculated because the rates of all relevant reactions are well known. The relative baryon number density is the only free parameter. This parameter can be determined from comparison of the results of calculations with astronomical observations of the abundance of relic nuclides.

Figure 1a shows the change in nuclide composition with time t and temperature T, which at that time were related as $t = AT^{-2}$, where the coefficient A is determined by the effective number of the degrees of freedom of ultrarelativistic

D A Varshalovich, A V Ivanchik Ioffe Physical Technical Institute, Russian Academy of Sciences, St. Petersburg, Russian Federation
E-mail: varsh@astro.ioffe.ru
Saint Petersburg State Polytechnic University, St. Petersburg, Russian Federation
S A Balashev Ioffe Physical Technical Institute, Russian Academy of Sciences, St. Petersburg, Russian Federation
P Petitjean Institut d'Astrophysique de Paris, France

Uspekhi Fizicheskikh Nauk **180** (4) 415 – 420 (2010)
DOI: 10.3367/UFNr.0180.201004f.0415
Translated by K A Postnov; edited by A M Semikhatov

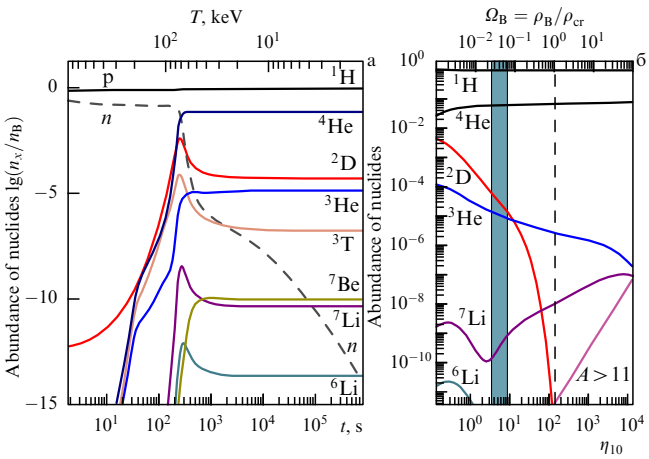

Figure 1. Primordial nucleosynthesis. (a) Change in the composition of matter with time t and, accordingly, with the temperature T. (b) The relative abundance of the relic nuclides eventually formed as a function of the baryon-to-photon number density ratio $\eta_{10} = 10^{10}(n_B/n_\gamma)$. The solid lines show the theoretical calculations, the dark vertical strip indicates the range of η_{10} obtained from different astronomical observations of the relic nuclide abundance, and the dashed line corresponds to $\rho_B = \rho_{cr}$.

particles (photons, neutrinos, and antineutrinos of all flavors).

Figure 1b presents the results of calculations of the final abundance of relic nuclides as a function of the baryon-to-photon number density ratio n_B/n_γ. The dark strip in the figure shows the n_B/n_γ range corresponding to astronomical measurements of the relative abundance of relic nuclides. It can be seen from this figure that the relative abundance of primordial deuterium $(D/H)_p$ is the most sensitive indicator of the baryon-to-photon number density ratio n_B/n_γ. For example, if the density of baryonic matter were equal to the critical one, $\rho_B = \rho_{cr}$, we would not observe primordial deuterium at all, because its abundance would be eight orders of magnitude smaller than the present value.

The isotopic ratio D/H could only decrease in the course of the subsequent evolution, because deuterium rapidly burns out into helium in stellar interiors. Therefore, to determine $(D/H)_p$, the isotopic composition of the interstellar medium should be measured at as early a cosmological epoch as possible. The absorption spectra of high-redshift quasars are the most appropriate. Quasars are the most powerful sources of radiation, and can be observed from distances up to 10–12 bln light years. In other words, the spectra of quasars measured at the present time were formed 10–12 Gyr ago. The emission of quasars themselves is used to probe remote clouds of interstellar and intergalactic gas along the line of sight, which 'imprint' their absorption lines into the quasar spectrum. It can be said that quasars serve as an 'X-ray device' to shine through the Universe.

Until recently, the relative abundance of D/H was measured using only atomic lines of H I and D I in quasar absorption spectra. However, such measurements met with some difficulties. Optical spectra of H I and D I are virtually the same, with all wavelengths being slightly shifted by 0.027%. But the number densities of these atoms differ by 4–5 orders of magnitude, and hence if the column density along the line of sight of H I atoms is small, D I lines cannot be seen at all. If the hydrogen column density is too high, H I lines are saturated, broadened, and blended with D I lines. Moreover, the lines identified as D I lines can in principle originate from a small H I cloud moving relative to the cloud under study with a speed ~ 80 km s^{-1}; many such clouds moving with different speeds are indeed found along the line of sight (the so-called Lyman-alpha forest). The significant dispersion of values of the D/H ratio measured this way can possibly be explained by these facts (see Section 3, Fig. 5).

Difficulties in the line identification would not emerge in measurements of the relative abundance of HD and H_2 molecules because their spectra are significantly different and most absorption lines are unblended. This method has not been used until recently because no HD molecules have been found at high redshifts.

2. An HD/H_2 cloud with the redshift $z = 2.33771$

We were the first to discover hydrogen deuterium molecules at high redshifts. The HD lines were identified in the absorption spectrum of quasar Q 1232+082 redshifted such that $\lambda_{obs}/\lambda_{em} = 1 + z = 3.337714(3)$ [3]. In fact, a cloud containing HD molecules along with H_2 was first discovered at a large cosmological distance. This cloud existed about 12 Gyr ago, when neither the Solar System nor our Galaxy had formed.

Figures 2 and 3 show the scheme of energy levels and the Lyman series transitions Lv'' – 0 of H_2 and HD molecules and parts of the quasar spectra with the corresponding lines. For HD, R_0 lines corresponding to transitions from the ground state $J = 0$ are visible, while for H_2, lines corresponding to transitions from the ground state of para-H_2 ($J = 0$), ortho-H_2 ($J = 1$), and excited rotational states R_J and P_J with $J = 2$, 3, 4, 5 are also seen, because the lifetimes τ_J of H_2 molecules in these states are orders of magnitude longer than this time for HD molecules.

After additional careful measurements of these quasar spectra carried out by French colleagues on the 8.2 m VLT (Very Large Telescope) of the European Southern Observatory (ESO) in Chile, we have determined the column number densities of HD and H_2 molecules in this cloud: $N(\mathrm{HD}) = (3.4^{+1.6}_{-0.8}) \times 10^{15}$ cm^{-2} and $N(H_2) = (4.8 \pm 1.0) \times 10^{19}$ cm^{-2} [4]. The HD/H_2 column number density ratio turned out to be $(7.1^{+3.7}_{-2.2}) \times 10^{-5}$. This value significantly exceeds those measured in interstellar clouds of our Galaxy, $(0.4-4.0) \times 10^{-6}$ [5]. Figure 4 shows the obtained spectrum of quasar Q 1232+082 [4] in comparison with the absorption spectrum of ζ Oph [6], which is typical for an interstellar molecular cloud in our Galaxy. Both spectral fragments correspond to the same comoving wavelength range, but the quasar spectrum is observed redshifted 3.33771-fold. The comparison of spectra demonstrates that the relative abundance of HD/H_2 molecules in the cosmological cloud that existed 12 Gyr ago is much higher than the present-day value. A quantitative estimate gives a 200-fold difference.

The analysis of the obtained spectrum revealed that this cloud does contain a small amount of matter processed in stars, i.e., deuterium deficient. This is evidenced by the presence of certain spectral lines, in particular, the Ar I line shown in Fig. 4. The quantitative analysis in [7] demonstrated that the elemental abundance of such species as Si, P, S, Ar, and Ti is 30–40 times smaller than the solar one. For example, the abundance of Ar in the cloud amounts to $(2.4^{+1.9}_{-1.0})$% of the solar value. In the protosolar nebula, the D/H ratio was estimated to be 1.94×10^{-5} [8], i.e., differed from the primordial value by 30–50%, and hence there are grounds to believe that the D/H ratio in the

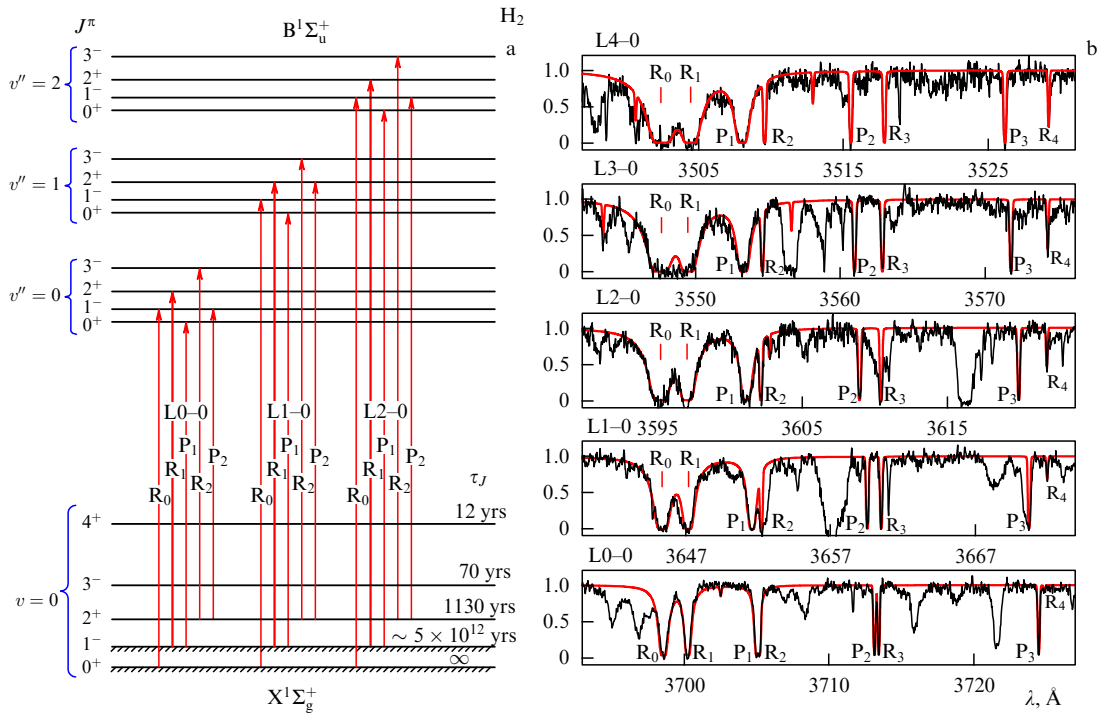

Figure 2. (a) Scheme of the Lyman series levels and $Lv''-v$ transitions in H_2 molecules. (b) Parts of the spectrum of Q 1232+082 quasar in which the corresponding H_2 absorption lines are seen, with wavelength cosmologically redshifted as $\lambda_{obs}/\lambda_{em} = 3.33771$.

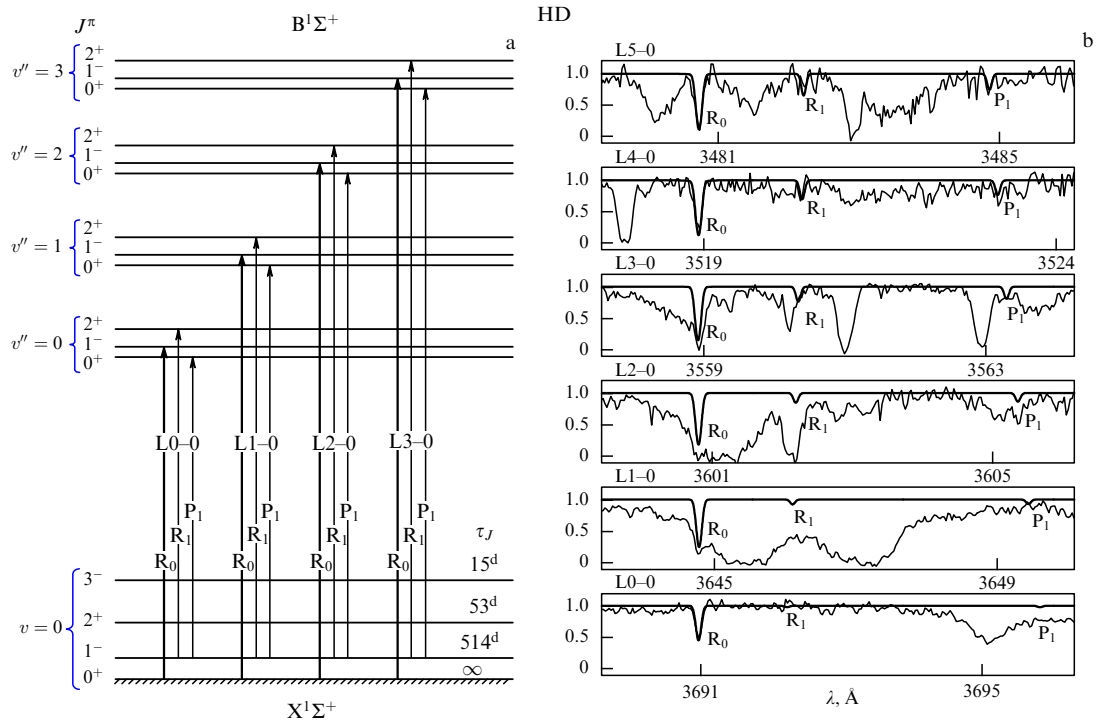

Figure 3. (a) Scheme of levels and transitions in HD molecules. (b) Parts of the spectrum in which HD lines are redhsifted as $\lambda_{obs}/\lambda_{em} = 3.33771$.

studied cosmological cloud differs from the primordial value $(D/H)_p$ by only 1–2%.

Therefore, the studied absorption system represents a cold rarefied cloud of interstellar gas in a remote galaxy or protogalaxy. The kinetic gas temperature inferred from the relative abundance of ortho- and para-H_2 is $T = 67 \pm 11$ K. The mean gas number density derived from the relative populations of the fine-structure levels of C I is (71 ± 28) cm^{-3}. The size of the cloud is estimated to be about one pc. Spectral analysis showed that O I, N I, and Ar I atoms with an ionization potential higher than 13.6 eV are present in the cloud only in neutral states, while C, Mg, Al, Si, P, S, Ti, Mn, and Fe with an ionization potential lower than 13.6 eV are present as neutral and single-ionized atoms (predominantly). This fact allows estimating the flux of ionizing ultraviolet (UV) radiation in the cloud. The radia-

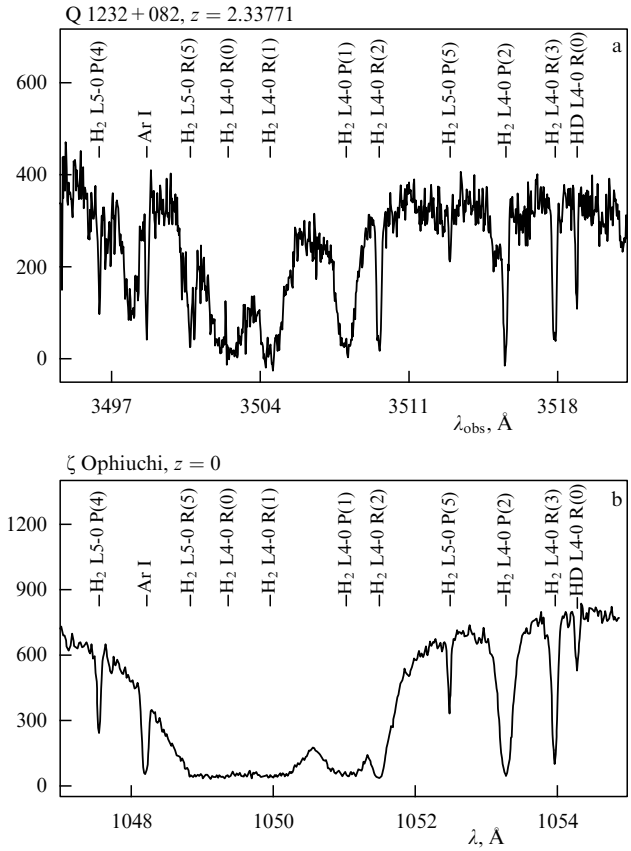

Figure 4. The spectrum of quasar Q 1232+078 [4] in comparison with the absorption spectrum of ζ Oph [6], which is typical for interstellar molecular clouds in our Galaxy. Both spectral fragments correspond to the same wavelength range in the comoving reference system, but the wavelength scale is redshifted 3.33771-fold for an astronomer on Earth. The comparison of spectra shows that the relative abundance of HD/H_2 molecules in the cloud that existed 12 Gyr ago was significantly higher than the present-day value.

tion transfer in H_2 lines in the studied cloud is treated in detail in paper [9].

Thus, detailed analysis of the obtained spectrum allowed us to reveal evolutionary changes in the chemical and isotopic composition of matter over 12 Gyr and to determine the physical conditions in the cloud at that time.

3. Relative abundance of D/H isotopes and baryonic content

To determine the abundance of the isotope ratio D/H from the relative abundance of HD/H_2 molecules, it is necessary in general to consider all channels of creation and destruction of these molecules. The dominating destruction channel of H_2 and HD molecules is due to dissociation by UV radiation in resonance lines. The UV radiation excites molecules from the ground state $X^1\Sigma_g^+$ to upper electronic states $B^1\Sigma_u^+$ and $C^1\Pi_u$, i.e., the excitation occurs in lines of the Lyman and Werner bands in the 1120–912 Å wavelength range (just these lines are observed in quasar spectra). About 87% of the molecules excited by UV radiation return to the ground electron state (to different vibration–rotational levels) and about 13% of the molecules dissociate. But radiation in the cloud is absorbed in lines of the Lyman and Werner series, and the lines saturate, and hence H_2 and HD molecules inside the cloud are shielded from the destroying action of UV radiation. The self-shielding of the cloud begins when the optical depth in resonance lines increases to unity, which corresponds to the molecular column density $\sim 4 \times 10^{14}\ \mathrm{cm}^{-2}$. In our case, the measured column density of H_2 and HD molecules is far above this threshold. Under such conditions, calculations of interstellar molecular clouds [10–12] show that virtually all gas must be molecularized, i.e., all D must be in the form of HD molecules and all H in the form of H_2 molecules. In this case, the universal relation $HD/H_2 = 2(D/H)$ sets in. This fact allowed us to determine the isotopic ratio in the cloud: $D/H = (3.6^{+1.9}_{-1.1}) \times 10^{-5}$ [4].

The obtained value of D/H is compared with results of other studies in Fig. 5, in which deuterium and other relic

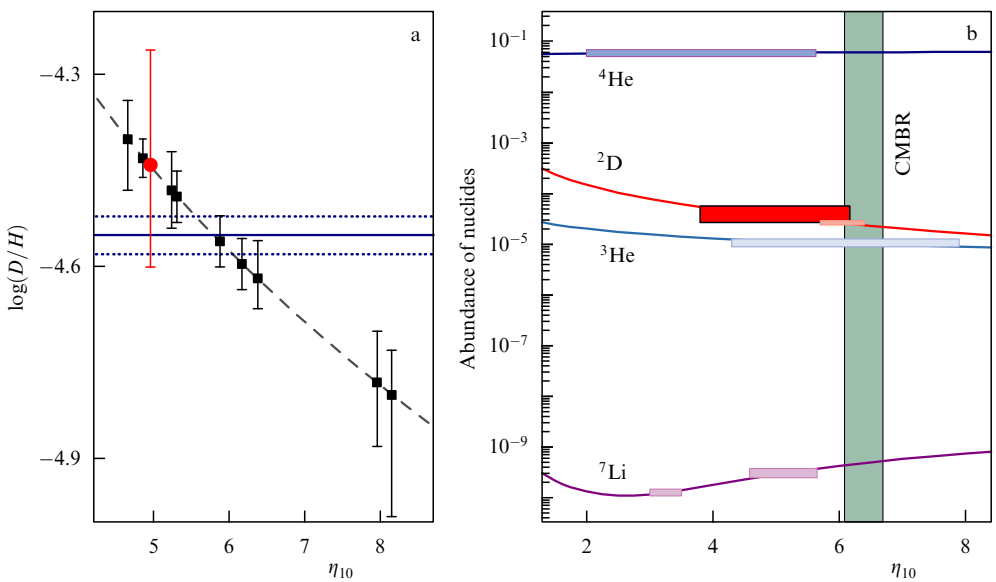

Figure 5. The relative abundance of deuterium and other relic nuclides as a function of the relative baryonic number density $\eta_{10} = 10^{10}(n_B/n_\gamma)$. (a) The results of measurements of D/H in quasar spectra using absorption lines of D I/H I atoms (the filled squares) and HD/H_2 molecular lines (the filled circle) [4]. (b) The results of measurements of the relative abundance of relic nuclides compared with theoretical curves and with the value n_B/n_γ obtained from a CMB anisotropy analysis [13].

nuclide abundance is plotted as a function of the relative baryon-to-photon number density ratio n_B/n_γ. Figure 5a demonstrates the value of (D/H) measured in quasar spectra from absorption lines of D I/H I atoms and from HD/H_2 molecular lines (see [4] for data and references). In Fig. 5b, the results of measurements of the relative abundance of relic nuclides ^{2}D, ^{3}He, ^{4}He, and ^{7}Li are compared with theoretical calculations and with the n_B/n_γ ratio derived from the analysis of angular fluctuations of the CMB temperature [13].

With the obtained value of (D/H), the BBN theory implies the baryon-to-photon number density ratio $n_B/n_\gamma = (5 \pm 1) \times 10^{-10}$ [4]. In calculating the kinetics of the primordial nucleosynthesis, we have used a modified code developed in [14, 15]. According to the standard cosmological model, the n_B/n_γ ratio should not have changed from BBN times to the present day, i.e., it is assumed that baryons have not decayed and CDM particles have not been converted into baryons. At present, the CMB photon number density corresponding to the temperature $T_0 = 2.726$ K is $n_\gamma = 411$ cm^{-3}. Therefore, the mean baryon number density must be $n_B(0) = (2.1 \pm 0.4) \times 10^{-7}$ cm^{-3}. But in contrast to the BBN epoch when the medium was almost homogeneous, the present-day space matter distribution is highly inhomogeneous, and hence $n_B(0)$ is the mean value in a large space region of the order of 100 Mpc. The obtained value $n_B(0)$ corresponds to $(3.7 \pm 0.8)\%$ of the critical density (for the assumed Hubble constant $H_0 = 70$ km s^{-1} Mpc^{-1}). This is consistent within error with the value obtained from the CMB temperature fluctuations [13].

4. Conclusion

For the first time, a cloud of cold rarefied gas containing H_2 and HD molecules was discovered at a large cosmological distance. The elemental abundance and physical conditions in the cloud that existed 12 Gyr ago have been studied. The column density ratio HD/H_2 in the cloud is shown to be much higher than the present-day value measured in interstellar clouds of our Galaxy [4].

The relative isotopic abundance D/H in this cloud was measured by a novel (independent) method from the column density ratio of HD/H_2 molecules using molecular cloud models [10–12]. The derived value $D/H = (3.6^{+1.9}_{-1.1}) \times 10^{-5}$ and the standard BBN model allowed an independent determination of the mean baryonic density at the present time $\Omega_B \equiv \rho_B/\rho_{cr} = (3.7 \pm 0.8)\%$ [4].

This work was supported by the Program of the Presidium of the Russian Academy of Sciences "Origin, Structure, and Evolution of Objects in the Universe," the Russian Foundation for Basic Research grant No. 09-02-01264a, and grant NSh-3769.2010.2.

References

1. Steigman G *Annu. Rev. Nucl. Part. Syst.* **57** 463 (2007)
2. Pettini M et al. *Mon. Not. R. Astron. Soc.* **391** 1499 (2008)
3. Varshalovich D A et al. *Pis'ma Astron. Zh.* **27** 803 (2001) [*Astron. Lett.* **27** 683 (2001)]
4. Ivanchik A V et al. *Mon. Not. R. Astron. Soc.* **404** 1583 (2010), doi: 10.111/j.1365-2966.2010.16383.x; arXiv:1002.2107
5. Lacour S et al. *Astron. Astrophys.* **430** 967 (2005)
6. Spitzer L (Jr.), Jenkins E B *Annu. Rev. Astron. Astrophys.* **13** 133 (1975)
7. Balashev S A et al, in preparation
8. Lodders K, Palme H, Gail H-P, in *Landolt-Börnstein, New Series. Astronomy and Astrophysics* (Berlin: Springer-Verlag) (in press); arXiv:0901.1149
9. Balashev S A, Varshalovich D A, Ivanchik A V *Pis'ma Astron. Zh.* **35** 171 (2009) [*Astron. Lett.* **35** 150 (2009)]
10. Abgrall H et al. *Astron. Astrophys.* **253** 525 (1992)
11. Le Bourlot J *Astron. Astrophys.* **360** 656 (2000)
12. Le Petit F, Roueff E, Le Bourlot J *Astron. Astrophys.* **390** 369 (2002)
13. Komatsu E et al. *Astrophys. J. Suppl.* **180** 330 (2009); arXiv: 0803.0547
14. Orlov A V, Varshalovich D A "Kinetika pervichnogo nukleosinteza: novyi paket programm" ("Kinetics of the primordial nucleosynthesis: a new code"), Preprint FTI No. 1719 (St. Petersburg: FTI RAN, 1999)
15. Ivanchik A V, Orlov A V, Varshalovich D A *Pis'ma Astron. Zh.* **27** 723 (2001) [*Astron. Lett.* **27** 615 (2001)]

PACS numbers: 95.55.Ka, 95.85.Pw, 98.70.Rz
DOI: 10.3367/UFNe.0180.201004g.0420

Cosmic gamma-ray bursts and gamma repeaters studies with Ioffe Institute Konus experiments

R L Aptekar, S V Golenetskii, E P Mazets, V D Pal'shin, D D Frederiks

This report presents the results of the study of cosmic gamma-ray transients, performed for many years by the Ioffe Institute onboard interplanetary space missions and near-Earth satellites. These transients have been found in several classes: many of the basic characteristics of cosmic gamma-ray bursts (GRBs), such as their temporal profiles and energy spectra, and the first all-sky map of their source distribution on the celestial sphere, were determined using early Konus experiments onboard the Venera 11, 12, 13, and 14 deep space vehicles. As a result of the observations of the giant flare on 5 March 1979 and the discovery of a series of short gamma-ray bursts, also determined to have a common source and to come from the same source direction as the giant flare, this new class of rare astrophysical phenomena was discovered. Somewhat later, these were named soft gamma repeaters (SGRs). In the subsequent Konus and Helicon experiments, studies of cosmic gamma-ray bursts and of soft gamma repeaters were continued, new SGR sources were found, and other giant SGR flares were also found. GRBs come from very distant cosmological sources, but SGR sources, analyzed in detail, have been localized in our Galaxy and in its satellite, the Large Magellanic Cloud (LMC). Recently and most notably, Konus measurements have been central in finding SGRs in the nearby galaxies M81 and M31, far outside our own Milky Way system. In the modern epoch of multi-wavelength studies of cosmic gamma-ray bursts, the joint Russian–American Konus–Wind experiment, which has already been operating for more than 15 years, provides important and often unique data regarding the various characteristics of these diverse types of transient gamma-ray emissions in the 10 keV to 10 MeV energy range.

R L Aptekar, S V Golenetskii, E P Mazets, V D Pal'shin, D D Frederiks
Ioffe Physical Technical Institute, Russian Academy of Sciences, St. Petersburg, Russian Federation. E-mail: aptekar@mail.ioffe.ru

Uspekhi Fizicheskikh Nauk **180** (4) 420 – 424 (2010)
DOI: 10.3367/UFNr.0180.201004g.0420
Translated by R L Aptekar; edited by A M Semikhatov

The nature of cosmic gamma-ray bursts as extremely explosive releases of electromagnetic energy has been the focus of astrophysicists' attention since soon after their discovery in 1967–1973 by the U.S. Vela satellites [1]. One of the earliest confirmations of the discovery of this new phenomenon was provided by observations collected by the Ioffe Institute using data from the Kosmos 461 satellite [2]. Later, an important breakthrough in the studies of gamma-ray bursts was made possible by the four Konus experiments carried out by the Ioffe Institute onboard the Venera 11 to 14 deep space missions in 1978 to 1983. These missions were among the first to make measurements of the various observational properties of gamma-ray bursts, which still constitute the basis of modern concepts in GRB research. In particular, Konus observations of the temporal structures of gamma-ray bursts were central to the discovery of the existence of a separate class of short bursts, demonstrating the so-called 'bimodal' duration distribution. In addition, the array of Konus detectors, each having anisotropic sensitivity, could be used along with the intersatellite triangulation method for a stand-alone determination of approximate source directions. It was for the first time shown with good statistical accuracy that their distribution over the celestial sphere is random (Fig. 1) [3, 4]. Later, this result was confirmed with an even larger number of events in the well-known BATSE experiment onboard NASA's Compton Gamma-Ray Observatory [5].

An extremely interesting and totally unexpected result was that a new class of gamma-ray transients was found using the Konus experiments onboard the early Venera deep-space missions. Occurring very rarely, yet found to repeat at random, these were later named soft gamma repeaters. These sources exhibit two kinds of bursting activity: the first is the emission of soft, repeated bursts with a duration ~ 0.1 s and an isotropic energy release $\sim 10^{39}$ to 10^{41} erg. The occurrences of the bursting activity of SGRs are highly nonuniform in time; the sources are predominantly in the quiescent state, which may continue for years, but is interrupted by periods of intense reactivation, with several to several hundred events at a time. The other, far more impressive kind of repeater activity consists of even more rare giant flares with the peak power comparable to the luminosity of quasars ($10^{45}-10^{47}$ erg s^{-1}). The relation of the flares to soft gamma repeaters was first manifested with the observation on 5 March 1979 of a giant outburst of hard X-ray and soft gamma radiation. Its source direction was shown by a network of international spacecraft to be that of a supernova remnant in the Large Magellanic Cloud at the distance 55 kpc; independently, its source was shown by Konus data to precisely agree with the source of the first soft gamma repeater series. The flare profile consisted of a narrow, exceedingly intense, initial emission peak with a harder energy spectrum, followed by a weaker decay with a softer spectrum, which exhibited pulsations having a period of 8 seconds for several minutes (Fig. 2a, [6, 7]). Already on the next day, the first of the repeated events from the same source direction was recorded by Konus instrumentation. Over several years, 17 repeated bursts of this kind were detected from SGR 0526-66 (Fig. 2b, [8]). The giant flare of 5 March 1979 had been a unique astrophysical event for nearly 20 years when, on 27 August 1998, a giant flare was detected by many spacecraft, with the source agreeing with that of SGR 1900+14. It was shown with the Konus–Wind experiment and the Ulysses mission to have the same specific

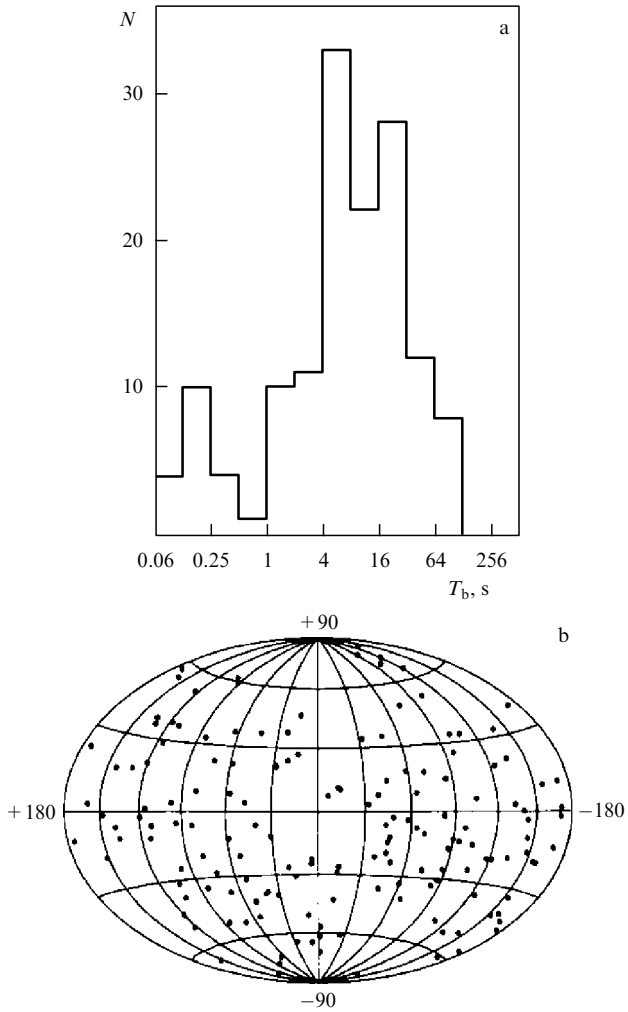

Figure 1. The duration distribution of gamma-ray bursts and the first statistically meaningful distribution of GRB sources over the celestial sphere, determined by the Konus experiments onboard the Venera 11 to 14 interplanetary missions from 1978 to 1983.

features as the first repeater, SGR 0526-66. A third, even stronger flare of this kind was observed on 27 December 2004 from SGR 1806-20. According to the commonly accepted concept, soft gamma repeaters are young ($\sim 10^3$ to 10^4 yr) neutron stars with rapidly slowing rotation rates and super-strong ($\sim 10^{15}$ G) magnetic fields.

A new and important chapter in the research of cosmic gamma-ray bursts and soft gamma repeaters at the Ioffe Institute is associated with a joint Russian–American experiment with the Russian Konus scientific instrument onboard the U.S. Wind spacecraft. The deep space trajectory of the Wind spacecraft is actively adjusted to have an apogee as great as a 5-light-second distance. Remote from Earth and the Moon, its exposure to the entire celestial sphere is exceedingly favorable for studies of unpredictable and transient sources. Two high-sensitivity scintillation detectors of the Konus–Wind gamma-spectrometer permanently view the entire celestial sphere, such that no event important to the astrophysics of gamma-ray bursts and gamma repeaters has yet been missed by the Konus–Wind experiment throughout the 15 years of its successful and uninterrupted operation. These advantages of the Konus–Wind experiment were clearly demonstrated by its observation in 1998 of a contrast

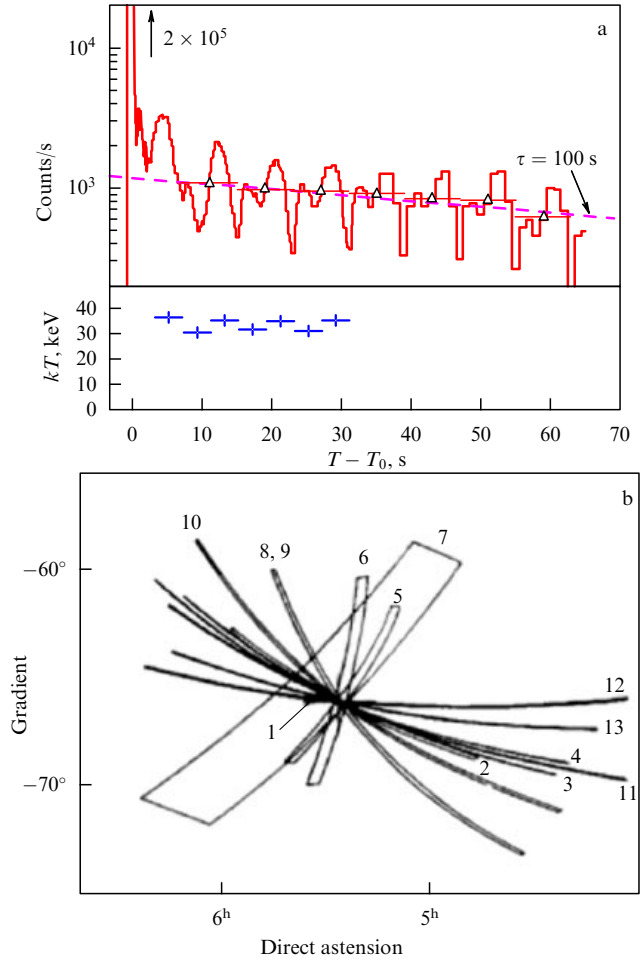

Figure 2. The giant flare on 5 March 1979 and repeated bursts from SGR 0526-66, measured by the Konus instrument onboard the Venera 11 and 12 interplanetary missions.

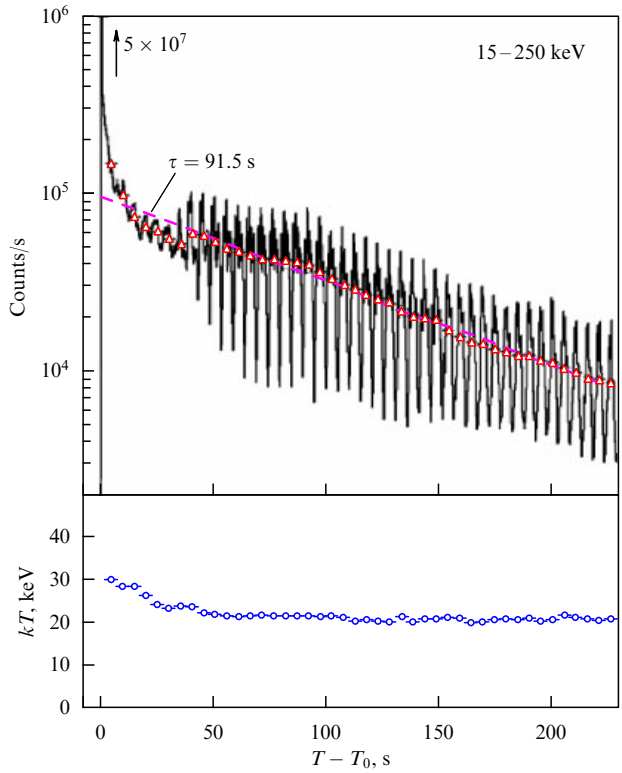

Figure 3. The giant flare from SGR 1900+14 on 27 August 1998, found by the Konus–Wind instrument.

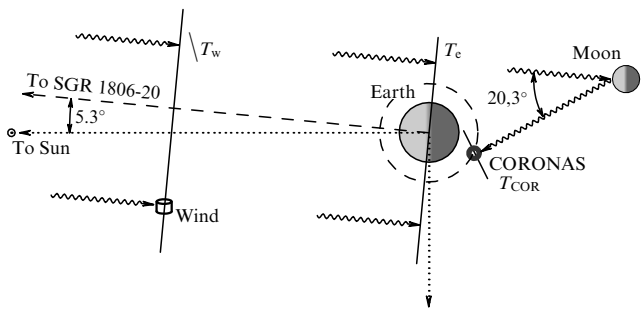

Figure 4. The reflection of the initial pulse of the giant flare from SGR 1806-20 by the Moon, and its detection by the Helicon instrument onboard the CORONAS-F spacecraft.

between two of the rarest celestial gamma-ray transients. The first — exceedingly powerful but having no pulsating tail — was observed on 18 June from the recently discovered SGR 1627-41 [9]. The second was a giant flare on 27 August from SGR 1900+14 (Fig. 3, [10]), in which all the typical indications of this phenomenon are observed: a short, exceedingly intense initial pulse of gamma radiation followed by a ≈ 500 s decay that pulsated with a period characteristic of a neutron star.

The photon counting rate of gamma rays in the initial pulse of a giant SGR flare is always so great that the sensitive detectors of the instrument are fully overloaded ('saturated'), such that precise measurements of the characteristics of the initial pulse become difficult and only rough lower-bound estimates are possible. For the most powerful of the recorded bursts, which came from the SGR 1806-20 source on 27 December 2004, the Konus–Wind detector was evidently saturated for over 1.5 s. However, the situation was fortunately quite favorable for the observation of a most unusual aspect of this event, due to the simultaneous measurements made using the Helicon spectrometer onboard the CORONAS-F orbiting solar observatory. This instrument, also from the Ioffe Institute, is identical to that of Konus–Wind. The positions of the spacecraft at the time of detection of this burst are shown schematically in Fig. 4 [11]. The detectors of the Helicon instrument were screened by Earth from direct exposure to the initial pulse of the giant burst from the gamma repeater, but clearly recorded its reflection from the Moon surface. This reduction in intensity allowed, for the first time, reliably reconstructing the temporal profile of the initial pulse of a giant burst from a gamma repeater (Fig. 5 [11]) and determining its energy parameters: the full isotropic energy release of 2.3×10^{46} erg and the peak luminosity 3.5×10^{47} erg s^{-1}. The research on the 27 December 2004 flare was the first example of studying Moon-reflected X-ray and gamma radiation coming from a source outside the Solar System.

The observations of giant flares from SGR 1900+14 and SGR 1806-20 have revived interest in a suggestion originally made in 1981–1982 by Mazets and Golenetskii [3] that some short, hard gamma-ray bursts that seem to be cosmic GRBs may instead be the initial pulses of giant flares from extragalactic SGRs. With the energy parameters of the

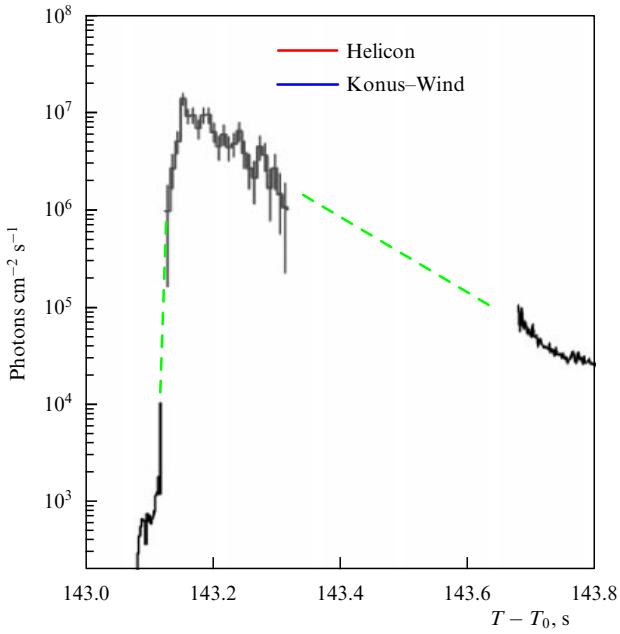

Figure 5. The temporal profile of the initial pulse of the giant flare from SGR 1806-20, reconstructed from Konus–Wind and Helicon data.

initial pulse of the giant flare from SGR 1806-20 having been determined, it became possible to estimate the maximum distance at which flares of this kind can be recorded: for modern detectors with wide field of view, this distance is 30 to 50 Mpc. The available data on giant flares from the known Galactic SGRs can be used to find the expected temporal and spectral characteristics of a short gamma-ray event that might instead be the detection of a flare from another galaxy. Such a burst must have the form of a single pulse with a steep leading edge and duration from approximately 5 to 15 ms, having an exponential decay with a time constant ≈ 50 to 70 ms and the total pulse width ~ 200 to 300 ms. The energy spectrum of the emission must be rather hard in the initial part, rapidly evolving to become noticeably softer by the end of the pulse. An analysis of the characteristics of short bursts on the basis of the Konus–Wind data demonstrated that the fraction of events with the above characteristics does not exceed several percent of the total number of short bursts.

On 3 November 2005, Konus–Wind recorded the short and hard bright gamma-ray burst GRB 051103. [12]. Localization of this event by the Interplanetary Network, which includes most of the experiments observing gamma-ray bursts, demonstrated that its source lay close to the M81 group of galaxies, situated at a distance ≈ 3.6 Mpc from Earth. The light curve of the event, recorded with a 2 ms time resolution, displayed a single pulse with a steep leading edge $\leqslant 6$ ms and exponential decay in approximately 55 ms. The event had the total duration 170 ms. The time variation of the radiation hardness pointed to a pronounced softening of the emission spectrum. Among the three detailed energy spectra recorded with the accumulation time 64 ms, only the first contained a high-intensity hard component with energies up to 10 MeV. The results of studies of matter distribution in the M81 group of galaxies, furnished by optical, X-ray, and radio astronomy, were compared with the localization data for the source of this gamma-ray burst. This source direction and the details of the event measurements support its identification as a giant flare from a gamma repeater located in the M81 group of galaxies. The flare energy would be $\sim 7 \times 10^{46}$ erg in this case, which is quite comparable to the energy of the giant flare in SGR 1806-20 [11].

Another intense gamma-ray event with a short duration and a hard spectrum was recorded by the Konus–Wind experiment on 1 February 2007 (Fig. 6a, [13]); it had the highest intensity in the entire history of gamma-ray extragalactic transient observations. The temporal profile of the event is a narrow pulse with a leading edge ≈ 20 ms and a more gently sloping decay, with the total duration ~ 180 ms. The burst exhibited a strong spectral evolution, with the hard emission component observed within the first 80 ms. The source localization region, determined by the Interplanetary Network and shown in Fig. 6b, coincides with the outer arms of the Andromeda Galaxy (M31), situated at the distance ≈ 780 kpc. A detailed analysis of the observations of the Andromeda Galaxy by IR, UV, and X-ray space telescopes and a comparison of the characteristics of GRB 070201 with those of the previously observed giant flares from gamma repeaters argue convincingly that this event is a giant flare from a gamma repeater in the Andromeda Galaxy. The sources of the events on 3 November 2005 and 1 February 2007, as the first soft gamma repeaters from beyond our local galaxy group, were designated SGR 0952+69 (M81) and SGR 0044+42 (M31) [13].

The Konus–Wind experiment revealed a strong spectral variability of a new source of soft repeated bursts, SGR 0501+4516, discovered in August 2008 by the BAT telescope onboard NASA's Swift spacecraft. Five bursts from the new gamma repeater were recorded and analyzed in detail by the Konus–Wind gamma spectrometer [14]. An important feature of this new repeater is the strong correlation between the emission hardness in the course of repeated bursts and the emission intensity. A correlation of this kind had been observed previously only for the SGR 1627-41 gamma repeater [9]. Among the recent results obtained in the Konus experiments, we note the discovery of a new source of soft repeated gamma-ray bursts, the SGR 0418+5729 gamma repeater, on 5 June 2009 [15]. This study was carried out in joint observations with GBM detectors of the Fermi observatory, the Konus-RF spectrometer of the CORONAS-Photon observatory, and the BAT telescope of the Swift observatory.

Owing to its high, omnidirectional sensitivity and location far from Earth, providing the optimum situation for observations of the entire sky at all times, the Konus–Wind experiment is a unique source of information about the temporal and spectral characteristics of gamma-ray events in the energy range from 10 keV to 10 MeV. These data constitute an essential element of modern multi-wavelength studies of gamma-ray transient events involving spacecraft with a network of ground-based optical and radio telescopes. Konus–Wind observations are thus in wide demand. Thanks to rapid and accurate localization of gamma-ray transient sources by the BAT telescope at the Swift observatory, studies by a global network of ground-based telescopes are commenced, in fact, within seconds to tens of seconds after the beginning of an event. Figure 7 shows the gamma-ray light curve from the Konus–Wind experiment and that obtained in the visible spectral range by TORTORA (Special Astrophysical Observatory, Russian Academy of Sciences) and the Polish Pi of the Sky telescopes for the famous GRB 080319B burst. Its luminosity in the optical spectral range reached the

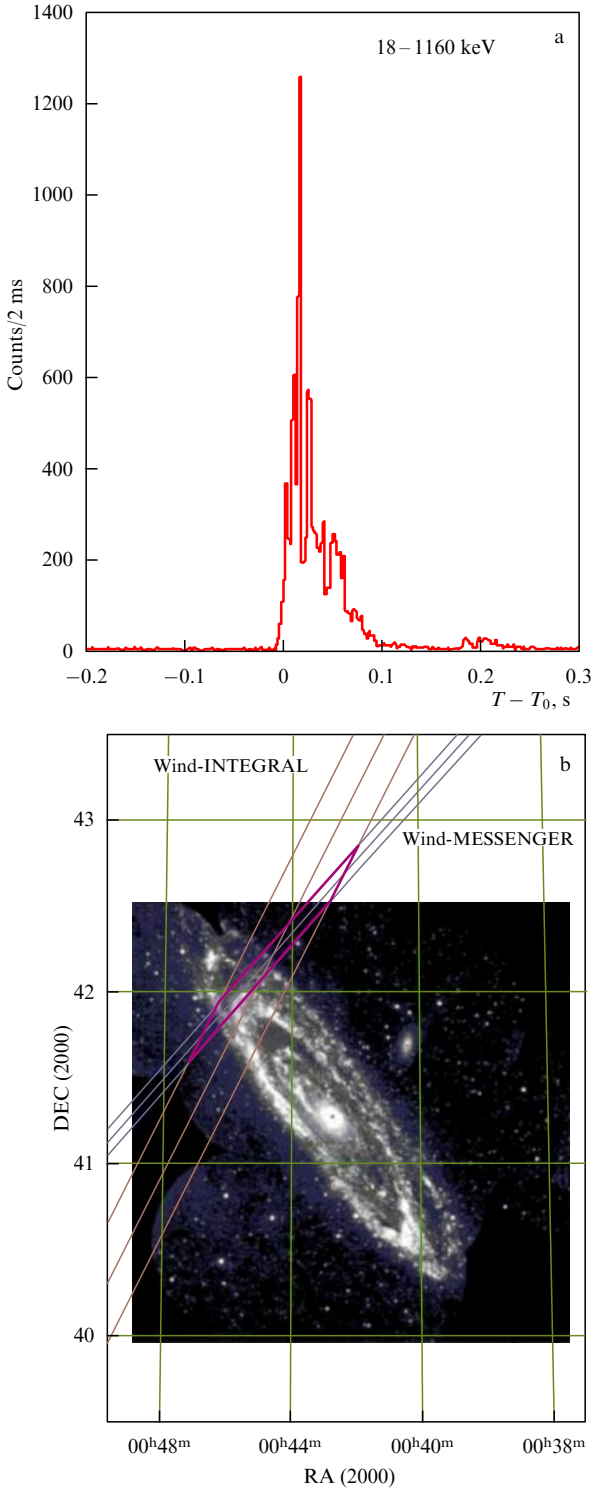

Figure 6. GRB 070201, a giant flare from the SGR 0044+42 gamma repeater in the Andromeda Nebula: the light curve from Konus–Wind data and the source localization by the Interplanetary Network.

apparent magnitude 5.5. Multi-wavelength studies demonstrate that the optical and gamma emissions of this burst begin and end at the same time, which strongly suggests that they originate from the same physical region [16].

In conclusion, we emphasize that the Konus instruments developed at the Ioffe Institute continue to be reliable and adequate to the task of studying cosmic gamma-ray bursts. Owing to the importance, quality, and completeness of the

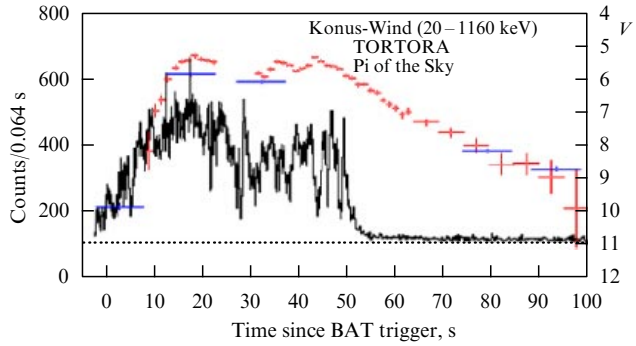

Figure 7. Optical and gamma emission of the GRB 080319B event, determined by Konus-Wind, TORTORA, and Pi of the Sky data [16].

information obtained, the Russian–American Konus–Wind experiment has advanced to the forefront of research into extremely explosive phenomena in the Universe. During the 15 years of uninterrupted observations, more than 3500 cosmic gamma-ray bursts have been detected and analyzed, and the bursting activity of all the known gamma repeaters has been studied. An earlier summary of data on gamma repeaters obtained with the Konus experiments can be found in "Konus Catalog of SGR Activity: 1978 to 2000" [17]. Printed and electronic versions of a second catalog of data obtained in observations of gamma repeaters in Konus–Wind, Konus, Helicon, and Konus-RF experiments are in preparation. This catalog covers the observation period from 1994 to 2009 and contains information about all the known soft gamma repeaters and giant SGR flares obtained from all the identical Konus instruments. Another summary of observational data regarding only the short gamma-ray bursts detected from 1994 to 2002 is contained in the electronic catalog at www.ioffe.ru/LEA/shortGRBs/Catalog/. Printed and electronic versions of the catalog of short bursts from 1994 to 2009 are also being prepared.

References

1. Klebesadel R W, Strong I B, Olson R A *Astrophys. J. Lett.* **182** L85 (1973)
2. Mazets E P, Golenetskii S V, Il'inskii V N *Pis'ma Zh. Eksp. Teor. Fiz.* **19** 126 (1973) [*JETP Lett.* **19** 77 (1973)]
3. Mazets E P, Golenetskii S V, in *Astrofizika i Kosmicheskaya Fizika* (Astrophysics and Space Physics) (Ed. R A Syunaev) (Moscow: Fizmatlit, 1982) p. 216 [Translated into English: *Soviet Scientific Reviews, Sec. E, Astrophysics and Space Physics Reviews* Vol. 1 (Ed. R A Syunyaev) (Chur: Harwood Acad. Publ., 1981) p. 205]
4. Mazets E P, Golenetskii S V, in *Astrofizika i Kosmicheskaya Fizika* (Astrophysics and Space Physics) (Itogi Nauki i Tekhniki. Ser. Astronomiya, Vyp. 32, Ed. R A Syunyaev) (Moscow: VINITI, 1987) p. 16 [Translated into English: in *Soviet Scientific Reviews, Sec. E, Astrophysics and Space Physics Reviews* Vol. 6 (Ed. R A Syunyaev) (Chur: Harwood Acad. Publ., 1988) Pt. 3, p. 283]
5. Paciesas W S et al. *Astrophys. J. Suppl.* **122** 465 (1999)
6. Mazets E P et al. *Pis'ma Astron. Zh.* **5** 307 (1979) [*Astron. Lett.* **5** 163 (1979)]
7. Mazets E P et al. *Nature* **282** 587 (1979)
8. Golenetskii S V, Ilyinskii V N, Mazets E P *Nature* **307** 41 (1984)
9. Mazets E P et al. *Astrophys. J. Lett.* **519** L151 (1999)
10. Hurley K et al. *Nature* **397** 41 (1999)
11. Frederiks D D et al. *Pis'ma Astron. Zh.* **33** 3 (2007) [*Astron. Lett.* **33** 1 (2007)]
12. Frederiks D D et al. *Pis'ma Astron. Zh.* **33** 22 (2007) [*Astron. Lett.* **33** 19 (2007)]

13. Mazets E P et al. *Astrophys. J.* **680** 545 (2008)
14. Aptekar R L et al. *Astrophys. J. Lett.* **698** L82 (2009)
15. Van der Horst A J et al. *Astrophys. J. Lett.* **711** L1 (2010)
16. Racusin J L et al. *Nature* **455** 183 (2008)
17. Aptekar R L et al. *Astrophys. J. Suppl.* **137** 227 (2001)

PACS numbers: 95.55.Ka, 95.85.Pw, 98.70.Rz
DOI: 10.3367/UFNe.0180.201004h.0424

Discovery of the fast optical variability of GRB 080319B and the prospects for wide-field optical monitoring with high time resolution

G M Beskin, S V Karpov, S V Bondar, V L Plokhotnichenko, A Guarnieri, C Bartolini, G Greco, A Piccioni

1. Introduction

The systematic investigation of the night sky variability on a subsecond time scale is an important but still largely unaddressed issue. That the observations of this kind are necessary for the search for transient objects with a priori unknown localization and for their study was already emphasized by Bondi [1]. Investigations have been pursued in this area [2, 3]; but owing to technical difficulties, they either were able to attain a high temporal resolution on the level of several dozen microseconds in monitoring $5'-10'$ fields or used a temporal resolution of 5–10 s for wide $(20^\circ-30^\circ)$ fields. The currently operating wide-field monitoring systems, like WIDGET [4], RAPTOR [5], BOOTES [6], and 'Pi of the Sky' [7], have large fields of view with relatively good detection limits, but low temporal resolution, which hinders their use for detecting fast transients.

We give several examples of these transients: UV Cet-type stellar flares with rise times 0.2–0.5 s [8], 30% of gamma-ray bursts that last less than 2 s, while individual details of their light curves may be as short as 1 ms [9]. Also of substantial interest are very fast meteors, which are conceivably produced beyond the Solar System [10].

One further problem that invites conducting regular wide-field observations with a high temporal resolution is the monitoring of circumterrestrial space. The trajectories of a large number of satellites, as well as of a great quantity of small particles of space debris, change rather quickly and the speeds of these objects are quite high, making their observation by conventional techniques a very difficult task.

Since the late 1990s, we are developing the strategy of optical monitoring with a high temporal resolution of celestial sphere regions comparable in size to the field of view of space-borne gamma-ray telescopes. Initially, we planned to use instruments with large mirrors of relatively low quality [11, 12], for instance, Cherenkov telescopes and solar concentrators involving photomultiplier arrays with a temporal resolution as high as several microseconds. But more recently, we decided in favor of a project involving a wide-field camera with a lens of a relatively small diameter, an image intensifier (II) for the effective shortening of the focal length, and a fast low-noise CCD (charge-coupled device). A prototype of this system, FAst Variability Optical Registrator (FAVOR), put into operation in 2003, is located near the 6-meter BTA (Big Telescope Alt-azimuthal) of the Special Astrophysical Observatory of the Russian Academy of Sciences [13, 14]. The TORTORA (Telescopio Ottimizzato per la Ricerca dei Transienti Ottici RApidi) camera of similar design [15], which was installed on the mounting of the Rapid Eye Mount (REM) robotic telescope in the La Silla Observatory (ESO, European Southern Observatory, Chile) in 2006, makes up a two-telescope TORTOREM complex together with REM [17]. It was precisely this camera that discovered and enabled a detailed study of the optical emission of the brightest ever gamma-ray burst GRB 080319B [18–20].

In this report, we describe the design and special implementation features of the TORTORA camera and outline several results of its operation, including a comprehensive analysis of the data on the GRB 080319B burst. Also discussed is the project of a next-generation wide-field monitoring system capable not only of discovering much weaker transients but also of performing their multicolor photometry and polarimetry.

2. Description of the wide-field cameras FAVOR and TORTORA

The parameters of the FAVOR and TORTORA cameras are compared with those of other presently existing wide-field monitoring systems in Table 1. It follows that only the FAVOR and TORTORA cameras have a high temporal resolution and a relatively large field of view.

The structure of the TORTORA camera is diagrammed in Fig. 1, its parameters are collected in Table 2, and its image is given in Fig. 2. The instrument consists of the main

Table 1. Main optical wide-field monitoring systems currently operating.

Designation	Field of view, deg	Resolution, s	Detection threshold
WIDGET	62×62	5	$10^{\rm m}$
RAPTOR A/B	40×40	60	$12^{\rm m}$
RAPTOR Q	180×180	10	$10^{\rm m}$
BOOTES	16×11	30	$12^{\rm m}$
BOOTES-AllSky	180×180	30	$10^{\rm m}$
'Pi of the Sky'	33×33	10	$11.5^{\rm m}$
AROMA-W	25×35	5–100	$10.5^{\rm m}-13^{\rm m}$
MASTER-VWF	20×21	5	$11.5^{\rm m}$
MASTER-Net	30×30	1	$9^{\rm m}$
FAVOR*	16×24	0.13	$10^{\rm m}-11.5^{\rm m}$
TORTORA*	24×32	0.13	$9^{\rm m}-10.5^{\rm m}$

* For the FAVOR and TORTORA cameras, the detection threshold corresponds to the detection of a transient at a $3\,\sigma$ level in a single frame and may be different from the real detection threshold of the difference technique being used.

G M Beskin, S V Karpov, V L Plokhotnichenko Special Astrophysical Observatory, Russian Academy ofSciences, Nizhnii Arkhyz, Karachaevo-Cherkesiya, Russian Federation. E-mail: beskin@sao.ru
S V Bondar Scientific Research Institute of Precision Instrument Making, Moscow, Russian Federation
A Guarnieri, C Bartolini, A Piccioni Astronomy Department, Bologna University, Bologna, Italy
G Greco Astronomical Observatory of Bologna, INAF, Italy

Uspekhi Fizicheskikh Nauk **180** (4) 424 – 434 (2010)
DOI: 10.3367/UFNr.0180.201004h.0424
Translated by E N Ragozin; edited by A M Semikhatov

Table 2. Technical parameters of the TORTORA camera.

Main objective		Image intensifier		CCD matrix	
Diameter	120 mm	Photocathode	S20	Model	VS-CTT285-2001
Focal length	150 mm	Diameter	90 mm	Size	1388 × 1036 pixels
D/F	1/1.2	Gain	150	Scale	81″/pixel
Field of view	24 × 32 deg	Scaling	7.7	Exposure time	0.13 – 10 s
		Quantum efficiency	10%	Pixel size	6.45 μm

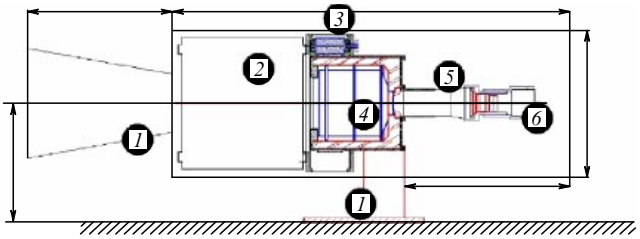

Figure 1. Schematic diagram of the TORTORA camera: *1* — viewing hood, *2* — main objective, *3* — main objective focusing unit, *4* — image intensifier, *5* — image-transfer optics and CCD-matrix focusing unit, *6* — fast CCD matrix.

Figure 2. Image of the TORTORA camera mounted on the REM robotic telescope in the La Silla Observatory (ESO, Chile).

objective *2*, its focusing unit *3*, an image intensifier *4* used for scaling and intensifying images, image-transfer optics *5*, and a fast low-noise CCD matrix *6*. The camera is mounted on a robotic REM telescope with an alt-azimuthal mount.

The fast CCD matrix records 7.5 frames per second for the exposure time 0.128 s; in this case, the intervals between individual exposures are negligible. The information from the matrix is distributed via a gigabit local area network and is stored in a 1 terabyte RAID disk array. The data stream is equal to about 20 Mb s^{-1}, and therefore the information obtained during one night of observations may be stored for only one day. Apart from that, the data are transferred to a dedicated computer for real-time processing with the aid of special-purpose software controlled by the Linux operating system. Various kinds of transient objects are detected and classified, and previously known objects are eliminated by comparing their related information with the data from satellite and stellar catalogues.

3. The method of detecting transients

Wide-field monitoring cameras with high temporal resolution can be used for the detection and investigation of different classes of transient events (variable stars, supernovae, active galactic nuclei, microlensing events, occultation of stars by planets) with constant, although a priori unknown coordinates. On the other hand, FAVOR- and TORTORA-type cameras are also able to track moving objects: artificial satellites, elements of space debris, comets, asteroids, and meteors. Special-purpose algorithms of data processing were elaborated for addressing these issues.

Because of the high intensity of the observational information data stream, it is hardly possible to use standard photometric packages for real-time information processing, and we therefore developed a special fast method for extracting transients. This method is based on studying the statistical behavior of the radiation intensity I in every image pixel in the course of time. The instantaneous value of I is compared with the current average

$$\langle I \rangle = \sum \frac{I}{N}\,, \tag{1}$$

which is evaluated for some number (e.g., $N = 100$, which corresponds to a temporal window of 13 s) of preceding frames, and with the sample variance

$$\sigma_I = \sqrt{\frac{\sum I^2 - \left(\sum I\right)^2/N}{N-1}} \tag{2}$$

to determine the significance of its deviation from the average

$$A = \frac{I - \langle I \rangle}{\sigma_I}\,. \tag{3}$$

All pixels with a high significance are then grouped into spatially connected clusters — observable objects. Some of them, e.g., single-pixel events, are rejected as noise.

When all objects in a given frame are revealed, their positions are compared with the trajectories of previously revealed transients (all objects are assumed to be moving, immobile objects are considered to be moving with zero velocity). The detection of an object in three sequential frames (on a temporal scale of 0.4 s) is sufficient for making an assertion that the object belongs to one of three possible classes: 'noise' events if the object disappears after a single frame, moving objects if it displays a statistically significant position variation, or immobile transients. The flare events of slowly moving high-altitude artificial satellites are identified by comparing the positions of the transients with the data from regularly updated catalogues of orbital satellite elements [21].

Revealing meteors, however, requires a somewhat different approach, because they are typically seen only in one or two frames. Furthermore, their velocities are substantially higher than those of artificial satellites. Events of this kind are distinguished by their high brightness and length.

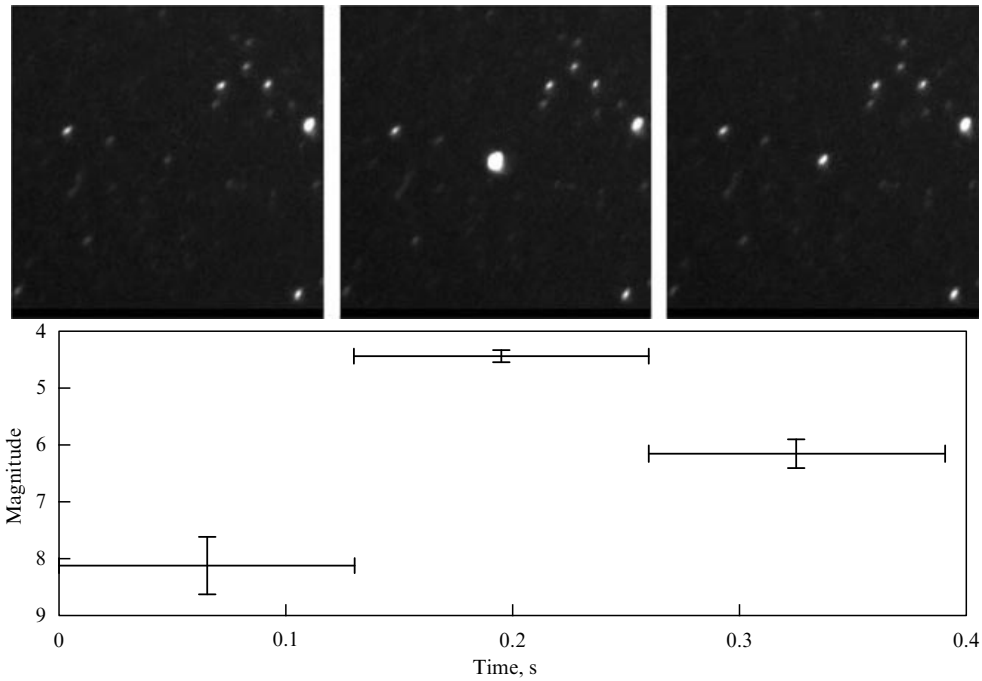

Figure 3. Example of a short flare (result of high-altitude satellite rotation) detected by the monitoring camera. The total duration of the event is 0.4 s; it is observed in three sequential frames.

Astrometric and photometric calibration is regularly performed (once a minute for the TORTORA camera, because the alt-azimuthal mount leads to a systematic rotation of the field of view) via a special procedure involving photometric measurements of all stars in a frame with the help of the SExtractor code [22] and their identification in the *Tycho-2* astrometric catalogue [23].

Therefore, the real-time data analysis system enables detecting and classifying any sufficiently bright optical transients 0.4 s after their emergence and up to the instant their brightness becomes lower than the difference method detection level. A rather short flare of this kind is exemplified in Fig. 3. Upon detection and classification, the information about an event can be transferred to the network for conducting an in-depth investigation. Furthermore, all real-time data about the transient, including its light curve, trajectory, and images of its neighborhood, may be stored for further studies.

4. Results obtained by the TORTORA camera

About half of the available observation time (when the REM telescope does not conduct routine observations), the TORTORA wide-field monitoring camera, which has been in service since June 2006, monitors celestial sphere regions observed at the moment by the Swift spacecraft, in accordance with the telemetric information distributed in real time through the GCN (Gamma-Ray Burst Coordination Network) [24].

Every observation night, the camera detects about 300 meteors and 150 transits of satellites with varying brightnesses.

4.1 Gamma-ray burst observations in the reactive regime

Owing to the automatic reaction of the REM telescope to reports of gamma-ray burst detection by spacecraft, the TORTORA camera has been able to observe the error boxes of three gamma-ray bursts shortly after their onset [25–27].

Table 3. Upper limits of the constant component of the optical flux and of the sinusoidal variable component of the brightness of possible optical transients.

Flare	Beginning of observations (Time since event), s	13 s limit (100 frames)	Frequency range, Hz	Upper limit
GRB 060719	59	12.4	0.01 – 3.5	15.3^{m}
GRB 061202	92	11.3	0.1 – 3.5	14.0^{m}
GRB 060719	118	11.3	0.01 – 3.5	16.4^{m}

The upper limits for the brightness of the corresponding optical transients obtained in these observations are collected in Table 3. The limits for the constant component of the flux were obtained by analyzing 100 sequential frames, which corresponds to the effective temporal resolution of 12.8 s.

4.2 GRB 080319B gamma-ray burst observations

The 19th and 20th of March 2008 turned out to be most fruitful for wide-field monitoring systems all over the world. At that time, five gamma-ray bursts were recorded during 24 hours. One of the bursts, GRB 080319B [28], was the brightest of all those observed in the gamma-ray range, as well as in the optical range, and was the first to be independently discovered by a ground-based monitoring system. Its position on the celestial sphere was observed by the Pi of the Sky [29], RAPTOR Q [30], and TORTORA [18] cameras prior to, during, and after the gamma-ray event.

Our TORTORA camera monitored the position of the GRB 080319B burst [18, 19] from 05:46:22 UT, approximately half an hour prior to its onset (the time of its discovery in the gamma-ray range: 06:12:49 UT), as well as during several dozen minutes after its termination. In the interval between 06:13:13 and 06:13:20 UT, the REM robotic telescope performed an automatic repointing on receiving

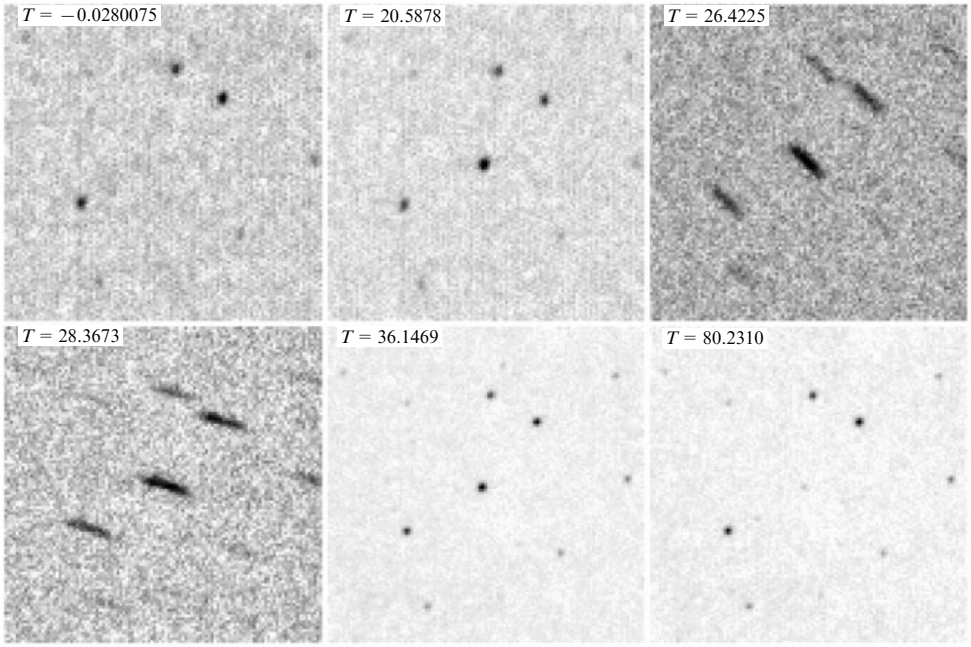

Figure 4. Images of the optical transient arising from the GRB 080319B gamma-ray burst as seen by the TORTORA camera at different stages of its development. Shown are the sums of 10 consecutive frames with the effective temporal resolution of 1.3 s for the onset of the gamma-ray event ($T \approx 0$), the instant of the highest brightness of the first peak ($T \approx 20.6$ s), two points in about the middle of the event ($T \approx 26.4$ s and $T \approx 28.4$ s), during the last peak ($T \approx 36$ s), and at the stage of early afterglow ($T \approx 80$ s). The field size is 2.5 by 2.5 degrees. The stellar profiles in the third and fourth frames were deformed in the course of repointing (from $T \approx 24$ s to $T \approx 31$ s) of the REM robotic telescope on which the camera was mounted, using the burst coordinates obtained from the Swift space telescope. Initially, the burst was located at the edge of the camera field of view; due to the repointing, it shifted to the center of the field of view, resulting in some improvement in the quality of the images.

the coordinates distributed by the Swift satellite [28], which moved the position of the burst from the edge of the camera field of view to its center. Figure 4 shows sample images of the burst region at different stages of the event.

The images acquired by the camera were processed using a conventional reduction procedure involving CCD-matrix noise reduction and flat-fielding. Flux measurements were made with the use of a custom version of the aperture photometry technique and the DAOPHOT code of the IRAF (Image Reduction and Analysis Facility) package for the entire dataset, with the exception of the repointing interval. At this phase, the images of the object and the neighboring stars were deformed because of their significant displacement during the exposure time, which substantially decreased the signal-to-noise ratio. As a result, the flux could not be accurately measured in individual frames. We summed nonoverlapping sequences of 10 frames with the corresponding shift to compensate stellar motion. For the track images thus obtained, the signal-to-noise ratio is in reasonable agreement with that in the remaining portions of the light curve (Fig. 5b). Next, from the summary frames, we measured the fluxes arising from the object and the neighboring stars, using aperture photometry with elliptic apertures and the method of the point scattering function approximation. Both techniques yielded mutually consistent results. We made a separate check of the behavior of the fluxes from comparison stars and found no features that might cast doubt on the accuracy of photometry and brightness variability for the object in the given interval. The effective temporal resolution was 1.3 s in this case; for all other phases of the flare, the photometry was performed with both high (0.13 s, individual frames) and low (1.3 s, sums of 10 consecutive images) temporal resolution. The

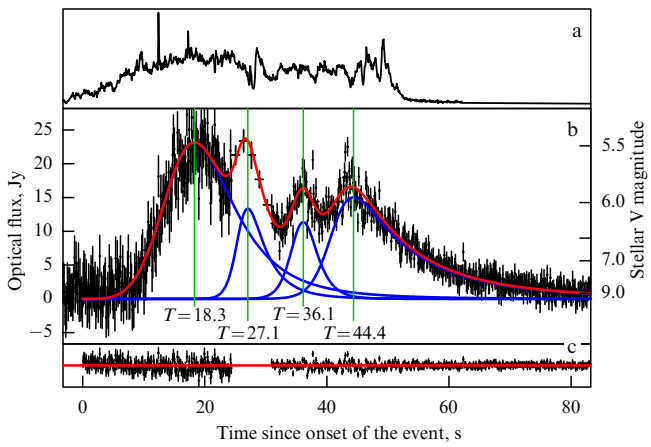

Figure 5. Light curve of the optical companion of the GRB 080319B gamma-ray burst, according to the data obtained from the wide-field TORTORA camera. (a) Gamma-ray radiation, which commences at the instant $T \approx -4$ s and decays at $T \approx 55$ s. (b) High-temporal-resolution data (0.13 s exposure, shown in gray) are available for the entire gamma-ray activity period except the interval of REM telescope repointing ($24.5 \text{ s} < T < 31$ s); low-resolution data obtained with the effective exposure time 13 s by summation of each 10 consecutive frames, available for the entire duration of the event. (c) Residuals of light curve approximation by four nearly equidistant peaks with parameters collected in Table 4.

resultant instrumental stellar magnitudes were then calibrated to the Johnson V photometric system by normalization to nearby stars of the Tycho-2 catalogue [23]. The light curve thus obtained (see Fig. 5) was found to agree nicely with the data of other monitoring cameras that also

Table 4. Optimal approximation parameters for the decomposition of the transient light curve into four peaks described by expression (4) and shown in Fig. 5.

T_0, c	F_0, Jy	r	d	ΔT, s
18.3 ± 0.3	23.2 ± 0.6	4.0 ± 0.4	-5.4 ± 4.1	8.7 ± 0.4
27.0 ± 0.3	13.4 ± 3.4	24.8 ± 8.3	-9.7 ± 4.9	9.1 ± 0.4
36.1 ± 0.2	11.4 ± 1.7	25.9 ± 7.6	-22.0 ± 17	8.3 ± 0.5
44.4 ± 0.5	15.1 ± 1.8	21.9 ± 3.3	-5.1 ± 0.2	

Note. ΔT is the distance between a given peak and the previous one. The probability of these interpeak distances occurring at random was estimated as 10^{-3} by calculating the $\Delta T_{1,2}\Delta T_{2,3}\Delta T_{3,4}/\Delta T_{1,4}^3$ statistics from 10^5 realizations of a group of four Poisson peaks.

observed this event, like Pi of the Sky [29] and RAPTOR (RAPid Telescope for Optical Response) [30].

According to the TORTORA data, the optical radiation of the transient is confidently detected beginning from about the tenth second from the onset of the gamma-ray burst. Its intensity increases as t^4 to attain a stellar magnitude $V \approx 5.5^{\mathrm{m}}$, then varies by a factor of one and a half to two on a time scale of several seconds, and finally decreases as $t^{-4.6}$, down to and below the detection threshold approximately 100 s later. In this case, the gamma-ray radiation terminates during the 57th second from the onset of the burst.

The light curve of the transient exhibits four peaks with similar amplitudes, durations, and shapes. We approximated them by a sum of four curves, each of which smoothly connects two power-law expressions describing the leading and trailing edges [31],

$$F = F_0\left(\frac{t}{T_0}\right)^r\left(\frac{d}{d+r}+\frac{r}{d+r}\left(\frac{t}{T_0}\right)^{r+1}\right)^{-(r+d)/(r+1)} . \quad (4)$$

Here, T_0 and F_0 are the location of the peak and its integral flux, and r and d are the respective exponents for the leading and trailing edges. The parameters of these curves are collected in Table 4. The intervals between the peak maxima turn out to be almost equal, about 8.5 s, which corresponds to 4.4 s for the GRB 190308B red shift $z = 0.937$ [19]. The probability that the combination of these intervals occurs randomly is 10^{-3} (see Table 4). In the power-density spectrum of the central part of the light curve, there also exists a spike at the frequency that corresponds to the interpeak distance at a 10^{-15} significance level (Fig. 6b). Therefore, it is valid to say that we have discovered periodic variations of the transient optical emission on a time scale of a few seconds.

The power-density spectrum of gamma-ray radiation observed by the BAT detector of the Swift space telescope, which is shown in Fig. 6a, does not exhibit clearly defined features at this frequency. This might be attributed to a significant contribution of stochastic variability, which has the form of shot noise, in the frequency interval from several tens of seconds to fractions of a second [32], which may conceal a periodic structure with a moderate amplitude for a multiplicative noise character.

To analyze the variability of the light curve over short periods of time, we subtracted the smooth approximation curve with four fitted peaks from the original data and studied the residual curve shown in Fig. 5c. The Fourier analysis of its different intervals revealed indications of periodic intensity variations during the last peak, in the interval from $T = 40$ s to $T = 50$ s (see Fig. 6). The remaining phases of the light curve exhibit no signs of significant variability at frequencies 0.1–3.5 Hz (0.3–10 s) with a power exceeding 15% prior to repointing and 10% after it. To exclude the instrumental nature of these periodic intensity variations, we performed a similar analysis for comparison stars and the background flux, which revealed no similar features.

The significance level of this peak in the power-density spectrum is equal to about 1%. The period and amplitude of the corresponding sinusoidal component, derived by means of nonlinear least-square fit, are 1.13 s (0.6 s in the reference system of the burst) and 9%.

To compare the temporal structure of optical and gamma-ray light curves, we performed a cross-correlation analysis of the central part of the burst, excluding the obviously correlated regions of the intensity increase and decay [33] (Fig. 7). The correlation between the high-temporal-resolution data does not exceed 0.5 due to a significant contribution of the noise component in the 0.1–1 s range, both to the optical light curve (measurement errors) and to the gamma-ray light curve (shot-noise-like high-frequency variability of physical origin [32]). For the low-resolution data (1.3 s binning), the correlation coefficient, by contrast, amounts to 0.82 when the optical light curve is shifted 2 seconds back relative to the gamma-ray one (see Fig. 7). The correspondingly rebinned data for the gamma-ray light curve exhibit the same four peaks spaced at nearly equal intervals as the optical light curve.

The $\Delta t \approx 2$ time lag of the optical emission relative to the gamma-ray emission is a solid indication that they are generated in different parts of the burst, with the optical photons emanating from regions whose central distance is greater by $\Delta R \approx 2c\Gamma^2\Delta t(1+z)^{-1} = 1.5 \times 10^{16}\Gamma_{500}^2$ cm, where Γ_{500} is the ejected-substance Lorentz factor in units of 500 [34, 35].

The burst emission characteristics that we discovered obviously contradict the emission generation models reliant on different kinds of interaction within one ensemble of electrons and their emitted photons (synchrotron and inverse-Compton mechanisms) [19, 36, 37], the model involving two internal shock waves, forward and backward [38], and the model with relativistic turbulence in the outflow [39]. On the other hand, the steepness of the leading edges and the nearly equal durations of all four optical flares are inconsistent with the model involving an external shock, both forward and backward, as the source of the optical radiation [40].

Two models have been proposed to date that rely on internal shocks and in which optical radiation and gamma rays are generated via synchrotron mechanism in different parts of the outflow: the higher the photon energy, the closer to the central source the site of its emission is. These are the model of 'residual collisions' [35] and the model with a significant neutron component in the outflow [41]. In these scenarios, gamma-ray photons are emitted at the distance $10^{14} - 10^{15}$ cm from the center due to the heating of electrons at the fronts of shock waves generated in the collision of proton shells ejected from the central source. In the framework of the first model, optical photons are produced in optically thin plasma in the collision of 'residual' shells (the results of the merging of separate groups of initial shells) at substantially greater distances, $\gtrsim 10^{16}$ cm [35]. In the second model, the one with a significant neutron component, optical radiation is generated by the electrons emitted in the β-decay of neutrons, which reach distances $R \sim 10^{16}$ cm without interacting with other components of the outflow. The

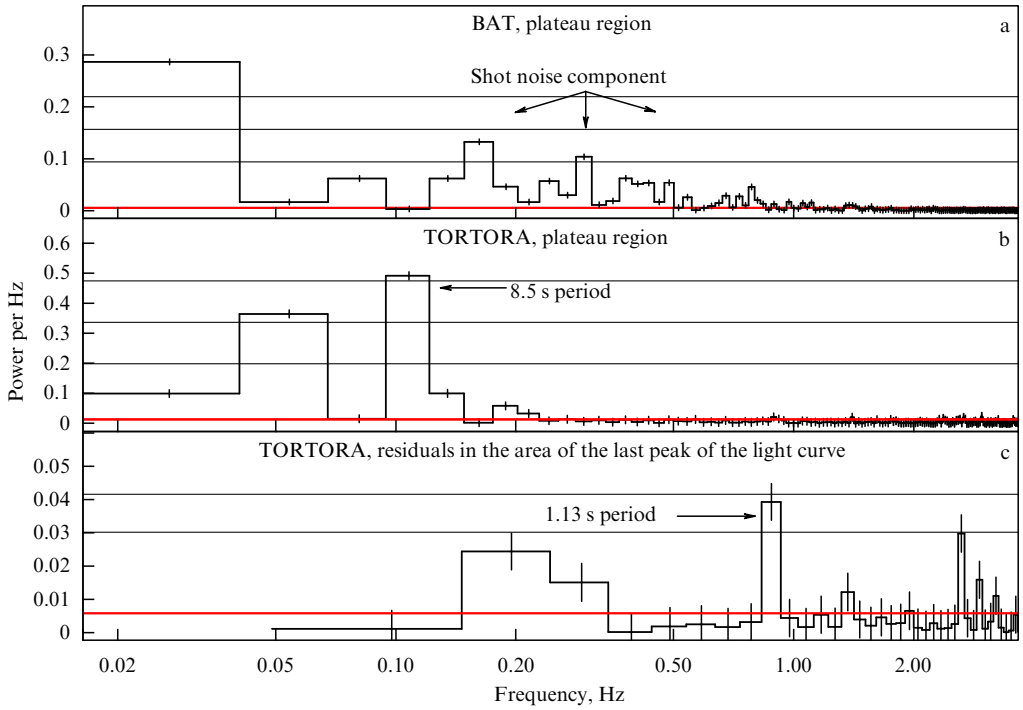

Figure 6. Power-density spectrum of the central part (from $T = 13$ s to $T = 50$ s) of the burst obtained from the data of the BAT (Burst Alert Telescope) gamma-ray telescope of the Swift spacecraft (the sum of all spectral channels) (a) and the TORTORA optical camera (b), and the power-density spectrum of the residual of the optical light curve after subtraction of the smooth approximation function, which was depicted in Fig. 5, for the interval of the last peak (from $T = 40$ s to $T = 50$ s) (c). The linear trend was subtracted from all curves. The missing part of the optical curve (the interval from $T = 24.5$ s to $T = 31$ s) with a high temporal resolution was filled with white Gaussian noise with the variance corresponding to that in the remaining part of the data and with the mean values lying on the smooth approximation curve. Errors and significance levels for the spectra were estimated by the bootstrap method, i.e., by generating a large number of synthetic light curves by randomly shuffling the intensities in the original light curve, which completely destroys its time structure but preserves the distribution of these values. The significance levels then correspond to the probability that a value exceeding the given one in any of the frequency bins randomly occurs for the power-density spectrum of a completely random process with a sampling distribution that coincides with the observed one. The feature near 9 s, which is clearly seen in the power-density spectrum of the middle part of the light curve at the significance level 10^{-15}, corresponds to the four peaks in the light curve spaced at almost equal intervals. The low-frequency components in the optical spectrum and in the gamma-ray spectrum correspond to two different intensity levels of light curves. The feature in the power-density spectrum of the difference curve with a significance level ~ 0.01 may correspond to the periodic component with the 10% amplitude and the period ≈ 1.13 s, which manifests itself during the last peak. The remaining intervals of the difference light curve do not exhibit features of this kind.

products of this decay, protons and electrons, collide with fast proton shells ejected from the central source later on, giving rise to secondary shock waves and heating the electrons, which generate synchrotron radiation. Both models explain the observed two-second lag of periodic peaks in the optical curve relative to those in the gamma-ray curve, as well as its significantly higher smoothness on the 0.1–1 s time scale in comparison with gamma-ray radiation, which exhibits high variability on this time scale [32]. On the other hand, the large quantitative difference between the optical and gamma-ray fluxes ($F_o/F_\gamma \sim 10^3$) [19] is easier to interpret in the model with the neutron component [41]. Furthermore, a significant neutron fraction in the outflow is practically inevitable in the case of bright gamma-ray bursts like GRB 080319B [42, 43]. This model is therefore preferred and our findings are strong evidence in favor of the existence of a significant neutron component in the outflow.

We note that the conclusion that the optical radiation and gamma rays are emitted at different distances from the central source directly follows from the structural similarity between the optical light curve and the gamma-ray curve, which we discovered. This conclusion is independent of both the specific mechanisms of transformation of the kinetic outflow energy to the internal energy of particles, and the particle radiation mechanisms. This effect may not stem from density or velocity variations of substance in the outflow, like those observed in the afterglow of some other bursts on a time scale of several tens of seconds [44, 45]. It is difficult to imagine such a behavior, particularly a periodic one, which coincides in different parts of the relativistic outflow spaced by about 10^{16} cm. Therefore, we are compelled to conclude that the variations we observe have a common origin: specifically, they are caused by a periodic activity of the central source (in this case, each optical peak corresponds to one activity episode).

The discovered variability of the radiation flux from outflowed matter may be a manifestation of transient accretion caused by the periodically developing gravitational instability [46] of the hot inner parts of a massive accretion disc (about one solar mass) rotating around a black hole with a mass of three solar masses, which resulted from the collapse of the nucleus of a massive star [47, 48]. This disc must contain a substantial fraction of neutrons [43]. The four peaks seen in the optical light curve reflect four episodes of accretion activity responsible for the jet outflow of matter. The gas in the internal parts of the disc is fragmented due to various instabilities and forms separate shells inside the outflow, whose collisions generate internal shocks. Furthermore, the half-second intensity variations seen at the last stages of the burst may result from the

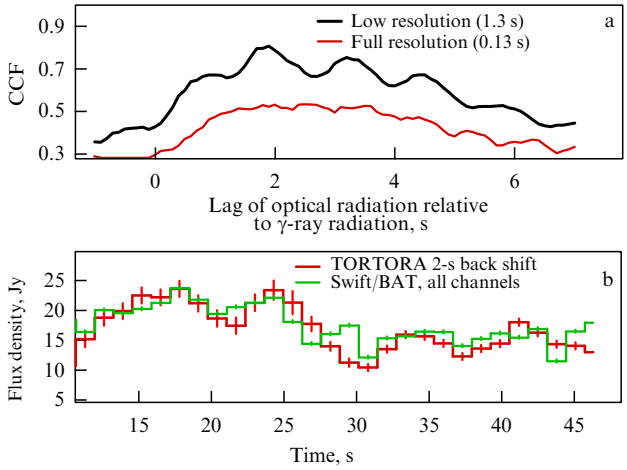

Figure 7. (a) Cross correlation (CCF) of the Swift/BAT gamma-ray curve (sum of all spectral channels) and the TORTORA optical light curve for the full (0.13 s) and low (1.3 s) temporal resolution. Only the central part of the burst was used for the analysis, with the exception of the radiation buildup and decay phases. For each value of the argument, the correlation coefficient was calculated by rebinning the gamma-ray curve, which was accordingly shifted to the optical one. The peak correlation for the high-resolution data (0.13 s) is substantially lower than for the low-resolution data (1.3 s) owing to their higher noisiness. Quasiperiodic variations of the cross correlation are due to the transition of the sharp peak of the gamma-ray curve for $T + 30$ s across the boundaries of individual bins. (b) The low-resolution light curve shifted 2 s back, and the correspondingly rebinned gamma-ray curve. In this case, the correlation coefficient r is close to 0.82 at the significance level 5×10^{-7}.

Lense–Thirring precession of the internal parts of the accretion disc.

5. Next-generation wide-field monitoring system

Obviously, it is important to further develop the methodology for the wide-field search for optical transients in two directions. First, it is required to improve the detection threshold of the system by at least 2–3 stellar magnitudes while retaining or even broadening its field of view. This may be attained with the use of multiobjective (or multitelescope) configurations by narrowing the field of view of an individual instrument and thereby improving its angular resolution [49]. The significant readout noise contribution of CCD matrices may be overcome either by improving their quantum efficiency and the gain of image intensifiers or by using low-noise electron-multiplying CCD matrices. The second important line of development is to measure the colors and polarization of the transients discovered.

In what follows, we outline one possible design of a multiobjective monitoring system based on internal-gain CCD matrices and capable of collecting multicolor and polarimetric information about transients.

5.1 Basic 3×3 unit

The proposed project has a modular structure and consists of separate base units that contain 9 objectives each and are placed on separate equatorial mounts (Fig. 8). Each objective is suspended in a gimbal controlled by two actuators and can be reoriented independently of the other ones. Furthermore, each objective has a set of color and polarization filters that may be inserted into the light beam in the course of observation. This enables a rapid move from a filter-free wide-field monitoring to narrow-band observations, with all objectives pointing towards the same region, for instance, one containing the newly discovered transient, and observing it using all possible combinations of color and polarization filters (Fig. 9). Simultaneous observation of a transient with all objectives with the same filter is also possible in order to improve the photometric accuracy by summing their data.

Each objective is equipped with a fast electron-multiplying CCD matrix with negligible readout noise even for a high frame rate. Possible versions of mass-produced CCD matrices and objectives are depicted in Fig. 10.

The data stream from each channel of this system, which amounts to about 20 megabytes per second, is collected by a dedicated computer, which stores it on a hard disc and analyzes it in real time with the use of a method similar to that described for the FAVOR and TORTORA cameras. The operation of the system as a whole is coordinated by the dedicated computer, which receives information about the uncovered transients from individual channels and controls the change in observation regimes.

In the wide-field monitoring regime, each base unit has the field of view about 260 square degrees (720 when using Canon objectives) and the B-band detection threshold

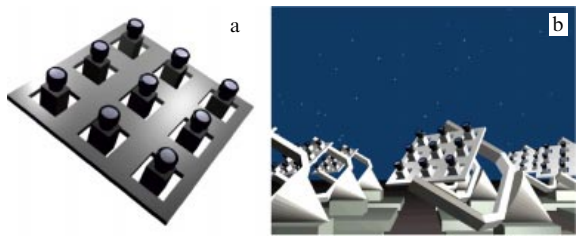

Figure 8. (a) Base 3×3 unit. Each objective is equipped with a set of installable color and polarization filters, which may be promptly inserted into the light beam and may be repointed independently of the other ones: B, V, anf R are photometric (for the B, V, and R bands) filters, and P_1, P_2, and P_3 are polarization filters for three different polarization orientations. (b) Artistic view of the complete monitoring system, which comprises an ensemble of base units placed on separate equatorial mountings.

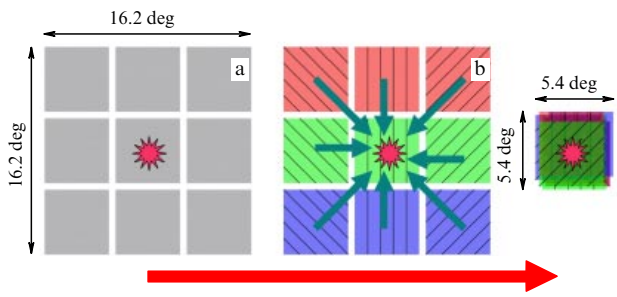

Figure 9. Different regimes of base unit operation. (a) Regime of wide-field monitoring in white light or in the transmission band of one of the color filters. (b) Insertion of color and polarization filters into the light beam as the first step following the discovery of an optical transient. (c) Repointing all objectives to the transient-containing field to simultaneously acquire information about the transient in three different photometric bands for three orientations of the polarization plane (indicated by different hatching directions). The exact time required for moving from one regime to another depends on the hardware configuration and is expected to be within 0.3 s.

Figure 10. Possible mass-produced components of the projected system: (a) Electron-multiplying Andor iXon^{EM}+888 CCD matrix with 1024 × 1024 pixels, each 13 μm in size; (b) Canon EF 85 f/1.2 L USM II objective capable of imaging a 9 × 9 degree field onto this matrix for a scale of 31″/pixel; (c) Marshall Electronics 140 mm f/1.0 objective, which affords a 5.4 × 5.4 degree field for a scale of 19″/pixel.

Table 5. Detection limits (in stellar magnitudes) of a base unit in the narrow-field regime for different combinations of photometric and polarization filters with the use of Marshall Electronics objectives. For Canon objectives, all threshold stellar magnitudes are 1.5^{m} lower.

Exposure time, s	Without filters/B	B + 3 polarizations	BVR	BVR + 3 polarizations
0.1	15.7	13.0	15.0	12.5
10	18.2	15.2	17.5	15.0
1000	20.7	17.9	20.0	17.5

$\sim 14.5^{m}$ in 0.1 s (13^{m} for Canon). Frame coaddition would increase the detection threshold up to 17^{m} for the effective exposure of 10 s and to 19.5^{m} for 1000 s (respectively to 15.5^{m} and 18^{m} with the use of Canon objectives). In the narrow-field regime, in the observation of individual objects, the field of view decreases to 30 square degrees (80 for Canon) and the detection threshold depends on the selected combination of spectral and polarization filters; these variants are summarized in Table 5. Furthermore, for bright events, the realization of a regime with high temporal resolution is possible when the CCD matrix supports the readout in a narrow window at a higher frame rate (for instance, for an Andor iXon^{EM}+888 matrix, a frame rate up to 65 Hz in the 128 × 128 field without binning or up to 310 Hz with an 8 × 8 binning is possible).

5.2 Complete system

The complete system is a set of basic units installed on separate mountings and operated in concert. Their number may be arbitrary: the greater the number is, the better the coverage of the celestial sphere and the lower the detection threshold in the narrow-field regime.

For instance, an assemblage configured of eight base units covers 2100 square degrees (5700 degrees with the use of Canon objectives) simultaneously in the sky in the wide-field monitoring regime, which allows surveying the attainable hemisphere twice per night while staying in each region for half an hour. When combining information from all 72 channels in the narrow-field regime, the detection threshold would range from 17.2^{m} to 19.7^{m} for effective exposure times from 0.1 to 10 s (the respective thresholds would be equal to 15.5^{m} and 18.2^{m} with the use of Canon objectives). The amount of data acquired by this system during a night of observations would be about 40 terabytes, which would be processed in real time as they arrive. The performance of this system in the observation of different classes of objects is shown in Figure 11 in comparison with other existing and projected wide-field monitoring systems.

Assuming that an objective costs 2 thousand euros, a CCD matrix 45 thousand euros, one computer 1 thousand euros, and a mounting 26 thousand euros, the cost of the base unit would amount to about 500 thousand euros, while the eight-unit configuration would cost about 5 million euros.

5.3 Prototype of the system

As a prototype of a multiobjective monitoring system, we are developing, with support from the Russian Foundation for Basic Research, an array consisting of nine Canon objectives with the aperture ratio 1/1.2 and the focal length 85 mm. The objectives will be equipped with a set of spectral and polarization filters, as well as with image intensifiers with GaAs photocathodes whose quantum efficiency is about 30% at the wavelength 4500 Å. The image intensifiers effectively shorten the focal length of the system by a factor of two and also permit suppressing the readout noise of fast Sony ICX285AL CCD matrices. As a result, the field of view of an individual channel is equal to about 100 square degrees (900 degrees for the complete system) for the detection threshold $B \approx 12.5^{m}$ in 0.1 s (the limiting value increases to $B \approx 15^{m}$ on summation of 100 consecutive frames, which corresponds to the effective temporal resolution of 10 s). Pointing all channels to one field permits attaining the limit $B \approx 13.5^{m}$ in 0.1 s and $B \approx 16^{m}$ in 10 s.

The prototype system will be completed and put into service in 2010–2011.

6. Conclusions

The discovery and comprehensive study of the optical companion of what is currently the brightest gamma-ray burst, GRB 080319B, made using the TORTORA camera emphasizes the importance of pursuing permanent monitoring of large regions of the celestial sphere, as well as of employing detectors and techniques with high temporal resolution in the search for fast optical transients of a priori unknown localization. The FAVOR and TORTORA cameras, which were developed in the framework of the general strategy of these observations, have demonstrated the success of this strategy, allowing us to further develop this research: the next step may be the proposed new-generation wide-field monitoring system.

Acknowledgements

This work was supported by the Russian Foundation for Basic Research grant Nos. 04-02-17555, 06-02-08313, and 09-02-12053, INTAS (04-78-7366), CRDF (RP1-2394-MO-02),

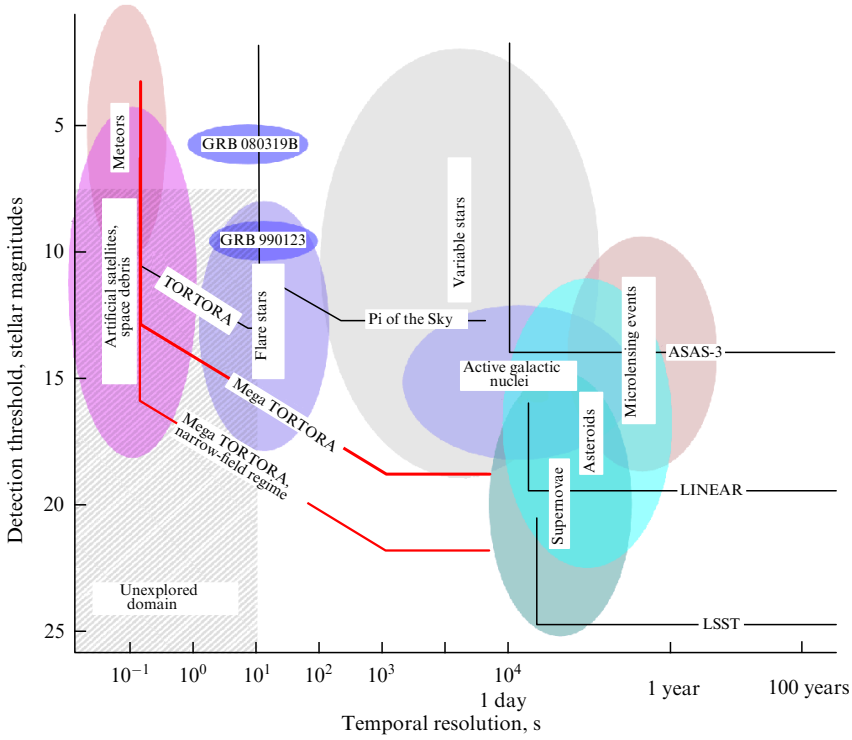

Figure 11. Efficiency of the monitoring system proposed for the observation of different classes of objects and its comparison with the efficiency of other instruments, such as the presently operating facilities (ASAS-3 (All sky Automated Survey), LINEAR (Lincoln Near-Earth Asteroid Research), Pi of the Sky, FAVOR/TORTORA), and those planned for the future [LSST (Large Synoptic Survey Telescope)].

the Progetti Pluriennali 2003 Grant of Bologna University, the Program of the Presidium of the Russian Academy of Sciences, and a grant from the President of the Russian Federation for the Support of Young Russian Scientists.

References

1. Bondi H *Q. J. R. Astron. Soc.* **11** 443 (1970)
2. Schaefer B E *Astron. J.* **90** 1363 (1985)
3. Schaefer B E et al. *Astron. Astrophys.* **174** 338 (1987)
4. Tamagawa T et al. *Nuovo Cimento C* **28** 771 (2005)
5. Borozdin K N et al. *Proc. SPIE* **4847** 344 (2002)
6. Castro-Tirado A J et al. *Astron. Astrophys. Suppl.* **138** 583 (1999)
7. Burd A et al. *New Astron.* **10** 409 (2005)
8. Shvartsman V F et al. *Pis'ma Astron. Zh.* **14** 233 (1988) [*Sov. Astron. Lett.* **14** 97 (1988)]
9. McBreen S et al. *Astron. Astrophys.* **380** L31 (2001)
10. Afanasiev V L, Kalenichenko V V, Karachentsev I D *Astrofiz. Byull.* **62** 319 (2007) [*Astrophys. Bull.* **62** 301 (2007)]
11. Beskin G M et al. *Astron. Astrophys. Suppl.* **138** 589 (1999)
12. Eichler D, Beskin G *Astrobiology* **1** 489 (2001)
13. Zolotukhin I et al. *Astron. Nachrichten* **325** 675 (2004)
14. Karpov S et al. *Nuovo Cimento C* **28** 747 (2005)
15. Molinari E et al. *Nuovo Cimento B* **121** 1525 (2006)
16. Zerbi F M et al., in *Gamma-ray Bursts in the Afterglow Era* (ESO Astrophys. Symp., Eds E Costa, F Frontera, J Hjorth) (Berlin: Springer-Verlag, 2001) p. 434
17. Beskin G et al. *Nuovo Cimento C* **28** 751 (2005)
18. Karpov S et al. *GRB Coordinates Network Circular* (7452) 1 (2008)
19. Racusin J L et al. *Nature* **455** 183 (2008)
20. Beskin G et al., arXiv:0905.4431
21. American department of defense satellite ephemerides database, http://www.spacetrack.org/
22. Bertin E, Arnouts S *Astron. Astrophys. Suppl.* **117** 393 (1996)
23. Høg E et al. *Astron. Astrophys.* **355** L27 (2000)
24. Barthelmy S D *AIP Conf. Proc.* **428** 129 (1998)
25. Guarnieri A et al. *GRB Coordinates Network Circular* (5372) 1 (2006)
26. Karpov S et al. *GRB Coordinates Network Circular* (5897) 1 (2006)
27. Karpov S et al. *GRB Coordinates Network Circular* (5941) 1 (2006)
28. Racusin J L et al. *GRB Coordinates Network Circular* (7427) 1 (2008)
29. Cwiok M et al. *GRB Coordinates Network Circular* (7439) 1 (2008)
30. Wozniak P et al. *GRB Coordinates Network Circular* (7464) 1 (2008)
31. Kocevski D, Ryde F, Liang E *Astrophys. J.* **596** 389 (2003)
32. Margutti R et al. *AIP Conf. Proc.* **1065** 259 (2008)
33. Beskin G et al. *AIP Conf. Proc.* **1065** 251 (2008)
34. Piran T *Rev. Mod. Phys.* **76** 1143 (2005)
35. Li Z, Waxman E *Astrophys. J.* **674** L65 (2008)
36. Kumar P, Panaitescu A *Mon. Not. R. Astron. Soc.* **391** L19 (2008)
37. Fan Y-Z, Piran T *Frontiers Phys. China* **3** 306 (2008)
38. Yu Y W, Wang X Y, Dai Z G *Astrophys. J.* **692** 1662 (2009)
39. Narayan R, Kumar P *Mon. Not. R. Astron. Soc.* **394** L117 (2009)
40. Zou Y-C, Piran T, Sari R *Astrophys. J. Lett.* **692** L92 (2009)
41. Fan Y-Z, Zhang B, Wei D-M *Phys. Rev. D* **79** 021301(R) (2009)
42. Derishev E V, Kocharovsky V V, Kocharovsky Vl V *Astrophys. J.* **521** 640 (1999)
43. Pruet J, Woosley S E, Hoffman R D *Astrophys. J.* **586** 1254 (2003)
44. Jakobsson P et al. *New Astron.* **9** 435 (2004)
45. Bersier D et al. *Astrophys. J. Lett.* **584** L43 (2003)
46. Masada Y et al. *Astrophys. J.* **663** 437 (2007)
47. Woosley S E *Astrophys. J.* **405** 273 (1993)
48. Zhang W, Woosley S E, Heger A *Astrophys. J.* **608** 365 (2004)
49. Beskin G et al. *Bull. Special Astrophys. Observatory* **60 – 61** 217 (2007)

PACS numbers: **95.35.+d**, 98.62.Hr, **98.65.–r**
DOI: 10.3367/UFNe.0180.201004i.0434

Galactic disks and their evolution

A V Zasov, O K Sil'chenko

We consider the key problems related to measuring the mass of stellar disks and dark halos in galaxies and to explaining the observed properties of disks formed in massive dark halos.

Rotating stellar-gas disks are the main structural elements of the majority of the observed galaxies. They contain mostly baryonic matter: stars and the diffuse interstellar medium (in the latter, cold gas dominates). Disks have a large angular momentum and rotate such that the local angular velocity decreases along the radius. The maximum rotational velocity of the disk depends on the total mass (luminosity) of the galaxy and is usually 100–300 km s^{-1}, which corresponds to the orbital period 200–300 Myr.

Galactic disks are inhomogeneous structures. They contain stars with different masses and ages, and the youngest stars are located near the equatorial plane of the disk: there, at the bottom of the disk gravitational potential well, the interstellar gas is stored. But because the entire disks are old structures, their spectra are dominated by very old stars with the age above 8 Gyr. The oldest stars form the so-called thick disk, which is two to three times thicker than the main stellar disk, but the mass of the thick disk is relatively small. In fact, the formation of disks in most galaxies has not yet been completed, because star formation is occurring there (predominantly in the spiral arms) even at present, but the observed rates of stellar population mass growth, except for rare cases, are very low, about 1–5 solar masses per year for an entire galaxy like our own.

The origin of galaxies and their disk formation mechanism remain unclear and are actively discussed in the literature. The shape of the disks clearly suggests that they were formed as a result of the evolution of a dissipative medium (gas), which had been losing its energy by radiation with the angular momentum conservation, and the age of the disks suggests that they already existed in the very early times of modern galaxies and had a very intensive star formation rate at that time. Disk galaxies, which frequently have big star-forming regions that appear from large distances like a collection of individual bright spots, are indeed abundantly present among galaxies with redshifts $z > 1$; we observe them at the time of their youth.

In modern galaxy formation theories, it is important that star formation in galaxies occurred in the gravitational field formed by dark matter, or the so-called hidden mass, which presently must form massive halos around galaxies extending far away from their visible parts. Numerical simulations showed that the role of the dark halo is decisive in both the disk formation and its later evolution. But it is not an easy task to divide the observed mass of a galaxy into gas, stellar, and dark matter components. This problem can be solved for individual galaxies by improving upon methods of measuring the kinematical characteristics of the disk, on the one hand, and models of evolution of stellar population spectra, on the other hand.

The scientific boom that gave rise to an avalanche of studies of hidden mass inside galaxies started when observational data on galaxy disk rotational velocity curves $V(R)$ as a function of radius R became available for sufficiently large distances from the center. In the optical band, either classical diffraction spectrographs with a long slit, or scanning Fabry–Perot interferometers in a high order of interference are typically used. The interferometers do not have slits, and the Doppler shifts of spectral lines can be measured simultaneously at many thousands of points on the galactic disk; using a complicated mathematical data processing then allows recovering the two-dimensional radial velocity field and obtaining the rotational curve. In Russia, such observations have been carried out on the 6-meter telescope at the Special Astrophysical Observatory, RAS. Radio observations have a lower angular resolution in the emission lines of atomic hydrogen or molecules than in the optical lines, while optical observations can be used to measure radial velocities of both gas and stars with high angular resolution. Nevertheless, radio velocity curves in gas-rich galaxies are traced to much longer distances than the optical ones, sometimes reaching far beyond the visible limits of a galaxy, because gas disks are frequently much larger than stellar ones. It was found that rotational velocities at large distances from galactic centers do not decrease, as a rule, but become almost flat (come to a plateau), or even increase with R.

That the rotational velocity reaches a plateau is frequently considered if not the proof, at least a decisive argument in favor of the existence in galaxies of a dark halo with a mass comparable with or even exceeding that of visible matter. In fact, this is not precisely the case because the form of the rotational velocity curve, for some reason, does not automatically imply the presence of dark matter. A rotational curve of any form, increasing or decreasing, can be explained by the presence of only one disk, and it can reflect only the peculiarity of mass distribution inside it. We illustrate this with simple examples.

If the disk density along the radius were conserved or decreased very slowly, the rotational velocity of the disk would increase without bound as R increases, even in the absence of a halo. Of course, the surface density is not constant in real galactic disks; it quite rapidly decreases with the distance from the center. But there is no major problem explaining the plateau on the rotational curve. The classical example is the so-called Mestel disk. This is a thin axially symmetric disk whose surface density $\Sigma(R)$ decreases from the center to the periphery as $1/R$. It can be shown theoretically that the circular velocity, which in the general case is determined by the radial gradient of the gravitational potential,

$$V^2(R) = R\,\frac{\partial\Phi(R)}{\partial R}\,,$$

is independent of R for the Mestel disk, and the rotational curve of the galaxy is a horizontal line from zero to infinity, and hence no dark halo is to be invoked! The mass of such a disk within any given radius R is $M(R) = V^2R/G$, i.e., the same as for a spherically symmetric density distribution. In

A V Zasov, O K Sil'chenko Shternberg State Astronomical Institute, Lomonosov Moscow State University, Moscow, Russian Federation. E-mail: zasov@sai.msu.ru

Uspekhi Fizicheskikh Nauk **180** (4) 434–439 (2010)
DOI: 10.3367/UFNr.0180.201004i.0434
Translated by K A Postnov; edited by A M Semikhatov

other words, it is impossible to distinguish the Mestel disk from a spherically symmetric galaxy using the form of the rotational velocity curve.

In that case, what indeed can be considered relevant to the presence of a hidden mass in disk galaxies, in particular, a massive dark halo? First, this is a discrepancy between the measured rotational velocity curve of the galactic disk (using both the form and the absolute value of velocities) and the expected curve calculated by assuming that the galaxy consists of only 'luminous' matter, i.e., of the directly observed components. Because most of the mass of the disk resides in stars, the brightness distribution of the stellar disk reflects its mass distribution, especially if the brightness is measured in the near-infrared spectral band, where emission from old stars dominates.

The brightness and hence the surface density in the broad range of distances from the center decreases with R exponentially as $I(R) \sim \exp(-R/R_0)$ (R_0 is the radial scale of brightness), i.e., more rapidly than $1/R$. This law results in a rotational curve reaching a maximum at $R \approx 2R_0$ and then slowly decreasing. However, the expected maximum is never observed in real galactic rotational curves. In a more rigorous approach, by adopting a certain mass–luminosity ratio M/L for the stellar population of the disk (this ratio can be estimated from the observed color indices using stellar population models), the brightness distribution can be easily transformed into the density distribution along the disk (not necessarily exponential) and the expected rotational velocity of the galaxy due to its baryonic components can be calculated. In this way, the calculated rotation velocity curve typically passes lower than the observed one (at least in the outer parts of the disk). This allows concluding that dark mass significantly contributes to the total mass of the galaxy. This contribution is most substantial in galaxies with low surface brightness of the disk: in these galaxies, the dark mass can exceed the baryonic mass by several times within the observed boundaries.

There are several other arguments in favor of high-mass dark halos in galaxies. Two main arguments follow from observations, although they are statistical.

The first argument is that the integral masses of galaxies and galaxy systems measured by other means (using the relative velocities of the motion of satellites) are found to be much larger than their visible mass. For example, the data in [1, 2] show that the mean ratio of the total mass of galaxies in pairs to their total infrared K-luminosity (2.2 μm), derived for more than 500 pairs, is very high: around 11 solar units, and for groups of galaxies, two times as high. For comparison, models of the purely stellar population yield the value of M/L_{K} below 1.5 [3]. Accounting for the mass of gas and the internal extinction only slightly increases this ratio. Consequently, in the region whose size exceeds the diameter of the galaxy and includes a pair or group of galaxies, the mass of dark matter exceeds the total mass of directly visible matter by many times.

The second argument is related to the stability condition of the disk with respect to gravitational perturbations. The disk is usually described as a massive strongly oblate stellar system in which oscillations can propagate and instabilities can develop on different scales. This inevitably leads to the dynamic heating of the disk up to a state with marginal stability under small perturbations. The higher is the radial velocity dispersion of stars constituting most of the disk mass and the faster it rotates and the lower its surface density is, the more stable the disk is. The critical (maximal) surface density at which the disk is still stable is determined both analytically (under several simplifying assumptions) and numerically. In the first approximation, the critical surface density is proportional to the radial velocity dispersion times the angular velocity of the disk at a given radius. Hence, after having obtained the rotational velocity curve and velocity dispersion of old stars of the disk from observations, it is possible to estimate the maximum admissible surface density of the disk and the corresponding M/L ratio, and then to find the upper limit of its mass from the total disk luminosity.

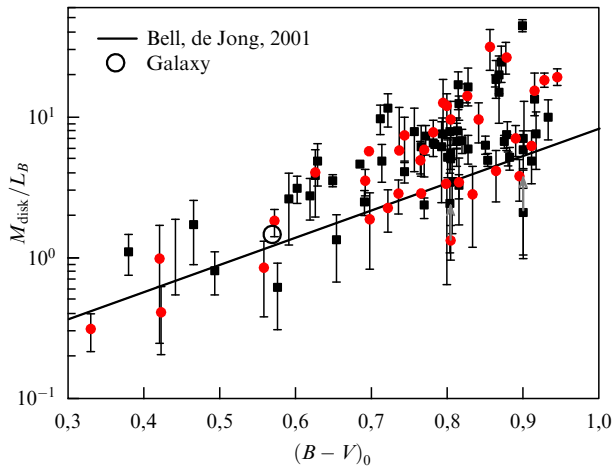

Figure 1. The upper limit on the mass–luminosity ratio in the blue spectral range M/L_B (in solar units) for galactic disks with different color indices, derived under the assumption of marginal disk stability [7]. The black circles show galaxies in pairs, the black squares mark isolated galaxies or members of groups, and the white circle stands for our Galaxy. The values of M/L_B for galaxies with active star formation ($(B-V)_0 < 0.7$) are close to those expected from the photometrical model of stellar systems evolution (the straight line according to [3]), which suggests the absence of strong dynamical heating of their disks.

The accuracy of this estimate for an individual galaxy is not very high, up to a factor of two, but data obtained for different galaxies allow making some general conclusions. Specifically, it has been confirmed that within the optical limits of a galaxy, the dark halo mass is typically comparable to that of the disk and often exceeds it [4], which is compatible with results derived from the analysis of rotational velocity curves. The same conclusion can be obtained from the photometric estimates of the thickness of stellar disks observed edge-on [5, 6]. There is another intriguing fact: because galactic disks are subjected to gravitational perturbations from both neighboring galaxies and massive dark halos (see below), it follows that, apparently, gravitational perturbations must have dynamically heated up the disk above the stability limit. Then the M/L ratios calculated under the assumption of marginally stable disks (Fig. 1) exceed the values derived from the color index of photometrical models of the evolution of galactic stellar disks (the straight line in Fig. 1). It turns out that such overheated systems do exist, but only some fraction of galaxies, predominantly those with high color indices (corrected for the disk inclination to the line of sight) $(B-V)_0 > 0.7$ relate to them. Such a color index corresponds to evolved disks whose luminosities are dominated by an old stellar population. In those galaxies, the star formation is very weak or totally absent; many of them are lenticular systems that contain almost no cold interstellar gas.

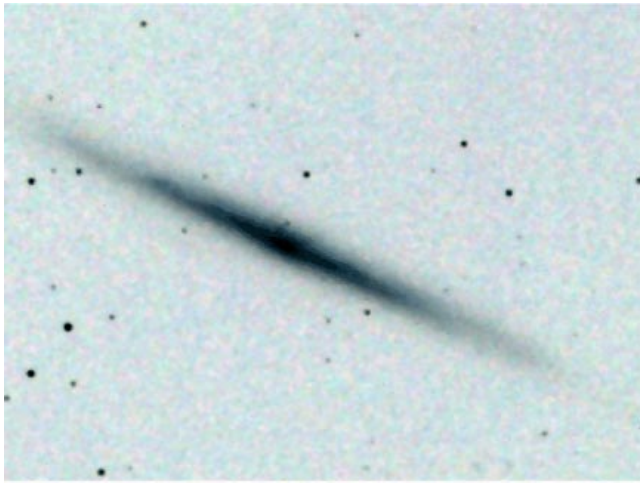

Figure 2. NGC 5907, an example of a galaxy with a thin disk observed edge-on (2MASS (Two Micron All-Sky Survey), 2 μm, near infrared).

Significantly 'overheated' disks are frequently observed in galaxies in pairs (circles in Fig. 1), obviously because they can experience a stronger gravitational perturbation from the companion. But among galaxies with 'overheated' disks, there are galaxies without close companions, and the increase in the velocity dispersion of their stars can then be due to the merging of the companions that are no longer observed as individual galaxies. But most important is that many lenticular and most spiral galaxies had no significant dynamical evolution over the several billion years of their lives, and their disks are kept weakly 'heated' in a state close to a marginally stable one. The presence of galaxies with enigmatically thin stellar disks points to the same fact. These galaxies show a low vertical velocity dispersion with respect to the rotational velocity and are frequently found among galaxies observed edge-on (Fig. 2).

Yet another argument favoring the existence of dark halos and dark matter in the Universe in general has a rather theoretical character: it is impossible to calculate the physical picture of galaxy formation within the standard theory of the expanding universe without assuming that most matter in the Universe is a nonbaryonic dark matter. The proper gravity of baryonic matter, which amounts to four percent of the critical density of the Universe, is far too insufficient to explain how minuscule primordial fluctuations could grow in a short time to form the observed galaxies and their systems.

In the framework of the so-called standard cosmological model, galaxies appear as a result of hierarchical clustering of numerous dark matter units (subhalos) in gravitational fields in which the primordial gas is concentrated, cools down, and then forms stellar galaxies. Later on, a long evolutionary path of the galaxy begins: the structure of the galaxy, the content and chemical abundance of gas and stars, and star formation rate can significantly change over several billion years. There are numerous problems from the theoretical standpoint: the role of processes such as the interaction of galaxies between themselves and with the intergalactic medium, galactic mergers, the activity of galactic nuclei, gas ejection from galaxies, and gas accretion on galactic disks that is capable of maintaining the current star formation rate for a long time remains unclear. In all these cases, the relative mass of dark matter in galaxies plays either a significant or a decisive role.

Numerical modeling of the galaxy formation process from dark matter and baryonic matter allowed finding, at least on the qualitative level, an explanation for the observed large-scale structure of the Universe and distribution of galactic mass. The galactic mass–size dependence [8] or mass concentration within the central kiloparsec, which is much lower than that predicted by numerical models (the so-called central cusp problem), are explained much more poorly. But one of the most relevant present-day problems of the hierarchical clustering model is the presence of a large number of *purely* disk galaxies in the nearby Universe (see, e.g., the statistics of the APM (Automated Plate Measuring) survey aimed at morphological classification of galaxies from their images [9]), i.e., galaxies without a significant central spheroid component, with only thin stellar disks. In many spiral galaxies, including our own, these thin stellar disks are also very *old*: the age of the oldest open clusters of our thin disk approaches 8–9 Gyr [10], and this means that starting from the time corresponding to the redshift $z = 1$, our Galaxy has not been seriously 'disturbed.'

The above proximity of the velocity dispersion of the disks of many galaxies to the minimum value required by gravitational stability, as well as the small velocity dispersion of disk stars in our Galaxy, also suggest the absence of strong dynamical heating of many galactic disks (see, e.g., [11]). This directly contradicts the hierarchical concept that predicts permanent galactic mergings during the evolution of the Universe. When small halos merge to produce a 10^{12} solar-mass halo (as in our Galaxy), the merged fragments preserve their identity inside the big halo for a long time. For example, recent GHALO (Galactic Halo) calculations [12] numerically modeling a small volume about 400 kpc in size with high spatial resolution have identified up to a hundred thousand 'subhalos' with continuous mass spectrum inside our dark halo. Dark matter clumps move inside the large halo in elongated orbits and inevitably cross the baryonic galactic disk. This 'bombardment' heats up the stellar disk, which gradually thickens due to an increase in the vertical velocity of stars. Moreover, dark matter clumps inside which stars emerged (dwarf satellites of our Galaxy) lose energy in encounters due to dynamical friction and finally fall on the disk to merge with it. According to recent calculations [13], a typical model galaxy similar to the Milky Way had to undergo about six 'minor' mergers during the last 8 Gyr (the time corresponding to $z = 1$), one of the mergers being with a satellite whose mass was about 10% of the mass of our Galaxy, which would thicken its disk by several times. In other words, thin stellar disks do not survive on a timescale of several billion years if they are plunged into dark matter halos as predicted by the theory; nevertheless, most nearby galaxies do have thin disks. This contradiction between theory and observations has not been resolved yet.

The key feature of the evolution of galactic disks is the permanent accretion of gas from outside, 'feeding' the star formation. The need for external gas accretion at a rate roughly comparable to that of star formation follows from many observational facts. In particular, the scenario of chemical evolution of the disk of our Galaxy cannot be built without considering an appreciable accretion of gas from outside (see, e.g., [14]). The star formation rate in the Galaxy disk over the last 9–10 Gyr is nearly constant (if averaged over 1–2 Gyr), which indirectly points to permanent gas accretion. Because stars synthesize all elements heavier than beryllium in the course of their evolution, the last generation of stars must

be much more metal-abundant than stars 8–10 Gyr old. However, no significant anticorrelation of metallicity with age is observed [15]. Moreover, there is the so-called G-dwarf problem: in a galaxy disk, stars of spectral class G with a mass around one solar mass or slightly lighter, among which stars of all ages are present, have almost the same metallicity within measurement errors [16]. It seems that the chemical evolution in the disk of our Galaxy over the last 8–10 Gyr 'did not go forth,' although nuclear reactions in the stars have undoubtedly occurred. This problem is resolved by introducing accretion from outside, i.e., by assuming that gas with minimal (and even better, zero) metal abundance has fallen onto the disk: such a low-metallicity gas 'dilutes' the gas enriched by the synthesized heavy elements and sustains the mean metallicity of the interstellar medium at a roughly constant level.

In spiral galaxy disks, including our own, there is a notable 'metallicity gradient': the mean heavy-element abundance in both stars and gas is higher at the center of the disk and decreases toward the disk periphery. Qualitatively, this is clear: in central parts of galaxies, star formation has already exhausted all the gas, i.e., was very effective, and on the periphery, a large amount of fresh gas remains, i.e., star formation has occurred very slowly. Naturally, star formation is then more effective at the disk center and more weak on its periphery. A model of the chemical evolution of the disk of our Galaxy with variable gas accretion along the radius was constructed in [17]. It was concluded there that the characteristic time of accretion in which the local disk density increases significantly (by e times) linearly increases along the radius: it is below 2 bln years at the center, about 8 Gyr near the Sun, and significantly exceeds the Hubble time on the far periphery (the disk only starts forming there). The characteristic times of large-scale star formation change correspondingly. This concept was dubbed 'inside-out,' meaning that the (disk) galaxy was formed inside out. This concept has been confirmed by many observational facts.

One of the most spectacular confirmations was obtained by the ultraviolet space telescope GALEX (Galaxy Evolution Explorer), which obtained images of a large sample of nearby galaxies with good sensitivity and reasonable space resolution in the far ($\lambda_{\rm eff} = 1516$ Å) and near ($\lambda_{\rm eff} = 2267$ Å) ultraviolet spectrum [18]. It turned out that many disk galaxies have much larger sizes in the ultraviolet band than in the visible range [19]. What does this mean? Young massive stars are the main 'contributors of ultraviolet' in galaxies: they have high temperatures well exceeding 10,000 K, and hence the energy is mainly released in the ultraviolet range. After the completion of the galaxy survey by GALEX, it turned out that star formation occurs in external regions of galaxy disks, where there are almost no old stellar populations and nothing is visible in the optical range. This means that the disks are indeed built up, or more precisely, built over in the outer parts, right in front of our eyes.

Recently, an interesting study was carried out to test this scenario [20]. The evolution of ultraviolet (i.e., star-forming) disks was observed by directly comparing the sizes of galaxies at different redshifts. How can this be done? Because of a finite speed of light, the more distant a galaxy is, the earlier epoch is observed. For example, a galaxy at the redshift $z = 0.5$ is seen 5 Gyr ago, while the time delay for $z = 1$ is 8 Gyr. This means that using modern large-aperture telescopes, we can directly probe two thirds of the age of the Universe and can observe the evolution of galaxies during most of their lives. Due to the redshift, the proper ultraviolet emission from a remote galaxy can be observed in the optical range. The authors of [20] inspected the change in the characteristic shape of the disk surface brightness radial distribution with z by fitting the observed wavelength to the redshift such that in the comoving frame of the galaxy, the measurements each time related to the same (ultraviolet) spectral range. They started from the GALEX galaxy survey carried out at $z \approx 0$. It turned out that indeed there is an evolution: at large redshifts, ultraviolet disks were observed to be more compact. At earlier times, star formation proceeded at the centers of the disks; has it now moved toward the periphery? This is exactly the evolution predicted by the inside-out scenario.

However, it turned out not to be so easy: when not only scales but also absolute levels of the ultraviolet surface brightness were compared, it was found that the peripheries of the disks appear almost identical at $z = 1$ and $z = 0$. The brightness profiles at $z = 1$ are more compact or, in other words, have larger slopes due to star formation in their centers at $z = 1$ occurring more intensively than at $z = 0$, while no evolution of the star formation rate at the disk peripheries is observed. That is, the disks have not 'grown' in the last 8 Gyr — they have simply been completing the star formation at the center and continued it at the periphery. The inside-out scenario clearly needed to be improved.

Another problem of the disk formation theory is that no real gas reservoirs for accretion have been discovered to date, although it is difficult to doubt the reality of gas accretion on galactic disks. In addition, it is important that the external gas also has a low metal abundance. At some period, it was thought that during the collapse of a dark matter halo, the primordial gas gravitationally bound to it gradually heats up during virialization and is preserved for a long time as a hot X-ray halo around the galaxy. By gradually cooling down, the hot X-ray halo could provide a long steady accretion of the primordial gas onto the entire galactic disk. Such hot gas halos are observed in galaxy clusters, but none has been found so far around a nearby spiral galaxy (with the possible exception of massive bulge regions in early-type galaxies). Moreover, a detailed gasdynamic model showed that even if such halos exist, the known mechanisms of thermal instability of hot virialized (i.e., equilibrated) gas are unable to provide the required amount of cold gas clouds near galactic disks and the steady accretion over billions of years is needed to build up a large-scale stellar disk [21].

For a long time, the outer gas source was 'nominated' to be high-velocity clouds of neutral hydrogen, which are actually observed outside the disk of our Galaxy. However, first, their number is too small to provide required accretion rates (at best, they give 0.1–0.2 solar masses per year, which is one order of magnitude smaller than required to sustain the modern star formation rate). Second, when the heavy-element abundance was estimated by absorption lines formed in clouds serendipitously located along the line of sight, the chemical composition of the gas of high-velocity clouds was close to the solar one, and hence this was not the primordial gas. Presently, most high-velocity clouds of neutral hydrogen are thought to consist of gas ejected from the Galaxy by so-called galactic fountains — gas outflows from active star formation regions, in which the gas is heated up by both stellar wind from massive stars and supernova explosions. This gas then cools to form clouds. However, first, this is not an 'addition' to the disk, but is originally a

proper part of the disk, and second, the chemical composition is not the primordial one but is instead enriched by the products of nucleosynthesis.

In recent years, important changes have been occurring in the theory of galaxy formation. Hot virialized gas halos of young galaxies are now 'disfavored' as sources of matter for stellar disk formation; theoreticians doubt that the gas virialization occurs in most of the collapsing halos. The formation of disks and bulges of galaxies does not necessarily have to occur via the merging of small-size subsystems only. A more and more important role in galaxy formation is probably played by cold filamentary gas flows directed to the inner part of a halo [22]. This is also a sort of flow accretion, but the accretion via gas streams that cannot occur on the entire disk and rather fuel its periphery. These cold streams pass without stopping through a hot gas halo and fall onto the disk. According to modern models [23], cold flows must dominate in low-mass (relative to the dark mass of clusters and groups of galaxies) halos at all redshifts starting from $z = 5-6$. This means that there has been no effective gas accretion from outside onto the *center* of the disk at any stage of galactic evolution. Therefore, the inside-out galaxy formation scenario in its classical formulation is now in conflict with both observations and the cosmological theory. Clearly, the time for its cardinal revision is coming.

Our studies of galaxy disks are supported in part by the RFBR grants 07-02-00792 and 07-02-00229.

References

1. Makarov D, Karachentsev I *Proc. Int. Astron. Union* **3** 370 (2008); arXiv:0801.0043
2. Karachentsev I D, Makarov D I *Astrofiz. Byull.* **63** 320 (2008) [*Astrophys. Bull.* **63** 299 (2008)]; arXiv:0812.0689
3. Bell E F, de Jong R S *Astrophys. J.* **550** 212 (2001)
4. Zasov A V, Khoperskov A V, Tyurina N V *Pis'ma Astron. Zh.* **30** 653 (2004) [*Astron. Lett.* **30** 593 (2004)]
5. Zasov A V et al. *Pis'ma Astron. Zh.* **28** 599 (2002) [*Astron. Lett.* **28** 527 (2002)]
6. Bizyaev D, Mitronova S *Astrophys. J.* **702** 1567 (2009)
7. Zasov A V, Khoperskov A V, Saburova A S, in preparation
8. Dutton A A et al. *Astrophys. J.* **654** 27 (2007)
9. Naim A et al. *Mon. Not. R. Astron. Soc.* **274** 1107 (1995)
10. Paunzen E, Netopil M *Mon. Not. R. Astron. Soc.* **371** 1641 (2006)
11. Wyse R F G *Proc. Int. Astron. Union* **4** 179 (2009); arXiv:0809.4516
12. Zemp M *Mod. Phys. Lett. A* **24** 2291 (2009)
13. Kazantzidis S et al. *Astrophys. J.* **700** 1896 (2009)
14. Tosi M, astro-ph/0308463
15. Feltzing S, Holmberg J, Hurley J R *Astron. Astrophys.* **377** 911 (2001)
16. Jørgensen B R *Astron. Astrophys.* **363** 947 (2000)
17. Chiappini C, Matteucci F, Gratton R *Astrophys. J.* **477** 765 (1997)
18. Gil de Paz A et al. *Astrophys. J. Suppl.* **173** 185 (2007)
19. Thilker D A et al. *Astrophys. J. Suppl.* **173** 538 (2007)
20. Azzollini R, Beckman J E, Trujillo I *Astron. Astrophys.* **501** 119 (2009)
21. Binney J, Nipoti C, Fraternali F *Mon. Not. R. Astron. Soc.* **397** 1804 (2009)
22. Dekel A, Birnboim Y *Mon. Not. R. Astron. Soc.* **368** 2 (2006)
23. Dekel A et al. *Nature* **457** 451 (2009)

PACS numbers: **95.35.+d**, **95.36.+x**, 98.80.Cq
DOI: 10.3367/UFNe.0180.201004j.0439

Dark components of the Universe

V V Burdyuzha

1. Dark energy

The multiverse, an internally growing fractal, is the new cosmological paradigm. The multiverse includes a large number of parts (universes) with different coupling constants, masses of fundamental particles, and other natural constants. Our Universe, whose age is about 14 Gyr, is one of them. During this time, the Universe has gone through a number of stages, including inflation, reheating, the radiation-dominated stage, and the matter-dominated stage, and is now in the vacuum-dominated stage. Starting from the redshift $z \sim 0.7$, the Universe is expanding with acceleration (at larger redshifts, the expansion decelerates).[1] The content of the Universe is also enigmatic. Baryons amounts to only 4% of the total density $\Omega_{\rm tot}$, dark matter ($\Omega_{\rm DM}$) contributes 23% to the total density, and the remaining 73% of the total density is due to dark energy ($\Omega_{\rm DE}$):

$$\Omega_{\rm tot} = \Omega_{\rm b} + \Omega_{\rm DM} + \Omega_{\rm DE} = 0.04 + 0.23 + 0.73\,,$$

$$\Omega_i = \frac{\rho_i}{\rho_{\rm cr}}\,, \qquad \rho_{\rm cr} = \frac{3H_0^2}{8\pi G_{\rm N}}\,,$$

where H_0 is the present-day value of the Hubble constant and $G_{\rm N}$ is the Newton gravitational constant.

Unfortunately, the nature of these components remains currently unknown, and they are therefore referred to as dark components, although more than a dozen models have been proposed for each of them. Very probably, dark energy is the vacuum. In this case, the cosmological constant, Λ-term, vacuum energy, and dark energy are identical notions. But in any case, it is better to start from the fundamental Einstein equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi G_{\rm N} T_{\mu\nu}\,, \qquad G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}\,R g_{\mu\nu}\,. \qquad (1)$$

Einstein introduced the cosmological constant Λ as a property of space. If the Λ term is placed in the right-hand side of the equations, then it can be treated as a form of energy, called dark energy (DE):

$$G_{\mu\nu} = -8\pi G_{\rm N} T_{\mu\nu} + \Lambda g_{\mu\nu}\,. \qquad (2)$$

The modern value of the DE density is

$$\rho_{\rm DE} = \rho_\Lambda \sim 10^{-47}\ {\rm GeV}^4 \approx 0.7\times 10^{-29}\ {\rm g\ cm}^{-3}\,,$$
$$\text{for}\quad H_0 = 70.5\ {\rm km\ s}^{-1}\ {\rm Mpc}^{-1}\,. \qquad (3)$$

[1] Presently, the redshift is $z = 0$, while at the time of the birth of the Universe, $z = \infty$.

V V Burdyuzha Astro Space Center, Lebedev Physical Institute RAS, Moscow, RF. E-mail: burdyuzh@asc.rssi.ru

Uspekhi Fizicheskikh Nauk **180** (4) 439–444 (2010)
DOI: 10.3367/UFNr.0180.201004j.0439
Translated by K A Postnov; edited by A M Semikhatov

In the Planck era, this energy density was

$$\rho_\Lambda \sim 2 \times 10^{76}\ \mathrm{GeV}^4\ (\approx 0.5 \times 10^{94}\ \mathrm{g\ cm}^{-3}) \quad \text{for} \quad M_{\mathrm{Pl}} = 1.2 \times 10^{19}\ \mathrm{GeV}\,, \tag{4}$$

which is 123 orders of magnitude larger than the DE density presently observed. This unexplained difference of 123 orders of magnitude gave rise to a crisis in physics, although, of course, many ways to overcome this problem have been proposed (see [1–5]).[2]

Here, it is relevant to recall the definition of the vacuum and its properties. In classical physics, the vacuum is the world without particles, and this world is flat. In quantum physics, the vacuum includes vacuum condensates resulting from relativistic phase transitions. In geometrical physics, the vacuum is a state in which the space–time geometry is nondeformed. A more general definition of the vacuum is as follows: a stable state of quantum fields without excitations of wave modes (nonwave modes represent condensates). The vacuum equation of state is $p = -\rho$. Hence, setting $w \equiv p/\rho$, we have the following cases:

if $w = -1$, the state is called the vacuum proper;

if $w > -1$, the state is called quintessence (scalar field);

if $w < -1$, the state is called phantom energy.

The last observational data obtained by the WMAP (Wilkinson Microwave Anisotropy Probe) satellite [6] suggest $-0.14 < 1 + w < 0.12$ at 95% confidence level (CL). A more precise value of the Hubble constant $H_0 \sim 70.5\ \mathrm{km\ s^{-1}\ Mpc^{-1}}$ was also inferred from these new data.

The vacuum in the Universe (as follows from its definition) is a combination of a large number of mutually dependent vacuum subsystems, including the gravitational condensate, the Higgs condensate, and the quark–gluon condensate. Other condensates with the energy in the range $265\ \mathrm{GeV} < E < 10^{19}\ \mathrm{GeV}$ are, unfortunately, poorly understood. The problem is how they are related and with what weight they contribute to the total vacuum energy, which can be defined as

$$\Lambda = \Lambda_{\mathrm{QF}} + \Lambda_{\mathrm{GVC}}\,, \tag{5}$$

where Λ_{QF} are quantum field condensates and Λ_{GVC} is the gravitational vacuum condensate — a new vacuum structure [7] including topological defects of different dimensions, such as wormholes (three-dimensional defects), micromembranes (two-dimensional defects), and point-like monopole defects. Of course, higher-dimensional microdefects are also present in this condensate. Three-dimensional microdefects (wormholes) explicitly contribute to the total vacuum energy density, i.e., renormalize the Λ term:

$$\Lambda = \Lambda_0 - \frac{\kappa \hbar^2 c_3^2}{768\pi^2}\,, \tag{6}$$

where κ is the gravitational constant in the system of units where $c = 1$ and $\hbar = 1$, and c_3 is a coefficient parameterizing the function $\mu(a)$ (see paper [7] for more details). Equation (6) gives the first indication of the presence of a compensation mechanism in the vacuum of our Universe, because three-dimensional topological defects decrease the Λ-term. As the temperature decreased, the condensates of other quantum fields also made negative contributions to the positive vacuum energy density (as the temperature was decreasing, the Universe lost its symmetry by forming condensates). The compensation hypothesis seems to have been first proposed by A D Dolgov.

We emphasize the profound meaning of the observed 'smallness' of the cosmological constant. A universe with a large negative Λ never becomes macroscopic (in order to give physical meaning to these statements, Λ-antigravity can be considered). In a universe with a large positive Λ, complex nuclear, chemical, and biological structures would be absent (because there would not be enough time for their formation).

But we continue discussing the compensation hypothesis and traces of relativistic phase transitions. Their plausible chain can be given by

$$\begin{aligned} \mathrm{P} &\underset{10^{19}\ \mathrm{GeV}}{\Rightarrow} \mathrm{D}_4 \times \left[\mathrm{SU}(5)\right]_{\mathrm{SUSY}} \\ &\underset{10^{16}\ \mathrm{GeV}}{\Rightarrow} \mathrm{D}_4 \times \left[\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)\right]_{\mathrm{SUSY}} \\ &\underset{\sim(10^5\text{–}10^{10})\ \mathrm{GeV}}{\Rightarrow} \mathrm{D}_4 \times \mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3) \\ &\underset{100\ \mathrm{GeV}}{\Rightarrow} \mathrm{D}_4 \times \mathrm{U}(1) \times \mathrm{SU}(3) \underset{0.15\ \mathrm{GeV}}{\Rightarrow} \mathrm{D}_4 \times \mathrm{U}(1)\,, \end{aligned} \tag{7}$$

where only the last two phase transitions can be specified in detail because they are well studied (these energies can be reached by current accelerators):

$$\Lambda_{\mathrm{QF}} = \Lambda_{\mathrm{EW}} + \Lambda_{\mathrm{QCD}}\,; \qquad \rho_{\mathrm{QF}} = -\rho_{\mathrm{EW}} - \rho_{\mathrm{QCD}}\,. \tag{8}$$

The electroweak (EW) phase transition occurred at the temperature about 100 GeV and was accompanied by the appearance of the Higgs condensate, which also contributed to decreasing the vacuum energy:

$$\rho_{\mathrm{EW}} = -\frac{m_{\mathrm{H}}^2 m_{\mathrm{W}}^2}{2g^2} - \frac{1}{128\pi^2}\left(m_{\mathrm{H}}^4 + 3m_{\mathrm{Z}}^4 + 6m_{\mathrm{W}}^4 - 12m_{\mathrm{t}}^4\right). \tag{9}$$

The first term in the right-hand side of Eqn (9) is the energy density of the semiclassical Higgs condensate, the second term is the vacuum polarization by quantum fields, m_{H}, m_{Z}, m_{W}, and m_{t} are masses of the Higgs boson, Z and W bosons, and the t quark, and g is the coupling constant. The boson contribution in formula (9) is negative, while the fermion one (t quark) is positive. Because the values of all constants (except the Higgs boson mass) are known, the vacuum stability condition can be derived: in the Standard Model (SM), the mutual compensation of positive and negative contributions to the vacuum energy density is prohibited by the stability condition! Therefore, the statement about a vacuum energy decrease by symmetry breaking in the evolution of the Universe due to relativistic phase transitions is related to the vacuum stability condition and apparently bears a universal character. For the Higgs boson mass $m_{\mathrm{H}} \sim 2m_{\mathrm{W}} \sim 160\ \mathrm{GeV}$,

$$\rho_{\mathrm{EW}} \sim -(120\ \mathrm{GeV})^4\,, \tag{10}$$

and there is little doubt that the Higgs boson will be discovered with the Large Hadron Collider.

The nonperturbative quark–gluon condensate is an element of theory incorporated into the SM. Studying quantum chromodynamics (QCD) equations has shown that the phenomenon of confinement occurs if quantum correla-

[2] A fresh look at the cosmological constant problem was suggested in recent review [5].

tors of quark–gluon fields are nonzero. The quark–gluon condensate is a system of mutually correlated nonperturbative fluctuations resulting from quantum topological tunnel transitions between degenerate states of the gluon vacuum [8]. The energy density of this condensate is

$$\rho_{\mathrm{QCD}} = -\frac{b}{32}\left\langle 0\left|\frac{\alpha_{\mathrm{s}}}{\pi}\, G^a_{ik}G^{ik}_a\right|0\right\rangle, \tag{11}$$

where $b = 9 + 8T_g(m_{\mathrm{u}} + m_{\mathrm{d}} + 0.8m_{\mathrm{s}}) \approx 9.6$, $T_g = (1.5\,\mathrm{GeV})^{-1}$ is the characteristic space–time scale of fluctuations, and m_{u}, m_{d}, and m_{s} are masses of u, d, and s quarks. The principal energy parameter of the quark–gluon condensate is

$$u^4 = \left\langle 0\left|\frac{\alpha_{\mathrm{s}}}{\pi}\, G^a_{ik}G^{ik}_a\right|0\right\rangle \approx (360\ \mathrm{MeV})^4\,.$$

According to the modern paradigm, the quark–gluon condensate has several phase states, in each of which fluctuations have a specific microstructure. As a result, we have

$$\rho_{\mathrm{QCD}} = -\frac{b}{32}\, u^4 \approx -(265\ \mathrm{MeV})^4\,. \tag{12}$$

The quark–hadron phase transition alone suppresses more than 10 orders ($120^4/0.265^4 \sim 4\times 10^{10}$) of the total vacuum energy decrease (more than 78 orders of magnitude) by vacuum condensates:

$$\left(\frac{M_{\mathrm{Pl}}}{M_{\mathrm{QCD}}}\right)^4 = \left(\frac{1.2\times 10^{19}}{0.265}\right)^4 \approx 4.5\times 10^{78}\,. \tag{13}$$

It is therefore quite plausible that the Universe lost more than 78 orders during the first 10^{-5} s of its evolution. The QCD phase transition was the last in the sequence of phase transitions and was a specific marker. The point is that the chiral $\mathrm{SU}(3)_{\mathrm{L}} \times \mathrm{SU}(3)_{\mathrm{R}}$ symmetry was not exact, and pseudo-Goldstone bosons are a physical realization of this symmetry breaking at $E \sim 265$ MeV.[3] Many years ago, D A Kirzhnits drew my attention to the fact that π-mesons are pseudo-Goldstone bosons. Therefore, π-mesons, being the lightest particles of the octet of pseudo-Goldstone states, characterize the ground state. In this case, the vacuum is the ground state. More than 40 years ago, from dimensional considerations, Zel'dovich [9] derived a formula for calculating the Λ-term (the formula was slightly modified by N S Kardashev), according to which the cosmological constant is the sum of zero oscillations of quantum fields, i.e., the vacuum energy:

$$\Lambda = 8\pi G_{\mathrm{N}}^2 m_\pi^6 h^{-4}\ [\mathrm{cm}^{-2}]\,, \qquad \rho_\Lambda = G_{\mathrm{N}} m_\pi^6 c^2 h^{-4}\ [\mathrm{g\ cm}^{-3}]\,, \tag{14}$$

and the vacuum condensate of the last phase transition can then be calculated as

$$\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_{\mathrm{cr}}} = \frac{\Lambda c^2}{3H_0^2}\,, \qquad \rho_{\mathrm{cr}} = \frac{3H_0^2}{8\pi G_{\mathrm{N}}}\,. \tag{15}$$

For the average π-meson mass $m_\pi = 138$ MeV and $H_0 = 70.5$ km s^{-1} Mpc^{-1}, we find $\Omega_\Lambda \approx 0.73$. The last value is almost equal to the observed dimensionless vacuum energy density $\Omega_\Lambda \approx 0.726 \pm 0.015$ obtained by the WMAP collaboration [6]. In other words, we can say that the vacuum energy at that time (10^{-5} s) was 'quenched' after the abrupt compensation by quantum field condensates.

But it remains to 'suppress' (during about 14 Gyr) almost 44 orders of magnitude to reach the present-day value of the vacuum energy (DE) density

$$\rho_{\mathrm{DE}} \sim (1.8\times 10^{-12}\ \mathrm{GeV})^4 \quad \left(\left(\frac{0.15}{1.8\times 10^{-12}}\right)^4 \sim 5\times 10^{43}\right). \tag{16}$$

Over this huge time period, the vacuum energy must have changed, because new quantum states had to be created in the expanding Universe at the expense of decreasing its energy. But during this period, the rate of the vacuum energy change was 10^{57} times smaller than during the quantum period of its evolution. To understand the 'recent' changes in the vacuum energy from 0.15 GeV to 1.8×10^{-12} GeV, we consider the holographic principle introduced by 't Hooft [10].[4] According to this principle, the 'physics' of a three-dimensional system can be described by a theory formulated on its two-dimensional boundary. Using the anti-de Sitter space–time, J Maldacena and E Witten showed that the description of the Universe by superstring theory corresponds to its description by quantum field theory formulated on its boundary. However, this example appears unconvincing, because our space–time is the de Sitter space–time. But we consider this problem at a greater depth, because there is a holographic limit on the number of the degrees of freedom that can exist inside a bounded surface. Bekenstein showed [11] that the entropy of a black hole is proportional to 1/4 of its horizon area expressed in Planck units. Therefore, if one bounds our Universe and wants to measure this 'boundary,' as proposed in [12], also in Planck units, then the vacuum energy density in the holographic limit is expressed by the simple formula

$$\rho_{\mathrm{DE}} = \frac{3M_{\mathrm{Pl}}^4}{8S}\,, \tag{17}$$

where $S \leqslant \pi R^2 M_{\mathrm{Pl}}^2$ is the entropy of the Universe. For $R = 10^{28}$ cm, $\rho_{\mathrm{DE}} \sim 10^{-57}$. In this formula, the entropy is proportional to 1/4 of the area of the 'surface' of the Universe. In fact, this derivation is called the Fischler–Susskind holographic conjecture [12].

It is also important to discuss the applicability of the holographic approach and to show how the remaining 44 orders of magnitude can be suppressed. As noted by 't Hooft [10], the entropy bound yields an upper limit on the mean energy density in the Universe. The physics here is as follows: new quantum degrees of freedom are generated as the area of the Hubble horizon increases, and their continuous creation requires some energy expenditure (see paper [12] for more details about the holographic approach in cosmology; it is noted there that general relativity (GR) is an illuminating example of the holographic theory). However, the holographic approach appears relevant as far as GR is applicable. Quantum theory in its present state is not a holographic theory. Using these 'arguments,' we make a numerical

[3] The QCD phase transition started at $E \sim 265$ MeV and ended at $E \sim 150$ MeV.

[4] All previous physical principles, including the Pauli principle, the equivalence principle, the relativity principle, and the Heisenberg uncertainty principle have led to significant progress in physics.

estimate. The holographic approach can possibly be used after a series of relativistic phase transitions, starting from the 'quenching' of the vacuum energy at $E \sim 150$ MeV, $t \sim 10^{-5}$ s, and $R \sim 3 \times 10^5$ cm (where R is the causality horizon at that time). The present-day size of the Universe is $R \sim 10^{28}$ cm, and the vacuum energy has lost around 45 orders of magnitude due to creation of new quantum states,

$$\left(\frac{10^{28}}{3 \times 10^5}\right)^2 \approx 10^{45}\,, \tag{18}$$

or even more if the causality horizon was smaller than 10^5 cm at the instant of quenching.

Summarizing the above arguments, we note that in the early Universe when the energy decreased from 10^{19} GeV to 150 MeV, quantum field condensates compensated 78 orders of magnitude of the initial vacuum energy density over the time period of only 10^{-5} s. Then in the next 14 Gyr ($\approx 4 \times 10^{17}$ s), the vacuum component lost another 45 orders of magnitude of its energy density due to the creation of new quantum states in the expanding Universe. This means that 123 orders of magnitude of the vacuum energy density were lost in ordinary physical processes.

However, we now discuss some considerations related to maverick statements that the black hole thermodynamics follow from the thermal nature of the Minkowski vacuum and the Einstein equations have a thermodynamic origin [13], i.e., these equations represent the equation of state of the Universe. The link established by Bekenstein between gravity and thermodynamics is extremely interesting because the Einstein equation is a second-order hyperbolic differential equation for the space–time metric, and thermodynamics is apparently irrelevant. But the point is that the Universe expands (with acceleration) and gradually cools down, and hence we here have a different physical situation, because nonequilibrium thermodynamics is applicable. The Klausius relation $\mathrm{d}S = \delta Q/T$ can be applied, where the entropy $\mathrm{d}S$ is equal to $1/4$ of the area of the horizon, δQ is the energy flux through the horizon, and T is the Unruh temperature seen by an accelerating observer inside the horizon [13]. The ideological foundation underlying these considerations is the statement that gravity on macroscopic scales is the manifestation of the thermodynamics of the vacuum. New quantum states are created in the Universe at the expense of the energy of the vacuum, and Eqn (17) is the Friedman equation.

Summarizing this part of the talk, we can argue that a satisfactory numerical difference between the vacuum energy density at the Planckian time and the present time could be realized if the compensation hypothesis and holographic approach are used to suppress 123 critical orders of magnitude due to phase transitions and the creation of new quantum states.

In addition, we also note that dark energy more and more 'gives way' to vacuum energy, as follows from the recent experimental results obtained by the WMAP collaboration [6] mentioned above.

2. Dark matter

The situation with dark matter (DM), another component of our Universe, with the density $\Omega_{\mathrm{DM}} \sim 0.23$, is equally intriguing. As early as 1933, Swiss astrophysicist Fritz Zwicky, who was working in the USA at a large telescope, recognized that a galaxy cluster in Coma Berenices (the Coma cluster) cannot be gravitationally bound unless an additional mass is present, which later was called dark mass. In recent decades, it has become clear that dark matter in our Universe is much more abundant than the visible matter ($\Omega_{\mathrm{stars}} \sim 0.005$) and baryonic matter ($\Omega_{\mathrm{b}} \sim 0.04$). A new field has even emerged, DM cosmology, whose main goal is to understand the nature of DM particles. Formally, this component of the Universe can have a baryonic nature. For example, it might consist of underformed 'Jupiter' stars, whose mass is too small for thermonuclear reactions to begin, or other baryonic structures, including black holes and white dwarfs. In particular, as stressed in review [14], more than half of dark matter can be baryonic. In addition, the Universe can harbor nonradiating remnants of black holes with the Planck mass (10^{-5} g), and, possibly, preonic stars with $M \sim 10^2\, M_{\oplus}$ (the Earth mass is $M_{\oplus} \sim 6 \times 10^{27}$ g) [15]. In the EROS-2 experiment (from the French Expérience pour la Recherche d'Objets Sombres) [16], the masses of massive astrophysical compact halo objects (MACHOs) were measured by microlensing 25 mln stars in Magellanic Clouds: $M_{\mathrm{MACHO}} \sim (10^{-2} - 10^{-6}) M_{\odot}$ (the mass of the Sun $M_{\odot} \approx 2 \times 10^{33}$ g). Omitting the details, we note that the baryonic component must necessarily be present in dark matter because $\Omega_{\mathrm{b}} \sim 0.04$ and $\Omega_{\mathrm{stars}} \sim 0.005$.

We note that a 'mad' hypothesis was proposed in [17] that the Standard Model can have an infinite number of replicas, and then the presence of baryons in hidden replicas can naturally explain dark matter. In that model, gravity is strong already at 1 TeV energies and the number of copies of dark baryons can be enormous:

$$10^{11} \leqslant N \leqslant 10^{32}\,. \tag{19}$$

This large number of copies became 'possible' by virtue of a new permutation symmetry introduced in [17]. This also involves another interesting point: inflation occurs due to the inflaton pertaining to our copy of the SM, while reheating after inflation is mediated by a modular field that is common for all copies. Nonbaryonic dark matter must consist of stable particles or these particles must have lifetimes exceeding the age of the Universe. Such particles primarily include neutrinos, neutralinos, and axions.

Of course, there are other models for dark matter, including the hypothetical Kaluza–Klein dark matter. Such dark matter can annihilate to produce charged leptons, which could be responsible for the electron–positron asymmetry in cosmic rays observed in PAMELA (Payload for Antimatter-Matter Exploration and Light-nuclei Astrophysics) and ATIC (Advanced Thin Ionization Calorimeter) experiments [18]. Nonbaryonic particles of dark matter can be cold (CDM) or hot (HDM) [19]:

$$\Omega_{\mathrm{CDM}} \sim 0.223 \pm 0.016\,, \qquad \Omega_{\mathrm{HDM}} \sim 0.0152\,. \tag{20}$$

We recall that if these particles at the time of formation (quenching) had relativistic or nonrelativistic velocities, then they are HDM or CDM particles, respectively. For example, a light neutrino is an HDM particle.[5] The discovery of neutrino oscillations was a watershed in our understanding of the nature of the neutrino. This discovery brought the physics of neutrinos into the focus of the physics community, because it became clear that neutrinos have mass (even the industry of neutrino oscillations emerged; see, e.g., http://

[5] Neutrinos with a mass above 1 GeV are already CDM particles.

neutrinooscillation.org). The small neutrino masses confirmed theoretical expectations of the early 1980s that the so-called 'see-saw' mechanism induces both light and superheavy neutrinos; however, some 'neutrino challenges' remain for the future. They include the CP-violation amplitude in neutrino oscillations, the existence of neutrinoless β-decay (in which case the neutrino is a Majorana particle), and the absolute scale of neutrino masses. The data on neutrino oscillations are summarized in [20] at the 3σ level:

$$\Delta m^2_{2,3} = (1.4-3.3)\times 10^{-3}\ \mathrm{eV}^2\,, \qquad \Delta m^2_{1,2} = (7.2-9.1)\times 10^{-5}\ \mathrm{eV}^2\,. \tag{21}$$

We also mention new long-based oscillation experiments, such as MINOS (Main Injector Neutrino Oscillation Search), CNGS (CERN Neutrinos to Gran Sasso), ICARUS (Imaging Cosmic And Rare Underground Signal), and OPERA (Oscillation Project with Emulsion-tRacking Apparatus), which will 'hunt' for $\nu_\mu \to \nu_e$ oscillations [21].

Cosmological constraints are sensitive to all three neutrino mass flavors:

$$\sum_i m_i \leqslant 0.2-1.7\ \mathrm{eV}\ (95\%\,\mathrm{CL}), \tag{22}$$

but cosmology does not provide values of mixing angles or possible CP-violations. Cosmological bounds can be obtained from the WMAP data on the CMB anisotropy, from the large-scale distribution of galaxies in the Sloan Digital Sky Survey (SDSS), from the Hubble Space Telescope (HST) observations, and from measurements of remote type-Ia supernovae. A new cosmological bound on neutrino masses was recently obtained in [23]: $m_\nu < 1.05$ eV. Experiments on the β-decay of tritium in [24] give the following limit on the electron antineutrino mass:

$$m_{\bar\nu} < 2.05\ \mathrm{eV}\,. \tag{23}$$

The KATRIN (KArlsruhe TRItium Neutrino) collaboration promises to measure the neutrino mass with a sensitivity of 0.25 eV [25]. There are interesting prospects in the neutrinoless β-decay ($\beta\beta_{0\nu}$), in which the effective Majorana neutrino mass is

$$m_{\mathrm{M}}^{\mathrm{eff}} \leqslant 0.3-1\ \mathrm{eV}\,. \tag{24}$$

We note that the lepton number conservation must be violated in the neutrinoless double β-decay, and therefore its detection will be a direct manifestation of supersymmetry, because the lepton number nonconservation (like the baryon number nonconservation) is a key prediction of supersymmetric (SUSY) theories.

All neutrino research groups are huge collaborations of different specialists. Neutrino fluxes from the Sun and even from supernovae are observed and measured by different methods in deep mines (Sudbury Neutrino Observatory in Canada, Baksan Neutrino Observatory in Russia, Boulby Mine Laboratory in the UK, National Laboratory Gran Sasso in Italy, Kamioka in Japan), in nuclear reactor experiments, and in secondary cosmic ray cascades. The general conclusion is that the main component of HDM,[6] neutrinos, provides a certain nonnegligible contribution to the dark matter density ($n_\nu \sim 112\ \mathrm{cm}^{-3}$ per neutrino flavor).

We now consider the principle CDM components of dark matter, such as the neutralino and axion (these particles come from an extension of the Standard Model). These are more exotic particles than neutrinos, but they should not stay in this category for a long time.

The neutralino χ is a weekly interacting massive particle (WIMP) that could have originated in the early Universe if supersymmetry took place. Supersymmetry can naturally solve the dark matter problem because in most minimal SUSY models, the lightest superpartner is absolutely stable due to the conservation of a multiplicative quantum number (the R-parity). Probably, superparticles born in pairs in the early Universe rapidly decayed to form the lightest supersymmetric particles in addition to ordinary particles. They must be noncharged and not strongly interacting in order to not violate the Big Bang Nucleosynthesis (BBN). These requirements are satisfied for the neutralino, which is described by a Majorana spinor. The wave function of the neutralino is given by a superposition of wave functions of four supersymmetric particles: two gauginos and two higgsinos. If neutralinos build up the halo of our Galaxy, then their number density is [26]

$$n_\chi \sim \frac{0{,}3}{m_\chi\ \mathrm{GeV}}\ [\mathrm{GeV\ cm}^{-3}]\,. \tag{25}$$

As noted in [14], neutralinos in the halo of our Galaxy (and, naturally, in halos of other galaxies) could form small-scale ($R \sim 10^{14}-10^{15}$ cm) hierarchical objects and even neutralino stars. Neutralinos could be observed by their decay products during annihilation. Seven (!) underground laboratories are searching out searches for the annihilation products. The neutralino annihilation signal falls in the energy range 100–200 GeV (energies available with the LHC), and the neutralino contribution to the total density of the Universe is

$$0.1 < \Omega_\chi < 0.3\,, \quad \text{if} \quad 5\times 10^{-8} < \sigma_\chi < 5\times 10^{-10}\ \mathrm{pb} \tag{26}$$

($1\ \mathrm{pb} \equiv 10^{-36}\ \mathrm{cm}^2$). Searches for neutralinos are being carried out by different neutrino research groups, such as SuperKamiokande, Baykal, Ananda, Baksan, and ANTARES (Astronomy with Neutrino Telescope and Abyss environmental RESearch) in the Mediterranean Sea.

The axion was postulated more than three decades ago to explain the P- and CP-symmetry conservation in strong interactions, although these symmetries are violated in the Standard Model (in the electroweak sector). Peccei and Queen [27] proposed solving the strong CP problem by introducing a new global symmetry U_{PQ}. Then axions appear as Nambu–Goldstone bosons associated with spontaneous breaking of this symmetry. The axion has zero spin, zero electric charge, and negative internal parity. The mass of the axion is

$$m_{\mathrm{a}} \sim 6\times 10^{-6}\,\frac{10^{12}}{f_{\mathrm{a}}}\ [\mathrm{eV}]\,, \tag{27}$$

and if the free coupling constant is $f_{\mathrm{a}} < 10^{12}$ GeV, the density of axions does not exceed the critical density in the Universe. In this case, $m_{\mathrm{a}} \sim 10^{-5}$ eV. This particle could have been

[6] Sterile neutrinos, as well as gravitinos, are warm dark matter.

formed during the QCD phase transition in the very early Universe. Axions can be detected in laboratory by stimulating their conversion into two microwave photons by a strong magnetic field [28]. The ADMX (Axion Dark Matter eXperiment) is aimed at registering relic axions. In a pilot search carried out by the ADMX collaboration, no axions were found in the mass range 1.98–2.17 μeV [29].

We note another important fact, which has no direct relation to the discussed problems, but is related to axions, more precisely, to familons [30]. If the next fundamental level of matter (preons) is discovered, the role of particle generations will be more transparent. The first particle generation composes our baryon world. The account for symmetry between generations (due to their mere existence) produces all dark matter. Therefore, particles (familons) that appear from symmetry between generations can explain structurization of dark matter and the subsequent structurization of baryons.

In a familon medium, a phase transition could occur that quenched the fractality (fractality is the prerogative of phase transitions only), and baryons would then reproduce the dark matter distribution. In that case, the fractality of the large-scale baryon structure can be naturally explained. We must also mention studies on the possible interaction of dark energy and dark matter (see [31] and the references therein), as well as $f(R)$ gravity studies (see [32] and the references therein), which are directly related to dark matter.

Other models for dark matter, which we did not discuss in detail here, include sterile (supersymmetric) neutrinos, gravitinos, axinos, light scalar particles, light Higgs bosons, Kaluza–Klein dark matter, superheavy dark matter (simpzillas), nontopological solitons (Q-balls), charged massive particles (CHAMPS), weakly interacting dark matter (SWIMPS), braneworld dark matter, heavy neutrinos of the fourth generation, and mirror particles. The list of exotic dark matter candidates was presented, for example, in the talk by J Colar at the Schramm memorial symposium in December 2005.

To conclude, we note that our main result is the explanation of the huge difference (123 orders of magnitude) between the vacuum energy density at the birth of the Universe and that at the present time, which initiated a long-lasting crisis in physics. In our paper [33], the vacuum density was calculated for redshifts ranging from $z = 0$ to $z = 10^{11}$ using the 'cosmological calculator' [34]. Apart from that, we have recognized why three particle generations are needed in our Universe. But such a 'recognition' invokes the next fundamental (preonic) level of matter. Then the first generation of particles forms the observed baryonic world, and the account of symmetry between generations yields all dark matter.

References

1. Weinberg S *Rev. Mod. Phys.* **61** 1 (1989)
2. Peebles P, Ratra B *Rev. Mod. Phys.* **75** 559 (2003)
3. Bousso R *Gen. Rel. Grav.* **40** 607 (2008); arXiv: 0708.4231
4. Frieman J A, Turner M S, Huterer D *Annu. Rev. Astron. Astrophys.* **46** 385 (2008); arXiv: 0803.0982
5. Burdyuzha V V *Astron. Zh.* **86** 419 (2009) [*Astron. Rep.* **53** 381 (2009)]
6. Komatsu E et al. *Astrophys. J. Suppl.* **180** 330 (2009); arXiv: 0803.0547
7. Burdyuzha V, Vereshkov G *Astrophys. Space Sci.* **305** 235 (2006)
8. Marochnik L, Usikov D, Vereshkov G, arXiv:0811.4484
9. Zel'dovich Ya B *Pis'ma Zh. Eksp. Teor. Fiz.* **6** 883 (1967) [*JETP Lett.* **6** 316 (1967)]
10. Stephens C R, 't Hooft G, Whiting B F *Class. Quant. Grav.* **11** 621 (1994); gr-qc/9310006
11. Bekenstein J D *Phys. Rev. D* **7** 2333 (1973)
12. Balazs C, Szapudi I, hep-th/0603133
13. Eling Ch, Guedens R, Jacobson T *Phys. Rev. Lett.* **96** 121301 (2006); gr-qc/0602001
14. Gurevich A V, Zybin K P, Sirota V A *Usp. Fiz. Nauk* **167** 913 (1997) [*Phys. Usp.* **40** 869 (1997)]
15. Hansson J, Sandin F *Phys. Lett. B* **616** 1 (2005); astro-ph/0410417
16. Renault C et al. *Astron. Astrophys.* **324** L69 (1997)
17. Dvali G, Sawicki I, Vicman A *JCAP* (08) 009 (2009); arXiv: 0903.0660
18. Hooper D, Zurek K M *Phys. Rev. D* **79** 103529 (2009); arXiv: 0902.0593
19. Frampton P H, astro-ph/0506676
20. Maltoni M et al. *New J. Phys.* **6** 122 (2004)
21. Balantekin A B *Phys. Lett. B* **613** 61 (2005); hep-ph/0405019
22. Dolgov A D *Phys. Rep.* **370** 333 (2002)
23. Malinovsky A M et al. *Pis'ma Astron. Zh.* **34** 490 (2008) [*Astron. Lett.* **34** 445 (2008)]
24. Lobashev V M *Nucl. Phys. A* **719** C153 (2003)
25. Karlsruhe Tritium Neutrino Experiment. KATRIN, http:// www-ik.fzk.de/ ˜ katrin
26. Majorana Collab., nucl-ex/0311013; SuperNEMO, http:// nemo.in2p3.fr/supernemo
27. Peccei R D, Quinn H R *Phys. Rev. Lett.* **38** 1440 (1977)
28. Sikivie P *Phys. Rev. Lett.* **48** 1156 (1982)
29. Duffy L D et al. *Phys. Rev. D* **74** 012006 (2006)
30. Burdyuzha V et al. *Astron. Astrophys. Trans.* **23** 453 (2004)
31. Micheletti S, Abdalla E, Wang B *Phys. Rev. D* **79** 123506 (2009)
32. Motohashi H, Starobinsky A A, Yokoyama J *Int. J. Mod. Phys. D* **18** 1731 (2009)
33. Burdyuzha V, arXiv:1003.1025; *Phys. Lett. B* (2010), submitted
34. Wright E R *Publ. Astron. Soc. Pacific* **118** 1711 (2006)