Astronomy Reports, Vol. 49, No. 11, 2005, pp. 859–862. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 963–966.

Variations in the Very-High-Energy Gamma-Ray Radiation from the Blazar H 1426+428

V. V. Fidelis, Yu. I. Neshpor, V. S. Eliseev, N. A. Zhogolev, E. M. Nekhaĭ, and Z. N. Skiruta

Crimean Astrophysical Observatory, Nauchnyĭ, Crimea, Ukraine

Received January 15, 2005; in final form, May 18, 2005

Abstract—Observations of the blazar H 1426+428 were carried with on the GT-48 Cerenkov telescope of the Crimean Astrophysical Observatory from April 15 to April 25, 2004. Analysis of the observational data shows an excess of very-high-energy photons ($\geq$1 TeV) in the direction of the source, with a statistical confidence of 5.8σ. A flash lasting no more than seven days was detected. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Blazars are objects of special interest in very-high-energy (VHE) astrophysics. Their radiation, which extends over more than 15 decades in energy, can be studied on various time scales. The infrared to X-ray emission is dominated by synchrotron radiation from the magnetized plasma associated with the relativistic jets that are ejected from the central regions of these objects and oriented roughly toward the observer (in the case of blazars). The radiation at higher energies, reaching the GeV and TeV ranges, may be due to the inverse Compton effect. The preferred model involves the scattering of the synchrotron photons on the same electrons that generated them [1].

It was discovered by the EGRET detector operating at high energies from 30 MeV to ≈20 GeV at the Compton Gamma-ray Observatory [2] that most extragalactic sources of high-energy gamma-rays are blazars—a discovery which has stimulated interest in studies of these objects. The gamma-ray radiation of blazars often dominates the remaining spectrum and is emitted by relatively small regions, consistent with the observed rapid variations of blazars (on time scales of less than an hour at TeV energies for Mrk 421 [3]). Therefore, long-continued observations of blazars at VHE can reveal characteristic time scales for their variability and the corresponding dimensions of emitting regions, and can also be used to identify mechanisms for the acceleration of charged particles (electrons, positrons, and protons) and the transfer of this energy into gamma-ray photons [4].

VHE gamma-rays can be detected only on ground-based instruments designed to detect the Cerenkov radiation associated with extensive air showers (EAS) generated by the gamma-rays as they travel through the atmosphere. The Cerenkov radiation from EAS initiated by VHE gamma-rays lasts for several nanoseconds and propagates in a narrow cone with an opening angle of about 1°, whose axis coincides with the direction of propagation of the primary gamma-ray. The intensity of this emission is proportional to the energy of the primary gamma-ray.

If the energy of the primary gamma-ray is 1 TeV, the Cerenkov light covers an area on the Earth's surface of about 10^4 m^2 and has a typical flux of 50–100 optical photons per square meter [5]. Once the radiation is measured using ground-based Cerenkov telescopes, its properties can be used to derive the direction of arrival and energy of the generating gamma-ray.

The first extragalactic objects for which fluxes of VHE gamma-rays were detected were the blazars Mrk 421 ($z = 0.031$) [6] and Mrk 501 ($z = 0.033$) [7]. These objects, as well as most other extragalactic sources of VHE gamma-rays, belong to the subclass of blazars known as BL Lac objects, which are characterized by weak or absent optical line emission. The blazar H 1426+428 ($z = 0.129$) is at a far greater distance than the two other objects indicated above, making observations of it interesting, especially for studies of the absorption of VHE gamma-rays in intergalactic space [8].

Analysis of observations of H 1426+428 using the HEGRA Cerenkov telescope in 1999, 2000, and 2002 showed the presence of fluxes of VHE gamma-rays at the 6.6σ level [9]. We present here the results of observations of this object carried out at the Crimean Astrophysical Observatory on the GT-48 gamma-ray telescope in 2004.

2. THE CERENKOV TELESCOPE

The GT-48 Cerenkov telescope is located 600 m above sea level, and consists of two identical sections

1063-7729/05/4911-0859$26.00

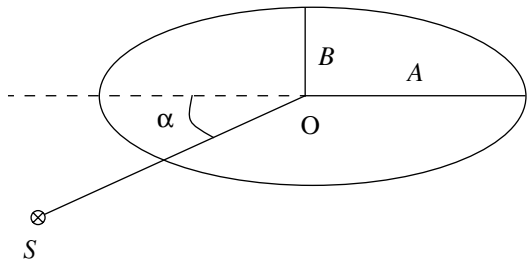

Fig. 1. Schematic representation of the main parameters of an image of a flash. The letter S marks the position of the source and the O the center of the flash brightness distribution. The segment OS corresponds to the parameter DIST.

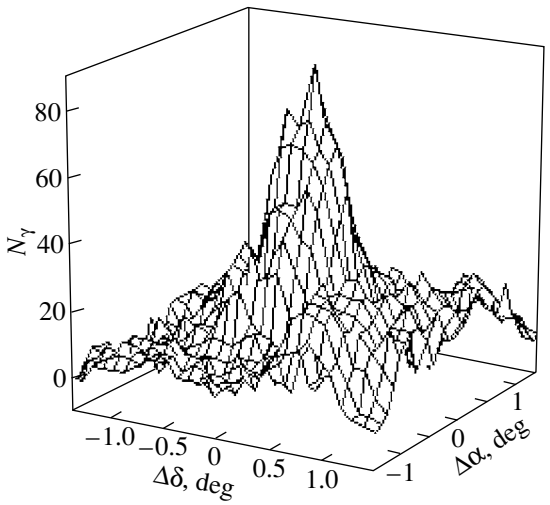

Fig. 2. Stereo imaging "maps" of the distribution of arrival directions of the gamma-rays. $\Delta\alpha$ is the deviation from the source position in right ascension and $\Delta\delta$ the corresponding deviation in declination (in degrees). The center of the map coincides with the coordinates of H 1426+428 (the gamma-ray source). N_γ is the number of gamma-rays detected.

separated by a distance of 20 m. Each section is equipped with four chambers (light receivers) containing 37 photomultipliers, which together with their conical light guides form 37 cells (channels) [10]. The Cerenkov flashes are recorded at visible wavelengths. The field of view of each cell is 0.4°, while the field of view of each chamber is 2.6°.

Each multielement camera is located in the focal plane of four parabolic mirrors 1.2 m in diameters and records an image of the Cerenkov flash. The signals from the cells of the four cameras in each section are summed channel by channel. Only flashes that coincide in both sections and exceed a specified threshold in any two of the 37 cells in each section are recorded.

The total surface area of the mirrors on both sections of the telescope is 36 m^2. The sections can be pointed with accuracy to within $\pm 0.05°$ using a guiding system. The effective energy threshold for the detection of gamma-rays is 1 TeV. A more detailed description of the facility can be found in [11].

3. OBSERVATIONS AND ANALYSIS

The blazar H 1426+428 (epoch 2004 coordinates: $\alpha = 14^h28^m39^s$ and $\delta = 42°39'42''$) was observed at the Crimean Astrophysical Observatory from April 15 to April 25, 2004.

Observations were carried out in a tracking (ON) regime with a duration of 45 min, followed by a recording of the background at a position shifted by 50^m in α with the same duration (the OFF regime). The observations in these two modes were carried out at the same zenith distance. During the observations in the ON mode, the source was located at the center of the field of view of the recording cameras. In all, 12 pairs of complete sessions were carried out. We excluded sessions carried out during distorted weather from our analysis. After excluding three pairs of sessions due to their large dispersions of the count rates during observations of the source or background, there remained nine pairs of observing sessions with total exposure times for the source and background of 405 min each.

These data were subject to a preliminary processing that excluded events observed with distorted guiding (deviation of the telescope direction from the specified coordinates by more than 0.05°), introduced a calibration coefficients, excluded flashes whose maximum amplitude was at the outer ring of the receivers, and excluded events in which the charge–digital transformer in at least one channel was saturated. After this preliminary data selection, there remained 5569 Cerenkov flashes during observations of the source and 5373 during observations of the background.

The detection of the Cerenkov radiation from EAS generated by VHE gamma-rays (gamma-ray showers) occurs against the dominant background of EAS generated by cosmic rays, primarily protons and nuclei of light elements (p-showers). The angular dimensions of the two types of showers, and accordingly of the Cerenkov flashes they generate, can easily be distinguished due to the properties they acquire when they are generated in the atmosphere. Flashes from gamma-ray showers have smaller angular sizes and more compact images that are extended in the direction toward the source, while flashes associated with p-showers have relatively large sizes and fragmented images without any well-defined direction. Consequently, gamma-ray showers can be distinguished from p-showers based on their angular dimensions and the shape and orientation of flash im-

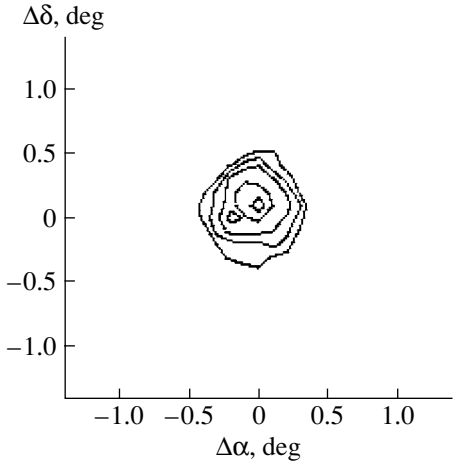

Fig. 3. Contours of the arrival directions of the gamma-rays. The outer contour corresponds to 40 events and the step between contour levels is 10 events.

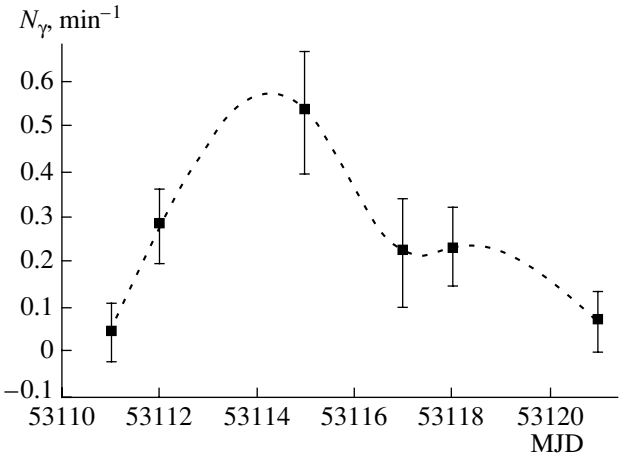

Fig. 4. Time run of the flux of gamma-rays with energies $\geq$1 TeV (mean nightly values). The statistical errors are shown.

ages. Since the parameters of low-energy flashes are determined with large uncertainties, restrictions on the amplitude of the detected signal are also imposed. After preliminary processing of the data, the sizes of the flash images along their major axes (A) and the perpendicular direction (B) axes were calculated. In addition to the dimensions of the flash images, we also used a parameter specifying their ellipticity, $E = 1 - B/A$. The selection of the flash images according to their shape was carried out using the Imaging Pattern Radio (IPR), which characterizes their degree of fragmentation. This parameter is assigned numerical values from zero to seven. A simple (ordered) flash corresponds to a value of zero, while increasingly less ordered flashes with growing degrees of fragmentation are assigned values from one to seven.

Parameters characterizing the energy (amplitude) of the flash and the dimensions, ellipticity, and shape of the images that do not depend on the position of the flash in the field of view of the camera are said to be "coordinate independent." The orientation of the flash images relative to the direction toward the source of the gamma-rays was derived from the deviation of the major axis of the flash image from the position of the source in the field of view (the angle ALPHA). The angular distance of the center of gravity of the flash image to the position of the source in the field of view was specified by the parameter DIST [12]. The parameters ALPHA and DIST depend on the position of the image in the field of view of the chamber, and are accordingly said to be "coordinate dependent."

The selection parameters are depicted schematically in Fig. 1.

The ranges of variation of the selection parameters were specified for each section of the telescope separately. Events having selection parameters that did not fall into the specified ranges were excluded from further consideration. The limiting values of the selection parameters were chosen such that the magnitude of the quantity $Q = (N_s - N_b)/\sqrt{N_s + N_b}$ was maximum, N_s and N_b denote the numbers of events for the observations of the source and background, respectively. The numerator is the number of gamma-ray events N_γ, while the denominator represents the error in this number.

The results of our selection according to the parameters indicated above are presented in the table. We also carried out a selection based on the coordinate dependent parameters ALPHA and DIST. The mean gamma-ray count rate over the observing interval was 0.215 ± 0.037 min^{-1}. A three-dimensional histogram of the number of selected gamma-rays over the field of view of the receiver constructed using the method of trial sources [13, 14] is shown in Fig. 2, and a contour "map" of this distribution in Fig. 3.

The maximum value of $N_\gamma = 87 \pm 15$ (the $5.8\,\sigma$ confidence level) is present at the coordinates $\Delta\alpha = 0.0°$, $\Delta\delta = 0.1°$ (Fig. 3), and coincides with the coordinates of H 1426+428 within the uncertainties in the coordinates ($\approx$0.1°).

Selection results

Selection method	N_s	N_b	$N_s - N_b$	Q
With no selection	5564	5374	190	1.82
Based on coordinate independent parameters	300	211	89	3.94
Based on coordinate dependent parameters	156	69	87	5.80

4. VARIATIONS IN THE EMISSION

Figure 4 presents a light curve of the mean nightly fluxes at energies $\geq$1 TeV for the blazar H 1426+428 over the observing interval. The left part of the light curve shows a smooth growth in the intensity of the TeV emission from the object.

A gamma-ray flux at a level of 3.9σ was detected on the third observing night, April 19 (MJD 53115).

The intensity of the gamma-ray emission fell on April 22, reaching a minimum on the following observing night, April 25. Based on the observed fluxes on April 16 (MJD 53112), April 19 (MJD 53115), and April 21 (MJD 53117), it appears that the maximum VHE gamma-ray flux occurred sometime in the interval MJD 53112–53117. We conclude from our analysis of this light curve that the gamma-ray flare lasted no more than seven days.

5. CONCLUSION

We have identified a source of gamma-rays with energies exceeding 1 TeV with the blazar H 1426+428. A flux of gamma-rays from this source was detected at the 5.8σ confidence level. A detected flare indicates the presence of variability in the gamma-ray emission of this object on time scales of a week.

ACKNOWLEDGMENTS

We thank A.A. Stepanyan for his useful comments during discussions of this work.

REFERENCES

1. G. Ghisellini, A. Celotti, G. Fosati, *et al.*, Mon. Not. R. Astron. Soc. **301**, 451 (1998).
2. D. J. Thompson, D. L. Bertsh, B. L. Dingus, *et al.*, Astrophys. J., Suppl. Ser. **101**, 259 (1995).
3. J. A. Gaidos, C. W. Akerlof, S. Biller, *et al.*, Nature **383**, 319 (1996).
4. F. Piron, A. Djannati-Ataï, M. Punch, *et al.*, Astron. Astrophys. **374**, 895 (2001).
5. K. E. Turver, Nucl. Phys. B **22**, 86 (1991).
6. M. Punch, C. W. Akerlof, M. F. Cawley, *et al.*, Nature **358**, 477 (1992).
7. J. Quinn, C. W. Akerlof, S. Biller, *et al.*, Astrophys. J. Lett. **456**, L83 (1996).
8. F. W. Stecker and O. C. De Jager, Astrophys. J. **476**, 712 (1997).
9. F. Aharonian, A. Akhperjanian, M. Beilicke, *et al.*, Astron. Astrophys. **421**, 529 (2004).
10. N. A. Andreeva, Yu. L. Zyskin, O. R. Kalekin, *et al.*, Pis'ma Astron. Zh. **26**, 243 (2000) [Astron. Lett. **26**, 199 (2000)].
11. B. M. Vladimirskii, Yu. L. Zyskin, A. P. Kornienko, *et al.*, Izv. Krymsk. Astrofiz. Obs. **91**, 74 (1994).
12. A. M. Hillas, in *Proc. 19th Int. Cosmic Ray Conf.* (1985), Vol. 3, p. 445.
13. C. W. Akerlof, M. F. Cawley, M. Chantell, *et al.*, Astrophys. J. **377**, L97 (1991).
14. Yu. I. Neshpor, A. P. Kornienko, A. A. Stepanian, and Yu. L. Zyskin, Exp. Astron. **5**, 405 (1994).

Translated by D. Gabuzda

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 863–870. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 967–975.

The Chemical Evolution of Globular Clusters

M. V. Kas'yanova and Yu. A. Shchekinov

Rostov State University, Rostov-on-Don, Russia

Received July 24, 2004; in final form, May 18, 2005

Abstract—The sequence of events determining the initial stages of star formation is analyzed in framework of the self-enrichment scenario. The computations are based on a single-zone chemo-dynamical model. It is shown that the first episode of star formation was characterized by an initial mass function shifted toward massive stars ($M \geq 8\,M_\odot$). We argue that the transition to a star formation with a normal (Salpeter) initial mass function was due to more efficient radiative cooling of the proto-globular cluster gas after its enrichment to a metallicity of $Z \sim 0.02\,Z_\odot$ in agreement with those observed in globular clusters.

1. INTRODUCTION

The chemical evolution of globular clusters is by no means a trivial problem. Indeed, the binding energy of a globular cluster is $E_G \sim GM^2/R \sim 10^{51}-10^{52}$ erg for $M \sim 3\times10^5-10^6\,M_\odot$ and $R \sim 10$ pc, and the explosion of one or several supernovae could be sufficient to completely destroy the gaseous cloud that could have produced the globular cluster [1]. This leads us to conclude that globular clusters formed from gas that was already enriched in heavy elements, implying that the formation of globular clusters must have been preceded by a preliminary stage of stellar nucleosynthesis in the Galaxy [2]. There is, however, some evidence of a mass–metallicity correlation for the old globular clusters [3], which clearly indicates that the enrichment is connected with the process of nucleosynthesis in clusters themselves—the so-called "self-enrichment" hypothesis.

From this viewpoint, the above estimate of the binding energy of a proto-globular cluster may indicate that the initial mass of the proto-globular-cluster clouds exceeded the observed masses of globular clusters by two to three orders of magnitude and that the star-formation efficiency (SFE) was about $0.001-0.03$ [4–6], so that the observed globular clusters contain only a small fraction of the initial mass. If the initial gaseous cloud had a shallow density distribution, most of its mass should have been lost from the periphery of the cloud without this loss affecting the gravitational binding of the central part, where active star formation took place. On the other hand, it cannot be ruled out that, if the proto-globular cluster has a sufficiently steep density distribution and supernova explosions occur mostly outside the central core, the remnant of an individual explosion might rapidly escape to infinity towards the sharp density decrease. In this case, only a small fraction of the supernova energy would contribute to the (thermal or mechanical) heating of the bulk of the mass of gas [6].

It is obvious that these scenarios differ from each other in their chemical evolution and the degree of homogeneity of the distribution of chemical elements. A complete analysis of these features clearly must be based on numerical modeling of a joint set of dynamical equations for the stellar and gaseous components and the equations of chemical evolution. However, even much simpler single-zone models for the chemical evolution of globular clusters (or models that can be reduced to single-zone models) yield fairly positive qualitative conclusions [6–8].

In this paper, we use a single-zone chemical–dynamical model to qualitatively describe the initial sequence of events determining the chemical evolution of globular clusters. Our model is based on the scheme suggested by Firmani and Tutukov [9], Shustov *et al.* [10], and Wiebe *et al.* [11] for modeling the evolution of spiral galaxies. The next section presents arguments supporting bimodal star formation and discusses its possible origins. Section 3 describes a simple single-zone chemo-dynamical model and the results of numerical simulations of chemical evolution of globular clusters and Section 4 gives our conclusions.

2. BIMODAL STAR FORMATION

2.1. Star-Formation Rate

The hypothesis that the chemical composition of globular clusters is determined by self enrichment naturally leads to the idea that globular clusters begin their evolution from gaseous clouds, i.e., proto-globular clusters. The star-formation rate (SFR) in

1063-7729/05/4911-0863$26.00

the initial state can be estimated using the quadratic Schmidt law usually assumed for galaxies—$\psi = f\rho^2 V$, where ρ and V are the gas density and volume of the system [10]. The star-formation efficiency f is adopted for spiral galaxies to be equal to $f = 2 \times 10^7$ cm^3 g^{-1} s^{-1}, implying that the total present-day SFR for our Galaxy is $\psi \sim 16\, M_\odot$ yr^{-1}, which is a factor of three higher than the observed value. The SFE f in the Schmidt law is determined by specific mechanisms that lead to a quadratic dependence of SFR on density, $\psi \propto \rho^2$, and to the assumption that this efficiency is, in general, a separate problem.

Cox [12] assumed that the star formation in our Galaxy is governed by the ionization of gas by ultraviolet photons produced by massive stars. He inferred a SFE of $f = 5 \times 10^7 q^{-1}$ cm^3 g^{-1} s^{-1}, where $q \ll 1$ is a factor characterizing the nonuniform distribution of gas. Assuming $q \leq 0.1$ for the cloud component of the interstellar medium, we obtain $f \geq 5 \times 10^8$ cm^3 g^{-1} s^{-1}, which implies a too high SFR for our Galaxy: $\psi \geq 100\, M_\odot$ yr^{-1} [12]. The actual SFR is determined not only by the ability of the gas to recombine, but also by mechanisms that can lead to its fragmentation: gravitational and thermal instability. Therefore Cox's [12] value for f can be viewed only as an upper limit. The value of $f = 2 \times 10^7$ cm^3 g^{-1} s^{-1} adopted by Shustov *et al.* [10] and Wiebe *et al.* [11] seems to be a good approximation for the SFE in the Schmidt law irrespective of the mechanisms that are responsible for this law. Note that the corresponding SFR in our Galaxy is close to the estimate $\psi \sim \epsilon M_g / t_J$, where $\epsilon = 0.05$ is the fraction of gas converted into stars in the process of gravitational instability, M_g the mass of the gaseous component of the Galaxy, and t_J the Jeans time scale. For $\rho \sim 10^{-21}$ g cm^{-3} and a proto-globular cluster size of 20 pc, the Schmidt law yields an SFR of $\psi = 1.28\, M_\odot$ yr^{-1}.

2.2. Evolution of the Initial Mass Function

After $t_{SN} \sim 3$ Myr, the first, most massive supernovae (SN II) begin to explode and enrich matter that has not yet been converted into stars. However, the number of stars formed by this time in the case of a Salpeter or a close-to-Salpeter initial mass function (IMF) is $\sim 3 \times 10^5 - 10^6$. In other words, under these assumptions, a considerable if not overwhelming fraction of stars in globular clusters should have metallicities equal to their initial metallicity. The situation appears even more serious if we take into account the late enrichment of matter of the proto-globular clusters due to type Ia supernovae, with time scales of $t_{SNIa} \gg t_{SN}$. In this connection, Zhang and Ma [13] concluded that the chemical evolution of globular clusters cannot be described in terms of a single episode of star formation.

Thus, the assumption that the measured metallicities of globular clusters represent the bulk of the cluster stars inevitably implies that the formation of low-mass stars lags behind the formation of massive stars, which are responsible for enriching the globular clusters in metals. In other words, the star-formation process must depend on the masses of stars, with low-mass stars being born only after the gas of the cluster has been enriched in metals. Parmentier *et al.* [7] and Jehin *et al.* [14, 15] have reached a similar conclusion expressed in a somewhat different form: they assumed that the initial stage of star formation in globular clusters involves type II supernovae, which can later trigger the birth of low-mass stars. In this paper, we report the main results of our studies of a star-formation model in which the IMF in the initial stage of star formation is shifted toward higher masses. More specifically, the first stage of star formation is characterized by a Salpeter IMF with $M_{min} = 8\, M_\odot$, whereas the distribution of stellar masses during the subsequent stages, starting from a time $t_{SN} + \Delta t$, is described by a Salpeter IMF with $M_{min} = 0.1\, M_\odot$. This assumption is quite compatible with modern views that stars forming from primordial matter are, as a whole, more massive than present-day stars [16]. In this case, to prevent an overproduction of heavy elements, we set the SFE in the initial stage equal to $f_1 = 2.6 \times 10^5$ cm^3 g^{-1} s^{-1}, more than two orders of magnitude lower than the value adopted for our Galaxy. Note that we did not link f_1 to the production of ionizing photons by massive stars, but treat this as a parameter of the model, which determines the total mass of metals produced in the initial stages of star formation. The corresponding SFR in the initial stages is $\psi_1 = 0.008\, M_\odot$ yr^{-1}.

Our assumption about the evolution of the IMF corresponds to a sequence of events resembling that predicted by the concept of star formation triggered by the first supernovae [7, 14, 15]: the initial episode of the formation of massive stars is converted to a normal star formation covering the entire interval of stellar masses. However, the underlying physical mechanisms for this change in the star formation regime in our model differs qualitatively from those supposed in the models of Parmentier [7] and Jehin *et al.* [14, 15]. Unlike these authors, we associate the transformation of the IMF with a change in thermal regime of the gas of the proto-globular cluster due to heavy-element enrichment of the cluster, so that the parameter Δt is set equal to the gas-cooling time scale (see Section 2.4). Moreover, we report below estimates showing that the idea of triggered star formation in globular cluster is, in certain ways, misleading.

In order for the evolution of the IMF to be internally consistent, the mixing of the metal-rich matter ejected by supernovae must be sufficiently rapid, or, more specifically, it must occur on time scales much shorter that the time scale for the radiative cooling of the ejected envelope, $t_{sp} \sim 4 \times 10^3$ yr, and the time scale for the formation of low-mass stars from the enriched matter. It is obvious that Δt must be greater than the mixing time scale.

As far as mixing in an individual supernova remnant is concerned, it is quite slow there—the mixing time scale is of the order of the dynamical age of the remnant [17] (see below). Therefore, the change in the behavior of the star formation in a globular cluster, i.e., the evolution of the IMF toward its current form, could be due to multiple supernovae: the mixing in the resulting remnants is much more efficient, since the remnants of different supernovae intersect, providing strong turbulent motions that lead to faster mixing.

2.3. Mixing of Metals

The hypothesis that the IMF is shifted toward massive stars in the initial stage of star formation implies a type-II supernova rate of the order of $\nu_{SN} \sim 10^{-3}$ yr^{-1} for an initial SFR of $\psi \sim 0.01\, M_\odot$ yr^{-1}. When the shock velocities of the supernova remnants reach $v_0 \simeq 4 \times 10^7$ cm s^{-1} (which corresponds to a temperature of $T_0 \simeq 10^6$ K behind the front), the radiative expansion phase starts, which is governed by the expansion law $R \propto t^{1/4}$—the so-called Oort's snow-plow phase (note that the exact expansion law $R \propto t^{0.286}$ [18] differs insignificantly from the Oort law that we use in this paper). The volume filling factor of the hot supernova remnants reaches unity if

$$R^3 \nu_{SN} t \sim R_c^3, \quad (1)$$

where $R = R(t)$ is the characteristic radius of the supernova remnants, t is their age, and R_c is the radius of the proto-globular cluster. The total number of supernovae that explode during this time interval is

$$N_{SN} = \nu_{SN} t_1 \simeq \left(\frac{5}{2}\frac{\rho_0}{E_0} v_0 \nu_{SN}\right)^{3/7} R_c^{12/7}, \quad (2)$$

where

$$t_1 = \left(\frac{5\rho_0 v_0}{2E_0}\right)^{3/7} \left(\frac{R_c^3}{\nu_{SN}}\right)^{4/7} \sim 2 \times 10^4 \text{ yr} \quad (3)$$

is determined by (1). For $\rho = 10^3 m_H$ g cm^{-1}, $E_0 = 10^{51}$ erg, and $R_c = 10$ pc, this relation yields $N_{SN} \simeq 30$. This number of supernovae is clearly insufficient to enrich the matter of the globular cluster to the observed metallicity of $Z \sim 0.01$–$0.03\, Z_\odot$. At the same time, it is clear that the system undergoes sufficiently efficient heavy-element mixing over a time interval of the order of several t_1, since several supernova shocks pass through each particular gas element during this time, providing sufficient mixing [19].

The change in the IMF, i.e., the transition of the star formation process to a new regime that enables the formation of low-mass stars, appears to result from a fundamental change in the thermal regime of the proto-globular cluster's gas. This change, in turn, is due to the ability of the already metal-enriched gas to cool to low temperatures ($T \lesssim 10^2$ K). The transition to this regime begins when $t \gg t_1$ (see below), so that the matter of the proto-globular cluster can be expected to be well mixed by the time formation of low-mass stars becomes possible. This may explain the high degree of homogeneity of the chemical compositions of globular clusters [7].

Note, in this connection, that this homogeneity of the chemical composition indicates that the effects of triggered star formation did not play an important role in globular clusters. Indeed, during triggered star formation, the passage of a supernova-produced shock through the molecular cloud is accompanied by the fragmentation of the shock-compressed matter, and the fragments formed in different layers behind the shock front have different chemical compositions, if the efficiency of mixing in the individual supernova remnants is low [17].

The inefficiency of metal mixing during the propagation of the ejected supernova shell through the interstellar medium becomes obvious from the following simple arguments. Mixing under these conditions can be provided primarily by the Rayleigh–Taylor instability [20]. The Rayleigh–Taylor instability increment of the shock-compressed gas layer is always positive only in a medium with rapidly decreasing density, where the shock front moves with acceleration [21]. However, this is possible only if the density decreases more rapidly than $\rho \propto r^{-3}$. Within a limited interval of shock Mach numbers and wavelengths of nonradial front perturbations, the instability increment can also be positive for decelerated fronts, such, as, for example, the radiative front in a constant-density medium [22].

However, in all these cases, the development of the instability is slow. Indeed, the Rayleigh–Taylor instability increment is of the order of

$$\gamma_{RT} \sim \sqrt{k|a|}, \quad (4)$$

where $a = \ddot{R}$ is the acceleration of the front and k the wavenumber of the cross-front perturbation. In the case of decelerated fronts, unstable perturbations are with wavenumbers $k < 1/6\Delta R \sim 10/R$, where ΔR is the thickness of the compressed gas layer

behind the front and R the radius of the front [22, 23]. Since $R \propto t^{1/4}$ during the radiation stage of the shock, we have $\gamma_{\rm RT} t < \sqrt{2}$. It follows that, even under the most favorable conditions, the perturbation amplitude cannot increase by more than $e^{\sqrt{2}}$ over the dynamical time scale of the supernova remnant.

Moreover, the very possibility of star formation triggered by supernova-induced shocks cannot be considered to be indubitable from the dynamical viewpoint. Indeed, the gravitational instability of a shock-compressed layer usually develops more slowly than its dynamical instability, except when the unperturbed ambient gas is gravitationally unstable on length scales smaller than the size of the supernova remnant [23]. The compressed layer can fragment under the action of thermal instability during the transition of the expansion from the Sedov phase to the radiative phase. However, the sizes of the fragments in this case are much smaller than the thickness of the shell [23], so that these fragments have different chemical compositions in a chemically nonuniform layer.

2.4. Change of the Thermal Regime of Gas

During the stages when the volume-filling factor of the supernova remnants reaches unity, the characteristic velocities of the shocks from supernovae can easily be shown to be a few times the virial velocity for the globular cluster:

$$v_1 \simeq 0.15 \left(\frac{E_0}{\rho_0 v_0}\right)^{4/7} \nu_{\rm SN}^{3/7} R_{\rm c}^{-9/7} \simeq 2 \times 10^6 \ {\rm cm/s}. \quad (5)$$

The temperature behind the front will settle at $T \sim 10^4$ K, since the radiative cooling of the gas behind the shock front in a hydrogen and helium medium is determined by recombination radiation and the Lyα radiation of hydrogen atoms. Further cooling of the shock-compressed matter can be only due to heavy elements. Cooling on hydrogen molecules under the conditions considered is less efficient due to the dissociation of H_2 and H^- ions by stellar ultraviolet radiation. For the above parameters, the time scale for the cooling of the matter behind the shock front is equal to

$$t_{\rm c} \sim \frac{kT}{\Lambda n x} \sim \frac{3 \times 10^{12}}{Z n x} \ {\rm s}, \quad (6)$$

where Λ is the cooling function, which we set equal to $\Lambda \sim 5 \times 10^{-25} Z$ erg cm^3 s^{-1} at $T \sim 10^4$ K; Z the relative abundance of heavy elements in units of the solar metallicity; n the number density of the gas behind the front; and x the degree of ionization of the gas. This value of Λ corresponds to cooling in lines of the fine structure of atomic and singly ionized carbon at $T \sim 10^4$ K [24, 25]. In the absence of external sources of gas ionization, such as cosmic rays or soft X-ray radiation, the degree of ionization is determined by freezing of recombination at $T \sim 10^4$ K; i.e., by the condition $t_r \sim 10^{13}/nx \sim t$, where t is the dynamical age of the supernova remnant. We adopt $t \sim 3t_1$ to find $x \sim 5/n$, which yields a cooling time of $t_{\rm c} \sim 6 \times 10^{11} Z^{-1}$ s.

Strictly speaking, the chemical evolution of globular clusters depends on the yield of heavy elements in supernovae. Supernova model A in the classification of Woosley and Weaver [26] is commonly used in chemical-evolution models. In this model, stars with the primordial composition exploding as supernovae eject no heavy elements into the ambient medium if their masses exceed $30\,M_\odot$. Thus, in model A [26], the first supernovae that are capable of enriching the metallicities of globular cluster are stars with masses $M < 30\,M_\odot$, which explode beginning from time $t = t_{\rm SN} = 3$ Myr. The total number of supernovae that have exploded in a proto-globular cluster during the cooling time $t_{\rm c}$ after this time is equal to $N_{\rm SN} = t_{\rm c}\nu_{\rm SN} \sim 4\,Z^{-1}$. Here, we have specified the supernova rate at time $t = t_{\rm SN}$ to be equal to the estimate $\nu_{\rm SN} \simeq 2 \times 10^{-4}$ yr^{-1} implied by the numerical model with a SFE of $f = 2.6 \times 10^5$ cm^3 g^{-1} s^{-1}.

The mass of metals ejected during a supernovae depends on the mass of the supernova. In model A of Woosley and Weaver [26], this mass is small for primordial-composition supernovae—the mean value averaged over the pre-supernova IMF in the mass interval $M = 15$–$30\,M_\odot$ is only $\langle \Delta M_Z \rangle = 0.9\,M_\odot$. The implied metal abundance in the gas of the proto-globular cluster by the time $t + t_{\rm c}$ is $Z = M_Z/MZ_\odot = N_{\rm SN}\langle \Delta M_Z \rangle / MZ_\odot \sim 2 \times 10^{-4}\,Z^{-1} \sim 0.014$, which is close to the metal abundance typically observed in globular clusters. Note that this result depends only slightly on the adopted model for the ejection of mass by supernovae. In model B of Woosley and Weaver [26], the mass of ejected metals is appreciably higher over the entire interval of pre-supernova masses, including for the primordial-composition pre-supernovae: $\Delta M_Z = 0.3\,M - 3.5$, where M is the mass of the pre-supernovae [7]. This model yields the value $\langle \Delta M_Z \rangle = 7\,M_\odot$ in the interval of pre-supernova masses $M = 20$–$100\,M_\odot$, almost an order of magnitude higher. At the same time, in this case, the initial contribution to the metal enrichment of globular cluster is provided by more massive stars (we restricted our computations to model B pre-supernova masses $M \leq 45\,M_\odot$), so that the supernova rate at the onset of enrichment is equal to 4×10^{-5} yr^{-1}, and the corresponding metallicity of

the gas a time t_c after the onset of the supernovae is equal to $Z \sim 0.017$ in solar units.

This leads us to conclude that the transition of the star formation to a regime with a Salpeter IMF may, indeed, be associated with the time when the shock-heated gas is cooled by heavy elements to lower temperatures, such that the Jeans mass decreases compared to the mass in the gas with the primordial composition. Note that the dependence of the resulting metallicity estimate on the mass of the proto-globular cluster is, to a large degree, determined by the star-formation law. Indeed, in the case of a power-law/star-formation law, the supernova rate is equal to $\nu_{SN} \propto M^n V^{-n+1}$. We now assume that the gaseous cloud of the proto-globular cluster is in a state of virial equilibrium at a temperature $T \sim 10^4$ K. Since $M \propto \rho^{-1/2}$ and $V \propto \rho^{-3/2}$ in equilibrium, we obtain in this case $Z \propto \nu_{SN} t_c / M \propto M^{-2n+2}/Z$, or $Z \propto M^{-n+1}$. Since $n = 2$ for the Schmidt law, the above relation yields $Z \propto M^{-1}$. We note in this connection that the observed globular-cluster metallicities are bounded from above by just such a dependence [3]. This may indicate that star formation in globular clusters obeyed a Schmidt law during the early stages of the evolution of these objects.

2.5. Late Enrichment

Type Ia pre-supernovae have long lifetimes (1–3 Gyr), and therefore can have a significant effect on the dynamics and heavy-element enrichment of the matter of proto-globular clusters only if a considerable fraction of the cluster stars formed during late stages, when the gaseous matter could have been enriched by the ejecta of type-Ia supernovae. Star formation in globular cluster is unlikely to have followed this scenario, since all the dynamical time scales in the gaseous proto-globular cluster are much shorter—$t_d \sim t_J \sim 1$ Myr. We would apparently have to assume that the presence of iron-peak elements in globular cluster stars is due to the accretion of enriched matter carried to the surfaces of stars during type-Ia supernovae [7]. This process is, in principle, possible if the ejecta of type-Ia supernovae are confined within the globular cluster and if the velocities (and the corresponding temperatures) of these ejecta are sufficiently low for accretion to be efficient. At the same time, simple estimates of the accretion rate onto the surface of a star with a nearly solar mass yield $\dot{M} \sim 10^{-14} n c_{s10}^{-3} M_\odot$ yr^{-1}, where c_{s10} is the sound speed of the accreting gas in units of 10 km s^{-1} (see, e.g., [27]). This accretion rate can be comparable to the mass-loss rate for solar-type stars, if $n \gtrsim 1$ cm^{-3} and $c_s \lesssim 10$ km s^{-1}. The same conclusion can be reached for s-element enrichment determined by stars on the asymptotic giant branch. This means that globular clusters must have contained fairly dense gas ($n \geq 1$ cm^{-3}) during late stages of their evolution, at $t \geq 100$ Myr.

3. A SIMPLE SINGLE-ZONE CHEMO-DYNAMICAL MODEL

3.1. Basic Equations

We describe here the chemo-dynamical evolution of globular clusters using equations similar to those presented by Firmani and Tutukov [9], Shustov *et al.* [10], and Wiebe *et al.* [11] to describe the evolution of disk galaxies. Our approach differs from that of Harris and Pudritz [5], who used equations describing the chemo-dynamical evolution of elliptical galaxies with well developed winds. The regime acting in disk galaxies appears to be more economical from the viewpoint of the chemical evolution of globular clusters during the initial stages of their lifetimes. The gaseous component of the proto-globular clusters can be described by the mass-conservation condition

$$\frac{dM_g}{dt} = -\Psi(t) + \int_{M_{min}}^{M_{max}} \Psi(t-\tau_M)(M - M_r)\phi(M)dM - \dot{M}_g^{ej}, \quad (7)$$

the energy-conservation equation [9]

$$\frac{dr}{dt} = \frac{1}{4\pi}\frac{M_g}{GM_0}\frac{\gamma f \epsilon R^3}{r^4} - \frac{1}{6}\frac{r}{\tau_d}, \quad (8)$$

and an equation describing the variation of the mass of chemical element Z_i

$$\frac{dZ_i M_g}{dt} = -Z_i\Psi(t) + \int_{M_{min}}^{M_{max}} \Psi(t-\tau_M)[Z_i(t-\tau_M)(M - M_r) + P_i(M)]\phi(M)dM - \dot{M}_i^{ej} + P_i^{SNIa}, \quad (9)$$

where $\log \tau_M = 9.9 - 3.8 \log M + \log^2 M$ is the main-sequence lifetime of a star (in years), $\Psi(t)$ the SFR, $\phi(M)$ the IMF, M_{min} and M_{max} the limiting masses of the IMF, $\dot{M}_g^{ej}$ the mass of gas lost by the globular cluster, M_0 the total mass of the globular cluster, R the radius of the stellar component, and r the radius of the gaseous component.

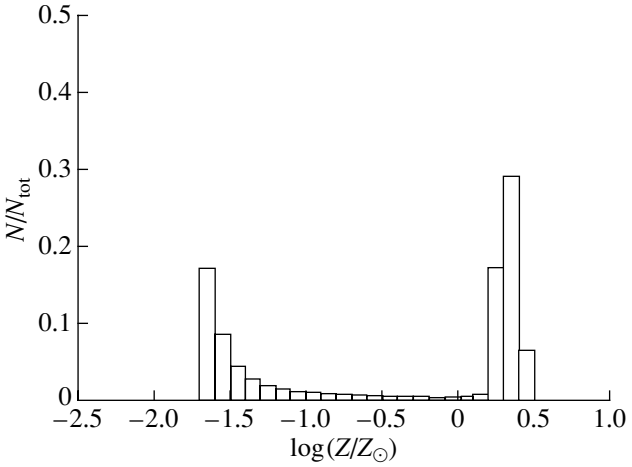

Fig. 1. Distribution of stellar metallicities in a model with a bimodal star-formation law with an initial episode with a Salpeter IMF and $M_{\rm min} = 8\,M_\odot$, followed by a change of the IMF to the standard law with $M_{\rm min} = 0.1\,M_\odot$, $M_{\rm max} = 100\,M_\odot$ for $\Delta t = t_{\rm c}$, $f_1 = 2.6 \times 10^5$ cm^3 g^{-1} s^{-1}, and $f_2 = 2 \times 10^7$ cm^3 g^{-1} s^{-1}.

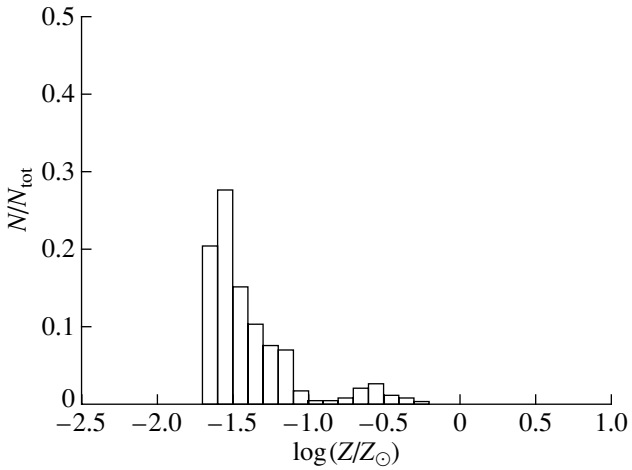

Fig. 2. Same as Fig. 1, but with the gaseous mass ejected at $t = 2t_{\rm SN}$.

3.2. Dissipation Time

One of the important parameters of the dynamical equation (8) is the dissipation time scale τ_d, which describes the loss rate of the supernova-shock energy [9–11]. For a shock propagating through a uniform medium, the dissipation time scale is determined by the time scale for radiative energy losses. We now assume that most of the thermal energy behind the shock front is radiated during the radiative stage. In this case, the dissipation time scale can be estimated as

$$\tau_d \sim \frac{E}{2\pi R^2 \rho u^3}, \qquad (10)$$

where R and u are the typical size and velocity of a supernova remnant. By the time when the volume filling factor reaches unity, these quantities are equal to (see above) $R \sim 3$ pc and $u \sim 20$ km s^{-1}, implying a dissipation time scale of $\tau_d \sim 3 \times 10^6$ yr in the gas of a proto-globular cluster with $n \sim 10^3$ cm^{-3}. If the remnants of individual supernovae overlap, their shocks collide inelastically, and the time scale for the dissipation of their energy can be estimated as the ratio of the size per an individual supernovae ($R \sim 3$ pc) to the characteristic velocity at the time of intersection ($u \sim 20$ km s^{-1}), or $\tau_d \sim 10^5$ yr. This means that the destructive efficiency of the supernovae can be appreciably lower than the values calculated based on the dynamics of individual remnants in a homogeneous medium. We adopt here for definiteness the value $\tau_d = 10^6$ yr.

3.3. Metallicity Function

We now assume that the IMF $\phi(M)$ is given by the Salpeter relation. This inevitably yields a metallicity distribution with an overwhelming majority of stars with zero metallicity. This is due to the fact that, according to this description, all stars—including low-mass stars—are born simultaneously and have the same (primordial) chemical composition; the gas is appreciably depleted by the time the metals are ejected by the first supernovae ($t \sim 3$ Myr), so that the total mass of stars with nonzero metallicities should be substantially smaller than the mass (and number) of stars with the primordial composition.

To avoid this conflict, we assume, based on the above arguments, that the first star-formation episode, which lasts $t \sim t_{\rm SN} + t_{\rm c}$, is characterized by a "top-heavy" IMF with $M_{\rm min} \geq 8\,M_\odot$. As we pointed out above, this is consistent with modern concepts suggesting that the stars born from primordial matter were massive. The birth of low-mass stars becomes possible only after the time interval necessary for the gaseous medium to be enriched in metals to $Z \sim 0.01$–0.02, when the thermal regime of the gas changes. For specificity, we assume that the lower boundary of the spectrum of stellar masses becomes equal to $0.1\,M_\odot$ after $t = t_{\rm SN} + t_{\rm c}$. Figure 1 shows the distribution of stellar metallicities in a model with such a bimodal IMF and with an SFE of $f_1 = 2.6 \times 10^5$ and $f_2 = 2 \times 10^7$ cm^3 g^{-1} s^{-1} during the first and subsequent stages. We can easily see that about half of all the stars are concentrated near a metallicity of $Z \sim 0.02$. At the same time, it is immediately obvious that an appreciable fraction of stars have metallicities comparable to or even higher than the solar value—this corresponds to the second

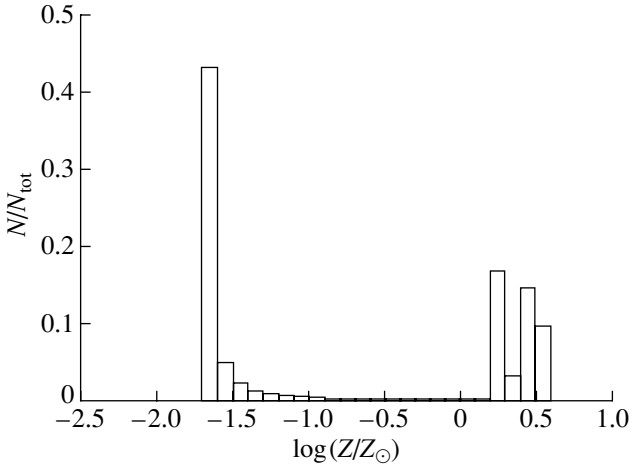

Fig. 3. Same as Fig. 1, but with the SFE in the second stage equal to $f_2 = 2 \times 10^8$ cm^3 g^{-1} s^{-1}.

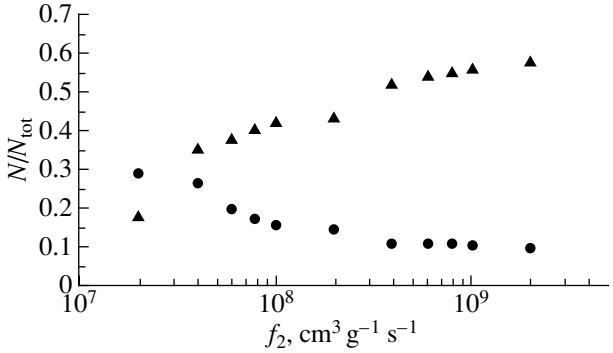

Fig. 4. Dependence of the metallicity of the first- and second-generation stars on the SFE f_2. The triangles and circles show the dependences for stars of the first and second peaks at $Z \simeq 0.02\,Z_\odot$ and $Z \gtrsim Z_\odot$, respectively.

peak in $N(M)$ distribution. This is due to the fact that, under the above assumptions, the mass M_g of the gas remaining in the proto-globular cluster by a certain time becomes too low, $M_g < M_Z/Z_\odot$, where M_Z is the total mass of metals ejected by type-II supernovae.

Such an overproduction of metals can be avoided if we allow the outer envelope and the overwhelming fraction of the mass in the region of active star formation to be ejected, so that the products of supernovae can freely escape from the globular cluster. This time obviously corresponds to the end of the formation of the globular cluster. The realization of this possibility may, in turn, be a natural result of the transition of the star formation to a regime with a Salpeter IMF after an initial stage of enrichment in metals. Indeed, assuming that the star formation in this regime involves fairly large masses and results in the birth of stars with a total mass comparable to the observed masses of globular cluster ($M \sim 10^6\,M_\odot$), this implies that about $\sim 4 \times 10^3$ type-II supernovae must explode in the globular cluster after $t \sim 2t_{\rm SN} \sim 6$ Myr, with a total energy of 4×10^{54} erg, which exceeds the cluster binding energy even if the mass of the proto-globular cluster is equal to $10^7\,M_\odot$. Figure 2 shows the distribution of stellar metallicities in a model globular cluster with a mass ejection at $t = 2t_{\rm SN}$. In this case, the maximum metallicity is $Z \simeq 0.2\,Z_\odot$ and the peak of the distribution is located at $Z \sim 0.01\,Z_\odot$, as in the case considered above.

However, we cannot rule out another possibility. If the SFR in the second stage (i.e., after the transition to a star-formation regime with the normal IMF) is sufficiently high that there remains little gas available for the formation of new stars by the time of the next episode of ejection of metals by supernovae, the fraction of high-metallicity stars should also be small proportional to the amount of remaining gas. Figure 3 illustrates this situation for the case when the SFE in the second stage is equal to $f_2 = 2 \times 10^8$ cm^3 g^{-1} s^{-1}. Figure 4 shows the dependence of the fraction of high-metallicity stars in the peak of the distribution on the SFE (the coefficient f_2). We can easily see that, for sufficiently high f_2, the fraction of high-metallicity stars is negligible. Note, however, that a high SFE f_2 implies triggered star formation. Indeed, for $f_2 = 2 \times 10^8$ cm^3 g^{-1} s^{-1}, $\rho \sim 10^{-21}$ g/cm^3, and $R_{\rm c} \simeq 20$ pc we have $\psi_2 \sim 2.6\,M_\odot$ yr^{-1}, which exceeds $\psi \sim M/t_J \sim 1\,M_\odot$ yr^{-1} by almost half an order of magnitude.

4. CONCLUSIONS

We have used a simple single-zone model to analyze the chemical evolution of globular cluster in framework of the self-enrichment scenario. We have shown that the star formation in globular clusters must involve two main episodes: an initial stage of star formation characterized by an IMF that is shifted toward massive stars ($M \geq 8\,M_\odot$)—in agreement with modern concepts suggesting that stars born from matter with the primordial composition should be massive—and a second stage of star formation in which the IMF transforms to its normal form with a predominance of low-mass stars.

We argue that the transition of the star formation to a regime with a normal IMF is associated with radiative cooling of the matter of the proto-globular cluster after its enrichment in heavy elements. This transition occurs when the abundance of

metals in the proto-globular cluster becomes equal to $Z \simeq 0.02\, Z_\odot$, which is close to the observed metallicities of globular clusters. It turns out that the metallicity of a globular cluster decreases with increasing mass as $Z \propto M^{-1}$ in the case of Schmidt star-formation law. This law agrees with the observed dependence.

ACKNOWLEDGMENTS

We are grateful to V.A. Marsakov for fruitful discussions of the results and to the referee for critical comments. This research has made use of the NASA Astrophysics Data System Bibliographic Services.

REFERENCES

1. G. Meylan and D. C. Heggie, Astron. Astrophys. Rev. **8**, 1 (1997).
2. S. Djorgovski and G. Meylan, Astron. J. **108**, 1292 (1994).
3. G. Parmentier and G. Gilmore, Astron. Astrophys. **378**, 97 (2001).
4. R. B. Larson, in *The Globular Clusters–Galaxy Connection,* Ed. by H. Graeme and J. P. Brodie (Astron. Soc. Pac., San Francisco, 1993); Astron. Soc. Pac. Conf. Ser. **48**, 675 (1993).
5. W. E. Harris and R. E. Pudritz, Astrophys. J. **429**, 177 (1994).
6. B. M. Shustov and D. S. Wiebe, Mon. Not. R. Astron. Soc. **319**, 1047 (2000).
7. G. Parmentier, E. Jehin, P. Magain, *et al.*, Astron. Astrophys. **352**, 138 (1999).
8. T. Tsujimoto and T. Shigeyama, Astrophys. J. **590**, 803 (2003).
9. C. Firmani and A. V. Tutukov, Astron. Astrophys. **264**, 37 (1992).
10. B. M. Shustov, D. S. Wiebe, and A. V. Tutukov, Astron. Astrophys. **317**, 397 (1997).
11. D. Z. Vibe, A. V. Tutukov, and B. M. Shustov, Astron. Zh. **75**, 3 (1998) [Astron. Rep. **42**, 1 (1998)].
12. D. P. Cox, Astrophys. J. **265**, L61 (1983).
13. M. Zhang and E. Ma, Chin. Astron. Astrophys. **17/1**, 43 (1993).
14. E. Jehin, P. Magain, C. Neuforge, *et al.*, Astron. Astrophys. **330**, L33 (1998).
15. E. Jehin, P. Magain, C. Neuforge, *et al.*, Astron. Astrophys. **341**, 241 (1999).
16. V. Bromm, P. S. Coppi, and R. B. Larson, Astrophys. J. **564**, 23 (2002).
17. E. O. Vasil'ev and Yu. A. Shchekinov, Astron. Zh. (in press).
18. J. P. Ostriker and C. F. McKee, Rev. Mod. Phys. **60**, 1 (1988).
19. M. A. de Avillez and M.-M. Mac Low, Astrophys. J. **581**, 1047 (2002).
20. D. Arnett, Astrophys. J., Suppl. Ser. **127**, 213 (2000).
21. I. B. Bernstein and D. L. Book, Astrophys. J. **225**, 633 (1978).
22. E. T. Vishniac and D. Ryu, Astrophys. J. **337**, 917 (1989).
23. E. T. Vishniac, Astrophys. J. **274**, 152 (1983).
24. M. V. Penston, Astrophys. J. **162**, 771 (1970).
25. M. G. Wolfire, D. Hollenbach, C. F. McKee, *et al.*, Astrophys. J. **443**, 152 (1995).
26. S. E. Woosley and T. A. Weaver, Astrophys. J., Suppl. Ser. **101**, 181 (1995).
27. Ya. B. Zel'dovich and I. D. Novikov, *Theory of Gravity and Evolution of Stars* (Nauka, Moscow, 1971).

Translated by A. Dambis

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 871–883. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 976–989.

A Model for the Population of Helium Stars in the Galaxy: Low-Mass Stars

L. R. Yungelson and A. V. Tutukov

Institute of Astronomy, Moscow, Russia

Received May 1, 2005; in final form, May 18, 2005

Abstract—We model the Galactic ensemble of helium stars using population synthesis techniques, assuming that all helium stars are formed in binaries. In this picture, single helium stars are produced by mergers of helium remnants of the components of close binaries (mainly, the merging of helium white dwarfs) or in the disruption of binaries with helium components during supernova explosions. The estimated total birthrate of helium stars in the Galaxy is 0.043 yr^{-1}; the total number is 4×10^6; and the binarity rate is 76%. We construct a subsample of low-mass ($M_{\mathrm{He}} \lesssim 2\,M_\odot$) helium stars defined by observational selection effects: the limiting magnitude ($V_{\mathrm{He}} \leq 16$), ratio of the magnitudes of the components in binaries ($V_{\mathrm{He}} \leq V_{\mathrm{comp}}$), and lower limit for the semiamplitude of the radial velocity required for detecting binarity ($K_{\mathrm{min}} = 30$ km s^{-1}). The parameters of this subsample are in satisfactory agreement with observations of helium subdwarfs. In particular, the binarity rate in the selection-limited sample is 58%. We analyze the relations between the orbital periods and masses of helium subdwarfs and their companions in systems with various combinations of components. We predict that the overwhelming majority (∼97%) of unobserved companions to helium stars will be white dwarfs, predominantly, carbon–oxygen white dwarfs.

1. INTRODUCTION

This study focuses on nondegenerate helium stars. The observed objects include representatives of two groups—helium subdwarfs and Wolf–Rayet (WR) stars. Interest in helium stars is enhanced by the fact that they are related to other groups of stars, such as AM CVn stars, RCrB stars, and double-degenerates (possible progenitors of type Ia supernovae). The most massive helium stars are the progenitors of type Ib and Ic supernovae. The optical component of Cyg X-3, one of the most remarkable X-ray sources, is a helium star. The most massive of stars with spectroscopically determined masses is the close binary WR 20a; both components of this system are helium stars with masses $M \sim 80\,M_\odot$ [1, 2].

Helium subdwarfs are subdivided into three groups: sdB stars with hydrogen-rich atmospheres, sdO stars with atmospheres dominated by helium, and an intermediate group of sdOB stars. The masses of helium subdwarfs derived from their locations in a diagram of effective temperature $\log T_{\mathrm{eff}}$ vs. effective gravity $\log g$ are usually close to $0.5\,M_\odot$ [3, 4]. These estimates are consistent with the masses determined for the components of eclipsing systems, which range from 0.48 to $0.54\,M_\odot$ (see [5] and references therein), and masses obtained from asteroseismological data: $0.49 \pm 0.02\,M_\odot$ for PG0014+067 [6] and $0.47\,M_\odot$ for Feige 48 [7]. However, the evolutionary tracks of numerous low-mass ($M_{\mathrm{WD}}/M_\odot \lesssim 0.3$) helium white dwarfs with contracting hydrogen envelopes also cross the region of the $\log T_{\mathrm{eff}} - \log g$ diagram populated by helium subdwarfs. Thus, some of the stars that are classified as subdwarfs may actually be relatively hot, helium white dwarfs rather than helium stars. For instance, the mass of HD 188112, which is classified as sdB, is $0.24\,M_\odot$; i.e., it is a precursor of a helium white dwarf [8]. Another example is HS 2333+3927: the position of this star in the $\log T_{\mathrm{eff}} - \log g$ diagram is consistent with the tracks of both a $0.29\,M_\odot$ white dwarf and a $0.47\,M_\odot$ helium star with a thin hydrogen envelope [9].

The lower limit to the masses of WR stars is $\simeq 7\,M_\odot$ [10]. However, according to modern theory on the evolution of close binary stars, the mass spectrum of helium stars should be continuous. It is possible that the WR phenomenon is observed only for stars with masses $\gtrsim 7\,M_\odot$, because the radial pulsations that are required to generate the shock waves that produce the dense radiative stellar winds of WR stars are not excited in lower mass stars, according to the mass-luminosity relationship typical for such stars (see, e.g., [11]).

There are several channels for the production of helium stars. Stars with zero-age main sequence (ZAMS) masses exceeding $\sim 50\,M_\odot$ become WR stars due to mass loss via their stellar winds [12]. It is possible that single stars and the components of

1063-7729/05/4911-0871$26.00

moderate-mass wide binaries may almost completely lose their hydrogen envelopes immediately before or during the helium flash [13, 14]. In close binaries, stars with helium cores and thin hydrogen envelopes form due to a mass loss from Roche lobe overflow [15, 16]. Single nondegenerate helium stars can form due to the merger of helium white dwarfs or nondegenerate helium stars in close binaries [17–19].

The main scenarios for the formation of low-mass helium stars in close binaries are considered [17, 18, 20–22]. In particular, it was shown that the locations of helium remnants of the components of close binaries, the products of the mergers of such remnants, and their evolutionary tracks in the Hertzsprung–Russell and $\log T_{\rm eff} - \log g$ diagrams are in good agreement with observations [18, 21, 22]. Detailed study of the efficiencies of various channels for the formation of low-mass helium stars led Han *et al.* [21, 22] to conclude that it is possible to explain the main parameters of observed helium subdwarfs when selection effects are taken into account.

In the present paper, we use a population-synthesis approach to construct a model for the ensemble of Galactic helium stars, and analyze the properties of the low-mass subsystem of this group of stars (i.e., of helium subdwarfs). We used the same population-synthesis code as in [23]. In contrast to Han *et al.* [21, 22], we compare our results with observations of a magnitude-limited subsample of model stars, since such a limit exists for the Palomar–Green catalog [24], which is the main catalog of hot subdwarfs. We are particularly interested in some aspects of the binarity of hot subdwarfs, especially the orbital-period distribution for subdwarfs in binaries with various companions.

Massive helium stars and the products of their evolution will be considered in a future paper in this series.

Section 2 describes the basic assumptions used in the population-synthesis code that are directly relevant to modeling the formation of helium stars. The main results are presented in Section 3, and a discussion follows in Section 4.

2. POPULATION SYNTHESIS FOR BINARY STARS

Both high- and low-mass helium stars are formed when one component of a close binary system overflows its Roche lobe in the hydrogen shell-burning stage [15, 16] (in so-called "case B" mass exchange). The population-synthesis code uses the following assumptions relevant to case B mass exchange.

Components of close binaries with ZAMS masses $M_{\rm i} < 2.8\,M_\odot$ form degenerate helium dwarfs. Components of close binaries that initially had masses higher than $2.8\,M_\odot$ (or accumulated that much mass as a result of mass exchange) produce nondegenerate helium stars. According to evolutionary computations [20, 25], the mass of these stars is

$$M_{\rm He}/M_\odot \approx \max(0.066(M_{\rm i}/M_\odot)^{1.54}, \quad (1)$$

$$0.082(M_{\rm i}/M_\odot)^{1.4}). \quad (2)$$

Equation (1) gives a minimum mass for helium stars close to $0.35\,M_\odot$, which is in reasonable agreement with the results of more detailed evolutionary computations [26]. We neglect the dependence of the mass of the helium remnant on the moment of Roche lobe overflow, which is not very important in case B mass exchange.

Based on the evolutionary computations of [20, 25], we approximated the lifetimes of stars in the region of the Hertzsprung–Russell diagram populated by helium stars as

$$\log T_{\rm He} = \begin{cases} 7.15 - 3.7\log(M_{\rm He}/M_\odot) \\ \text{if } M_{\rm He} \le 1\,M_\odot, \\ 7.15 - 3.7\log(M_{\rm He}/M_\odot) \\ \quad + 2.23[\log(M_{\rm He}/M_\odot)]^{1.37} \\ \text{if } M_{\rm He} > 1\,M_\odot. \end{cases} \quad (3)$$

Stellar-wind mass loss in the main-sequence stage is described by the difference between the initial and final masses [27]:

$$M_{\rm f} = \begin{cases} M_{\rm i} - 0.00128\,M_{\rm i}^{2.5} & \text{if } M_{\rm i} \le 30\,M_\odot, \\ 1.77\,M_{\rm i}^{0.763} & \text{if } M_{\rm i} > 30\,M_\odot. \end{cases} \quad (4)$$

It is assumed that stars with ZAMS masses in excess of $50\,M_\odot$ lose so much matter via their stellar winds in the core hydrogen-burning stage that they become helium stars immediately after the completion of the main-sequence stage, and never fill their Roche lobes. Thus, only helium stars with masses exceeding $\simeq 27\,M_\odot$ can be products of the evolution of single stars or of the components of wide binaries.

Mass loss by helium stars with masses $M \ge 5\,M_\odot$ is described by the following approximation to the observed mass-loss rates of WR stars (see [10]):

$$\dot{M}_{\rm He} = \min(1.38\times 10^{-8}(M_{\rm He}/M_\odot)^{2.87}, \quad (5)$$

$$10^{-4})M_\odot/{\rm yr}. \quad (6)$$

The analytical expression in (4) was suggested by Nelemans and van den Heuvel [28]. The slope of the $\dot{M}_{\rm He}(M_{\rm He})$ relation changes at $M_{\rm He} \approx 22.1\,M_\odot$.

Some signs of the presence of a stellar wind are suspected only for a few hot subdwarfs [29]; numerical estimates of the mass-loss rates in a radiative stellar wind model yield $\dot{M} \sim 10^{-11}\,M_\odot/{\rm yr}$ [30]. Given the lifetimes of helium stars, the only effect that such

mass loss should have on the evolution of a hot subdwarf is the loss of the remnant of its hydrogen envelope, as well as, possibly, changes in the chemical composition of the subdwarf atmosphere.

Helium stars with progenitor masses $M_\mathrm{i} = (2.8{-}5)\,M_\odot$ do not expand after the depletion of the helium in their cores, and become compact carbon–oxygen white dwarfs. Helium stars with $M_\mathrm{i} = (5{-}11.6)\,M_\odot$ finish their evolution as carbon–oxygen white dwarfs after their expansion in the helium shell-burning stage and second Roche lobe overflow, which is accompanied by only an insignificant mass loss.

Stars with ZAMS masses $M_\mathrm{i} = (11.6{-}30)\,M_\odot$ produce neutron stars with $M_\mathrm{NS} = 1.4\,M_\odot$. Initially more massive stars end up as black holes. The formation of black holes is accompanied by the loss of 20% of the presupernova mass. It is assumed that supernovae explosions are spherically symmetric; i.e., nascent neutron stars and black holes do not acquire a kick velocity (see the detailed discussion on this topic in the following paper of this series).

Stars with masses below $11.6\,M_\odot$ that overflow their Roche lobes in the AGB stage (case C mass exchange) produce white dwarfs, whose masses depend on the moment of Roche-lobe overflow. More massive stars produce neutron stars or black holes, as in case B mass exchange.

The relation between the initial mass of a star, its mass at the end of the main-sequence stage, the mass of the helium star it produces, and the masses of the final products of the star's evolution is plotted in Fig. 1.

We made the following assumptions about the mass-exchange process, based on the results of numerical modeling [31]. If both the donor and accretor are main-sequence stars, the donor has a radiative envelope, and accretor-to-donor mass ratio is $q \geq 0.8$, mass exchange is conservative and proceeds on the time scale for the nuclear evolution of the donor. If $0.5 \leq q < 0.8$, the donor loses mass on the thermal time scale t_th1, but accretion by the main-sequence star is limited to the rate $\dot{M}_\mathrm{th2}$, which corresponds to the thermal time scale for the secondary. The mass that can be accreted by the companion is limited to $\Delta M = \dot{M}_\mathrm{th2} t_\mathrm{th1}$. The variation of the component separation a is calculated assuming that a mass ΔM is accreted conservatively, and that the rest of the donor envelope, with mass $\Delta M_t = M_\mathrm{i} - M_\mathrm{f} - \Delta M$, is lost by the system; this latter process is described by the angular-momentum balance equation

$$\frac{\Delta J}{J} = \gamma \frac{\Delta M_t}{M_t}, \qquad (7)$$

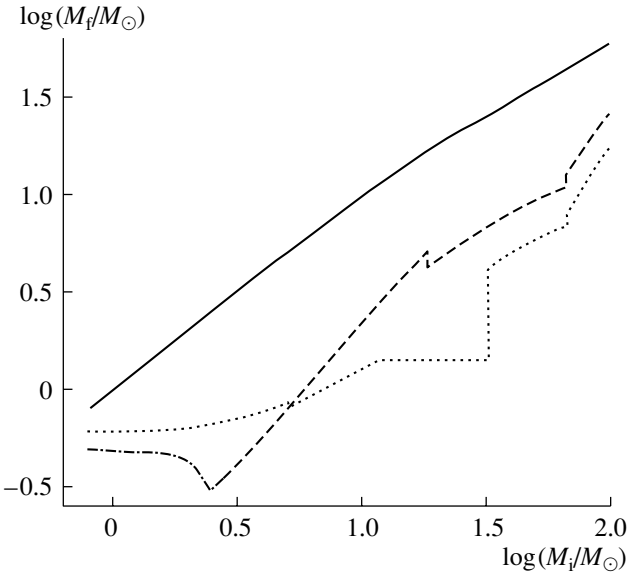

Fig. 1. Relation between the ZAMS mass of stars M_i, their mass at the end of the main-sequence stage (solid curve), the mass of helium stars produced by these stars (dashed curve), and the mass of the final products of their evolution: helium white dwarfs (dash-dotted curve), carbon–oxygen and oxygen–neon white dwarfs, neutron stars, and black holes (dotted curve). For stars with $M_\mathrm{i} \lesssim 5\,M_\odot$, the upper limit of the white dwarf mass attained in case C mass exchange is shown.

where J is the angular momentum of the binary, $M_t = M_1 + M_2$, and $\gamma = 1.5$. The validity of (5) with $\gamma = 1.5$ is based on the reconstruction of the mass-exchange history in observed close binaries containing white dwarfs and helium subdwarfs [32, 33].

When the mass ratio of components $q < 0.5$ or the donor has a convective envelope that extends down to more than 50% of the stellar radius, mass loss occurs on the dynamical time scale, and a common envelope is formed. Variation of the component separation is computed based on the balance of the orbital energy of the binary and the binding energy of the donor envelope [34]:

$$\frac{(M_\mathrm{i} + m)(M_\mathrm{i} - M_\mathrm{f})}{a_\mathrm{i}} = \alpha_\mathrm{ce} \left(\frac{M_\mathrm{f} m}{a_\mathrm{f}} - \frac{M_\mathrm{i} m}{a_\mathrm{i}} \right). \qquad (8)$$

The subscripts i and f in (6) refer to the initial and final component masses and separations; m is the accretor mass, and α_ce the so-called "common-envelope parameter" describing the efficiency with which the released energy is deposited in the common envelope. Trial computations indicate that the observations are best fit if $\alpha_\mathrm{ce} = 2$, independent of the evolutionary state of the donor.

Equation (6) is always used if the accretor is a white dwarf, neutron star, or a black hole.

Birthrate and number of helium stars

Type of star	Birth rate, yr^{-1}	Total number	Subsample with $V \leq 16$
Single star	0.67×10^{-2}	0.96×10^{6}	0.64×10^{4}
MS companion	0.21×10^{-1}	0.13×10^{7}	0.24×10^{3}
He white-dwarf companion	0.50×10^{-2}	0.29×10^{6}	0.18×10^{4}
CO white-dwarf companion	0.66×10^{-2}	0.13×10^{7}	0.64×10^{4}
ONe white-dwarf companion	0.80×10^{-3}	0.31×10^{5}	0.15×10^{3}
Neutron-star companion	0.19×10^{-2}	0.31×10^{5}	0.21×10^{3}
Black-hole companion	0.14×10^{-3}	0.96×10^{4}	0.66×10^{2}
Binary helium star	0.34×10^{-3}	0.61×10^{4}	0.58×10^{2}

Nondegenerate single helium stars can be produced by close binaries. We considered three channels for the formation of single helium stars: the merger of two helium white dwarfs, of a helium white dwarf and a carbon–oxygen one, or of a helium white dwarf and a nondegenerate helium star. In the first two cases, the merger results from the loss of systemic angular momentum via gravitational-wave radiation. A helium star and helium white dwarf can merge due to momentum loss either via gravitational-wave radiation or in the common envelope that forms after the depletion of helium in the core of the nondegenerate star and its expansion (if $M_{\rm He} \approx 0.8$–$2.8\,M_\odot$). In a merging pair of white dwarfs, the less massive component fills its Roche lobe first and, if $M_2/M_1 < 2/3$, the donor is disrupted on the dynamical time scale, probably forming a disk around its companion [35, 36]. Accretion in such systems has not been studied, but we expect that the rate of accretion from the disk is determined primarily by the viscosity of the disk material. The outcome of the accretion depends on the accretor mass and temperature, and the accretion rate. The minimum total mass $M_{\rm He}^{\rm min}$ for the formation of a nondegenerate helium star via the merger of two white dwarfs is, essentially, still a free parameter. It is clear only that it must exceed the minimum mass of nondegenerate helium stars $\sim 0.3\,M_\odot$ (see, for instance, [18]). Our computations assumed $M_{\rm He}^{\rm min} = 0.4\,M_\odot$, which is in good agreement with the results of computations [18, 21]; see also the discussion in Section 3. The lifetimes of the products of mergers of two helium white dwarfs or a helium star and helium dwarf were also calculated using (2), although this equation may overestimate somewhat the lifetimes of the merger products as helium stars (by $\lesssim$20%, depending on the mass of the merger product, as follows from a comparison with Eq. (1) in [18]. The lifetimes of the products of mergers of helium and carbon-oxygen dwarfs are determined by a "competition" of two processes: shell helium burning and stellar-wind mass loss. Note that mergers of helium dwarfs account for the formation of about 90% of single helium stars in our model.

Based on studies of eclipsing, spectroscopic, and visual binaries [37–41], we assumed that all stars are born in binaries with orbital separations a ranging from $6(M_1/M_\odot)^{1/3}\,R_\odot$ to $10^6\,R_\odot$. The binary birth rate was taken to be [38]:

$$\frac{dN}{dt} = 0.2 d(\log a) \frac{dM_1}{M_1^{2.5}} f(q) dq, \qquad (9)$$

where M_1 is the mass of the initially more massive component of the system, and $f(q)$ is the distribution of the mass ratio of components $q = M_2/M_1$ normalized to unity. We assumed that $f(q) = 1$ for close binaries. The normalization of (7) assumes that one binary system with $M_1 \geq 0.8\,M_\odot$ is born per year in the Galaxy. According to (7), about 40% of all binaries are "close." The star-formation rate was specified to be constant over 13.5 Gyr. Obviously, this assumption does not influence the results for helium stars in binary systems, since the lifetimes of even the least massive helium stars do not exceed $\sim$1 Gyr. However, taking into account that mergers of binary white dwarfs due to momentum loss via gravitational-wave radiation occurs in the Hubble time scale, the increased star formation rate can influence formation of single helium stars at early stages of Galaxy life.

Since our model assumes that all stars are formed in binaries, we did not consider the possible formation of low-mass helium stars during the evolution of single stars or of the components of wide binaries (see, e.g., [14]). In the latter case, it is supposed that the precursors of helium stars are red giants who lost nearly their entire hydrogen envelopes just before or during the helium flash. The mass-loss mechanism is unknown, and the mass-loss rate is treated like a free parameter. The real binarity rate for hot subdwarfs is

difficult to determine, since, with commonly applied observational methods, selection effects preclude the detection of binary subdwarfs with orbital periods longer than 200 to 300 days [42, 43]. It is obvious that we cannot exclude *a priori* the possible formation of low-mass helium stars from single stars or the components of wide binaries, but, as we shall show below, mergers of helium white dwarfs and/or nondegenerate helium stars are able to explain the formation of a sufficiently large number of single helium stars.

3. MAIN RESULTS

3.1. General Properties of the Population of Low-Mass Helium Stars

The main aim of our study is to model the population of helium stars and the distributions of their main observational parameters—the masses and orbital periods of the binaries—as well as to study relations between the parameters of close binaries with helium-star components.

As was noted in Section 2, helium stars in binaries form after Roche-lobe overflow by stars that are in the hydrogen shell-burning stage. These stars can be either the primary or secondary components of the systems. In the first case, the companions to the nascent helium stars are main-sequence stars, while, in the second case, they are white dwarfs, neutron stars, black holes, or, on very rare occasions, also helium stars. The table lists the total birth rate and current number of helium stars in the Galaxy for the assumptions described in Section 2. The dominant constituents of the population of helium stars are binaries with main-sequence and carbon–oxygen white-dwarf companions. The number of objects in the third main group, i.e., single stars, depends on assumptions about the minimum mass necessary for helium ignition in the product of the merger of two helium white dwarfs: if this critical mass $M_{\rm He}^{\rm min}$ is reduced from 0.4 to $0.35\,M_\odot$, the number of single helium stars increases by approximately 8%. About 0.8% of all helium stars have neutron-star companions. The number of such binaries may be overestimated by an order of magnitude (see, e.g., [44–47]), since the present model does not take into account the possibility that the neutron stars acquire a "kick" when they are born, which could result in the disruption of the system (see the more detailed discussion in our next paper in this series).

The total model Galactic birth rate of helium stars is $0.043\,{\rm yr}^{-1}$, and the total number of helium stars is 4×10^6. The binarity rate for helium stars is 76%. Since the progenitors of helium stars descend from moderate- and high-mass main-sequence stars ($M\geq2.8\,M_\odot$), the semithickness of the Galactic disk can be taken to be 200 pc for these progenitors (in agreement with the empirical scale height z for the field helium subdwarfs [3, 48, 49]); then, the volume of the disk component of the Galaxy is 2.8×10^{11} pc^3 and the spatial density of helium stars is 1.4×10^{-5} pc^{-3}. This latter estimate is considerably higher than observational estimates: 2×10^{-6} pc^{-3} [3], 4×10^{-6} pc^{-3} [48]. However, as we shall show further on, a considerable fraction of low-mass helium stars may be "lost" from the observations due to selection effects.

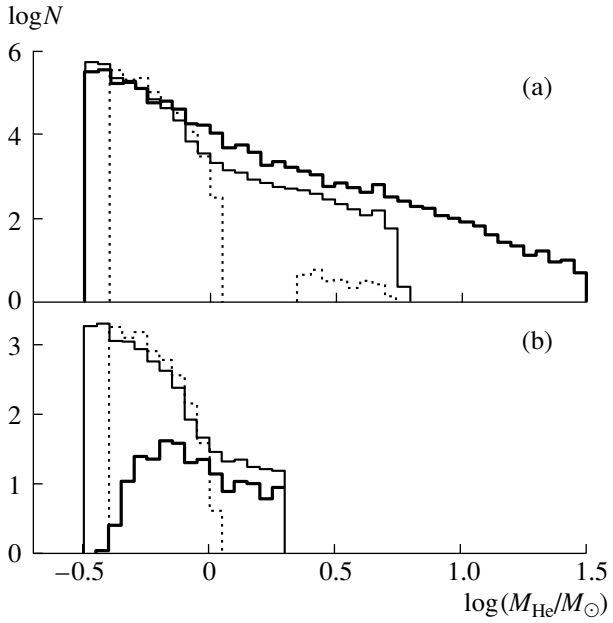

Fig. 2. Mass distribution of the helium stars (a) in the total sample and (b) in the sample limited by selection effects. The bold solid line shows the distribution for helium stars with main-sequence companions; the thin solid line, the distribution for stars with white-dwarf companions; and the dotted line, the distribution for single helium stars.

Figure 2 shows the mass distributions of helium stars with various companions (for the total population and for the subset limited by selection effects; see below). The masses of the helium stars are strongly concentrated toward the minimum value ($\simeq0.4\,M_\odot$) in all groups. The main group of single objects consists of products of mergers of white dwarfs and of white dwarfs with helium stars; very rare single stars with masses of 2.2–$5.6\,M_\odot$ are produced by mergers of helium stars and helium dwarfs or the former components of binaries that were disrupted by supernova explosions. The masses of the helium stars are limited to $\simeq0.9\,M_\odot$ in systems with helium-dwarf companions. This maximum mass corresponds to the case of a completely conservative first mass exchange and completely nonconservative second mass exchange, when the Roche lobe is

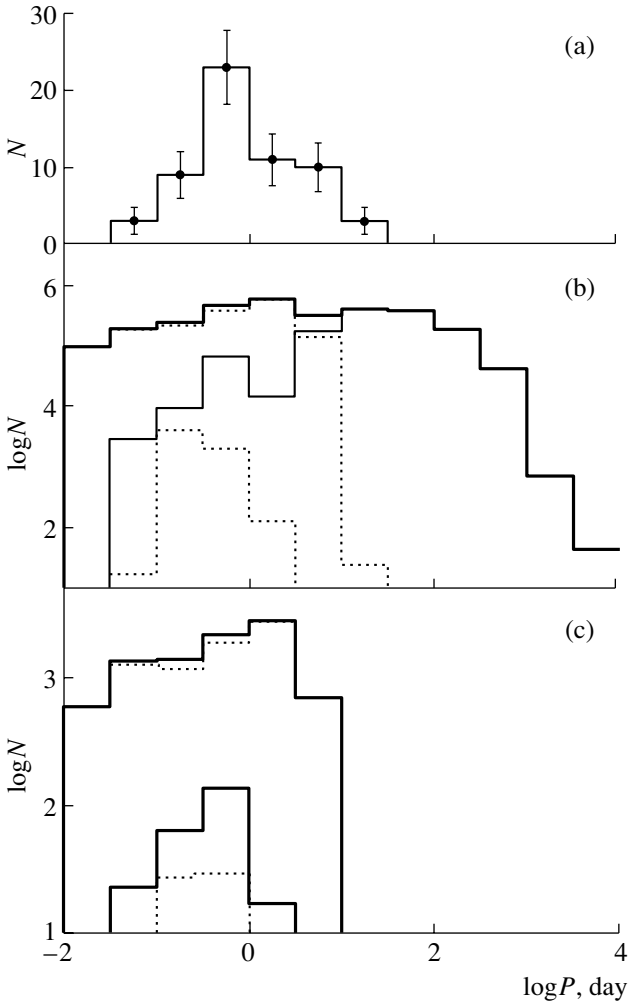

Fig. 3. Distribution of orbital periods for helium stars. (a) Observed subdwarfs (L. Morales-Rueda, priv. com.). (b) Model sample for the total population. The upper bold line shows the total distribution, and the thin solid line and upper and lower dashed lines show the contributions of stars with main-sequence, white-dwarf, and helium secondaries to the total population. (c) Period distribution in the sample is limited by selection effects; the meaning of the lines is the same as in (b).

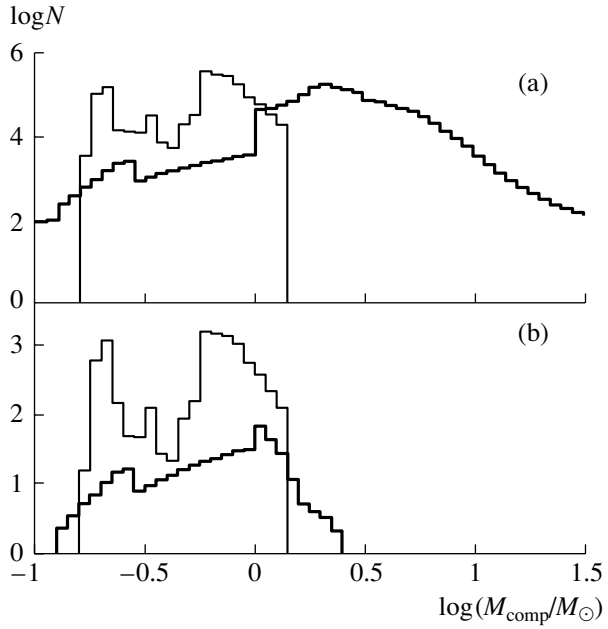

Fig. 4. Model mass distribution for the companions to helium stars. (a) Total model sample. The bold line shows the distribution for main-sequence companions, and the thin line, the distribution of white-dwarf companions. (b) Model sample is limited by selection effects; the meaning of the lines is as in (a).

filled by the initially less-massive component of the system; i.e., to the sequence of combinations of component masses: $(M_1,\ M_2) = (2.8\,M_\odot, 2.8\,M_\odot) \rightarrow (0.4\,M_\odot, 5.2\,M_\odot) \rightarrow (0.4\,M_\odot, 0.9\,M_\odot)$. The upper limit of the helium-star mass in systems with CO or ONe companions is associated with a similar scenario, and leads to an upper limit for the mass of the progenitors of white dwarfs: $(M_1,\ M_2) = (11.4\,M_\odot, 11.4\,M_\odot) \rightarrow (1.35\,M_\odot, 21.45\,M_\odot) \rightarrow (1.35\,M_\odot, 5.7\,M_\odot)$. In the latter systems, the white dwarf's companion explodes in the course of the subsequent evolution, disrupting the system.

Figure 3b shows the orbital-period distribution for helium stars. The lower limit of the orbital periods is determined by the conditions of the merger of the components in common envelopes. For systems with main-sequence companions, the upper limit of the period is the maximum period that can be reached in the course of conservative mass exchange. For systems with white-dwarf companions, the upper limit of the period is determined by the reduction in orbital separation in the common envelope. The model sample contains some systems with periods $\sim 0.01^{\rm d}$. The shortest orbital period among observed systems is $P = 0.073^{\rm d}$, for PG 1017−086. The existence of model systems with shorter orbital periods may be due to the fact that, in the formal procedure in the population-synthesis code that checks whether merger of components in the common envelope is possible, the radius of the helium star used is that of a helium star with a mass equal to the mass of the helium core of the donor; a more realistic procedure would employ the much larger radius of the inert helium core itself. It is also possible that the common-envelope parameter $\alpha_{\rm ce}$ is somewhat larger than we have assumed. Another reason for the presence of systems with shorter orbital periods may be underestimation of the radii of the helium stars, which depend on the masses of their hydrogen envelopes. Thus, the systems with the shortest orbital periods in Fig. 3 should, probably, be classified as

single helium stars. Their number should then be lower than is shown in Fig. 3, due to the shorter lifetimes of merger products of helium white dwarfs and helium stars.

The mass distribution of the companions to the helium stars is shown in Fig. 4a. The two discontinuities in the distribution for stars with main-sequence companions near 0.3 and 1.0 $M_\odot$ are due to the loss of angular momentum by main-sequence stars with $0.3 \lesssim M/M_\odot \lesssim 1.0$ via magnetically-coupled stellar winds. For some systems, their lifetime until contact is appreciably shorter than the lifetimes of helium stars with masses $\sim 0.4\,M_\odot$. The upper limit of the range of masses of stars that are subject to magnetic stellar winds is not known with certainty. If this limit is raised to $1.5\,M_\odot$, the number of systems with main-sequence companions is reduced by 10%, and the discontinuity shifts to $1.5\,M_\odot$ and becomes less pronounced.

The low-mass group of white dwarfs are helium white dwarfs, while the group of massive white dwarfs contains carbon–oxygen and oxygen–neon white dwarfs. The two groups of white dwarfs overlap due to the existence of low-mass ($M \approx 0.35$–$0.40\,M_\odot$) hybrid CO dwarfs; these stars descend from helium stars of the same mass.

3.2. Selection Effects

In the $\log T_{\rm eff} - \log g$ diagram, the helium remnants of stars with thin hydrogen envelopes spend their whole core helium-burning time in the region occupied by subdwarfs, since they have $\log T_{\rm eff} = 4.4$–4.5 and $\log g = 5.65$–5.90. For comparison with observations, we constructed a subset of model helium stars that is limited in magnitude to $V = 16$, since the main observational surveys of stars with UV excesses are limited to $B \approx 16$ [24] and $B \approx 15.3$ [48]. The SPY survey of white dwarfs [50, 51] is limited to $B \approx 16.5$; this catalog also contains a large number of hot subdwarfs due to errors in the spectral classification in the input catalog. Since our model is fairly primitive and depends on a number of poorly restricted parameters, we neglected the small color excess $B-V = -0.25 \pm 0.1$ [3, 52–54] that is typical for helium subdwarfs. For systems with main-sequence companions, we applied the condition $V_{\rm He} < V_{\rm MS}$ as an observational selection effect. The use of this rather restrictive condition essentially yields a lower limit of the number of binaries for which detection of a companion via spectroscopic methods is possible. The magnitude-limited sample contains only stars with masses lower than $2\,M_\odot$, since only for these was it possible to construct a bolometric-correction scale for V, based on the computations of Iben and Tutukov [20]. The interstellar absorption was taken to be 1.6^m/kpc [55]. For all binaries, we assumed that velocity semiamplitude must be $K_{\rm min} \geq 30$ km s^{-1}; this corresponds to the lower limit of K in known binaries with helium subdwarfs [56].

In contrast to [21, 22], we did not exclude model systems with G and K main-sequence companions from our study. Han *et al.* [21, 22] suggested that objects displaying the K CaII absorption line that were rejected as candidates to the PG catalog [24], are, in reality, helium subdwarfs with F, G, and K main-sequence companions. However, a detailed analysis of the rejected candidates [57] showed that they actually are low-metallicity sdF/sdG subdwarfs, but not sdB/sdO subdwarfs.

Since we were interested in *all* helium stars, we did not take into account the so-called "strip effect"—the restriction of the model sample to stars with $\log g$ and $\log T_{\rm eff}$ values that are typical for core helium-burning model helium stars with thin hydrogen envelopes; in fact, some of the stars that are classified as hot subdwarfs are located outside of this strip [4, 58].

The model sample of stars limited by selection effects contains 15 210 objects (including systems with relativistic companions, see the table). This number comprises ~0.4% of all helium stars. The Palomar–Green catalog [24], which is limited to $B \lesssim 16$ contains about 700 objects classified as helium subdwarfs distributed over a 10 714 square degree area, corresponding to a total of about 2800 "observed" stars. Bearing in mind all the uncertainties in the model, the relatively simple treatment of selection effects, and misclassifications in the PG catalog, we consider the agreement in the numbers of stars in our model and in the catalog (within ~5) to be satisfactory. The most uncertain parameters of the model are the threshold mass for the helium-star progenitors and the minimum mass for helium ignition in the single products of white-dwarf mergers (which constitute about a quarter of all helium stars). Increasing both these masses may improve the agreement between the model and observations.

Figure 2b shows the mass distribution of the model helium stars in the sample limited by selection effects. The reduction in the number of stars in the sample compared to the total sample is due, in the first instance, to the reduction of the volume of space in which these stars might be observed (by a factor of ~150–200). Since we exclude systems that have main-sequence stars brighter than their companions, the number of helium stars with main-sequence companions is further reduced by a factor of ~20. As a result, the binary rate in the selection-effect-limited sample is reduced to 58%. Restricting the sample based on the limiting radial-velocity semiamplitude

necessary for detecting binarity leaves the number of stars in the sample virtually unaffected.

Observational estimates of the fraction of helium subdwarfs that are binary depend on the sample of stars under examination and the observing strategy, which determines the efficiency of detecting the binaries. For example, Maxted *et al.* [42] estimate the binary fraction among sdB stars with periods from $0.03^{\rm d}$ to $10.0^{\rm d}$ to be $69 \pm 9\%$; Napiwotzki *et al.* [43] give 42–45%, depending on adopted assumptions about the true period distribution in the range $0.03^{\rm d}$ to $30.0^{\rm d}$ and the component masses; Saffer *et al.* [59] estimate an upper limit for the fraction of single helium stars of 35%. Possible reasons for the discrepancies in the observational estimates of the fraction of binary helium subdwarfs include differences in the samples of stars that were studied (they could belong to populations with different ages and metallicities) and differences in the observational methods used. As was shown by Han *et al.* [22], the fraction of binaries decreases with decreasing metallicity; as Napiwotzki *et al.* [43] admit, their sample may contain a significant admixture of halo and thick disk stars. The model estimate of the fraction of "observed" binary helium stars agrees with the range of the observational estimates.

Note that reducing the common envelope parameter $\alpha_{\rm ce}$ to unity increases the rate of mergers in common envelopes; this reduces the fraction of binary helium stars to 35%, clearly contradicting the observations.

The fraction of binaries with main-sequence components in the total sample of model helium stars is 38%. Magnitude-selection effects reduce this fraction to 3%; these binaries may have composite spectra. The presence of a late-type companion may be revealed by an infrared excess. For example, according to [53], the fraction of systems with unresolved late-type companions among hot subdwarfs may be ~30–40%, close to their fraction in the total model sample. In some cases, a main-sequence companion may be detected via the reflection effect, due to heating by hot radiation from the subdwarf. However, this effect can be appreciable only in the closest systems whose companions are an M star or brown dwarf [60]. Note, however, that, for some subdwarfs with composite spectra that enable estimation of the parameters of their companions, the masses of the companion are confined to the range 0.8–$1.2\,M_\odot$ [52], close to the maximum of the distribution for the sample limited by selection effects (Fig. 4b). The remaining subdwarfs that are classified as single stars may be genuinely single or have white-dwarf companions.

While the total model population is dominated by helium stars with main-sequence companions, systems with white-dwarf companions begin to dominate in the selection-effect-limited sample. The maximum of the mass distribution of the helium stars in the total "limited" sample is shifted toward somewhat higher masses (into the range $-0.40 \leq \log(M/M_\odot) \leq -0.35$ in Fig. 2), improving the agreement with the "canonical" mass of subdwarfs of $0.5\,M_\odot$.

The distribution of 59 observed sdB/sdO subdwarfs with known orbital periods (Fig. 3a, data from L. Morales-Rueda, priv. comm.) varies within a factor of ~2 in the interval $0.1^{\rm d} \leq P_{\rm orb} \leq 10.0^{\rm d}$. In the model "observed" sample, the number of stars likewise varies within a factor of about two in the period range $0.3^{\rm d} \leq P_{\rm orb} \leq 3.0^{\rm d}$. The relative decrease in the number of model systems toward short periods is comparable to the observed behavior. As $P_{\rm orb}$ is increased relative to the maximum of the distribution, the number of stars in the model sample declines much more rapidly than in the observed sample. There are no stars with $P_{\rm orb} > 10.0^{\rm d}$ in the model of observed sample. This may be related to the fact that some stars that form common envelopes in the first episode of mass exchange in our model, in fact, evolve stably toward possessing longer periods, for instance, due to the loss of some of their mass prior to Roche-lobe overflow or due to stabilization of the mass exchange via momentum loss [26]. This idea should be tested via computations of grids of evolutionary models.

The accumulation of long series of observations will likely result in the discovery of new subdwarfs with long orbital periods. However, we should bear in mind that the efficiency of binarity detection declines with increasing $P_{\rm orb}$: this efficiency exceeds 80% for $P_{\rm orb} \leq 10.0^{\rm d}$, decreases to 50% for $P_{\rm orb} \sim 20.0^{\rm d}$–$30.0^{\rm d}$ [42, 43], and then rapidly declines to zero. We expect that observations with high spectral resolution should result in the detection of subdwarfs with relatively long orbital periods. However, for example, among nine stars detected in the SPY survey, seven have $P \lesssim 1.0^{\rm d}$, while only two have $P = 5.87^{\rm d}$ and $7.45^{\rm d}$ [43].

3.3. Relations between Parameters of Binaries with Nondegenerate Helium Components

Figure 5 shows the distribution of binaries with helium components and various companions in the selection-limited model sample in a diagram of the logarithm of the orbital period $\log P_{\rm orb}$ vs. the logarithm of helium star mass $\log M_{\rm He}$. As we noted before, after taking observational selection effects into account, systems with white-dwarf companions dominate the "observed" helium-star binaries

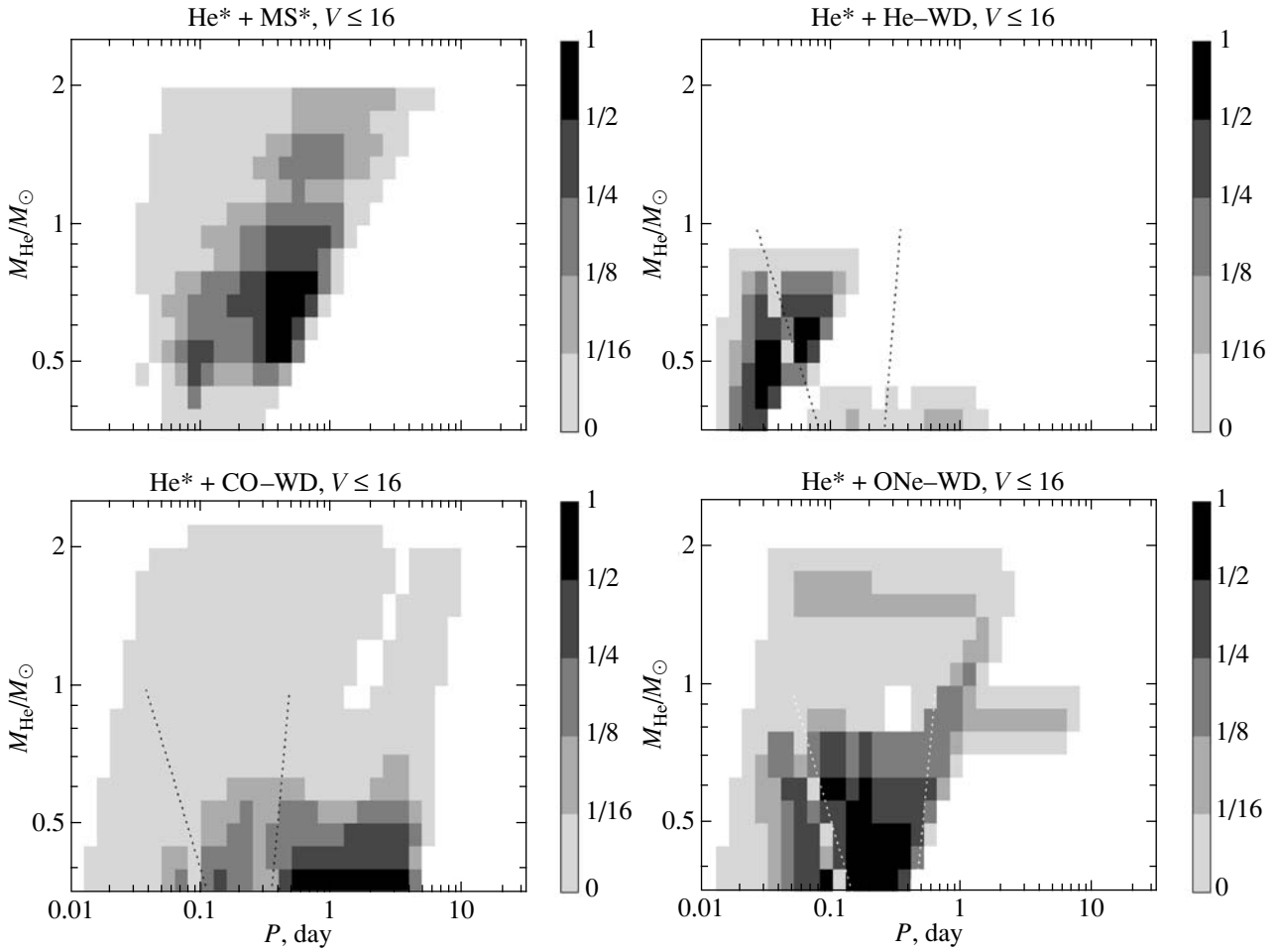

Fig. 5. Relations between the orbital periods and masses of helium stars with different companions. The sample limited by selection effects is shown. The five shades of grey scale show regions where the number density of systems is within a factor of one-half of the corresponding maximum of $\frac{\partial^2 N}{\partial \log P \partial \log M}$; or is confined to the ranges 1/4–1/2, 1/8–1/4, 1/16–1/8, and 0–1/16 of the overall maximum. Each panel is scaled separately. Binaries are absent in blank regions of the diagrams. In the panels for binaries with white-dwarf companions, systems that have orbital periods at birth which are shorter than those marked by the left dotted line merge during the lifetimes of the helium stars. Systems with periods at birth which are shorter than those marked by the right dotted line merge in less than the Hubble time.

(∼90%). The masses of the helium stars in these binaries must be confined between a minimum mass of $0.35\,M_\odot$ and a maximum mass of $0.5\,M_\odot$. Somewhat more massive helium stars with masses reaching 0.8–$1.0\,M_\odot$ can be observed in rare systems with main-sequence components.

The period distributions of the systems with different white dwarfs differ: when $P \lesssim 1.0^{\rm d}$, it is more likely to expect the presence of a helium white dwarf than a carbon–oxygen or oxygen–neon white dwarf. Despite the fact that we know from observations only the minimum masses of the white-dwarf companions to helium stars, we note that in only 4 of 12 known sdB + WD systems with $P_{\rm orb} \leq 1.0^{\rm d}$ can we exclude the presence of a helium white dwarf ($M_{\rm WD}^{\rm min} \geq 0.4\,M_\odot$). In the two known systems with $P_{\rm orb} > 1.0^{\rm d}$, the white dwarfs most probably belong to the carbon–oxygen/oxygen–neon family (see Table 5 in [56]).

Figure 6 shows the distributions of systems with helium stars and main-sequence or white-dwarf companions in the logarithm of the orbital period $\log P_{\rm orb}$ vs. logarithm of the companion mass $\log M_2$ plane for the same model sample limited by selection effects. The figure shows the masses of the companions to helium stars known from observations, as well as estimates of lower limits to the masses of such companions if only the mass function $f_{\rm m}$ is known:

$$f_{\rm m} = \frac{M_2^3 \sin^3 i}{(M_1 + M_2)^2} = \frac{P_{\rm orb} K_1^3}{2\pi G}.$$

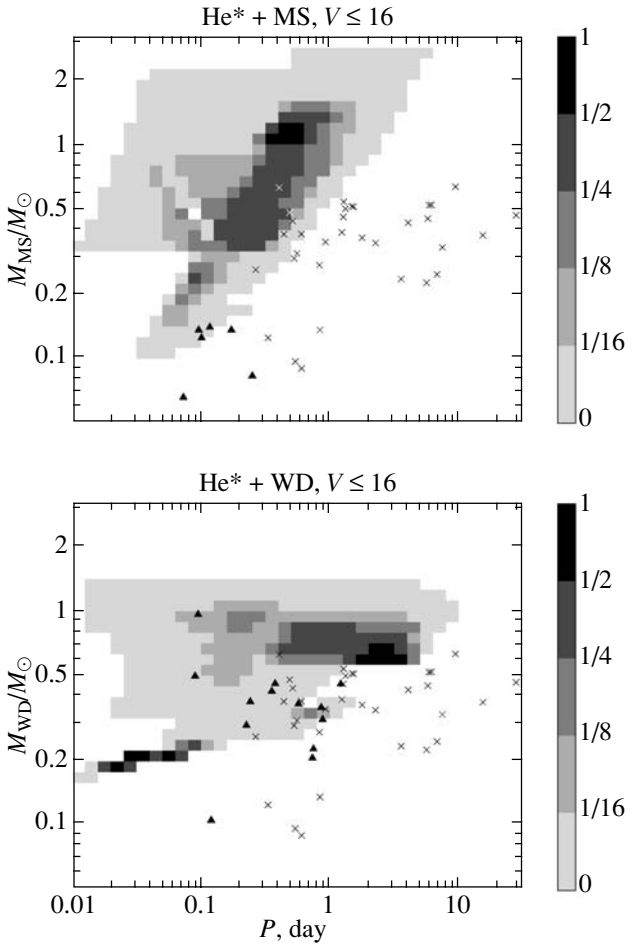

Fig. 6. Relation between the orbital periods of binaries and masses of the main-sequence and white-dwarf companions of the helium stars in these binaries. The grey scale is similar to that in Fig. 5. The triangles in the upper panel show known masses or lower limits for main-sequence companions, while those in the lower panel show masses or lower limits for known white-dwarf companions. In both panels, the crosses show lower limits to the masses of companions of an unknown type.

We use the data on the masses, radial-velocity semi-amplitudes K_1, and orbital periods for 52 objects listed in [43, 56] together with unpublished data provided to us by L. Morales-Rueda. When estimating $M_2^{\min}$, we assumed that $M_{He} = 0.5\,M_\odot$, $\sin i = 1$. Reducing M_{He} to $0.35\,M_\odot$ only weakly influences the positions of the stars in Fig. 6. Varying the orbital inclination from 90° to the most probable value, ∼60°, increases M_2 by about a factor of two. As a rule, shifting all the observed objects in Fig. 6 into the region populated by the model systems would require us to assign the observed systems quite probable orbital inclinations of 20°–30°.

Several binary sdB stars with M companions were discovered due to eclipses or reflection effects in the closest systems with $P \lesssim 0.25^d$. As we noted above, companions with masses of $1.0 \pm 0.2\,M_\odot$ are suspected in some systems with composite spectra, but are absent among systems with known orbital periods.

Since the number of model stars with white-dwarf components is about a factor of 200 higher than the number of systems with main-sequence companions, we expect the overwhelming majority of new identified and detected companions to helium subdwarfs to be carbon–oxygen or helium white dwarfs. As we noted above, the deficit of long-period model binaries may be due to underestimation of the fraction of binaries with stable first mass exchanges.

About 40% of the stars in the subsample of helium stars limited by observational selection effects are single. The mass spectrum of these stars is shifted slightly toward higher masses compared to the spectrum of the binary subdwarfs (Fig. 2). These stars are probably related to sdB/sdO subdwarfs whose atmospheres are enriched in He, and are slightly hotter than sdB subdwarfs with hydrogen-rich atmospheres. The location of these stars in the $\log T_{eff} - \log g$ diagram is more similar to the tracks of the merger products of helium dwarfs than to those of single low-mass stars with helium cores and thin hydrogen envelopes [61]. Note that the only sample of 23 sdO stars that has been systematically explored for binarity contains only one binary star [43]! The dedicated study of binarity detection probability [43] showed that the binary fraction in this sample is actually close to only ∼5%.

3.4. Final Evolutionary Stages of Close Binary Stars with Low-Mass Helium Components

As we noted above, in the closest systems with helium components, these stars may fill their Roche lobes before the depletion of helium in their cores, due to the loss of systemic angular momentum via gravitational-wave radiation. The corresponding orbital-period limits are plotted in Fig. 5 for the typical white-dwarf masses $M_{He} = 0.2\,M_\odot$, $M_{CO} = 0.6\,M_\odot$, and $M_{ONe} = 1.3\,M_\odot$. If the conditions for stable mass loss are fulfilled, AM CVn systems are produced [62, 63]. The birth rate of such systems in our model is $\simeq 5 \times 10^{-4}$ yr^{-1}. A similar estimate ($\simeq 3 \times 10^{-4}$ yr^{-1}) was obtained in [64] for somewhat different assumptions about the star-formation history and stellar evolution, but likewise using Eq. (5) to describe the mass exchange in systems of main-sequence stars with comparable component masses. The evolution of AM CVn stars and their observed features are considered in detail in [62–64]. If the conditions for stable mass exchange are not fulfilled, the merger of helium stars with carbon–oxygen or

oxygen–neon white dwarfs may result in the formation of R CrB stars [65].

After the depletion of the helium in their cores, helium stars with masses $\lesssim 0.8\,M_\odot$ turn into carbon–oxygen white dwarfs of the same mass, almost without any expansion. Close binary white dwarfs are formed in this way (another scenario leads to the formation of close binary white dwarfs through case C mass exchange). If binary white dwarfs have short enough orbital periods, they can merge due to the loss of angular momentum via gravitational-wave radiation on time scales less than Hubble time. The corresponding limiting periods are shown in Fig. 5. At present (March 2005), data on the orbital periods of 19 close binary white dwarfs have been published. Six of these systems (see, e.g., [66] and Table 11 in [67]) have periods short enough for them to merge over Hubble time, and one has a total mass larger than the Chandrasekhar mass. If both components of this system are carbon–oxygen white dwarfs, the system may turn out to be a precursor of a type Ia supernova [68–70]. Mergers of carbon–oxygen and helium white dwarfs probably result, not in SN Ia, but in "helium novae" [71, 72] or the formation of R CrB stars [65]. Note that our model predicts a merger rate for carbon–oxygen white dwarfs with super-Chandrasekhar total masses of $0.0013\,\mathrm{yr}^{-1}$ slightly lower than the predicted Galactic rate of SN Ia, which is equal to $\nu_{\mathrm{SN\,Ia}} = 0.004 \pm 0.002\ \mathrm{yr}^{-1}$ [73, 74] (based on the assumption that the Milky Way has a value of $\nu_{\mathrm{SN\,Ia}}$ equal to the average for galaxies with morphological type Sb–Sbc and a blue luminosity of $L_B = 2.3 \times 10^{10} L_{B\odot}$). The theoretical and observational estimates of ν_{SNIa} can be brought into better agreement by reducing α_{ce} in (6) [23]. Of course, it is quite possible that white-dwarf mergers are not the only mechanism for producing SNe Ia (see, e.g., the review [75]): the numerical relations between different evolutionary scenarios that can potentially lead to SNe Ia explosions remains a function of numerous population-synthesis parameters. However, our trial computations have shown that reducing α_{ce} shifts the maximum of the distribution of binary helium stars to orbital periods that are "too low" when compared to the observational values (Fig. 3). The situation will become clearer when enough data on close binary helium stars and white dwarfs have been accumulated.

4. CONCLUSION

We have modeled the population of low-mass helium stars in the Galaxy assuming that all are formed in close binary systems. We estimate their birth rate to be $0.043\,\mathrm{yr}^{-1}$, their total number to be 4×10^6, and the fraction that are binary to be 76%. These estimates are in a good agreement with our analytical estimates [17] and the estimates of Han *et al.* [22], which were also obtained via a population-synthesis analysis.

Low-mass helium stars are identified with observed sdB/sdO subdwarfs. Analysis of observational selection effects shows that the dominant factor in formation of the observed ensemble of sdB/sdO stars is the fact that the observed samples are limited to relatively bright magnitudes ($V \lesssim 16$). The number of helium stars with main-sequence companions accessible to observation is also limited by the fact that the overwhelming majority of subdwarfs are far brighter than their companions. All this reduces the binarity rate in the model observed sample to 58%, in agreement with observations. According to our computations, in the observed population of binary helium subdwarfs, the vast majority of the companions to helium stars are carbon–oxygen or helium white dwarfs.

In our model, single low-mass helium stars are predominantly produced by mergers of helium white dwarfs. Since mainly low-mass dwarfs merge, single helium subdwarfs have masses that are close to those of the components of binaries. It is possible that some fraction of merger products can be identified with sdO subdwarfs, which have a very low rate of binarity, according to the rather scarce existing data.

Our model assumes that all stars are formed in binaries, but, despite this, we are able to explain quite naturally the formation of helium stars as a consequence of evolution in *close* binaries. In addition, we can explain the binarity rate of low-mass helium stars. That being said, Saffer *et al.* [59] separate these stars into three groups based on observations of radial velocities of subdwarfs. Features associated with companions are absent from the spectra of stars in the first group, and the radial velocities of these stars are virtually constant. Features associated with companions are also absent from the spectra of stars in the second group, but radial-velocity variability indicating orbital periods of $\sim 1.0^{\mathrm{d}}$ is observed. Features of cool companions are present in the spectra of stars in the third group, with the radial velocities indicating orbital periods of months or years. The percentages of stars belonging to these three groups are 35, 45, and 20%, respectively. The following phenomenological interpretation has been proposed [59, 76]. The stars of the first group are intrinsically single objects that lost some of their mass not long before the ignition of helium in their cores (see the scenario suggested in [14]). The stars in the second group are binaries that have passed through a common-envelope stage and have invisible companions—M stars or white dwarfs. The third group contains stars that have passed through an interval of stable mass exchange

in the red-giant stage. If the model proposed by Saffer *et al.* [59, 76] obtains additional confirmation, based on the derived orbital-period distribution of the stars, estimated companion masses and selection effects that define the ratios of different groups of stars, our model for the population of low-mass helium stars and analysis of the selection effects leading to the observed sample of stars will require some revision.

ACKNOWLEDGMENTS

The authors acknowledge L. Morales-Rueda for providing a compilation of known orbital periods of subdwarfs and unpublished data. We are indebted to R. Napiwotzki and G. Nelemans for useful remarks and to E.M. Green for providing electronic files of papers otherwise unavailable in Russia. This study was partially supported by the Russian Foundation for Basic Research (project no. 03-02-16254), the Section of General Physics and Astronomy of the Russian Academy of Sciences (the Extended Objects in the Universe program), the Program of Support for Leading Scientific Schools of Russia (grant no. NSh-162.2003.2), and the program of the Presidium of the Russian Academy of Sciences "Nonstationary Processes in the Universe."

REFERENCES

1. G. Rauw, M. De Becker, Y. Nazé, *et al.*, Astron. Astrophys. **420**, L9 (2004).
2. G. Rauw, P. A. Crowther, M. de Becker, *et al.*, Astron. Astrophys. **432**, 985 (2005).
3. U. Heber, Astron. Astrophys. **155**, 33 (1986).
4. R. A. Saffer, P. Bergeron, D. Koester, and J. Liebert, Astrophys. J. **432**, 351 (1994).
5. H. Drechsel, U. Heber, R. Napiwotzki, *et al.*, Astron. Astrophys. **379**, 893 (2001).
6. P. Brassard, G. Fontaine, M. Billeres, *et al.*, Astrophys. J. **563**, 1013 (2001).
7. M. D. Reed, S. D. Kawaler, S. Zola, *et al.*, Mon. Not. R. Astron. Soc. **348**, 1164 (2004).
8. U. Heber, H. Edelmann, T. Lisker, and R. Napiwotzki, Astron. Astrophys. **411**, L477 (2003).
9. U. Heber, H. Drechsel, R. Østensen, *et al.*, Astron. Astrophys. **420**, 251 (2004).
10. T. Nugis and H. J. G. L. M. Lamers, Astron. Astrophys. **360**, 227 (2000).
11. Yu. F. Fadeev and M. F. Novikova, Pis'ma Astron. Zh. **29**, 592 (2003) [Astron. Lett. **29**, 592 (2003)].
12. P. S. Conti, in *IAU Symp. No. 83: Mass Loss and Evolution of O-Type Stars,* Ed. by C. de Loore and P. S. Conti (Reidel, Dordrecht, 1979), p. 431.
13. J. P. Cox and E. E. Salpeter, Astrophys. J. **133**, 764 (1961).
14. N. L. D'Cruz, B. Dorman, R. T. Rood, and R. W. O'Connell, Astrophys. J. **466**, 359 (1996).
15. R. Kippenhahn and A. Weigert, Z. Astrophys. **65**, 21 (1967).
16. B. Paczyński, Acta Astron. **17**, 355 (1967).
17. A. V. Tutukov and L. R. Yungel'son, Astron. Zh. **67**, 109 (1990) [Sov. Astron. **34**, 57 (1990)].
18. I. Iben, Jr., Astrophys. J. **353**, 215 (1990).
19. H. Saio and C. S. Jeffery, Mon. Not. R. Astron. Soc. **313**, 671 (2000).
20. I. Iben, Jr. and A. V. Tutukov, Astrophys. J., Suppl. Ser. **58**, 661 (1985).
21. Z. Han, P. Podsiadlowski, P. F. L. Maxted, *et al.*, Mon. Not. R. Astron. Soc. **336**, 449 (2002).
22. Z. Han, P. Podsiadlowski, P. F. L. Maxted, and T. R. Marsh, Mon. Not. R. Astron. Soc. **341**, 669 (2003).
23. A. V. Tutukov and L. R. Yungel'son, Astron. Zh. **79**, 738 (2002) [Astron. Rep. **46**, 667 (2002)].
24. R. F. Green, M. Schmidt, and J. Liebert, Astrophys. J., Suppl. Ser. **61**, 305 (1986).
25. A. V. Tutukov and L. R. Yungel'son, Nauch. Inform. Astron. Sovet Akad. Nauk SSSR **27**, 3 (1973).
26. Z. Han, C. A. Tout, and P. P. Eggleton, Mon. Not. R. Astron. Soc. **319**, 215 (2000).
27. D. Vanbeveren, Space Sci. Rev. **56**, 249 (1991).
28. G. Nelemans and E. P. J. van den Heuvel, Astron. Astrophys. **376**, 950 (2001).
29. U. Heber, P. F. L. Maxted, T. R. Marsh, *et al.*, in *Stellar Atmosphere Modeling,* Ed. by I. Hubeny, D. Mihalas, and K. Werner (Astron. Soc. Pac., San Francisco, 2003), Astron. Soc. Pac. Conf. Proc., Vol. 288, p. 251.
30. J. S. Vink, Astrophys. Space Sci. **291**, 239 (2004).
31. A. V. Tutukov, A. V. Fedorova, and L. R. Yungel'son, Pis'ma Astron. Zh. **8**, 365 (1982) [Sov. Astron. Lett. **8**, 198 (1982)].
32. G. Nelemans, F. Verbunt, L. R. Yungelson, and S. F. Portegies Zwart, Astron. Astrophys. **360**, 1011 (2000).
33. G. Nelemans and C. A. Tout, Mon. Not. R. Astron. Soc. **356**, 753 (2005).
34. A. V. Tutukov and L. R. Yungelson, in *IAU Symp. No. 83: Mass Loss and Evolution of O-Type Stars,* Ed. by C. de Loore and P. S. Conti (Reidel, Dordrecht, 1979), p. 401.
35. J. E. Pringle and R. F. Webbink, Mon. Not. R. Astron. Soc. **172**, 493 (1975).
36. A. V. Tutukov and L. R. Yungelson, Acta Astron. **29**, 665 (1979).
37. Z. E. Kraĭcheva, E. I. Popova, A. V. Tutukov, and L. R. Yungel'son, Pis'ma Astron. Zh. **7**, 488 (1981) [Sov. Astron. Lett. **7**, 269 (1981)].
38. E. I. Popova, A. V. Tutukov, and L. R. Yungelson, Astrophys. Space Sci. **88**, 55 (1982).
39. Z. E. Kraĭcheva, E. I. Popova, A. V. Tutukov, and L. R. Yungel'son, Astrofizika **22**, 105 (1985) [Astrophys. **21**, 63 (1985)].
40. S. Vereshchagin, A. Tutukov, L. Yungelson, *et al.*, Astrophys. Space Sci. **142**, 245 (1988).
41. Z. E. Kraĭcheva, E. I. Popova, A. V. Tutukov, and L. R. Yungel'son, Astrofizika **30**, 524 (1989).
42. P. F. L. Maxted, U. Heber, T. R. Marsh, and R. C. North, Mon. Not. R. Astron. Soc. **326**, 1391 (2001).

43. R. Napiwotzki, C. A. Karl, T. Lisker, *et al.*, Astrophys. Space Sci. **291**, 321 (2004).
44. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Astron. Astrophys. **310**, 489 (1996).
45. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Mon. Not. R. Astron. Soc. **288**, 245 (1997).
46. S. F. Portegies Zwart and L. R. Yungelson, Astron. Astrophys. **332**, 173 (1998).
47. D. Lommen, L. R. Yungelson, E. P. J. van den Heuvel, *et al.*, Astron. Astrophys. (2005) (in press).
48. R. A. Downes, Astrophys. J., Suppl. Ser. **61**, 569 (1986).
49. K. S. de Boer, M. Geffert, M. Odenkirchen, *et al.*, in *IAU Symp. No. 164: Stellar Populations,* Ed. by C. van der Kruit and G. Gilmore (Kluwer, Dordrecht, 1994), p. 393.
50. R. Napiwotzki, N. Christlieb, H. Drechsel, *et al.*, Astron. Nachr. **322**, 411 (2001).
51. R. Napiwotzki, N. Christlieb, H. Drechsel, *et al.*, Messenger **112**, 25 (2003).
52. R. Aznar Cuadrado and C. F. Jeffery, Astron. Astrophys. **385**, 131 (2002).
53. M. A. Stark and R. A. Wade, Astron. J. **126**, 1455 (2003).
54. M. Altmann, H. Edelmann, and K. S. de Boer, Astron. Astrophys. **414**, 181 (2004).
55. C. W. Allen, *Astrophysical Quantities* (Athlone, London, 1976; Mir, Moscow, 1977).
56. L. Morales-Rueda, P. F. L. Maxted, T. R. Marsh, *et al.*, Mon. Not. R. Astron. Soc. **338**, 752 (2003).
57. R. A. Wade, M. A. Stark, and R. F. Green, astro-ph/0410287 (2004).
58. T. Lisker, U. Heber, R. Napiwotzki, *et al.*, Astron. Astrophys. **430**, 223 (2005).
59. R. A. Saffer, E. M. Green, and T. Bowers, in *Proceedings of the 12th European Workshop on White Dwarfs,* Ed. by J. L. Provencal, H. L. Shipman, J. McDonald, and S. Goodchild (Astron. Soc. Pac., San Francisco, 2001); Astron. Soc. Pac. Conf. Ser. **226**, 408 (2001).
60. P. F. L. Maxted, T. R. Marsh, U. Heber, *et al.*, Mon. Not. R. Astron. Soc. **333**, 231 (2001).
61. A. Ahmad and C. S. Jeffery, Astron. Astrophys. **402**, 335 (2003).
62. A. V. Tutukov and L. R. Yungelson, Mon. Not. R. Astron. Soc. **280**, 1035 (1996).
63. G. Nelemans, S. F. Portegies Zwart, F. Verbunt, and L. R. Yungelson, Astron. Astrophys. **368**, 939 (2001).
64. G. Nelemans, L. R. Yungelson, and S. F. Portegies Zwart, Mon. Not. R. Astron. Soc. **349**, 181 (2004).
65. I. Iben, Jr., A. V. Tutukov, and L. R. Yungelson, Astrophys. J. **456**, 750 (1996).
66. R. Napiwotzki, L. Yungelson, G. Nelemans, *et al.*, in *Spectroscopically and Spatially Resolving the Components of the Close Binary Stars,* Ed. by W. Hilditch, H. Hensberge, and K. Pavlovski (Astron. Soc. Pac., San Francisco, 2004); Astron. Soc. Pac. Conf. Ser. **318**, 402 (2004).
67. L. Morales-Rueda, T. R. Marsh, P. F. L. Maxted, *et al.*, astro-ph/0502408 (2005).
68. A. V. Tutukov and L. R. Yungel'son, Nauch. Inform. Astron. Sovet Akad. Nauk SSSR **49**, 3 (1981).
69. I. Iben, Jr. and A. V. Tutukov, Astrophys. J., Suppl. Ser. **54**, 331 (1984).
70. R. F. Webbink, Astrophys. J. **277**, 355 (1984).
71. E. Ergma, A. V. Fedorova, and L. R. Yungelson, Astron. Astrophys. **376**, L9 (2001).
72. S.-C. Yoon and N. Langer, Astron. Astrophys. **419**, 645 (2004).
73. E. Cappellaro, R. Evans, and M. Turatto, Astron. Astrophys. **351**, 459 (1999).
74. E. Cappellaro, Mem. Soc. Astron. Ital. **72**, 863 (2001).
75. L. R. Yungelson, in *White Dwarfs: Galactic and Cosmological Probes,* Ed. by E. M. Sion, H. L. Shipman, and S. Vennes (Kluwer, Dordrecht, 2005); astro-ph/0409677.
76. E. M. Green, J. Liebert, and R. A. Saffer, in *Proceedings of the 12th European Workshop on White Dwarfs,* Ed. by J. L. Provencal, H. L. Shipman, J. McDonald, and S. Goodchild (Astron. Soc. Pac., San Francisco, 2001); Astron. Soc. Pac. Conf. Ser. **226**, 192 (2001).

Translated by L. Yungel'son

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 884–891. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 990–998.

Three-Dimensional Gas-Dynamical Modeling of Changes in the Flow Structure during the Transition from the Quiescent to the Active State in Symbiotic Stars

M. Mitsumoto[1], B. Jahanara[1], T. Matsuda[1], K. Oka[2], D. V. Bisikalo[3], E. Yu. Kilpio[3], H. M. J. Boffin[4], A. A. Boyarchuk[3], and O. A. Kuznetsov[3,5]

[1]*Department of Earth and Planetary Sciences, Kobe University, Kobe 657-8501, Japan*

[2]*Misuho Research Institute, Tokyo 101-8443, Japan*

[3]*Institute of Astronomy, Moscow, Russia*

[4]*European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85738 Garching, Germany*

[5]*Keldysh Institute for Applied Mathematics, Moscow, Russia*

Received May 1, 2005; in final form, May 18, 2005

Abstract—The results of three-dimensional modeling of the flow structure in the classical symbiotic system Z Andromedae are presented. Outbursts in systems of this type occur when the accretion rate exceeds the upper limit of the steady-burning range. Therefore, in order to realize the transition from a quiescent to an active state, it is necessary to find a mechanism capable of sufficiently increasing the accretion rate on the time scales typical for outburst development. Our calculations provide support for a mechanism for the transition from quiescence to outburst in classical symbiotic systems suggested earlier based on two-dimensional calculations (Bisikalo *et al.*, 2002). Our results show that an accretion disk forms in the system for a wind velocity of 20 km s^{-1}. The accretion rate for the solution with the disk is $\sim$22.5–25% of the mass-loss rate of the donor, which is $\sim 4.5-5 \times 10^{-8}\, M_\odot$ yr^{-1} for Z And. This value is in agreement with the steady-burning range for the white-dwarf masses usually accepted for this system. When the wind velocity increases from 20 to 30 km s^{-1}, the accretion disk is destroyed and the disk material falls onto the accretor surface. This process is followed by an approximately twofold jump in the accretion rate. The resulting growth in the accretion rate is sufficient so as to exceed the upper limit of the steady-burning range, thus bringing the system into an active state. The time during which the accretion rate is above the steady-burning value is in very good agreement with observations. Our analysis leads us to conclude that small variations in the donor wind velocity can lead to the transition from disk accretion to wind accretion and, as a consequence, to the transition from a quiescent to an active state in classical symbiotic stars. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Symbiotic stars are characterized by peculiar spectra in which molecular absorption bands—features characteristic of cool giants—are present, together with emission lines corresponding to high excitation levels. The red and infrared spectra of symbiotic stars are typical of cool giants, while they are characterized by a very hot continuum in the UV region. It is widely assumed that most symbiotic stars are detached binaries consisting of a cool giant and a white dwarf surrounded by a nebulosity [1]. The mass exchange in these systems is driven by the stellar wind.

The goal of this paper is to examine classical symbiotic systems using of the best-studied representatives of this class as an example—Z And. According to analyses of its spectral energy distribution over a wide range [2], the components of Z And have the following characteristics. The cool M3.5III giant has a mass of $\sim 2\, M_\odot$ and a radius of $\sim 100\, R_\odot$, while the white dwarf has a mass of $\sim 0.6\, M_\odot$, a radius of $\sim 0.07\, R_\odot$, and a temperature of $\sim 10^5$ K. The giant loses mass at a rate of $\sim 2 \times 10^{-7}\, M_\odot$ yr^{-1}. The gas in the circum-binary envelope has an electron density of $\sim 2 \times 10^{10}$ cm^{-3} and a temperature of $\sim 1.5-8 \times 10^4$ K. The separation of the system is $482\, R_\odot$, and the orbital period is 758 days. More than 100 years of available observations demonstrate that this star displays outbursts. The last of these took place in 2000, and has been actively observed at various wavelengths.

Based on the characteristics of their outbursts, symbiotic stars can be divided into two types [3, 4]. The first includes so-called "classical symbiotics"

1063-7729/05/4911-0884$26.00

(e.g., Z And, AG Peg, AG Dra, CI Cyg, AX Per), whose outbursts typically last a few months and have amplitudes of $2-4^m$. A typical feature of such outbursts is the weakening of high excitation lines as brightness increases. A detailed description of the behavior of these stars is given in [5, 6]. Symbiotic novae (e.g., V1016 Cyg, V1329 Cyg, RS Oph, HM Sge) show the second type of outburst, which are longer and brighter; their energetics can even exceed those of novae outbursts. The principal difference of these outbursts from those of the first type is the increase in ionization degree as brightness increases. The behavior of these stars during outburst is described in [7–10]. Outbursts of the second type are similar to slow classical novae outbursts [11, 12] (for differences between symbiotic novae and slow classical novae outbursts see [13]).

The most probable mechanism describing the observational manifestations of classical symbiotics and symbiotic novae is thermonuclear burning on the accretor's surface (see, e.g., [6, 13]). It has been shown [3, 14–17] that the process of thermonuclear burning of hydrogen on a white-dwarf surface depends strongly on the accretion rate $\dot{M}^{\rm accr}$ [3]. Only, in a narrow range of $\dot{M}^{\rm accr}$ is steady hydrogen burning possible [15, 18, 19]. The lower limit of this range is given by the expression [16, 20]

$$\dot{M}_{\rm steady,min} = 1.3\times10^{-7}\left(\frac{M}{M_\odot}\right)^{3.57} M_\odot\ {\rm yr}^{-1},$$

while the higher limit is given by [21, 22]

$$\dot{M}_{\rm steady,max} = 6\times10^{-7}\left(\frac{M}{M_\odot}-0.522\right) M_\odot\ {\rm yr}^{-1}.$$

For Z And, with $M = 0.6\,M_\odot$, steady burning is possible in the range $2.1\times10^{-8}\,M_\odot\ {\rm yr}^{-1} \lesssim \dot{M}^{\rm accr} \lesssim 4.7\times10^{-8}\,M_\odot\ {\rm yr}^{-1}$. If the accretion rate is below this interval, thermonuclear burning depletes matter faster than accretion can replenish it, bringing the burning to a halt. Following this, hydrogen begins to accumulate on the white-dwarf surface until the pressure near the base of the hydrogen envelope reaches a critical value and a hydrogen-shell flash occurs. The bolometric luminosity increases by a factor of 10–100 over ~1 year, and the star remains in an active state for ~10 years. The total duration of the outburst is tens years, which corresponds to the nuclear time scale. This is observed in symbiotic novae.

If for some reason $\dot{M}^{\rm accr}$ becomes larger than $\dot{M}_{\rm steady,max}$, the accreted matter accumulates above the burning shell and expands to giant dimensions. This behavior is typical for classical symbiotic stars. Though the bolometric luminosity stays constant, the visual brightness increases by one to two magnitudes and the effective temperature decreases. An optical outburst develops on the thermal timescale. In accordance with existing models for thermonuclear burning on white-dwarf surfaces [15, 20], small variations in the accretion rate can lead to significant changes in temperature with practically constant bolometric luminosity. For a white dwarf with a mass of $1.2\,M_\odot$, a change in the envelope mass of $4\times10^{-7}\,M_\odot$ can result in a 3^m brightness increase [3]. This value corresponds to the typical amplitude of the outbursts in Z And.

Z And belongs to the classical symbiotics and, in accordance with the accepted model, its accretion rate in the quiescent state is in the steady-burning range. For an outburst to develop, the accretion rate must become large enough to reach a value above the steady-burning range. Note that the time scale for the increase in accretion rate should correspond to the characteristic time for the development of the outburst (~100 days). Such behavior of the accretion rate in a binary system is possible in the framework of the mechanism proposed in [23] based on the results of two-dimensional (2D) gas-dynamical modeling. The main idea of this mechanism is that even small variations in donor-wind velocity can lead to a change in the accretion regime from disk to wind accretion. In the transition period, when the disk is being destroyed, the accretion rate increases abruptly and exceeds the upper limit of the steady-burning range. The goal of our study was to attempt to realize computations of such a mechanism using more a realistic three-dimensional (3D) model.

2. THE MODEL

We carried out 3D numerical simulations in order to study the gas-flow structure in the symbiotic system Z And. The zero point of the coordinate system was placed at the center of the accretor, and the x-axis was directed along the line connecting the centers of the components away from the mass-losing star, the y-axis along the accretor's orbital motion, and the z-axis along the axis of rotation of the binary system. The flow was described by a system of Euler equations in the corotating coordinate frame:

$$\frac{\partial\rho}{\partial t}+\frac{\partial\rho u}{\partial x}+\frac{\partial\rho v}{\partial y}+\frac{\partial\rho w}{\partial z}=0,$$

$$\frac{\partial\rho u}{\partial t}+\frac{\partial(\rho u^2+P)}{\partial x}+\frac{\partial\rho uv}{\partial y}+\frac{\partial\rho uw}{\partial z}$$
$$=-\rho\frac{\partial\Phi}{\partial x}+2\Omega v\rho,$$

$$\frac{\partial \rho v}{\partial t} + \frac{\partial \rho uv}{\partial x} + \frac{\partial (\rho v^2 + P)}{\partial y} + \frac{\partial \rho vw}{\partial z} = -\rho \frac{\partial \Phi}{\partial y} - 2\Omega u \rho,$$

$$\frac{\partial \rho w}{\partial t} + \frac{\partial \rho uw}{\partial x} + \frac{\partial \rho vw}{\partial y} + \frac{\partial (\rho w^2 + P)}{\partial z} = -\rho \frac{\partial \Phi}{\partial z},$$

$$\frac{\partial \rho E}{\partial t} + \frac{\partial \rho uh}{\partial x} + \frac{\partial \rho vh}{\partial y} + \frac{\partial \rho wh}{\partial z} = -\rho u \frac{\partial \Phi}{\partial x} - \rho v \frac{\partial \Phi}{\partial y} - \rho w \frac{\partial \Phi}{\partial z}.$$

Here, $\mathbf{u} = (u, v, w)$ is the velocity vector, P the pressure, ρ the density, $h = \varepsilon + P/\rho + |\mathbf{u}|^2/2$ the specific total enthalpy, $E = \varepsilon + |\mathbf{u}|^2/2$ the specific total energy, ε the specific internal energy, Ω the angular velocity of binary's rotation, and $\Phi(\mathbf{r})$ the force potential.

In the standard definition, when only the gravitational forces from point-mass components and centrifugal force are taken into account, the force potential is given by

$$\Phi(\mathbf{r}) = -\frac{GM_1}{|\mathbf{r} - \mathbf{r}_1|} - \frac{GM_2}{|\mathbf{r} - \mathbf{r}_2|} - \frac{1}{2}\Omega^2 (\mathbf{r} - \mathbf{r}_c)^2.$$

Here, M_1 is the mass of the accretor, M_2 the mass of the donor, $\mathbf{r}_1$ and $\mathbf{r}_2$ the radius vectors of the component centers, and $\mathbf{r}_c$ the radius vector of the center of mass of the system. This is the so-called "Roche potential." However, in our case, the additional force responsible for accelerating the donor wind must also be taken into account, thereby changing the form of the potential.

Previous studies (e.g., [24–26]) have shown that the general flow structure in a system whose components do not fill their Roche lobes is defined primarily by stellar-wind parameters. Since the gas-acceleration mechanism is poorly known for cool giants, we imitate the radiation pressure by reducing the gravitational attraction force of the donor by a factor of $(1 - \Gamma)$.

Thus, the modified force potential becomes

$$\Phi(\mathbf{r}) = -(1 - \Gamma)\frac{GM_2}{|\mathbf{r} - \mathbf{r}_2|} - \frac{GM_1}{|\mathbf{r} - \mathbf{r}_1|} - \frac{1}{2}\Omega^2 (\mathbf{r} - \mathbf{r}_c)^2.$$

We closed the system using the perfect gas equation of state:

$$P = (\gamma - 1)\rho\varepsilon.$$

The adiabatic index was taken to be $\gamma = 1.01$, which corresponds to a value close to the isothermal case [27–29].

We adopted the parameters of Z And for the binary system. In our numerical simulations, all the variables were put in dimensionless form by normalizing the length to the separation of the two stars, A, and the time to Ω^{-1}, so that velocities were normalized by $A\Omega$. The density was normalized to the value at the surface of the mass-losing star.

The method of computational fluid dynamics is described in [30] and details are explained in [31, 32]. We use a simplified flux-vector splitting (SFS) finite volume scheme to discretize the Euler equation (the description of the SFS is given in the Appendix of [31]).

We assumed symmetry about the orbital plane, and therefore only computed the upper half-space of the computational domain, which is $-2 < x < 1$, $-1.5 < y < 1.5$, $0 < z < 1.0$. The region is divided into $253 \times 253 \times 85$ cells. The inner numerical boundary surrounding the mass-losing star is represented by an equipotential surface with mean radius R_{2n} (we set the numerical boundary at a slightly larger radius, $R_{2n} = 101\,R_\odot$, rather than R_2). The radius of the accreting star is too small to fit in our numerical grid, and so is represented by one cell.

The cells just inside the inner numerical boundary around the mass-losing star are assumed to be filled by gas with dimensionless density $\rho_2 = 1$, dimensionless sound speed $c_s = 0.178$ (corresponding to $T_2 = 3200$ K), and a dimensionless speed normal to the surface V. The real normal velocity of the gas at the numerical boundary is determined by solving the Riemann problem between the cells adjacent to the inner boundary surface. The cell representing the accreting star and just outside the outer numerical boundary are filled by gas with density $\rho_0 = 10^{-9}$, pressure $p_0 = 10^{-8}/\gamma$, and three velocity components $u_0 = v_0 = w_0 = 0$. This assumption does not mean that accretion onto the accreting star or escape from the computational domain do not occur. The velocity of the gas at the boundaries is computed by solving Riemann problems. At the initial stage, $t = 0$, apart from the region inside the inner boundaries, the computational domain is assumed to be filled by gas with $\rho_0 = 10^{-9}$, $p_0 = 10^{-8}/\gamma$, and $u_0 = v_0 = w_0 = 0$. This gas is gradually replaced by gas supplied from the mass-losing star. We computed up to a few rotational periods, generally long enough for the system to reach a quasi-steady state.

3. RESULTS

The velocity of the donor wind in Z And is $\sim$25 km s^{-1} [2]. It is supposed in the considered mechanism that, an accretion disk will form in the case of slightly lower velocities, while the disk will be

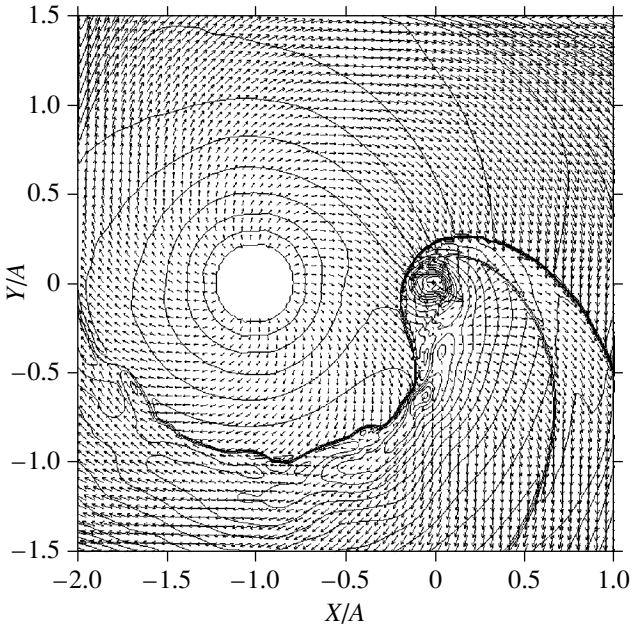

Fig. 1. Density contours and velocity vectors in the equatorial plane of the system for the case $V_w = 20$ km s^{-1} and $\Gamma = 0.94$. The empty circle centered at $(-1, 0)$ corresponds to the donor (radius of the circle is equal to the donor's radius), and the diamond at $(0, 0)$ marks the accretor.

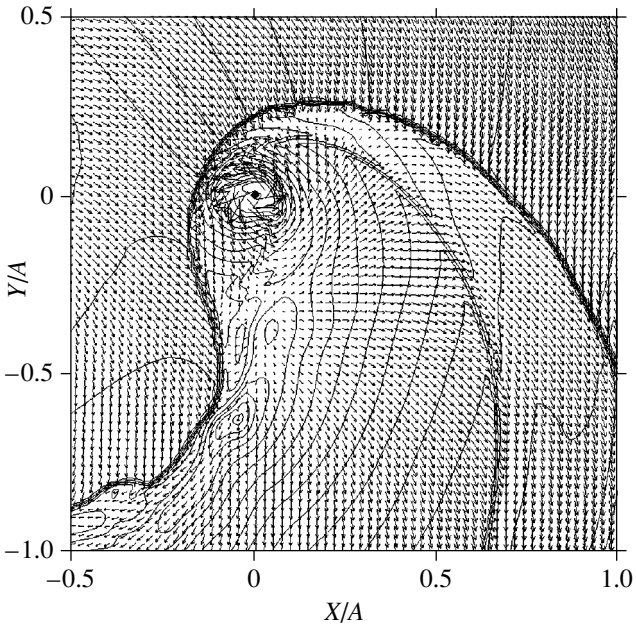

Fig. 2. Same as Fig. 1 in the vicinity of the accretor.

destroyed and a wind-accretion type of flow will occur for slightly higher values. To check the applicability of the proposed mechanism, numerical simulations for velocities of 25 ± 5 km s^{-1} were carried out.[1)] Calculations for a wind velocity of $V_w = 20$ km s^{-1} were necessary to study the possible formation of a disk in the system, to estimate the accretion rate, and to check if the accretion rate was indeed within the limits of the steady-burning range. Calculations for $V_w = 30$ km s^{-1} were conducted in order to obtain the parameters of the flow structure in the system without a disk. To answer questions concerning processes taking place after the change in wind velocity—such as whether the disk will be destroyed, how the accretion rate will grow, whether the accretion rate will exceed the upper limit of the steady-burning range—calculations with the wind-velocity jump $20 \to 30$ km s^{-1} were carried out.

The results of our numerical modeling showed that, for a wind velocity of $V_w = 20$ km s^{-1} and values of the parameter $\Gamma \in [0.85; 0.95]$, a steady accretion disk forms in the system. Figures 1 and 2 show the flow structure for the case $V_w = 20$ km s^{-1} and $\Gamma = 0.94$. Density contours and velocity vectors in the equatorial plane of the system are presented for the entire computational domain (Fig. 1) and for the area near the accretor $[-0.5\,A \ldots A] \times [-A \ldots 0.5\,A]$ (Fig. 2). The results of the calculations are presented at the time $t \sim 5\,P_{orb}$, when a steady regime is already established in the system. These results show that, with this wind velocity, a bow shock and an accretion disk with radius approximately 50–$60\,R_\odot$ form in the system. Note that two spiral density waves are seen in the accretion disk. In this solution, the accretion rate is $\sim$22.5–25% of the matter lost by the donor. Similar values of the accretion efficiency for a solution with an accretion disk were obtained by Mastrodemos and Morris [33] for detached binaries. For Z And, whose mass-loss rate is estimated to be $\sim 2 \times 10^{-7}\,M_\odot$ yr^{-1}, the accretion rate will be ~ 4.5–$5 \times 10^{-8}\,M_\odot$ yr^{-1}, corresponding to the steady-burning range (near the upper limit).

As we can see, due to the influence of the pressure gradient and the presence of spiral shocks, the disk structure is far from a conventional Keplerian disk. Since we used the full set of Euler equations, including an advection term in the energy equation, the obtained flow structure is more similar to advection-dominated disks, as discussed by Walder and Folini [34]. Of course, the system of Euler equations with the adiabatic energy equation does not incorporate viscous heating or radiative cooling, but the numerical viscosity and the choice of an adiabatic index γ close to unity do provide an implicit account of these processes. Thus, we believe that our

[1)] Small variations of the wind velocity within ± 5 km s^{-1} can be easily explained by the activity of the giant [4].

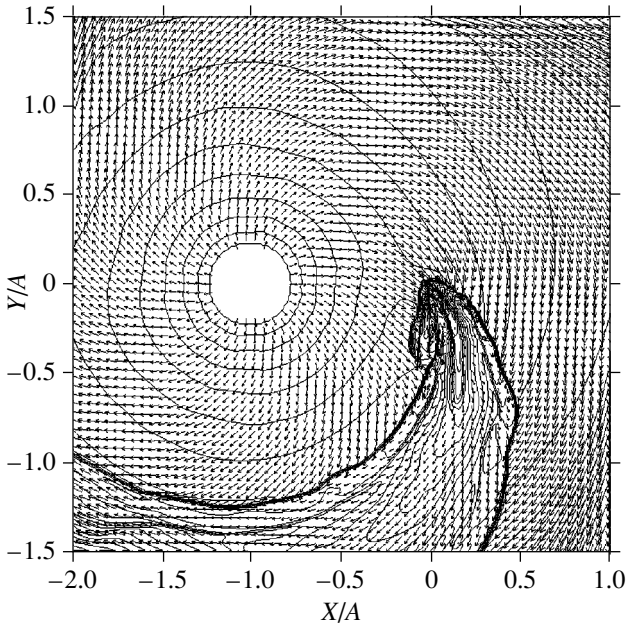

Fig. 3. Density contours and velocity vectors in the equatorial plane of the system for the case $V_w = 30$ km s^{-1} and $\Gamma = 0.94$. All notate is the same as in Fig. 1.

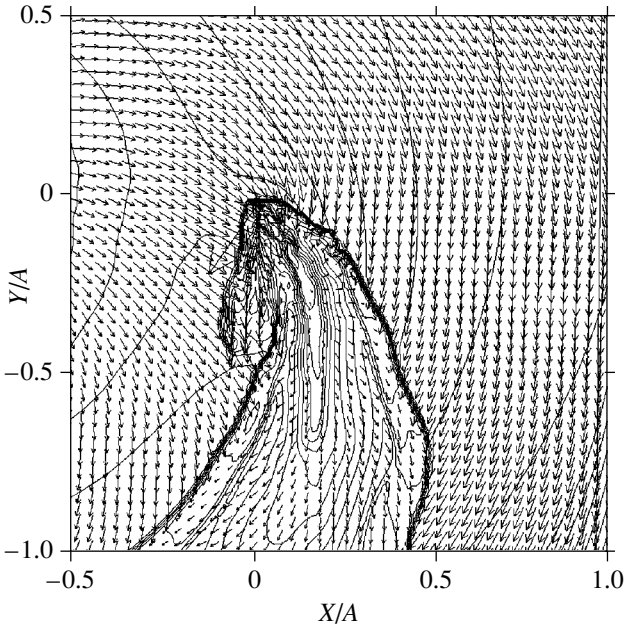

Fig. 4. Same as Fig. 3 in the vicinity of the accretor.

mathematical model adequately describes the physics of accretion disks.

The solution with wind velocity $V_w = 30$ km s^{-1} is presented in Figs. 3 and 4. Just as in Fig. 1, density contours and velocity vectors are presented for the entire computational domain (Fig. 3) and for the area near the accretor $[-0.5\,A \ldots A] \times [-A \ldots 0.5\,A]$ (Fig. 4). The situation presented corresponds to a time when the steady regime has already been established. Analysis of these results shows that, when the wind velocity equals 30 km s^{-1}, the conical shock is close to the accretor, and does not leave any room for the formation of a disk. In this solution, a wind-accretion type of flow rather than disk accretion occurs, with the accretion rate being $\sim$11–13% of the matter leaving the donor surface. Similar flow structures were obtained in 3D simulations performed by Dumm *et al.* [35]. According to their results, the accretion efficiency was 6% in the case of wind accretion.

Analysis of the results presented above leads us to conclude that even a small change in the donor-wind velocity (within ± 5 km s^{-1} about the observed value, 25 km s^{-1}) leads to changes in the flow structure and accretion regime; namely, to a transition from disk- to wind-accretion flow. To consider the transition between these two regimes after the wind-velocity increase, we took the stationary solution for the case the velocity of 20 km s^{-1} after $\sim 5\,P_{orb}$ from the beginning of the calculations (the situation presented in Figs. 1 and 2) then raised the velocity to 30 km s^{-1}. After the increase in wind velocity at the inner boundary, the flow structure changes, evolving from a state with an accretion disk to the state with a conical shock (presented in Figs. 3 and 4). Let us consider the flow restructuring: at time $0.17\,P_{orb}$ ($\sim$133 days) after the wind-velocity change, matter moving with the increased velocity reaches the vicinity of the accretor, namely, the bow shock located in front of it. The wind then continues to move further, crushing the accretion disk and making the matter of the disk fall onto the accretor surface. A snapshot of this flow rearrangement is presented in Fig. 5. The area size and all the notations are the same as in Figs. 2 and 4. This situation corresponds to $\sim$180 days after the wind-velocity increase. Note that studies of the transition period are limited in the framework of this model. After the accretion-rate jump, the adopted boundary conditions at the accretor change, and the model no longer describes the real physical situation. Accordingly, the presented calculations of the flow-rearrangement period are correct only in the first stages.

The behavior of the accretion rate with time is shown in Figs. 6 and 7. The time that has passed from the beginning of the calculations in units of $t = P_{orb}/2\pi$ is shown on the horizontal axis. The accretion rate is given in dimensionless units.[2)] Starting

[2)] In these units the donor mass-loss rate is equal to 0.355 for the solution with a disk and 0.535 for the solution with a conical shock.

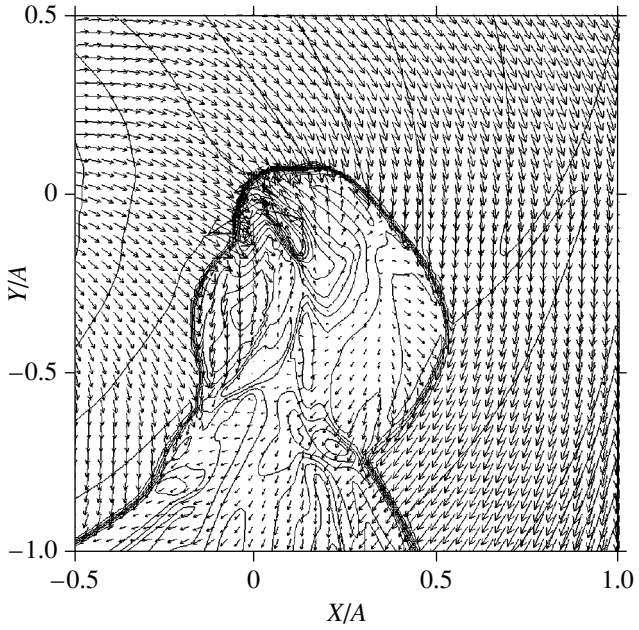

Fig. 5. Density contours and velocity vectors in the area near the accretor ∼180 days after the wind-velocity increase from 20 to 30 km s^{-1}.

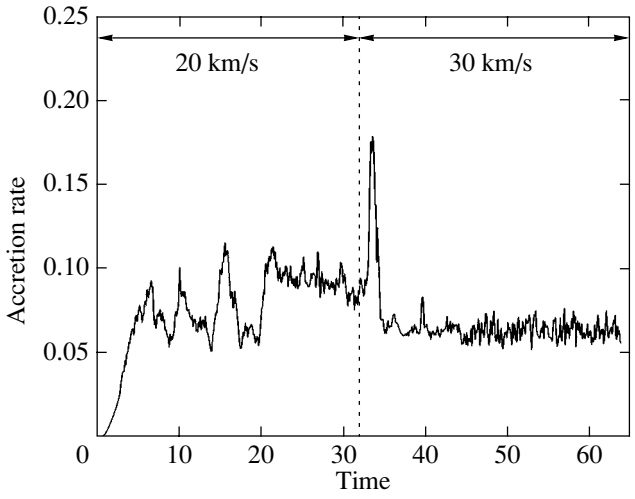

Fig. 6. Accretion-rate change for the solution where the wind velocity is increased from 20 to 30 km s^{-1}. The vertical line marks the instant of wind-velocity change.

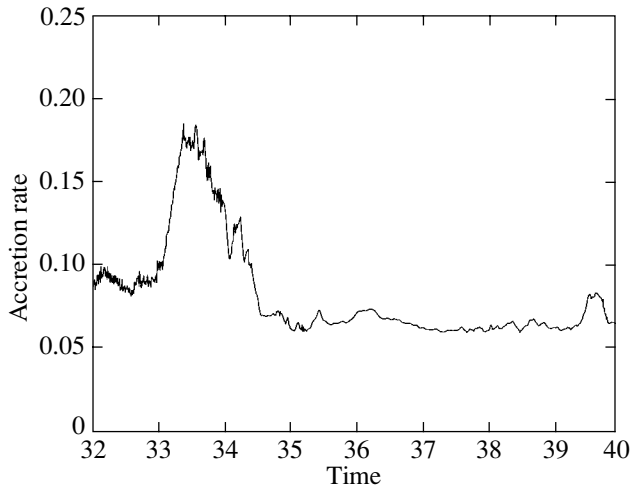

Fig. 7. Same as Fig. 6 for a short time interval beginning from the moment of wind-velocity increase from 20 to 30 km s^{-1}.

from the time when the matter with increased velocity reaches the vicinity of the accretor, the accretion rate begins to grow, and reaches its maximum after approximately $0.06\,P_{\rm orb}$ (∼47 days). The maximum accretion rate is a factor of 2–2.2 higher than in the case of disk accretion.

As is shown above, for an accretor with mass $0.6\,M_\odot$, the ratio of the upper and lower limits of the steady-burning range is $\dot{M}_{\rm steady,min}/\dot{M}_{\rm steady,max} \approx 2.2$. For a white dwarf with mass $0.55\,M_\odot$ (the value given in [4]) an increase of only 10% should be enough to exceed the upper limit. Thus, the approximately twofold accretion-rate growth obtained in our calculations is large enough to change the system from a quiescent to active state.

Analysis of the data presented in Figs. 6 and 7 shows that the time for the full destruction of the disk is ∼180 days. The time during which the accretion rate exceeds the upper limit of the steady-burning range should correspond to the time scale for the development of the outburst. If we suppose that the limit of the steady-burning interval is exceeded when the accretion rate increases by a factor of 1.5, the time during which accretion rate is above this value will be approximately 100 days, and $\sim 2 \times 10^{-8}\,M_\odot$ will be accreted during this time.

4. CONCLUSIONS

Outbursts in classical symbiotic stars occur when the accretion rate becomes greater than the upper limit of the steady-burning range. A transition from the quiescent to active state requires an increase in the accretion rate over a fairly long time interval, corresponding to the characteristic time required for development of an outburst (∼100 days). Earlier [23], a mechanism providing the required accretion-rate increase in such a system was proposed based on 2D calculations. In this mechanism, even small changes in the donor wind velocity are sufficient

to bring about a change in the accretion regime. During the transition from disk to wind accretion, the accretion disk is destroyed and the wind with increased velocity causes the disk matter to fall onto the accretor surface. Analysis of the 2D calculations showed that, during this process, the growth of the accretion rate was enough for the accretion rate to exceed the limits of steady burning, leading to the development of an outburst.

To elucidate whether this mechanism can operate in observed astrophysical objects, we carried out gasdynamical modeling of the flow structure in the classical symbiotic system Z And using a more realistic 3D model. The obtained results lead to the following conclusions.

(1) With a donor-wind velocity of 20 km s^{-1} and $\Gamma = 0.85-0.95$, an accretion disk forms in the system. The accretion rate for the solution with a disk is $\sim 22.5-25\%$ of the matter lost by the donor.

(2) If the wind velocity is 30 km s^{-1}, the accretion disk disappears and a conical shock forms. The accretion rate for this case is $\sim 11-13\%$ of the matter lost by the donor.

(3) These two solutions show that even a small change in the donor-wind velocity (± 5 km s^{-1} from the observed value 25 km s^{-1} [2]) leads to changes in the flow structure and the accretion regime; namely, to a transition from disk accretion to wind accretion.

(4) In the proposed mechanism, the solution with an accretion disk should correspond to the quiescent state of the system. According to our results, the accretion rate in this case is $\sim 22.5-25\%$. Thus, for Z And, whose mass-loss rate is estimated to be $\sim 2 \times 10^{-7}\, M_\odot$ yr^{-1}, the accretion rate will be $\sim 4.5-5 \times 10^{-8}\, M_\odot$ yr^{-1}, in good agreement with the steady-burning interval for the most probable for Z And white-dwarf masses.

(5) In our solution where the wind velocity increases from 20 to 30 km s^{-1}, the accretion disk is destroyed and the matter it contained falls onto the accretor surface.

(6) A detailed study of the accretion-regime change after the wind velocity increase indicates that the destruction of the disk is followed by a jump in the accretion rate. The maximum accretion rate is a factor of $\sim 2-2.2$ higher than the initial value. As was noted above, the steady-burning range is fairly narrow, and, if the accretor mass is $0.6\, M_\odot$, an increase of the accretion rate by a factor of 2.2 is, clearly, enough for this rate to exceed the upper limit of this range. The results of our calculations yield an accretion-rate increase that is large enough to put the system outside the steady-burning range and move it into the active state.

(7) The time during which the accretion rate is above the steady-burning range should correspond to the outburst development time, which is approximately 100 days for Z And. According to our results, the time for the complete destruction of the disk is approximately ~ 180 days. If we assume that the system leaves the steady-burning range after the accretion rate has increased by a factor of 1.5, the corresponding time will be ~ 100 days, in good agreement with observations.

(8) The typical amplitude of the outbursts in Z And is $\sim 3^m$. For a white dwarf with mass $1.2\, M_\odot$, such a brightness increase is provided by an envelope mass of $\sim 4 \times 10^{-7}\, M_\odot$ [3]. The envelope mass in our calculations is smaller, $\sim 10^{-8}\, M_\odot$.[3)] It is obvious that a correct comparison of the envelope masses, calculations of thermonuclear burning on the surface of a white dwarf with a mass equal to that of the accretor in Z And are required. Moreover, it is necessary to take into account the asymmetry of the accretion. Unfortunately, we are not aware of any studies where such estimations were made, and the question of the envelope mass remains open. Given an accretion rate corresponding to the upper limit of the steady-burning range and a typical time for an outburst of $\sim 1/3$ year, the envelope mass should not significantly exceed $\sim 10^{-8}\, M_\odot$, in good agreement with our results.

Analysis of the results presented here leads us to our main conclusion that the transition from the quiescent to the active state in symbiotic stars may be associated with a change in accretion regime (a transition from disk to wind accretion), as a result of modest variations in the donor wind velocity.

ACKNOWLEDGMENTS

T.M., D.V.B., and H.B. acknowledge financial support from the 21st Century COE Program of the Origin and Evolution of Planetary Systems of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan. T.M. was supported by a scientific research grant from the Japan Society for the Promotion of Science (no. 13640241). The calculations were performed on the ISAS/JAXA NEC SX-6 supercomputer in Japan. This work was also supported by the Russian Foundation for Basic Research (project nos. 03-02-16622, 03-01-00311, 05-02-16123, 05-02-17070, and 05-02-17874), the Program of Support for Leading Scientific Schools of Russia (project no. NSh-162.2003.2), and the

[3)] Note that, in the solution obtained, the mass of the accreted envelope depends on the model parameters (wind velocity, disk temperature, density etc.) and is, therefore, not a fixed value.

programs of the Presidium of the Russian Academy of Sciences "Mathematical Modeling and Intellectual Systems" and "Nonstationary Phenomena in Astronomy."

REFERENCES

1. A. A. Boyarchuk, Izv. Krymsk. Astrofiz. Obs. **38**, 155 (1967).
2. T. Fernandez-Castro, A. Cassatella, A. Gimenez, *et al.*, Astrophys. J. **324**, 1016 (1988).
3. B. Paczyński and B. Rudak, Astron. Astrophys. **82**, 349 (1980).
4. J. Mikołajewska and S. J. Kenyon, Mon. Not. R. Astron. Soc. **256**, 177 (1992).
5. A. A. Boyarchuk, Izv. Krymsk. Astrofiz. Obs. **39**, 124 (1969).
6. S. J. Kenyon, *The Symbiotic Stars* (Cambridge Univ. Press, Cambridge, 1986).
7. G. B. Baratta, A. Cassatella, and R. Viotti, Astrophys. J. **187**, 651 (1974).
8. A. Mammano and F. Ciatti, Astron. Astrophys. **39**, 405 (1975).
9. F. Ciatti, A. Mammano, and A. Vittone, Astron. Astrophys. **68**, 251 (1978).
10. R. C. Puetter, R. W. Russell, B. T. Soifer, and S. P. Willner, Astrophys. J. **223**, L93 (1978).
11. D. A. Allen, Mon. Not. R. Astron. Soc. **192**, 521 (1980).
12. S. J. Kenyon and J. W. Truran, Astrophys. J. **273**, 280 (1983).
13. I. Iben, Jr. and A. V. Tutukov, Astrophys. J., Suppl. Ser. **105**, 145 (1996).
14. A. V. Tutukov and L. R. Yungel'son, Astrofizika **12**, 521 (1976).
15. B. Paczyński and A. Żytkow, Astrophys. J. **222**, 604 (1978).
16. I. Iben, Jr., Astrophys. J. **259**, 244 (1982).
17. I. Iben, Jr. and A. V. Tutukov, Astrophys. J. **342**, 430 (1989).
18. E. M. Sion, M. J. Acierno, and S. Tomczyk, Astrophys. J. **230**, 832 (1979).
19. M. Y. Fujimoto, Astrophys. J. **257**, 767 (1982).
20. R. Sienkiewicz and W. Dziembowski, in *IAU Colloq. No. 42: The Interaction of Variable Stars with Their Environment,* Ed. by R. Kippenhahn, J. Rahe, and W. Strohmeier (Remeis-Sternwarte, Bamberg, 1977), p. 327.
21. B. Paczyński, Acta Astron. **20**, 47 (1970).
22. U. Üüs, Nauch. Inform. Astron. Sovet Akad. Nauk SSSR **17**, 32 (1970).
23. D. V. Bisikalo, A. A. Boyarchuk, E. Yu. Kil'pio, and O. A. Kuznetsov, Astron. Zh. **79**, 1131 (2002) [Astron. Rep. **46**, 1022 (2002)].
24. G. S. Bisnovatyi-Kogan, Ya. M. Kazhdan, A. A. Klypin, and A. E. Lutskii, Astron. Zh. **56**, 359 (1979) [Sov. Astron. **23**, 201 (1979)].
25. D. V. Bisikalo, A. A. Boyarchuk, O. A. Kuznetsov, *et al.*, Astron. Zh. **71**, 560 (1994) [Astron. Rep. **38**, 494 (1994)].
26. D. V. Bisikalo, A. A. Boyarchuk, O. A. Kuznetsov, and V. M. Chechetkin, Astron. Zh. **73**, 727 (1996) [Astron. Rep. **40**, 662 (1996)].
27. K. Sawada, T. Matsuda, and Hachisu, Mon. Not. R. Astron. Soc. **219**, 75 (1986).
28. D. Molteni, G. Belvedere, and G. Lanzafame, Mon. Not. R. Astron. Soc. **249**, 748 (1991).
29. D. V. Bisikalo, A. A. Boyarchuk, O. A. Kuznetsov, *et al.*, Astron. Zh. **72**, 367 (1995) [Astron. Rep. **39**, 325 (1995)].
30. T. Nagae, K. Oka, T. Matsuda, *et al.*, Astron. Astrophys. **419**, 335 (2004).
31. M. Makita, K. Miyawaki, and T. Matsuda, Mon. Not. R. Astron. Soc. **316**, 906 (2000).
32. H. Fujiwara, M. Makita, and T. Matsuda, Prog. Theor. Phys. **106**, 729 (2001).
33. N. Mastrodemos and M. Morris, Astrophys. J. **497**, 303 (1998).
34. R. Walder and D. Folini, in *Thermal and Ionization Aspects of Flows from Hot Stars*, Ed. by H. Lamers and A. Sapar; Astron. Soc. Pac. Conf. Ser. **204**, 331 (2000).
35. T. Dumm, D. Folini, H. Nussbaumer, *et al.*, Astron. Astrophys. **354**, 1014 (2000).

Translated by D. Gabuzda

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 892–904. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 999–1012.

Hydrodynamical Modeling of Circularization in Close Binary Systems in Early Stages of Their Evolution on the Dynamical Time Scale

V. G. Karetnikov and F. V. Sirotkin

Astronomical Observatory, National University, Shevchenko Park, Odessa, 270014 Ukraine

Received November 25, 2004; in final form, May 18, 2005

Abstract—We have carried out three-dimensional hydrodynamical modeling of the process of circularization in close binary systems with fully convective components that are close to filling their Roche lobes. The computations are based on the smooth-particle (SPH) method and were carried out for the case of eccentricities less than 0.1 on the dynamical time scale. We find that the circularization time in such systems should be of the order of hundreds of orbital periods. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The observations of close binary systems generalized and interpreted in [1, 2] provide evidence for the following relationships.

(1) The eccentricities e of close binaries increase in proportion to their orbital periods P and decrease with an increase in the ratio of their equivalent radii to semimajor axes.

(2) Close binaries consisting of low-mass stars (with the total mass of the components lower than $4\,M_\odot$) that are younger than 10^9 years and have ratios of their equivalent radii and semimajor axes $\sim$0.2 display zero eccentricity.

(3) For all close binaries, the transition period is $P_{\rm tr} \sim 3^{\rm d}$: a system with $P < 3^{\rm d}$ has eccentricity $e = 0$, while a system with $P > 3^{\rm d}$ has eccentricity $e \neq 0$.

In basic studies dedicated to circularization mechanisms (see, for example, [3–5]), it is suggested that the frequency σ of the perturbation function is small compared to the frequency associated with the free-fall time for the star [3]. We denote the ratio of these frequencies

$$K_{ZN} = \frac{\sigma^2 R^3}{GM} \ll 1, \qquad (1)$$

where M is the mass, G the gravitational constant, and R the radius of the star. Thus, errors in the method increase in proportion to R^3 (or in inverse proportion to the average density).

Here, we derive an independent estimate for the circularization time for close binaries with low-mass components that are close to filling their Roche lobes and periods close to the transition period, assuming that the components are fully convective. This assumption is completely justified for low-mass protostars [6]. We chose the parameters of the binary so that $K_{ZN} \sim 0.1$; in this case, calculations made with classical methods [3–5] can yield large errors.

In the early stages of their evolution, stars display dramatic variations in size and structure. It is obvious that these variations should affect the parameters of a binary, in particular, its orbital angular velocity. Let us consider close binaries in which the mass of the primary component is $M_1 = 0.80\,M_\odot$, the mass of the secondary is $M_2 = 0.57\,M_\odot$, the effective radius of the primary is $R_1 = 4.33\,R_\odot$, the effective radius of the secondary is $R_2 = 3.80\,R_\odot$, the distance between the centers of mass of the components is $A_{\rm K} = 11\,R_\odot$, and the Keplerian period is $P_{\rm K} = 3.615^{\rm d}$. Using this example, we will estimate the effect of small variations of the orbital angular velocity on the system's evolution.

2. MATHEMATICAL TECHNIQUE

Essentially, numerical methods of hydrodynamics may be divided into two groups, according to their description of motions. The first group is represented by the methods developed by Euler, and the second group, by the method of Lagrange. Both groups display certain advantages and disadvantages. Lagrangian methods are physical, but are not correct in the case of large deformations. Eulerian methods are good for large deformations, but display substantial errors in mass-transfer calculations. Methods combining elements of both the Euler and Lagrange approaches also exist, but are very complicated due to the need to transform the Euler parameters into their Lagrangian equivalent. Here, we used the purely Lagrange-type method of smoothed particle hydrodynamics (SPH) [7, 8], which is able to describe

1063-7729/05/4911-0892$26.00

processes involving large deformations. The SPH method does not require any grid, which makes it convenient for modeling asymmetrical events in astrophysics. In some cases, the SPH method takes a smaller computational time than other techniques.

The basic idea of the SPH method is that any physical parameter can be expressed through a set of its values at distributed points (particles). A parameter can be specified as

$$A_s(\mathbf{r}) = \int A(r')W(r-r',h)dr', \qquad (2)$$

where the integration is carried out over the entire volume, h is the carrier, and $W(r,h)$ is the core, which satisfies the conditions

$$\int W(r-r',h)dr' = 1,$$

$$\lim_{h\to 0} W(r-r',h) = \delta(r-r').$$

In practice, the integral (2) is replaced by the summation [9]

$$A_s(\mathbf{r}) = \sum_{i=1}^{N} m_i \frac{A(\mathbf{r}_i)}{\rho(\mathbf{r}_i)} W(|\mathbf{r}-\mathbf{r}_i|,h),$$

where m_i is the mass of the ith particle and $\rho(\mathbf{r_i})$ is the density at the point $\mathbf{r_i}$.

The basic equations of hydrodynamics for an ideal media are the mass-conservation equation (or continuity equation)

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0, \qquad (3)$$

the internal-energy equation

$$\rho\frac{\partial u}{\partial t} + \rho(\mathbf{v}\cdot\nabla)u + P\nabla\cdot\mathbf{v} = -\mathfrak{F}, \qquad (4)$$

and the equation of motion

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{1}{\rho}\nabla P - \mathbf{F}, \qquad (5)$$

where $\mathbf{v}$ is velocity, P is pressure, u is internal energy, $\mathbf{F}$ is outer forces, and $\mathfrak{F}$ is a function taking into account all possible nonadiabatic processes of absorption and energy release.

It is convenient to use the following equation of state to complete the system (3)–(5):

$$P = K\rho^{\gamma},$$

where γ is the adiabatic index and K is the entropy constant. The energy equation can then be replaced by the equation for the entropy constant:

$$\rho\frac{\partial K}{\partial t} + \rho(\mathbf{v}\cdot\nabla)K = -\frac{\gamma-1}{\rho^{\gamma}}\mathfrak{F}\,.$$

We will assume that the only origin of variations of the entropy constant is viscosity. Assuming that the components are fully convective and their matter is fully ionized, we can use the adiabatic index $\gamma = 5/3$.

Let us now present approximate forms of the hydrodynamical equations in the SPH method, where the continuity equation can be written in two ways [9]:

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} m_i W(|\mathbf{r}-\mathbf{r}_i|,h) \qquad (6)$$

or

$$\frac{d\rho(\mathbf{r})}{dt} = \sum_{i=1}^{N} m_i(\mathbf{v}-\mathbf{v}_i)\cdot\nabla W(|\mathbf{r}-\mathbf{r}_i|,h). \qquad (7)$$

Here, we have used (6), since (7) does not ensure exact conservation of mass.

The equation of motion for a particle is [9]

$$\frac{d\mathbf{v_i}}{dt} = -\sum_{j=1}^{N} m_j\left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2}\right)\cdot\nabla W(|\mathbf{r}_i-\mathbf{r}_j|,h) \qquad (8)$$

$$- G\sum_{j=1}^{N}\frac{M(r_{ij})}{r_{ij}^3}\hat{r}_{ij} + \mathbf{F}_{\mathrm{visc},i},$$

where $M(r_{ij})$ is the mass within a sphere with its center at $\mathbf{r}_j$ and with radius r_{ij}:

$$M(r_{ij}) = 4\pi m_i\int_0^{r_{ij}} r^2 W(r,h)dr,$$

$$r_{ij} = |\mathbf{r}_i-\mathbf{r}_j|, \qquad \hat{r}_{ij} = \frac{\mathbf{r}_i-\mathbf{r}_j}{r_{ij}}.$$

The third term in (8) accounts for the forces of viscosity. We use the following form for the artificial viscosity [10]:

$$\mathbf{F}_{\mathrm{visc},i} = \sum_{j=1}^{N}\Pi_{ij}m_j\nabla_i W(r_{ij},h)dr,$$

$$\begin{cases}\Pi_{ij} = -\dfrac{\alpha\mu_{ij}c_{ij}+\beta\mu_{ij}^2}{\rho_{ij}}, (\mathbf{v}_i-\mathbf{v}_j)\cdot(\mathbf{r}_i-\mathbf{r}_j) > 0,\\ \Pi_{ij} = 0, \quad (\mathbf{v}_i-\mathbf{v}_j)\cdot(\mathbf{r}_i-\mathbf{r}_j) < 0,\end{cases} \qquad (9)$$

$$\mu_{ij} = \frac{h(\mathbf{v}_i-\mathbf{v}_j)\cdot(\mathbf{r}_i-\mathbf{r}_j)}{r_{ij}^2+\varepsilon^2 h},$$

$$c_{ij} = 0.5(c_i+c_j),$$

$$\rho_{ij} = 0.5(\rho_i+\rho_j),$$

where c_i is the speed of sound at point $\mathbf{r}_i$, and α, β, and ε are constants matched in solutions of standard

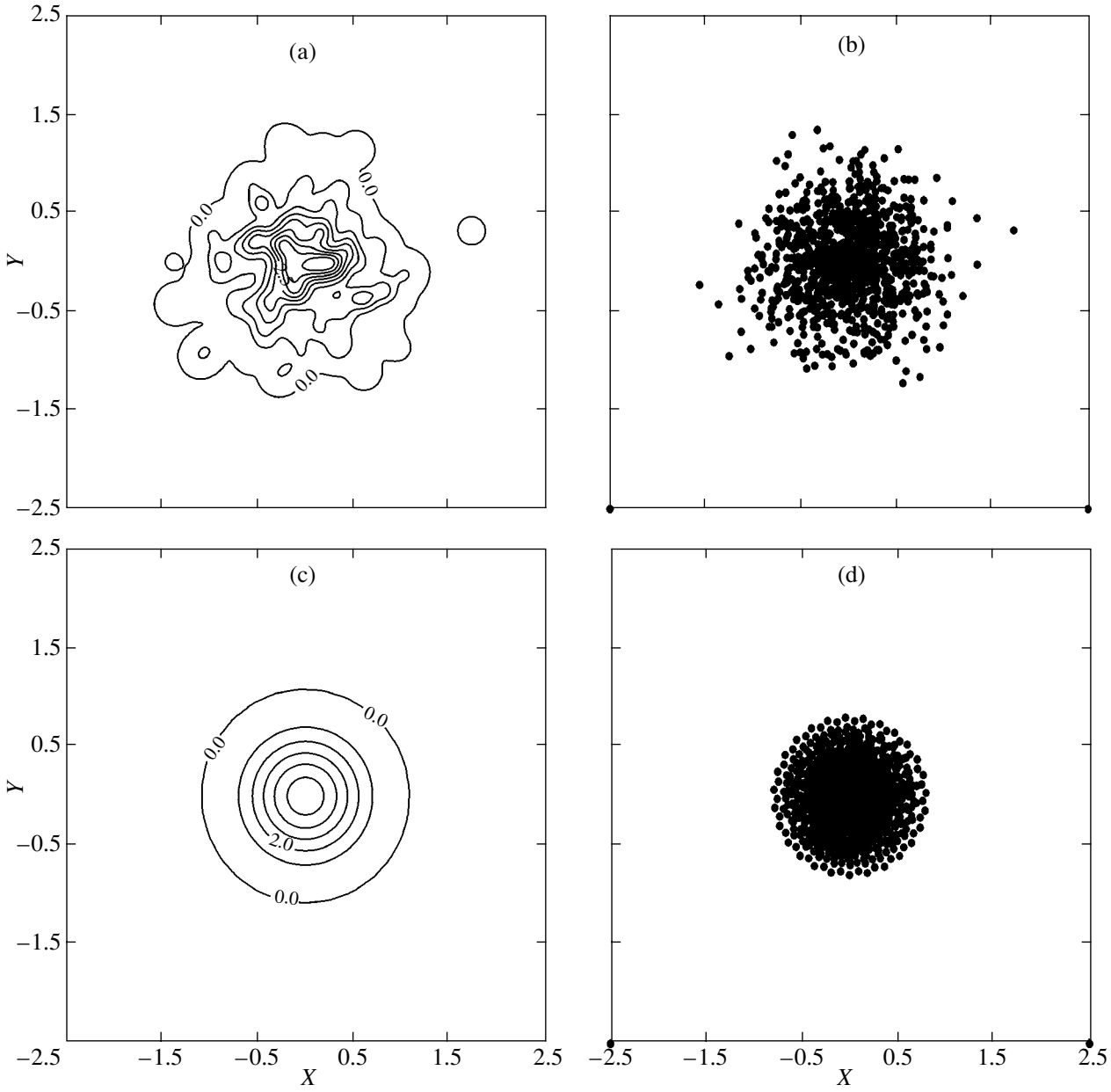

Fig. 1. Initial (a, b) and final (c, d) distributions for the density and particles, respectively. The distance is given in units of $A_{\rm K}$, and the density in units of the average density; $\gamma = 5/3$, $N = 500$.

problems ($\alpha = 0.5, \beta = 1$, and the parameter ε, which is needed to prevent division by zero in (9), is usually taken to be small; in this study, $\varepsilon = 0.001$). We use this expression for the artificial viscosity because it yields zero viscosity for solid-body rotation and ensures conservation of angular momentum. In the SPH method, the equation for the entropy constant can be written in the form [11]

$$\frac{dK_i}{dt} = \frac{\gamma - 1}{2\rho_i^{\gamma-1}} \sum_{j=1}^{N} \Pi_{ij} m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}.$$

Using different cores in the SPH is analogous to using various difference schemes in finite-difference methods. We use the Gaussian core

$$W(r,h) = \frac{\exp(-r^2/h^2)}{\pi^{3/2} h^3},$$

since it provides SPH equations with a simple physical interpretation. In addition, this core makes it possible to achieve higher accuracy compared to other cores [9].

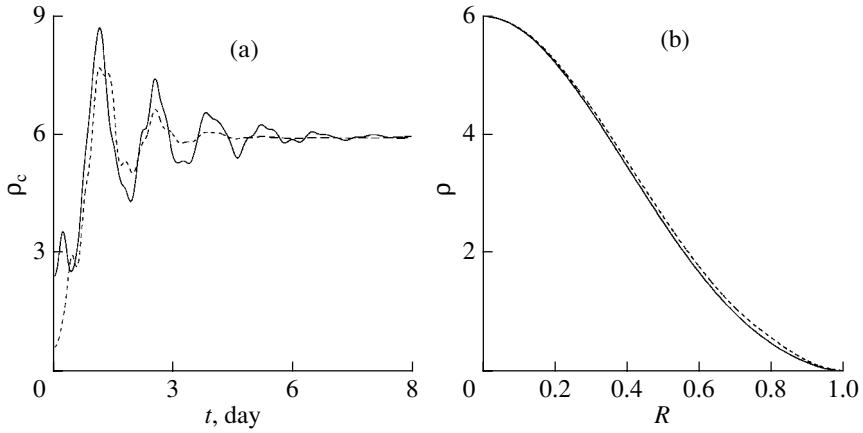

Fig. 2. (a) Central density as a function of time for $\Gamma = 0.95$ (dashed) and $\Gamma = 0.98$ (solid). (b) Density as a function of the radius for a polytrope calculated using the SPH method (solid) and for the exact solution (dashed). Density is given in units of average density.

Thus, the system of equations to be solved is

$$\begin{cases} \rho_i = \sum_{j=1}^{N} m_j W(r_{ij}, h), \\ \dfrac{d\mathbf{v}_i}{dt} = -\sum_{j=1}^{N} m_j \left(\dfrac{P_i}{\rho_i^2} + \dfrac{P_j}{\rho_j^2} \right) \\ \times \nabla W(|\mathbf{r}_i - \mathbf{r}_j|, h) \\ - G \sum_{j=1}^{N} \dfrac{M(r_{ij})}{r_{ij}^2} \hat{r}_{ij} + \mathbf{F}_{\mathrm{visc},i}, \\ \dfrac{dK_i}{dt} = \dfrac{\gamma - 1}{2\rho_i^{\gamma-1}} \sum_{j=1}^{N} \Pi_{ij} m_j (\mathbf{v}_i - \mathbf{v}_j) \cdot \nabla_i W_{ij}, \\ P_i = K_i \rho_i^{\gamma}. \end{cases} \tag{10}$$

System (10) was integrated using the scheme

$$\tilde{v}_{1/2} = v_0 + \frac{\Delta t}{2} f_0, \quad \tilde{r}_{1/2} = r_0 + \frac{\Delta t}{2} v_0,$$
$$\rho_{1/2} = \rho(\tilde{r}_{1/2}),$$
$$f_{1/2} = f(\tilde{r}_{1/2}, \tilde{v}_{1/2}, \rho_{1/2}, \ldots),$$
$$v_{1/2} = v_0 + \frac{\Delta t}{2} f_{1/2}, \quad r_{1/2} = r_0 + \frac{\Delta t}{2} v_{1/2},$$
$$r_1 = 2r_{1/2} - r_0, \quad v_1 = 2v_{1/2} - v_0,$$

where the subscripts denote the number of the integration step; f is the total acceleration of the particle.

The time step was selected so that it was smaller than the time needed for a sound wave to cover a distance equal to the carrier of the core (the Currant–Friedrichs–Levy condition):

$$\frac{h_i}{\Delta t} \geq c_i,$$

where c_i is the speed of sound. The optimum time step for SPH is [9, 10]:

$$\Delta t_2 = \min_i \left(\frac{h_i}{c_i + \alpha c_i + \beta \max_i |\mu_{ij}|} \right),$$

$$\Delta t_1 = \min_i \left(h_i / f_i \right)^{1/2},$$

$$\Delta t = 0.25 \ \min(\Delta t_1, \Delta t_2),$$

where f_i is the acceleration of the ith particle.

3. PREPARATION OF THE INITIAL DATA FOR CALCULATIONS

Solution of the differential equations (10) is specified by the selection of initial conditions. We specified all models at the initial time so that the close-binary components have the same distance between their centers of mass, display the same internal structure, and are close to hydrostatic equilibrium. The only parameter that varies from model to model is the initial velocity of the components. A restriction on the number of varying parameters is required due to the complexity of the problem.

The specification of the initial conditions may be tentatively divided into three stages.

(1) Modeling the internal structure for each component treated as a polytropic star in equilibrium, without taking into account rotation and tidal deformations.

(2) Taking into account the intrinsic rotation and tidal deformations for each component assuming that the other component can be treated as a point mass.

(3) Construction of initial conditions using the model components obtained at previous stages.

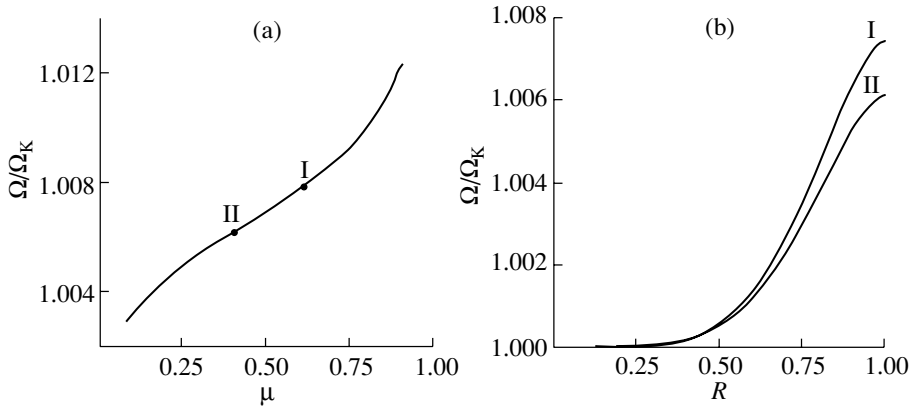

Fig. 3. $\Omega/\Omega_{\rm K}$ as a function of the (a) reduced mass and (b) effective radius of a component filling its Roche lobe. I marks the primary and II the secondary components.

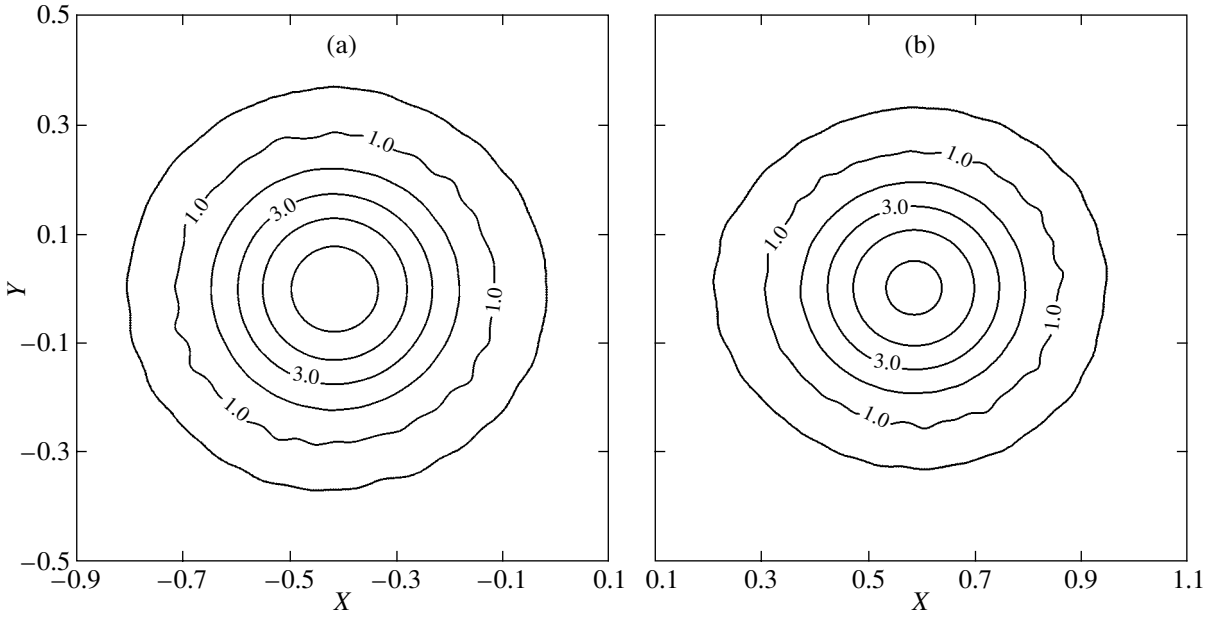

Fig. 4. The density distribution in the XOY plane for the (a) secondary and (b) primary components at the initial time. In the shaded areas, $t_\pi/P_{\rm K} > 1$. The distance is in units of $A_{\rm K}$, and the density in units of average density of the primary.

In the first stage, we must find the mutual locations of the particles for which the conditions of hydrostatic equilibrium are fulfilled for a specified core carrier h, entropy constant K, mass M, radius R, and adiabatic index γ. The entropy constant may be taken, for example, from [12], or calculated using the well-known expression [13]

$$K = N_n G M^{2-\gamma} R^{3\gamma-4},$$

where R is the radius of the polytropic star (or its effective radius if the star is a binary component). Since the entropy constant is the same for the total volume, the equation for the entropy constant can be excluded from the system (10), and the viscosity terms can be excluded from the equation of motion. In addition, in order to model the dissipation of kinetic energy, a damping constant should be introduced into the equation of motion [8]. Equation (10) can then be written in the form

$$\begin{cases} \rho_i = \dfrac{M}{N}\sum_{j=1}^{N} W(r_{ij}, h), \\ P_i = K\rho_i^{\gamma}, \\ \dfrac{d\mathbf{v}_i}{dt} = -\Gamma\dfrac{d\mathbf{r}_i}{dt} - K\gamma\rho_i^{\gamma-2}\nabla\rho \\ \quad - G\sum_{j=1}^{N}\dfrac{M(r_{ij})}{r_{ij}^2}\hat{r}_{ij}, \end{cases} \tag{11}$$

where Γ is the damping constant. In this case, all particles have the same mass and core carrier, which

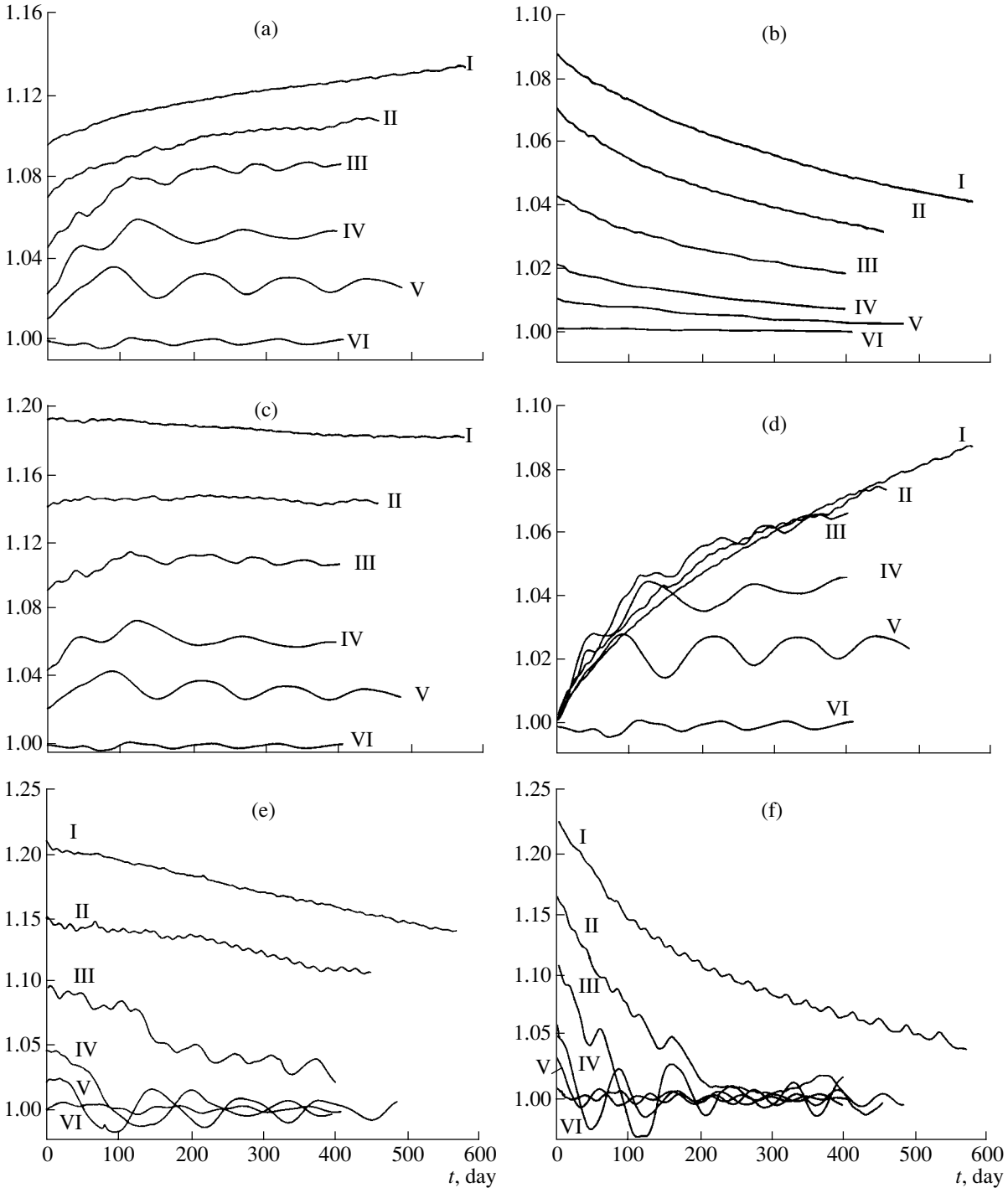

Fig. 5. Time variations of the (a) semimajor axis, (b) eccentricity, distances between the centers of mass of the components at (c) apastron and (d) periastron, and ratio of the angular velocity of axial rotation and the orbital angular velocity for the (e) primary and (f) secondary components, for models with $k_\omega \geq 1.010$: $k_\omega = 1.050$ (curve I), 1.040 (II), 1.030 (III), 1.020 (IV), 1.015 (V), 1.010 (VI). The vertical axes are in units of A_K.

somewhat degrades approximation, but substantially decreases calculation time.

In the first stage, calculations begin with specifying the initial conditions, which can be chosen arbitrarily. We assumed zero initial velocities for the particles and assigned them random positions with the dispersion $\sigma_r = 0.5$ (Fig. 1). The damping parameter was selected to be in the interval $0.5 < \Gamma < 1$. Figure 1 displays the result of integrating system (11), while Fig. 2 shows how hydrostatic

Table 1. Eccentricity e_0 and semimajor axis A_0 at time $t = 0$, together with the factor B in (14) and circularization time $t_{\rm circ}$ for models from the first and second groups

k_ω	e_0	$A_0/A_{\rm K}$	B	$t_{\rm circ}$, days
1.050	0.088	1.092	0.0013	757
1.040	0.066	1.070	0.0017	578
1.030	0.040	1.046	0.0021	474
1.020	0.019	1.023	0.0026	379
1.015	0.009	1.011	0.0030	338
1.008	0.007	0.994	0.0072	138
1.005	0.001	0.965	0.0113	89
1.005*	0.001	0.965	0.0075	133

* Calculated parameters of the artificial viscosity ($\alpha = 0.05$, $\beta = 0.1$) are an order of magnitude smaller than those in the other models ($\alpha = 0.5$, $\beta = 1$).

equilibrium is reached for two different damping parameters Γ.

In the second stage, the effects of the stellar rotation and tidal deformations are taken into account. Mutual particle positions for which the conditions of hydrostatic equilibrium are fulfilled must be found, taking into account external forces and rotation. We assumed solid-body rotation and took the gravitational potential of one of the components (which we will call the companion) to be that of a point mass.

Let the coordinate origin be at the center of mass of the distributed component M. The mass of the companion is M_p, the distance between their centers of mass is A, the OX axis is directed towards the companion, and the OZ axis is directed in the direction of rotation. In this case, system (10) can be written

$$\begin{cases} \rho_i = \dfrac{M}{N}\sum_{j=1}^{N} W(r_{ij}, h), \\ P_i = K\rho_i^{\gamma}, \\ \dfrac{d\mathbf{v}_i}{dt} = -\Gamma\dfrac{d\mathbf{r}_i}{dt} - K\gamma\rho_i^{\gamma-2}\nabla\rho \\ - G\sum_{j=1}^{N}\dfrac{M(r_{ij})}{r_{ij}^2}\hat{r}_{ij} \\ - \boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times(\mathbf{r}_i - \mathbf{r}_{\rm cm})) \\ - GM_p\dfrac{\mathbf{r}_i - \mathbf{r}_{m_p}}{|\mathbf{r}_i - \mathbf{r}_{m_p}|^3}, \\ \dfrac{d}{dx}\dfrac{d\mathbf{v}}{dt}\Big|_{(0,0,0)} = 0, \end{cases} \tag{12}$$

where $\mathbf{r}_{\rm cm}$ is the position vector of the system's center of mass and $\mathbf{r}_{m_p}$ is the position vector of the companion.

The last equation of system (12) is used to determine the rotational angular velocity Ω, since Kepler's third law is not valid for a distributed mass. Figure 3 presents the ratio of the angular velocity derived from (12) and the Keplerian angular velocity (which we will denote $\Omega_{\rm K}$) as a function of the reduced mass $M/(M + M_p)$ and the Roche-lobe filling factor. We used the solution resulting from the integration of system (11) as the initial conditions for integrating system (12) (Fig. 1). The damping constant can be the same as in the solution of (11). This procedure is applied to both components. Figure 4 presents the result of integrating system (12) for the case considered.

In the third stage, we construct model close binaries from the models of the components described above, transforming coordinates and specifying the initial velocity for each particle. For the secondary component,

$$\mathbf{r}_i = \mathbf{r}_i + \mathbf{r}_{\rm cm_1}, \quad \mathbf{v}_i = \boldsymbol{\Omega}\times\mathbf{r}_i, \quad i = 1\ldots N,$$

and for the primary

$$\mathbf{r}_i = \mathbf{r}_i + \mathbf{r}_{\rm cm_2}, \quad \mathbf{v}_i = \boldsymbol{\Omega}\times\mathbf{r}_i, \quad i = N+1\ldots 2N,$$

where $\mathbf{r}_{\rm cm_1}$ and $\mathbf{r}_{\rm cm_2}$ are the centers of mass of the primary and secondary. The initial angular velocity of rotation Ω was specified relative to $\Omega_{\rm K}$:

$$\Omega = k_\omega\Omega_{\rm K}, \tag{13}$$

where k_ω is a parameter specifying the initial velocity of the components in units of $\Omega_{\rm K}$. Further, we will use the parameter k_ω instead of the initial angular velocity.

4. CALCULATIONS AND RESULTS

In total, we constructed 14 models, with the parameter k_ω varied in the interval $0.95 < k_\omega < 1.05$. In all the models, the distance between the centers of mass was $A = 11\,R_\odot$ at the initial time. The ratio of the frequency of the perturbing function (using the Keplerian period) and the frequency associated with the free-fall time was $K_{ZN} = 0.107$ for the primary and $K_{ZN} = 0.104$ for the secondary.

The value $k_\omega \simeq 1.01$ corresponds to initial conditions for which the state of the binary components is close to hydrostatic equilibrium, and their orbital parameters do not vary substantially with time. This value of k_ω corresponds to an angular velocity exceeding that velocity derived from the solution of system (12) for each of the components. For the primary, the ratio of the angular velocity derived from

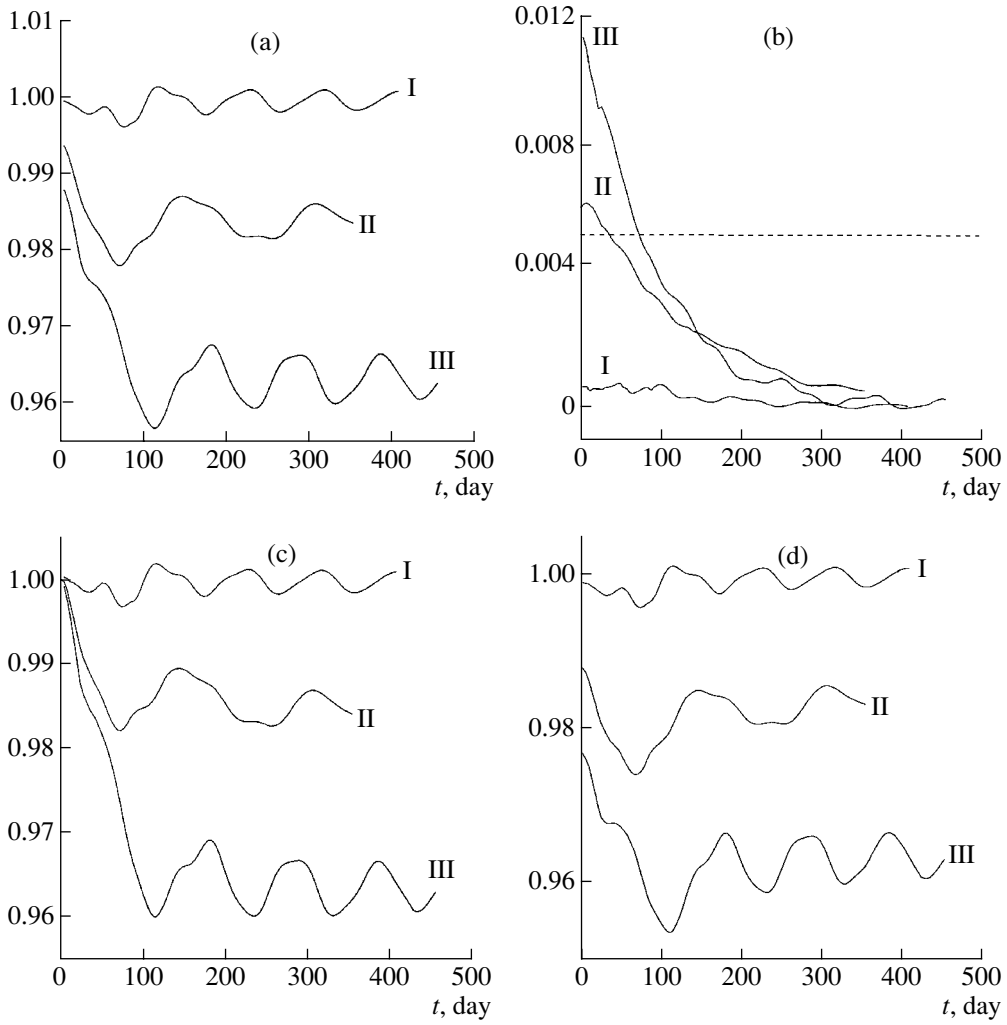

Fig. 6. (a) Semimajor axis, (b) eccentricity, and distances between the centers of mass of the components at (c) apastron and (d) periastron as functions of time for models with $1.005 \leq k_\omega \leq 1.010$: $k_\omega = 1.010$ (curve I), 1.0075 (II), 1.005 (III). In graph (b), the dotted curve corresponds to the limiting eccentricity obtained from observations [14]. The vertical axes are in units of A_K.

the solution of (12) and the Keplerian angular velocity of rotation is $\Omega/\Omega_K = 1.005$, while $\Omega/\Omega_K = 1.008$ for the secondary (Fig. 3). This is due to the fact that the solution of (12) treated one of the components as a point mass.

Models of the selected binary can be divided into four groups, according to the characteristic variation of their basic parameters with time. Within each group, small variations of k_ω result in smooth variations in the solutions of (10), providing evidence for the correctness of the calculations.

The first group includes models with $k_\omega > 1.01$, which corresponds to the initial condition for which the binary components move along elliptical orbits as they pass through periastron. For all models from the selected interval of k_ω, both the semimajor and semiminor axes increase with the time, while eccentricity decreases (Fig. 5). For both components, the intrinsic angular momenta decrease, while the orbital angular momentum increases.

Table 1 presents the approximation for the eccentricity as a function of time using the formula

$$e = e_0 \exp(-Bt) \tag{14}$$

The advantages of (14) are its simplicity and transparency. Physically, factor B in (14) is simply the damping coefficient. Using the definition of the circularization time [4]

$$\frac{1}{t_{\rm circ}} = -\frac{1}{e}\frac{de}{dt}$$

and (14), we obtain for the circularization time

$$t_{\rm circ} = 1/B. \tag{15}$$

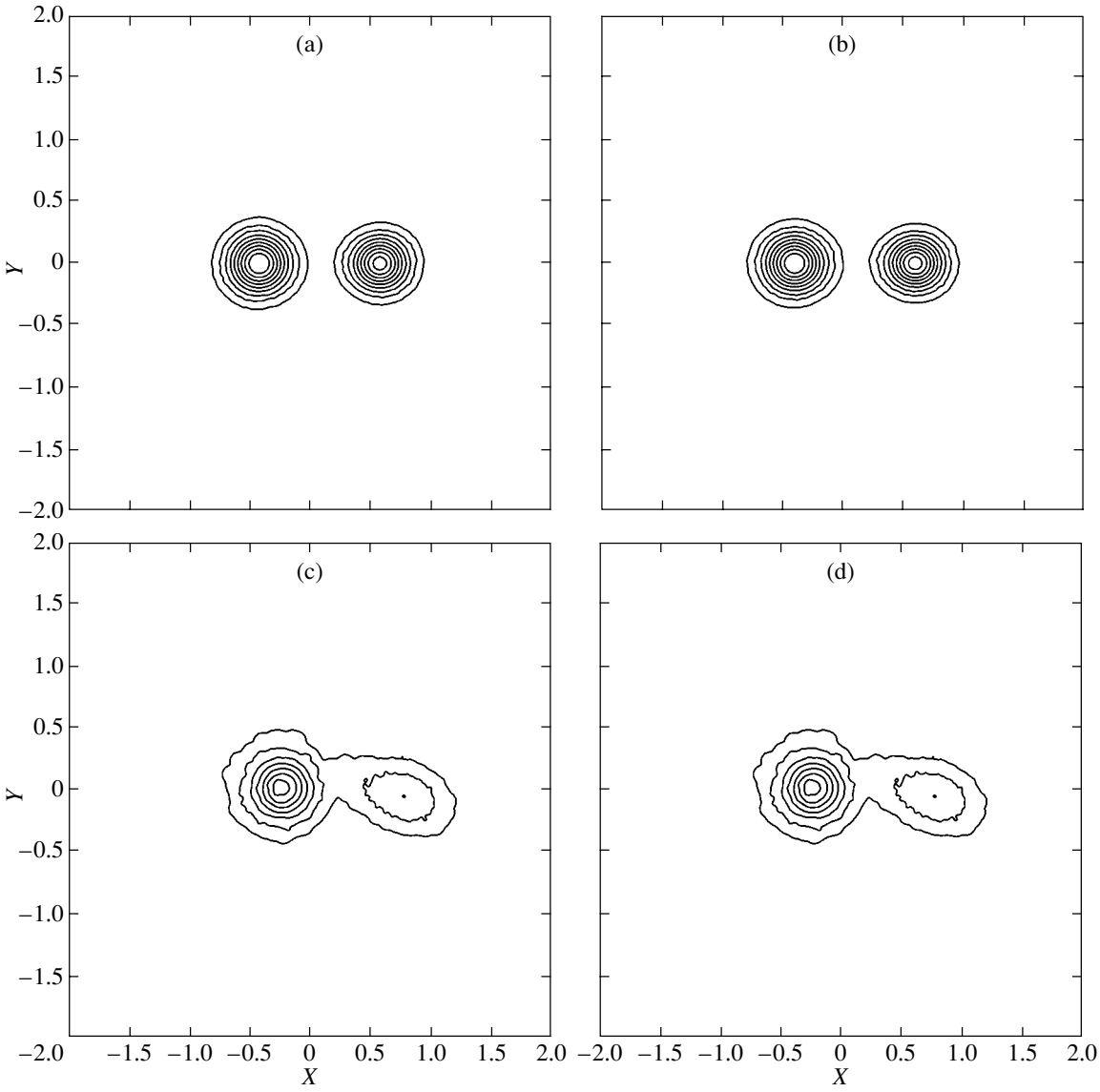

Fig. 7. Initial density distribution in the (a) XOY and (b) XOZ planes for all models (the distance between the centers of mass of the components is 11 $R_\odot$), and the density distribution in the (c) XOY and (d) XOZ planes for the model with $k_\omega = 1.0025$ at time $t = 181^{\mathrm{d}}$ (the distance between the centers of mass of the components is 1.16 A_{K}, $e = 0.01$). The axes are in units of A_{K}.

Table 1 presents the circularization times calculated from (15).

We estimated the circularization time for the considered system using [4]:

$$\frac{1}{t_{\mathrm{circ}}} = 21\frac{\lambda_{\mathrm{circ}}}{t_f}q(1+q)\left(\frac{R}{A}\right)^8,$$
$$\lambda_{\mathrm{circ}} = 0.048\sqrt{\frac{320}{320+(\Omega t_f/\pi)^2}}, \qquad (16)$$
$$t_f = \left[\frac{MR^2}{L}\right]^{1/3},$$

where $q = M/M_{\mathrm{c}}$ is the mass ratio, M the mass of the companion (treated as a point mass), M_{c} the mass of the considered component, and L the luminosity. If the secondary is assumed to be a point mass, the circularization time calculated from (16) is $t_{\mathrm{circ}} = 2.7 \times 10^6$ days. When the primary is treated as a point mass, $t_{\mathrm{circ}} = 1.9 \times 10^6$ days.

The second group includes models with $1.005 < k_\omega < 1.01$, which corresponds to initial conditions for which the binary components move along elliptical orbits as they pass through apastron. For models with initial conditions in this interval, the semimajor axis and eccentricity gradually decrease (Fig. 6).

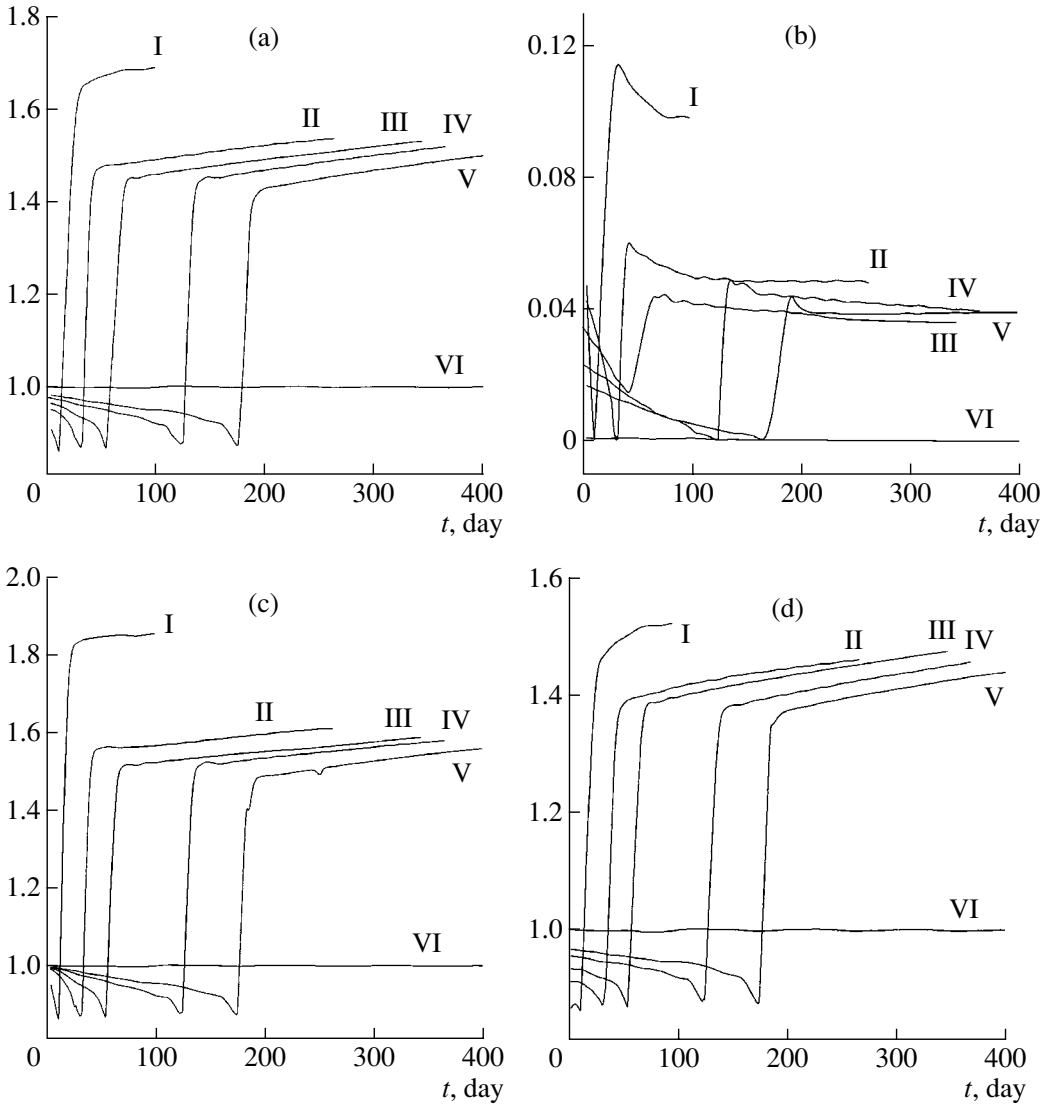

Fig. 8. Same as Fig. 6 for models with $k_\omega < 1.005$: $k_\omega = 0.980$ (curve I), 0.990 (II), 0.995 (III), 1.000 (IV), 1.003 (V).

The intrinsic angular momenta of both the secondary and primary increase, as well as the orbital angular momentum. Against the background of the general variations of the semimajor axis, the amplitude of the periodic variations exceeds that of the models with $k_\omega > 1.01$. This is due to the increase in the ratio of the effective radius of the components and the semimajor axis; i.e., to the increase of the contribution of the asphericity of the components.

The third group contains models with $0.96 < k_\omega < 1.005$, which corresponds to initial conditions for which the binary components move along elliptical orbits as they pass through apastron. For these models, the decrease in the semimajor axis results in overflow of the secondary's Roche lobe and mass exchange. In the stage preceding the mass exchange, the semimajor axis and eccentricity decrease. The orbital angular momentum decreases, while the intrinsic angular momenta of the components increase.

In order to calculate the system parameters in the stage of mass exchange, it was necessary to categorize the matter according to the component to which it belonged. To this end, at the initial time, particles were distinguished according to the component to which they were ascribed. For each particle, during the integration, we calculated the number of adjacent particles within a volume corresponding to a sphere with radius h. If the number of particles ascribed to the same component as the current particle was smaller than the number of particles ascribed to the other component, the "affiliation" of the current particle was changed.

This calculation procedure is very simple; however, as a result, the basic curves become piecewise

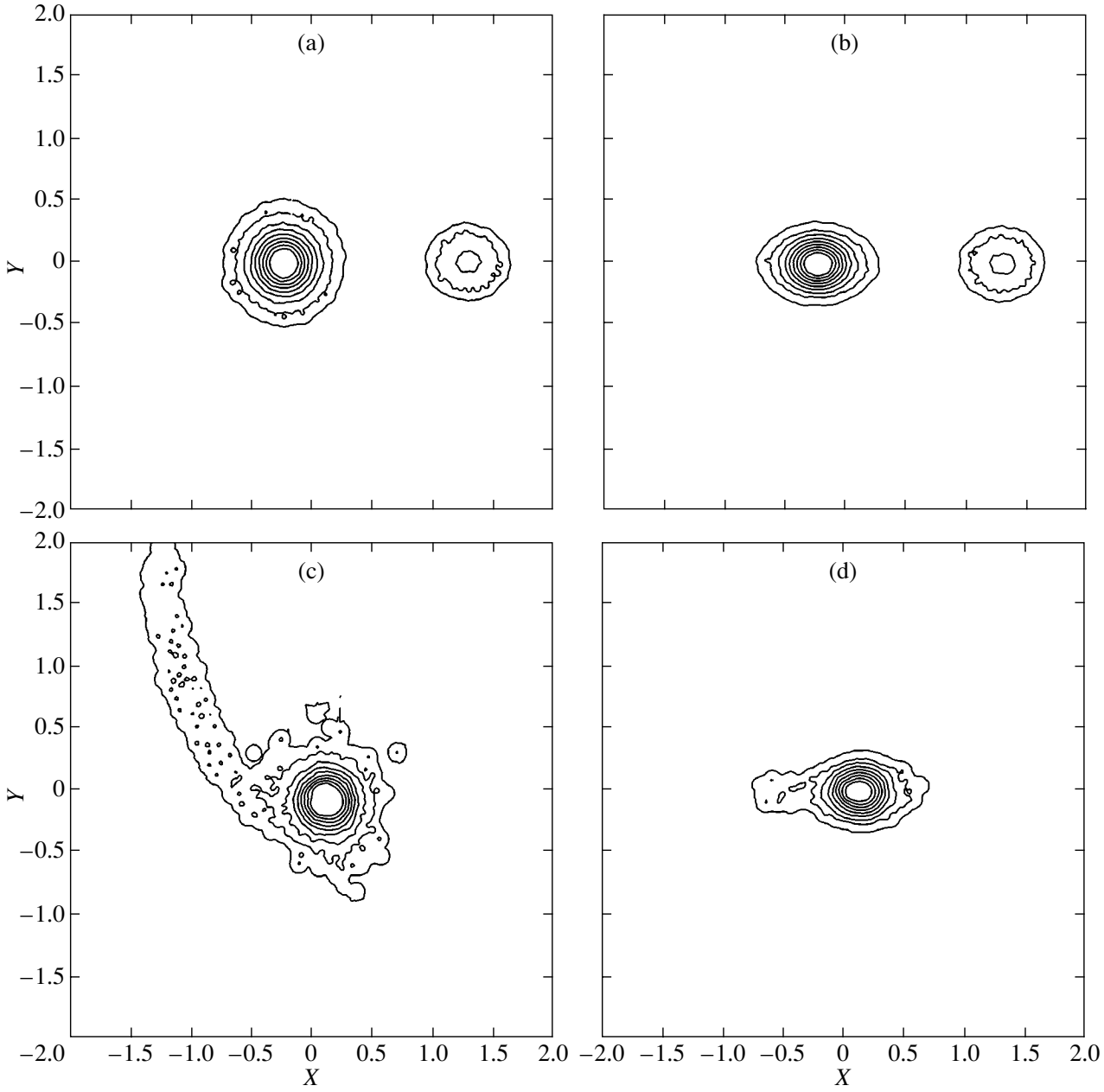

Fig. 9. Density distribution in the (a, c) XOY and (b, d) XOZ planes for the models with (a, b) $k_\omega = 1$ at time $t = 271^{\mathrm{d}}$ (the distance between the centers of mass of the components is $1.52\,A_{\mathrm{K}}$, $e = 4 \times 10^{-2}$) and (c, d) $k_\omega = 0.95$ at time $t = 2.5^{\mathrm{d}}$.

smooth. In addition, this method can be used to correctly model mass exchange with $\Delta M > (M_1 + M_2)/2N$, which restricts the range of problems that can be solved (this method was initially suggested as a Lagrange technique making it possible to describe large deformations [9]).

We do not consider the mass transfer in detail here; however, the results obtained can be used in detailed studies of binaries with mass exchange on the dynamical time scale.

The mass exchange lasts for fewer than ten initial orbital periods. The width of the flow is comparable to the size of the secondary (Fig. 7). The mass exchange results in an increase of the eccentricity and semimajor axis (Figs. 8, 9). The intrinsic angular momentum of the secondary decreases, while that of the primary increases, and the orbital angular momentum decreases. The eccentricity can double relative to its initial value in the course of the mass transfer (Fig. 8, Table 2). After the mass exchange, the rotational angular velocity of the primary exceeds the orbital angular velocity. As a result, the primary begins to trans-

Table 2. A_0, A_{ts}, and A_{te} at $t = 0$ and at the onset of and after the mass exchange; e_0 and e_{te} at $t = 0$ and after the mass exchange; the reduced mass μ_{te} at the end of the mass exchange; the onset of the mass exchange t_s and the end of the mass exchange t_e for the third group of models

k_ω	A_0/A_K	A_{ts}/A_K	A_{te}/A_K	e_0	e_{te}	μ_{te}	t_s, days	t_e, days
1.003	0.982	0.847	1.458	0.017	0.048	0.842	167.99	184.47
1.000	0.976	0.861	1.500	0.022	0.056	0.845	118.54	130.88
0.995	0.964	0.871	1.494	0.034	0.046	0.848	48.24	62.58
0.990	0.951	0.867	1.540	0.046	0.063	0.852	24.81	37.99
0.980	0.908	0.877	1.806	0.047	0.115	0.877	2.09	18.99
0.970	0.880	0.873	2.050	–	0.132	0.899	1.65	10.58

form its energy into orbital motion, which can lead to an increase in the eccentricity (Fig. 8, curve V). For all models from this interval of initial conditions, after the mass exchange, the semimajor axis increases, the intrinsic angular momentum of the primary decreases, while that of the secondary increases, and the orbital angular momentum increases.

When $k_\omega < 0.96$, mass exchange occurs, resulting in disruption of the secondary. A disklike, rapidly rotating structure with a spiral arm (a remnant of the secondary) is formed (Fig. 9).

For all models, the total momentum of the system remains constant (this is ensured by the selected form of artificial viscosity), as well as the total mass of the system (due to the selected form of the continuity equation). For models with $k_\omega > 1.005$, the total energy E varied within $\Delta E = \pm 10^{-6}$ relative to its initial value; for models with $0.960 < k_\omega < 1.005$, the energy varied within $\Delta E = \pm 10^{-3}$. The increase in the integration error for systems with mass exchange is due to the selection of the approximation for the constant of entropy.

5. BASIC CONCLUSIONS

Analyzing calculation results, we arrived at the following conclusions.

(1) The presented method may be used to model orbital circularization and synchronization of the orbital and axial rotations in close binaries whose components are close to filling their Roche lobes.

(2) For the studied system, equilibrium can be reached in the case of synchronous axial and orbital rotation for an orbital angular velocity of $\Omega \simeq 1.01\Omega_K$.

(3) Small deviations from equilibrium towards decreasing rotational angular velocity can lead to mass exchange on the dynamical time scale.

(4) Further decrease of the initial angular velocity results in merging of the secondary and the primary. Rapid variations of the internal structure of the close binaries components, which are inherent in the evolutionary stage considered, can result in mass exchange on the dynamical time scale and, provided these variations are substantial, in merging of the components.

(5) The circularization times (Table 1) derived using formula (15) are approximately three orders of magnitude smaller than those obtained using the data of [4]. This difference is due to the fact that formula (16) was derived using approximation (1), which is not applicable to the considered system. To illustrate this statement, let us subdivide a close-binary component into spherical layers with their center at the center of mass of the component. For each layer with thickness $r + \delta r$, we introduce the parameter

$$t_\pi = 2\pi r/c_s,$$

where $c_s = \sqrt{\gamma p/\rho}$ is the average speed of sound in the layer, while p and ρ are the average pressure and density in the layer, respectively. Physically, t_π is the time needed for the disturbance to cover a distance equal to the circumference of a circle with radius r moving with the speed of sound. If this time exceeds the period of the system, tidal peaks may deviate from the line of the centers, even if the viscosity is small (in Fig. 4, the zone with $t_\pi/P_K > 1$ is shaded). This may provide a supplementary mechanism for orbital circularization and synchronization of the rotational and orbital motions.

(6) To estimate the effect of viscosity on the synchronization time, we calculated a sequence of models in which the artificial-viscosity parameters were varied, all other factors remaining the same (see (9), the parameters α and β). When the artificial viscosity is decreased by an order of magnitude, the circularization time increases only slightly (Table 1, Fig. 10). Thus, viscosity is not a crucial factor in circularization for such systems.

(7) The efficiency of the supplementary circularization mechanism depends strongly on the average density of the components. Since the average density

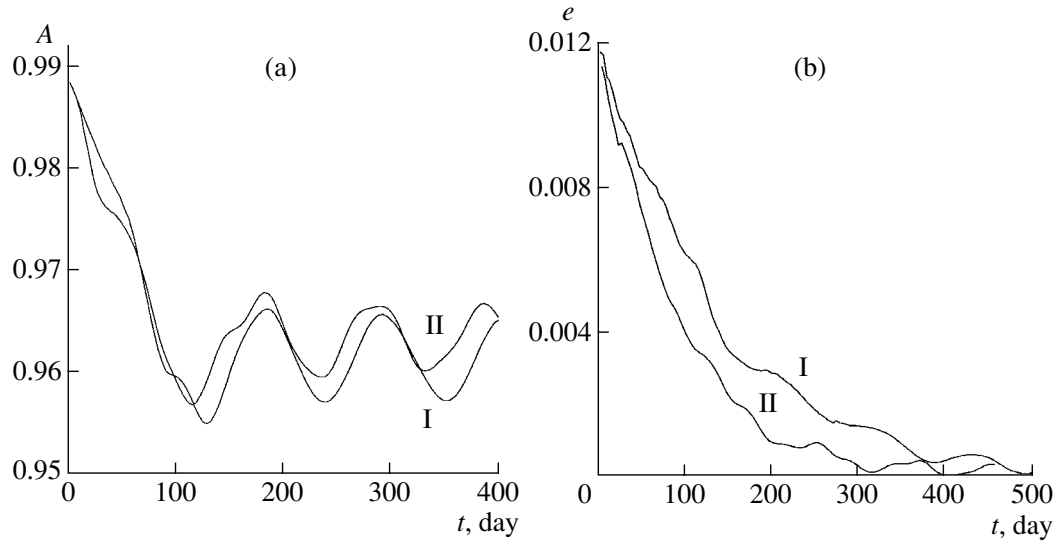

Fig. 10. Time variations of the (a) semimajor axis and (b) eccentricity for models with $k_\omega = 1.005$ and the artificial-viscosity parameters ($\alpha = 0.5$, $\beta = 1$) (curve I), ($\alpha = 0.05$, $\beta = 0.1$) (curve II). The semimajor axis is given in units of A_K.

of the primary exceeds that of the secondary, this mechanism should be more efficient for the secondary. We conclude from the curves showing the ratio of the rotational angular velocity of the components and the orbital angular velocity (Fig. 5) that the synchronization time t_{sync} for the secondary is smaller than that for the primary. Regardless of the type of binary, the synchronization time obtained using the methods of [3–5] is always smaller than ($t_{sync} \sim (R/A)^6$). In our case, the synchronization time for the secondary is comparable to the circularization time: $t_{sync} \approx t_{circ}$ (Fig. 5, Table 1).

(8) Thus, for the considered close binaries and similar systems consisting of fully convective stars close to filling their Roche lobes, the methods of [4, 5] may cause large errors in the estimated circularization times. A criterion for applicability of these methods can be given as follows:

$$t_\pi / P \ll 1,$$

where the average density of the close-binary components can be used to calculate t_π.

ACKNOWLEDGMENTS

The authors thank Prof. A.M. Cherepashchuk for his useful discussions.

REFERENCES

1. V. G. Karetnikov and A. M. Cherepashchuk, Astron. Zh. **47**, 27 (2003) [Astron. Rep. **47**, 38 (2003)].
2. V. G. Karetnikov and A. M. Cherepashchuk, Odessa Astron. Publ. **16**, 27 (2003).
3. J.-P. Zahn, Astron. Astrophys. **41**, 329 (1975).
4. J.-P. Zahn and L. Bouchet, Astron. Astrophys. **223**, 112 (1989).
5. J.-P. Zahn, Astron. Astrophys. **220**, 112 (1989).
6. S. W. Stahler, Publ. Astron. Soc. Pac. **100**, 1474 (1988).
7. L. B. Lucy, Astron. J. **82**, 1013 (1977).
8. R. A. Gingold and J. J. Monaghan, Mon. Not. R. Astron. Soc. **181**, 375 (1977).
9. J. J. Monaghan, Annu. Rev. Astron. Astrophys. **30**, 543 (1992).
10. J. J. Monaghan and R. A. Gingold, J. Comp. Phys. **46**, 429 (1983).
11. L. Hernquist, Astrophys. J. **404**, 717 (1993).
12. O. A. Kuznetsov, Astron. Zh. **72**, 508 (1995) [Astron. Rep. **39**, 450 (1995)].
13. S. Chandrasekar, *An Introduction to the Study of Stellar Structure* (University of Chicago, Chicago, 1939).
14. L. B. Lucy and M. A. Sweeney, Astrophys. J. **76**, 544 (1971).

Translated by K. Maslennikov

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 905–915. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 1013–1024.

Asymmetry Function of Interstellar Scintillations of Pulsars

V. I. Shishov and T. V. Smirnova

Pushchino Radio Astronomy Observatory, Astrospace Center of the Lebedev Institute of Physics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia

Received February 27, 2004; in final form, May 18, 2005

Abstract—A new method for separating intensity variations of a source's radio emission having various physical natures is proposed. The method is based on the joint analysis of the structure function of intensity variations and the asymmetry function, which is a generalization of the asymmetry coefficient and which characterizes the asymmetry of the distribution function of intensity fluctuations on various scales for inhomogeneities in the diffractive scintillation pattern. Relationships for the asymmetry function in the cases of a logarithmic normal distribution of the intensity fluctuations and a normal distribution of the field fluctuations are derived. Theoretical relationships and observational data on interstellar scintillations of pulsars (refractive, diffractive, and weak scintillations) are compared. The data for PSR B0329+54, B1133+16, B1642−03, and B1933+16 pulsars were used for comparison. Pulsar scintillations match the behavior expected for a normal distribution of field fluctuations (diffractive scintillation) or logarithmic normal distribution of intensity fluctuations (refractive and weak scintillation). Analysis of the asymmetry function is a good test for distinguishing scintillations against the background of variations that have different origins. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In analyses of interstellar scintillations of intrinsically variable radio sources, the problem of separating variations of source emission with different origins arises. In particular, this is true of separating intrinsic variations of pulsar radio emission, interstellar scintillations, and additive and modulational noise. A similar problem appears in the analysis of rapid variability of extragalactic sources. In [1–5], it was proposed that one used the asymmetry coefficient of the distribution function of the radio-emission intensity fluctuations as a test for determining the nature of these variations. The asymmetry coefficient is a measure of the deviation of the distribution function of intensity fluctuations from a normal distribution, and is defined as

$$\gamma = \frac{\langle (I - \langle I \rangle)^3 \rangle}{[\langle (I - \langle I \rangle)^2 \rangle]^{3/2}} = \frac{M_3}{M_2^{3/2}}. \qquad (1)$$

Here, $\langle I \rangle$ is the mean intensity of the source's radio emission, and M_3 and M_2 are the third and second central moments of the distribution of the intensity fluctuations. In the case of scintillations, γ is positive, and is related to the scintillation index by a linear relationship

$$\gamma = Am, \qquad (2)$$

where A is a numerical factor and m is the scintillation index, which is defined as

$$m^2 = \frac{\langle (I - \langle I \rangle)^2 \rangle}{\langle I \rangle^2} = \frac{M_2}{\langle I \rangle^2}. \qquad (3)$$

If we are dealing with an extragalactic source consisting of a compact scintillating component (core) and an extended nonscintillating component (halo), I corresponds to the flux from the scintillating component. The factor A depends on the form of the turbulence spectrum and the scintillation regime. In particular, in the case of weak scintillations of a pointlike source in the Fraunhofer zone relative to the outer turbulence scale (i.e., the characteristic size of the largest inhomogeneities), the flux fluctuations are distributed according to a Rice–Nakagami law [6, 7], and $A = 3/2$. In the case of weak scintillations of a pointlike source in the Fresnel zone relative to the inner scale (i.e., the characteristic size of the smallest inhomogeneities), the flux fluctuations are distributed according to a lognormal law [8], and $A = 3$. As is shown in [1], relationship (2) is a good test for isolating interstellar scintillations that give rise to flux variations of extragalactic radio sources.

However, the asymmetry coefficient is an integrated parameter, which describes the summed fluctuations on all scales. Interstellar scintillations can dominate on some time scales, while intrinsic variations or noise can dominate on others. To analyze the nature of the fluctuations on various scales, it

1063-7729/05/4911-0905$26.00

is desirable to introduce some function that would yield the value of the asymmetry coefficient for a given time scale in place of the asymmetry coefficient itself. Below, we will introduce such an asymmetry function, $\gamma_2(\tau)$, which is a generalization of the asymmetry coefficient, in the same way that the structure function is a generalization of the variance of the temporal process. We will also investigate the application of the asymmetry function to the analysis of interstellar scintillations of pulsars.

2. DEFINITION OF THE ASYMMETRY FUNCTION

Let us introduce the first and second differences

$$\Delta_1(\tau) = I(t+\tau) - I(t), \quad (4)$$
$$\Delta_2(\tau) = I(t+\tau) - 2I(t) + I(t-\tau),$$

where t is the time and τ is the time shift from instant t. The structure function of the flux fluctuations is defined as

$$D_I(\tau) = \langle[\Delta_1(\tau)]^2\rangle = 2[\langle I^2\rangle - \langle I(t+\tau)I(t)\rangle] \quad (5)$$

and the mean square of the second difference is

$$\langle[\Delta_2(\tau)]^2\rangle = \langle\{[I(t+\tau) - I(t)] \quad (6)$$
$$- [I(t) - I(t-\tau)]\}^2\rangle = \langle\{[I(t+\tau) - I(t)]^2$$
$$+ [I(t) - I(t-\tau)]^2 - 2[I(t+\tau) - I(t)]$$
$$\times [I(t) - I(t-\tau)]\}\rangle = 4D_I(\tau) - D_I(2\tau).$$

We also introduce the third moment of the second difference $\Delta_2(\tau)$ using the formula

$$\langle[\Delta_2(\tau)]^3\rangle = -6\langle I^3\rangle + 6\langle I^2(t+\tau)I(t-\tau)\rangle \quad (7)$$
$$- 12\langle I(t+\tau)I(t)I(t-\tau)\rangle + 12\langle I(t+\tau)I^2(t)\rangle$$

and define the second-order asymmetry function as

$$\gamma_2(\tau) = -\frac{\langle[\Delta_2(\tau)]^3\rangle}{\langle[\Delta_2(\tau)]^2\rangle^{3/2}}. \quad (8)$$

The definition of the function $\gamma_2(\tau)$ formally corresponds to the definition of the asymmetry coefficient for a random quantity, $\Delta_2(\tau)$. Since the third moment of $\Delta_2(\tau)$ is negative in the case of scintillations, we have added a minus sign in (8), in order for the function $\gamma_2(\tau)$ to be positive for intensity fluctuations due to scintillation.

We also introduce a somewhat different asymmetry function

$$\gamma_{2,1}(\tau) = -2\frac{\langle[\Delta_2(\tau)]^3\rangle}{\langle[\Delta_2(\tau)]^2\rangle\langle[\Delta_1(2\tau)]^2\rangle^{1/2}} \quad (9)$$
$$= -2\frac{\langle[\Delta_2(\tau)]^3\rangle}{[4D_I(\tau) - D_I(2\tau)][D_I(2\tau)]^{1/2}}.$$

As will be shown below in Appendices A and B, in the case of intensity fluctuations distributed according to a Rice–Nakagami law or lognormal law, the function $\gamma_{2,1}(\tau)$ for small values of τ is related to the function $[D_I(2\tau)]^{1/2}/\langle I\rangle$ by a formula of the form (2):

$$\gamma_{2,1}(\tau) = \frac{A[D_I(2\tau)]^{1/2}}{\langle I\rangle}. \quad (10)$$

In general, the factor A is a function of τ, and we introduce the function $A_{2,1}(\tau)$:

$$A_{2,1}(\tau) = \langle I\rangle\frac{\gamma_{2,1}(\tau)}{[D_I(2\tau)]^{1/2}}. \quad (11)$$

The form of the functions $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ for the two distribution laws (a normal distribution for wave-field fluctuations and lognormal law for intensity fluctuations) is determined in Appendices A and B.

3. WEAK SCINTILLATIONS

We will consider the properties of the asymmetry function in the case of scintillations of radio sources on electron-density inhomogeneities of plasma with a power-law turbulence spectrum. It is known that the scintillation regimes are determined by the structure function of the phase fluctuations on a scale that is equal to the scale of the first Fresnel zone $D_S(b_{\rm Fr})$, where $b_{\rm Fr} = (r_{\rm eff}/k)^{1/2}$, $r_{\rm eff}$ is the effective distance from the observer or the source to the layer containing the turbulent medium (in the case of a statistically uniform medium, this is the distance from the observer to the source), and $k = 2\pi/\lambda$ is the wavenumber. If

$$D_S(b_{\rm Fr}) \ll 1, \quad (12)$$

the scintillations are weak, the characteristic spatial scale giving rise to the intensity fluctuations is equal to $b_{\rm Fr}$, and the scintillation index can be found from the formula [9]

$$m^2 \cong K(n)D_S(b_{\rm Fr}), \quad (13)$$

where $K(n)$ is a numerical factor of the order of unity that depends on the exponent n of the turbulence spectrum.

Numerical calculations of the one-dimensional distribution function of the flux fluctuations $f(I)$ were carried out in [10, 11], and the second and third moments of the intensity fluctuations in the case of a Kolmogorov spectrum for the inhomogeneities of the index of refraction for various values of the inner turbulence scale were determined. It was shown that, in the weak scintillation regime, as well in the strong scintillation regime with values of $D_S(b_{\rm Fr}) < 3$, the distribution function $f(I)$ is close to a lognormal

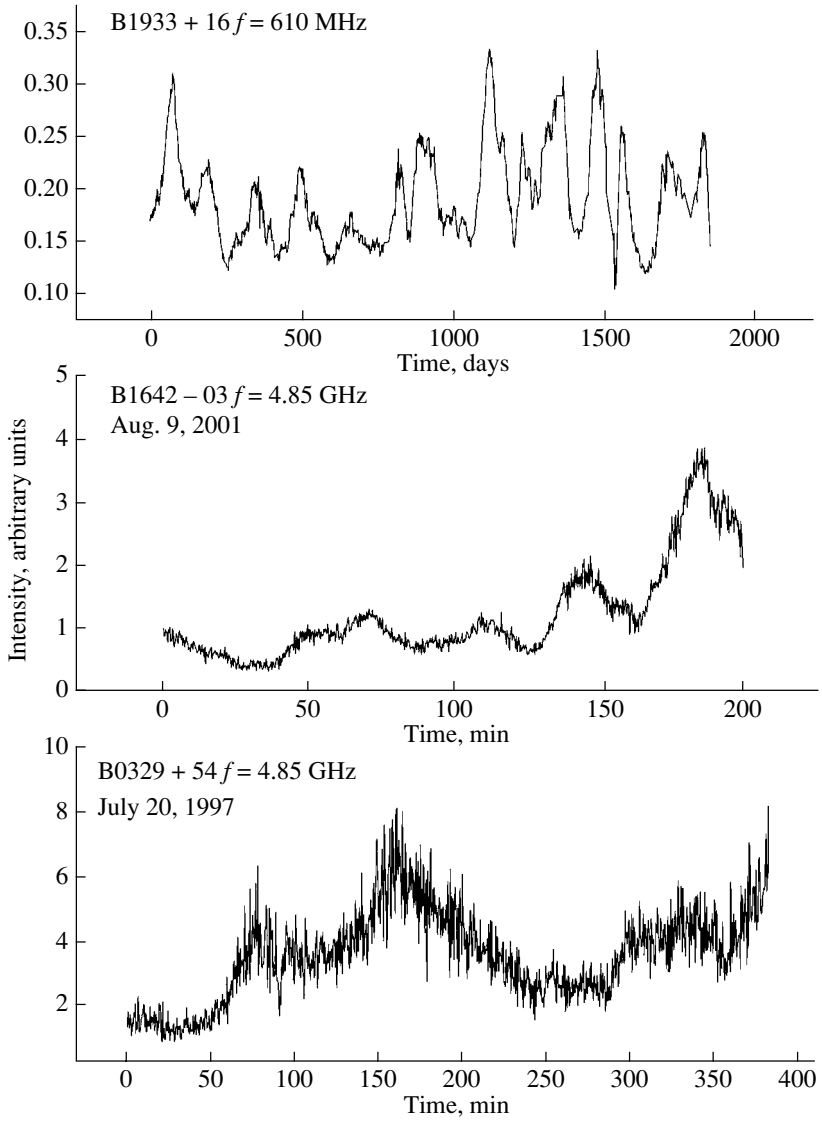

Fig. 1. Time variations of the emission intensity for PSR B1933+16 at 610 MHz (top), PSR B1642−03 at 4.85 GHz (center), and PSR B0329+54 at 4.85 GHz (bottom).

function. Using the results of these studies, we find $A = 2.78$ for a Kolmogorov spectrum without a turnover and $A = 2.86$ for a Kolmogorov spectrum with an inner scale l equal to the scale b_F of the first Fresnel zone ($l/b_F = 1$). Thus, with increase in the inner scale of the turbulence spectrum, the factor A increases and approaches three, which corresponds to a lognormal distribution for the flux fluctuations. The calculations show that relationship (2) is valid up to a value of m that approaches unity from the side of the weak (unsaturated) scintillation regime.

Note that the one-dimensional distribution function $f(I)$ and asymmetry coefficient γ describe the flux fluctuations on the main scale of the spatial pattern of the scintillations and, accordingly, on the main temporal scale τ_0 of the structure function of the flux fluctuations. When $\tau \geq \tau_0$, the functions $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ will be determined by asymptotic relationships that correspond to a lognormal distribution law for the intensity fluctuations (see Appendix B):

$$\gamma_2(\tau) \cong \frac{3}{2}\frac{D_I(2\tau)}{[4D_I(\tau) - D_I(2\tau)]^{1/2}\langle I\rangle}, \qquad (14)$$

$$\gamma_{2,1}(\tau) \cong 3[D_I(2\tau)]^{1/2}/\langle I\rangle, \qquad (15)$$

$$A_{2,1}(\tau) \cong 3. \qquad (16)$$

Flux fluctuations with small spatial and temporal scales are determined by the diffraction of the radio waves on inhomogeneities in the index of refraction with sizes that are much smaller than the first

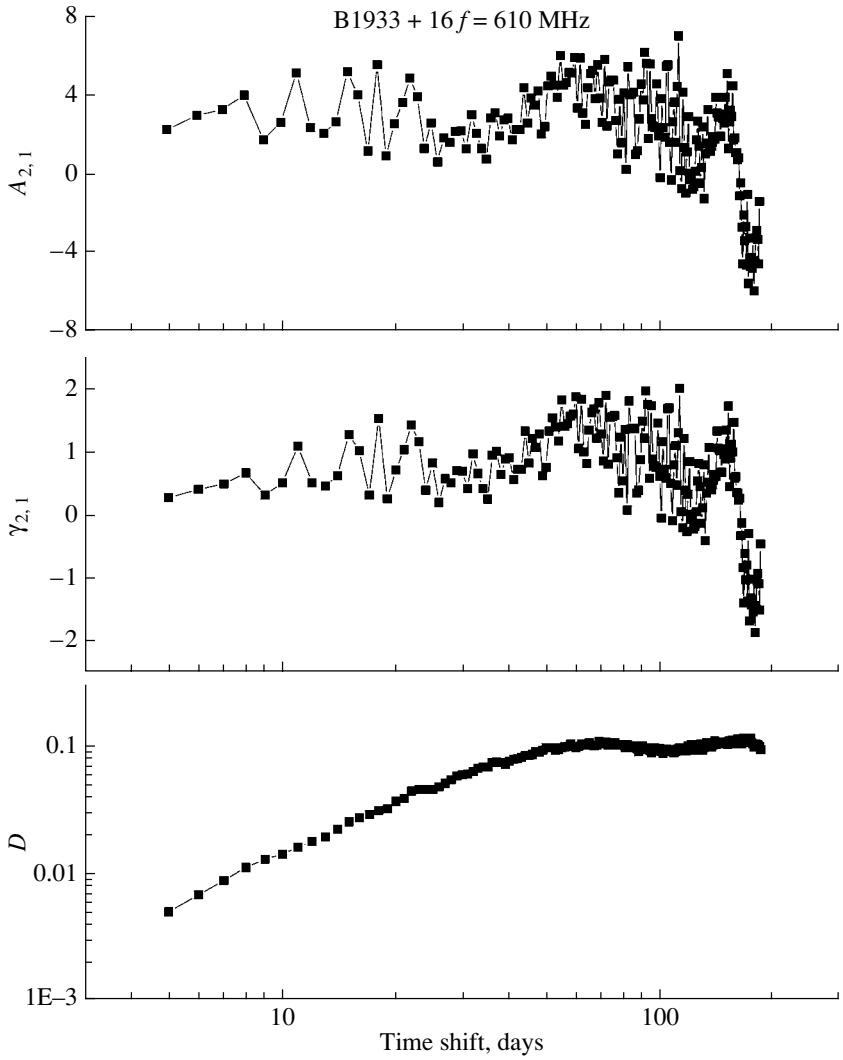

Fig. 2. Structure function, asymmetry function, and coefficient $A_{2,1}$ versus the time shift for refractive scintillations of PSR B1933+16.

Fresnel zone. Therefore, the distribution of small-scale flux fluctuations will more closely fit a normal distribution law, and the functions $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ will better match the asymptotic relationships (A.8)–(A.10) of Appendix A.

4. DIFFRACTIVE SCINTILLATIONS

For large values of the parameter $D_S(b_{\rm Fr})$, the intensity fluctuations consist of two components: small-scale (diffractive) and large-scale (refractive). With increasing $D_S(b_{\rm Fr})$, the distribution function of the field of the diffractive component tends to a normal distribution [9], the scintillation index tends to unity, and the characteristic spatial scale of the diffraction pattern corresponds to the field coherence scale $1/k\theta_{\rm scat}$, where $\theta_{\rm scat}$ is the characteristics scattering angle. When $\tau \le \tau_0$, the functions $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ will correspond to the asymptotic expressions (Appendix A)

$$\gamma_2(\tau) \cong \frac{3}{4}\frac{D_I(2\tau)}{[4D_I(\tau) - D_I(2\tau)]^{1/2}\langle I\rangle}, \qquad (17)$$

$$\gamma_{2,1}(\tau) \cong \frac{3}{2}\frac{[D_I(2\tau)]^{1/2}}{\langle I\rangle}, \qquad (18)$$

$$A_{2,1}(\tau) = \frac{\gamma_{2,1}(\tau)}{[D_I(2\tau)]^{1/2}} \cong \frac{3}{2}, \qquad (19)$$

and, when $\tau \gg \tau_0$, will be determined as $\gamma_2(\tau) \cong 2/\sqrt{6}$, $\gamma_{2,1}(\tau) \cong 2\sqrt{2}$, $A_{2,1}(\tau) \cong 2$.

5. REFRACTIVE SCINTILLATIONS

The refractive component is determined by weak focusing of the radiation on inhomogeneities with

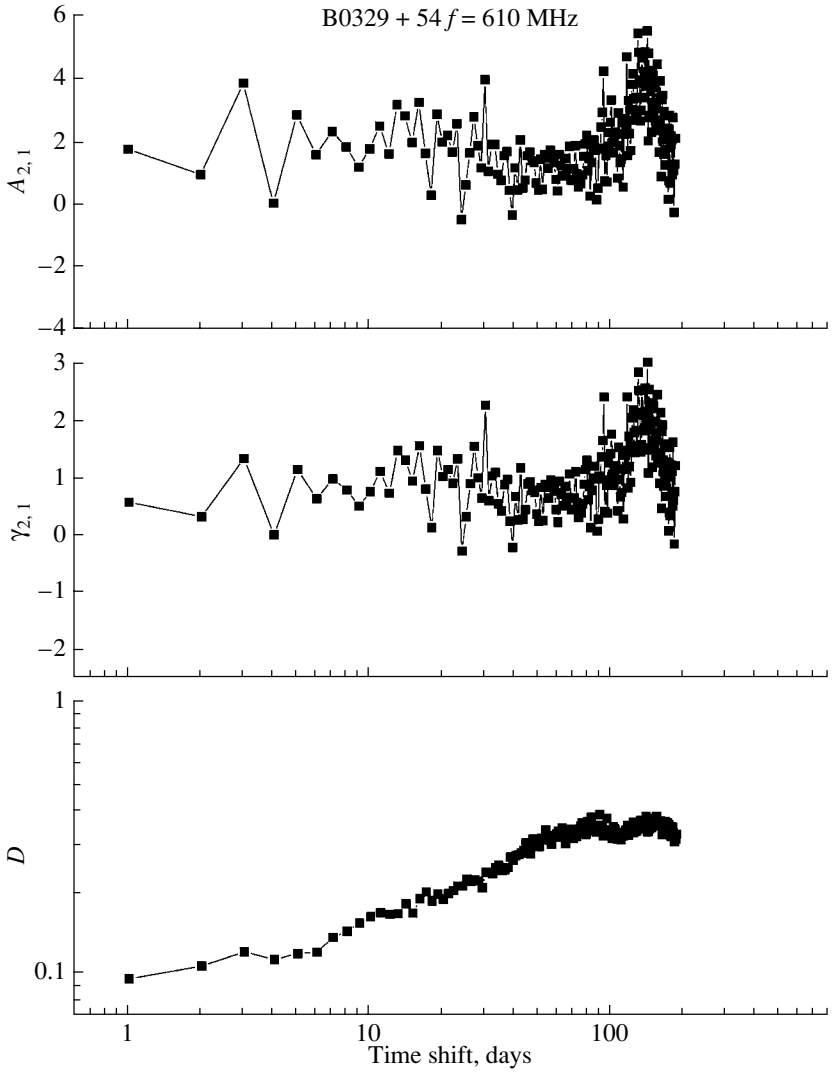

Fig. 3. Same as Fig. 2 for PSR B0329+54.

characteristic scales of the order of the scattering disk: $r_{\text{eff}}\theta_{\text{scat}}$ [9]. Refractive scintillations can be treated as weak scintillations on large-scale inhomogeneities, with sizes much larger than the scale of the first Fresnel zone. Therefore, we expect a lognormal distribution law for the flux fluctuations. When $\tau \geq \tau_0$, the functions $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ will be determined by the asymptotic relationships (14)–(16).

6. COMPARISON WITH OBSERVATIONS

We used data on diffractive, refractive, and weak scintillations of several pulsars to analyze the applicability of the above relationships for the asymmetry function to actual observations of interstellar scintillations of pulsars.

Refractive and Weak Scintillations

We used the five-year monitoring data for PSR B1933+16 and B0329+54 (1990–1995) at 610 MHz [12] (NRAO, Green Bank) to analyze refractive scintillations. The observations were carried out in a 16-MHz band; diffractive scintillations were considerably smoothed by averaging the signal over about an hour of observations. Figure 1 (top) shows the time variations of the emission intensity for PSR B1933+16. The time scale of these variations is 33 days [12]. We used relationships (5), (9), and (11) to calculate the structure function of flux fluctuations $D(\tau)$, asymmetry function $\gamma_{2,1}(\tau)$, and coefficient $A_{2,1}(\tau)$ for all the types of scintillations considered below. Before calculating these functions, we have normalized the time series to their average values.

Figures 2 and 3 show $D(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ as functions of the time shift in days for PSR B1933+16

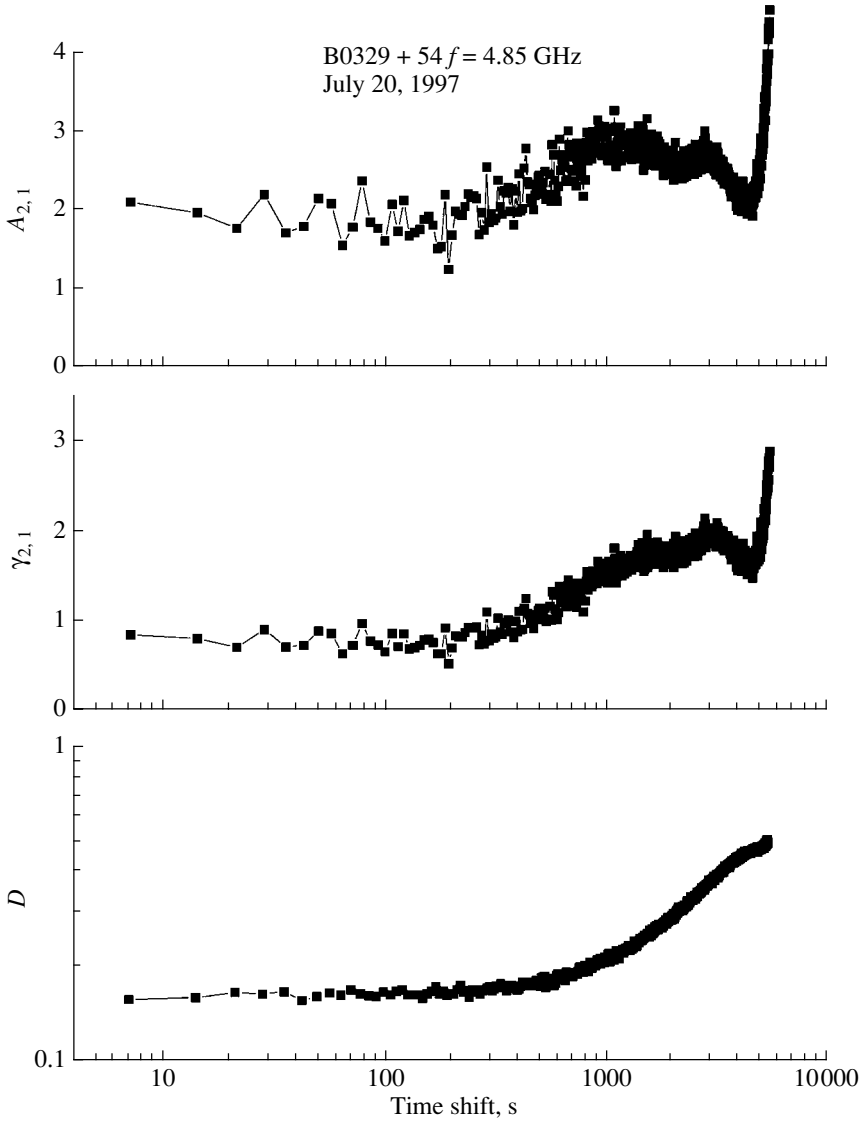

Fig. 4. Same as Fig. 3 but for weak scintillations.

and B0329+54, respectively. The structure function reaches a constant level equal to $2m^2$, where m is the modulation index of the intensity variations. The time shift at which $D(\tau)$ decreases by a factor or two gives the time scale of these variations. In $\gamma_{2,1}(\tau)$ and $A_{2,1}(\tau)$, this time scale corresponds to a shift that is a factor of two smaller [see formula (10)]. We can see from Figs. 2 and 3 that $A_{2,1}(\tau)$ is approximately constant for shifts of less than 40 days, and this average level is $A_{2,1} \approx 3$ for PSR B1933+16, based on formula (5) for refractive scintillations, and $A_{2,1} \approx 2$ for PSR B0329+54. The difference of $A_{2,1}$ from three for PSR B0329+54 may be due to the fact that the index of the electron-density fluctuation spectrum toward this pulsar differs from the Kolmogorov index [13]: it is $n = 3.5$. It seems reasonable to suppose that the decrease of the spectral index will result in a decrease in $A_{2,1}$ at small time shifts.

We used observations of the pulsars PSR B0329+54 and B1642−03 carried out at 4.85 GHz on the 100-m Effelsberg telescope for our analysis of weak scintillations. The input data and observing parameters are given in [13, 14]. Before recording the signal, we averaged the pulsed emission over $10\,P_1$ for PSR B0329+54 and $39\,P_1$ for PSR B1642−03 (where P_1 is the pulsar period). The time resolutions were 7.15 and 15 s, respectively. Figure 1 shows the time series for these pulsars (central and bottom graphs). We used the fourth channel (the observations were carried out in four frequency channels of 60 MHz each), which was the most free of interference, for PSR B0329+54. The modulation indices for the intensity variations for

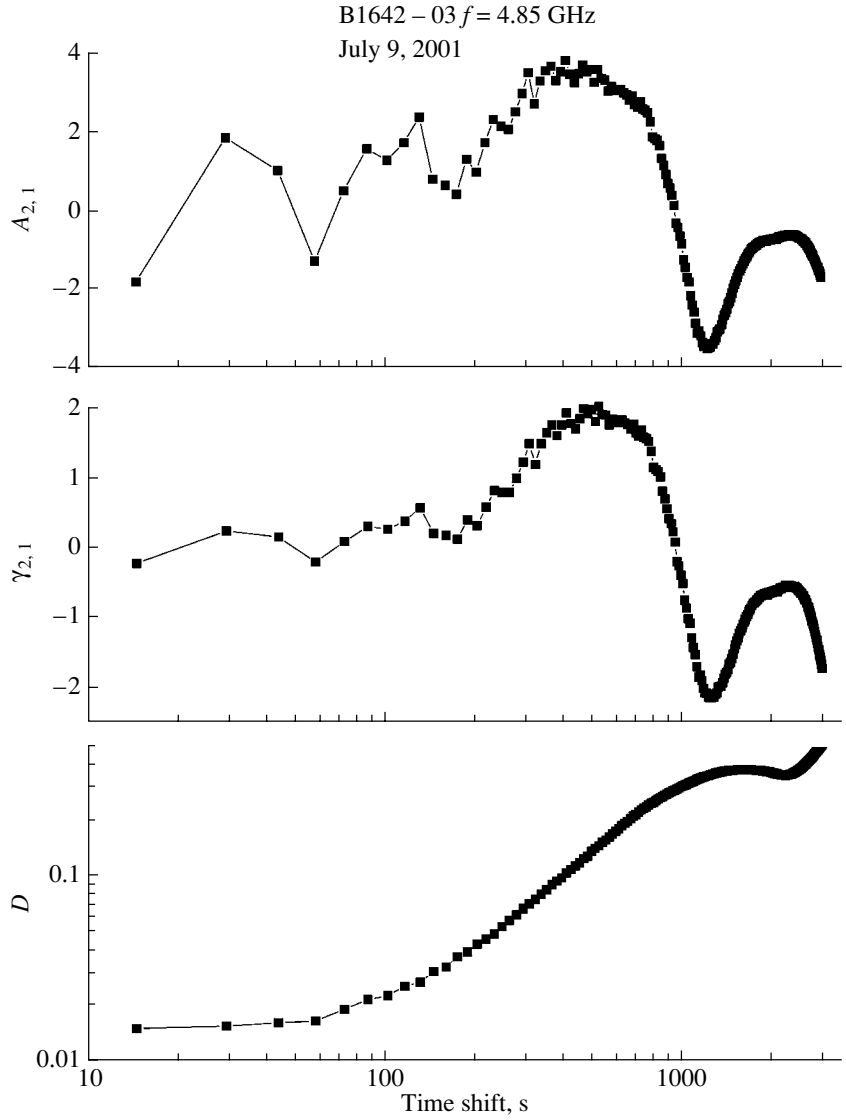

Fig. 5. Same as Fig. 4 but for PSR B1642−03.

PSR B0329+54 and B1642−03 were 0.45 and 0.64, respectively.

Figures 4 and 5 show $D(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$ as functions of the time shift in seconds. The maximum time shift for which these functions were calculated was one-fourth of the total time interval of the observations. The scintillation time scales for PSR B0329+54 and B1642−03 are 2200 and 1170 s, respectively (at the $1/e$ level of their autocorrelation functions). On time scales of 1000−2000 s for PSR B0329+54 and 400−700 s for PSR B1642−03, the values of $A_{2,1}(\tau)$ are 2.8 and 3.3, respectively. These agree well with the value $A_{2,1} = 3$ predicted by the theory for weak pulsations [formula (16)]. The constant level at the shortest time shifts in the structure function $D(\tau)$ is due to uncorrelated noise.

We did not correct the calculated functions for noise, because, as was shown by analysis, their third moments are nonzero, and this correction would introduce strong distortions to $\gamma_{2,1}(\tau)$ and $A_{2,1}(\tau)$. The constant level $A_{2,1} \approx 2$ at time shifts $\tau \leq 200$ s for PSR B0329+54 is probably due to diffractive scintillations, whose time scale at 610 MHz is $t_d \approx 90$ s [13]. For PSR B1642−03, $t_d \approx 130$ s at the same frequency [14]. For this pulsar there are strong fluctuations in $A_{2,1}(\tau)$ at shifts shorter than 200 s, due to the effect of noise; however, the mean level of about 1.5 is probably associated with diffractive scintillations on small-scale inhomogeneities.

Diffractive Scintillations

We used observations of PSR B1133+16 carried out on September 27 and 29, 2001, on the Large Phased Array of the Lebedev Physical Institute in

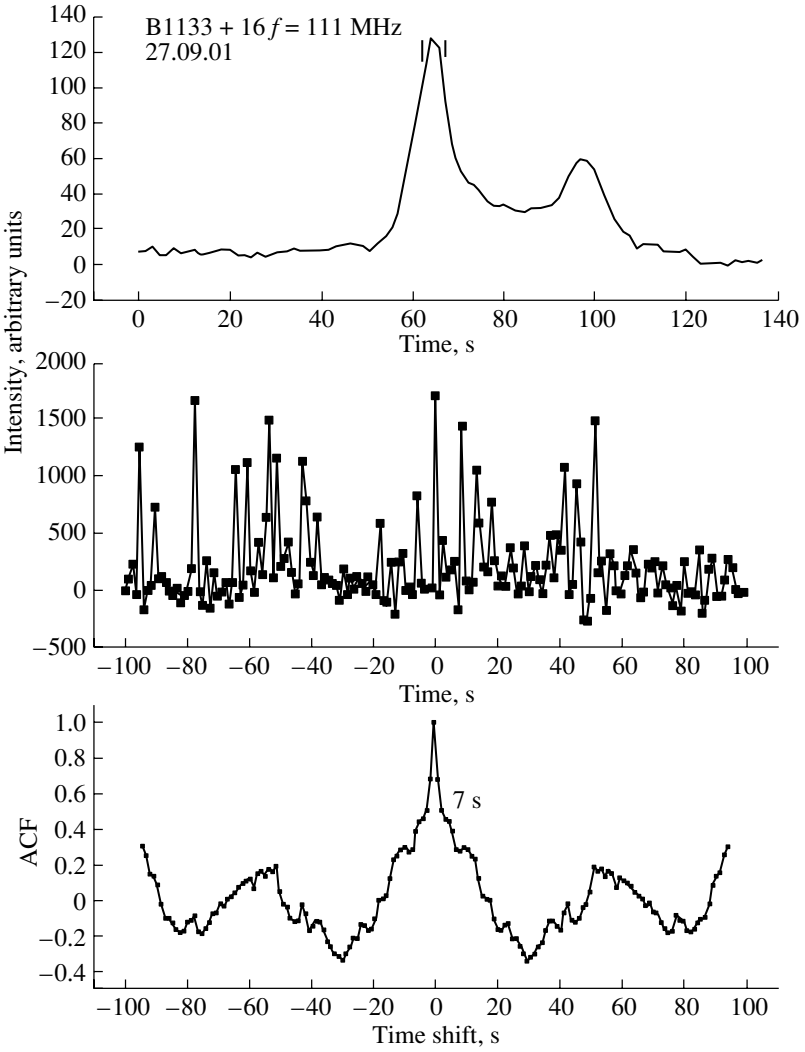

Fig. 6. Top: average profile of PSR B1133+16 at 111 MHz. Center: pulse-to-pulse intensity variations in a single 1.25-kHz channel (for each pulse the intensity was obtained by integrating within the longitude interval shown with dashes on the average profile). Bottom: autocorrelation function of these intensity variations.

Pushchino at 111 MHz to analyze the intensity variations of this pulsar using the asymmetry function. Individual pulses of the pulsar were recorded in 96 channels of the receiver (the bandwidth of each channel was 1.25 kHz) with a time resolution of 1.56 ms during 200 s, and were stored on a computer disk. We performed the structural analysis after removing the dispersion shift in all frequency channels and integrating the spectrum in the selected interval of pulse longitudes. Figure 6 (top) shows the mean pulse of PSR B1133+16, obtained by accumulating pulses during 200 s in a 120-kHz band. The same figure (center) presents the pulse-to-pulse intensity variations in one of the channels (1.25 kHz). The intensity was obtained for each pulse in all the channels by integrating the signal in the longitude interval of the first component of the mean profile, which is denoted in Fig. 6 with dashes. The same figure (bottom) shows the autocorrelation function (ACF) calculated for the intensity fluctuations $I(t)$ in this channel. The ACF has three time scales: the first is due to pulse-to-pulse variations within one pulsar period ($P_1 =$ 1.188 s); the second, which is about 7 s, is due to diffractive scintillations; and the third represents slower variations on a time scale of $\sim$20 s. The fact that the 7-s time scale is due to diffractive scintillations follows from the fact that the decorrelation of the spectra (96 channels 1.25 kHz each) occurs precisely on this time scale (with pulses separated by $6\,P_1$). To reduce the effect of intrinsic rapid variations of the pulsar, we smoothed the input $I(t)$ data in all the channels on a three-point interval ($3\,P_1$). When calculating $D(\tau)$, $\gamma_{2,1}(\tau)$, and $A_{2,1}(\tau)$, we averaged the second and third moments of $I(t)$ over all frequency

channels. Figure 7 shows the obtained functions averaged over two observational days. We can see that the different time scales are well separated in the functions $\gamma_{2,1}(\tau)$ and $A_{2,1}(\tau)$. The value of $A_{2,1}(\tau)$ on scales $\leq$7 s is 1.4, in good agreement with the theoretical value of 1.5, corresponding to diffractive scintillations.

7. CONCLUSION

Our analysis of observed pulsar intensity variations using the asymmetry function $\gamma_{2,1}(\tau)$ and function $A_{2,1}(\tau)$ has shown that these variations match the behavior expected for a normal distribution for field fluctuations (diffractive scintillations) or a log-normal distribution for intensity fluctuations (refractive scintillations, weak scintillations), in accordance with the derived theoretical relationships. For all scintillation regimes, there is a functional relationship between the asymmetry function and the structure function of the intensity fluctuations. This relationship has the simplest form for the asymmetry function $\gamma_{2,1}(\tau)$ at small values of τ: $\gamma_{2,1}(\tau) \propto \sqrt{D_{2,1}(2\tau)}$. Accordingly, the function $A_{2,1}(\tau) = \langle I \rangle \dfrac{\gamma_{2,1}(\tau)}{\sqrt{D_I(2\tau)}}$ is a constant. This indicates that analysis of the function $A_{2,1}(\tau)$ could provide a good test for isolating scintillations against the background of variations that have other origins, and that the proposed method can be efficiently used to study and separate out intensity variations that have different physical natures. In particular, it can be applied to the analysis of flux variations of extragalactic sources, in order to distinguish variations due to intrinsic variability of the source and effects associated with the propagation of radiation in interstellar plasma.

We note also that there are prospects for further studies in this direction. It is necessary to study the dependence of the form of the asymmetry function on the spectrum of inhomogeneities in the index of refraction, in particular, for a power-law spectrum, and the dependence of the asymmetry function on the spectral index. Studies of the dependence of this form on the source structure are also required. If such theoretical calculations confirmed by experimental data are available, measurements of the asymmetry function can yield additional information about both the medium and source.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 03-02-16522 and 03-02-16509), INTAS (grant no. 00-849), the NSF (grant no. AST 0098685), the Federal Science and Technology Program in Astronomy, and the Program of the Division of Physical Sciences of the Russian Academy of Sciences on Fundamental Studies of Extended Objects in the Universe.

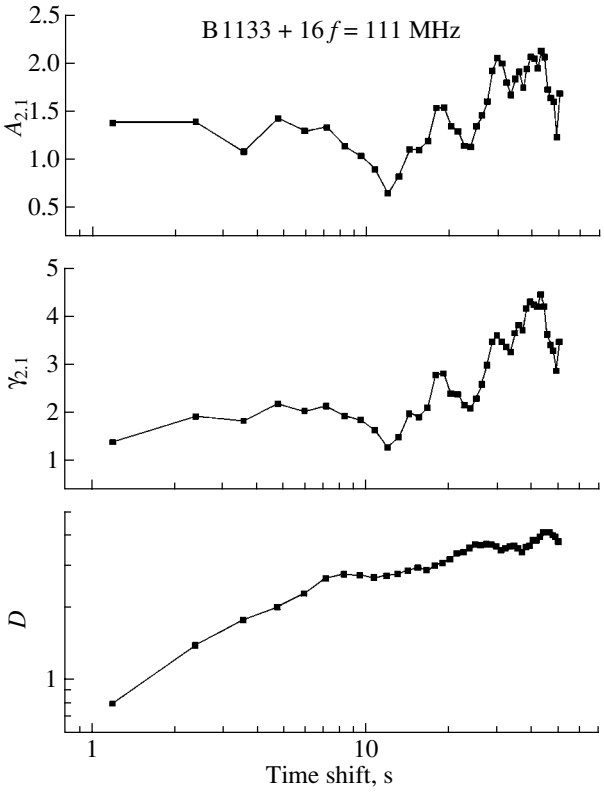

Fig. 7. Structure function, asymmetry function, and coefficient $A_{2,1}$ versus time shift for PSR B1133+16 at 111 MHz averaged over two observational days (diffractive scintillations).

APPENDIX A

THEORETICAL RELATIONSHIP BETWEEN THE ASYMMETRY FUNCTIONS $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$ AND THE STRUCTURE FUNCTION $D_I(\tau)$ FOR A NORMAL DISTRIBUTION FOR THE FIELD FLUCTUATIONS

Let us assume that the wave-field fluctuations have a normal distribution and that the average value of the field is zero:

$$\langle E(t) \rangle = 0. \qquad \text{(A.1)}$$

This equality implies that second-order moments of the form $\langle E^2(t_1) \rangle$ are also equal to zero [9]:

$$\langle E^2(t_1) \rangle = \langle [E^*(t_1)]^2 \rangle = 0. \qquad \text{(A.2)}$$

Formulas (A.1) and (A.2) correspond to diffractive scintillations in the saturated scintillation regime [9].

In this case, the third moment of the intensity or the sixth moment of the field can be represented as a sum of products of second-order field-coherence functions [15]:

$$\begin{aligned}\langle I(t_1)I(t_2)I(t_3)\rangle &= \langle E(t_1)E^*(t_1)E(t_2)E^*(t_2)E(t_3)E^*(t_3)\rangle \\ &= \langle E(t_1)E^*(t_1)\rangle\langle E(t_2)E^*(t_2)\rangle\langle E(t_3)E^*(t_3)\rangle \\ &+ \langle E(t_1)E^*(t_2)\rangle\langle E(t_2)E^*(t_1)\rangle\langle E(t_3)E^*(t_3)\rangle \\ &+ \langle E(t_1)E^*(t_3)\rangle\langle E(t_2)E^*(t_2)\rangle\langle E(t_3)E^*(t_1)\rangle \\ &+ \langle E(t_1)E^*(t_1)\rangle\langle E(t_2)E^*(t_3)\rangle\langle E(t_3)E^*(t_2)\rangle \\ &+ \langle E(t_1)E^*(t_2)\rangle\langle E(t_2)E^*(t_3)\rangle\langle E(t_3)E^*(t_2)\rangle \\ &+ \langle E(t_1)E^*(t_3)\rangle\langle E(t_2)E^*(t_1)\rangle\langle E(t_3)E^*(t_2)\rangle \\ &= \langle I\rangle^3 + \langle I\rangle[B_I(t_1-t_2)+B_I(t_1-t_3) \\ &+ B_I(t_2-t_3)] + B_E(t_1-t_2)B_E(t_2-t_3) \\ &\times B_E(t_3-t_2) + B_E(t_1-t_3)B_E(t_2-t_1) \\ &\times B_E(t_3-t_2).\end{aligned} \tag{A.3}$$

Here,

$$B_E(t_i-t_k) = \langle E(t_i)E^*(t_k)\rangle \tag{A.4}$$

is the field-coherence function and $B_I(t_i-t_k) = \langle[I(t_i)-\langle I(t_i)\rangle][I(t_k)-\langle I(t_k)\rangle]\rangle = B_E(t_i-t_k)\times B_E^*(t_i-t_k)$ is the correlation function of the intensity fluctuations. Using formula (A.3) in the calculation in (7), we obtain a relationship for the third moment of the second difference:

$$\begin{aligned}\langle[\Delta_2(\tau)]^3\rangle &= -12\langle I\rangle^3 + 12\langle I\rangle B_I(2\tau) \\ &+ 24\langle I\rangle B_I(\tau) - 24B_E(2\tau)B_I(\tau) = -6\langle I\rangle D_I(2\tau) \\ &+ 24\langle I\rangle\left\{\langle I\rangle - \left[\langle I\rangle^2 - \frac{1}{2}D_I(2\tau)\right]^{1/2}\right\} \\ &- 12D_I(2\tau)\left\{\langle I\rangle - \left[\langle I\rangle^2 - \frac{1}{2}D_I(2\tau)\right]^{1/2}\right\}.\end{aligned} \tag{A.5}$$

Let τ_0 be the characteristic scale of the structure function of the intensity fluctuations $D_I(\tau)$. For small values, $\tau \le \tau_0$, we obtain the asymptotic formula

$$\langle[\Delta_2(\tau)]^3\rangle \cong \frac{3}{4}D_I(2\tau)\frac{4D_I(\tau)-D_I(2\tau)}{\langle I\rangle}. \tag{A.6}$$

For large values, $\tau \gg \tau_0$, using equality $D_I(\tau) = 2\langle(I-\langle I\rangle)^2\rangle = 2\langle I\rangle^2$ yields

$$\langle[\Delta_2(\tau)]^3\rangle \cong -12\langle I\rangle^3. \tag{A.7}$$

Using these asymptotic relationships for $\langle[\Delta_2(\tau)]^3\rangle$ and formulas (6) and (8), we derive for the function $\gamma_2(\tau)$ the asymptotic expressions

$$\gamma_2(\tau) \cong \begin{cases}\dfrac{3}{4}\dfrac{D_I(2\tau)}{[4D_I(\tau)-D_I(2\tau)]^{1/2}\langle I\rangle}, & \tau \le \tau_0, \\ \dfrac{2}{\sqrt{6}}, & \tau \gg \tau_0.\end{cases} \tag{A.8}$$

Similarly, we have

$$\gamma_{2,1}(\tau) \cong \begin{cases}\dfrac{3}{2}[D_I(2\tau)]^{1/2}/\langle I\rangle, & \tau \le \tau_0, \\ 2\sqrt{2}, & \tau \gg \tau_0.\end{cases} \tag{A.9}$$

We see that, for small values of τ, $\gamma_{2,1}(\tau)$ is related to $[D_I(2\tau)]^{1/2}/\langle I\rangle$ by the same formula as that relating the asymmetry coefficient to the scintillation index.

We obtain for $A_{2,1}(\tau)$

$$\begin{aligned}A_{2,1}(\tau) = \langle I\rangle\frac{\gamma_{2,1}(\tau)}{[D_I(2\tau)]^{1/2}} \cong \frac{3}{2}, \quad \tau \le \tau_0, \\ A_{2,1}(\tau) \cong 2, \quad \tau \gg \tau_0.\end{aligned} \tag{A.10}$$

APPENDIX B

THEORETICAL RELATIONSHIP BETWEEN THE ASYMMETRY FUNCTIONS $\gamma_2(\tau)$, $\gamma_{2,1}(\tau)$ AND THE STRUCTURE FUNCTION $D_I(\tau)$ FOR A LOGNORMAL DISTRIBUTION FOR INTENSITY FLUCTUATIONS

Let us represent the intensity $I(t)$ as

$$I(t) = I_0\exp[\chi(t)] \tag{B.1}$$

and assume that the fluctuation distribution law $\chi(t)$ is normal. For a normally distributed quantity, the following equality is valid [8]:

$$I_0\langle\exp[\chi(t)]\rangle = I_0\exp\left[\langle\chi(t)\rangle + \frac{1}{2}\langle(\delta\chi(t))^2\rangle\right]. \tag{B.2}$$

Taking into account the fact that the mean intensity does not change during the propagation of the wave in a turbulent medium, i.e., $\langle I\rangle = I_0$, we obtain

$$\langle\chi(t)\rangle = -\frac{1}{2}\langle(\delta\chi(t))^2\rangle. \tag{B.3}$$

We find with (B.2) and (B.3)

$$\begin{aligned}\langle I^2\rangle &= \langle I_0^2\exp[2\chi(t)]\rangle \\ &= I_0^2\exp\left[2\langle\chi(t)\rangle + 2\langle(\delta\chi(t))^2\rangle\right] \\ &= I_0^2\exp\left[\langle(\delta\chi(t))^2\rangle\right],\end{aligned} \tag{B.4}$$

$$\langle I(t)I(t+\tau)\rangle = \exp\left[\langle\delta\chi(t)\delta\chi(t+\tau)\rangle\right] = \langle I^2\rangle \exp\left[-\frac{1}{2}D_\chi(\tau)\right], \quad \text{(B.5)}$$

where

$$D_\chi(\tau) = \langle[\chi(t+\tau)-\chi(t)]^2\rangle = 2\langle[\delta\chi(t)]^2\rangle - 2\langle\delta\chi(t)\delta\chi(t+\tau)\rangle \quad \text{(B.6)}$$

is the structure function of the fluctuations, $\chi(t)$.

Formulas (B.4) and (B.5) enable us to express the quantities $\exp\left[\langle(\delta\chi(t))^2\rangle\right]$ and $\exp\left[-\frac{1}{2}D_\chi(\tau)\right]$ in terms of the scintillation index and the structure function of the intensity fluctuations:

$$\exp[\langle(\delta\chi(t))^2\rangle] = 1 + m^2, \quad \text{(B.7)}$$

$$\exp\left[-\frac{1}{2}D_\chi(\tau)\right] = 1 - \frac{D_I(\tau)}{2I_0^2(1+m^2)}. \quad \text{(B.8)}$$

Then, using the formula

$$\langle I(t_1)I(t_2)I(t_3)\rangle = I_0^3 \exp\left[3\langle(\delta\chi)^2\rangle - \frac{1}{2}D_\chi(t_1-t_2) - \frac{1}{2}D_\chi(t_1-t_3) - \frac{1}{2}D_\chi(t_2-t_3)\right] = I_0^3(1+m^2)^3 \times \exp\left[-\frac{1}{2}[D_\chi(t_1-t_2) + D_\chi(t_1-t_3) + D_\chi(t_2-t_3)]\right], \quad \text{(B.9)}$$

we obtain for the third moment of the second difference $\Delta_2(\tau)$

$$\langle[\Delta_2(\tau)]^3\rangle = -6I_0^3 \exp\left[3\langle(\delta\chi)^2\rangle\right] \times \left\{1 - \exp[-D_\chi(2\tau)] - 2\exp[-D_\chi(\tau)] + 2\exp\left[-D_\chi(\tau) - \frac{1}{2}D_\chi(2\tau)\right]\right\}. \quad \text{(B.10)}$$

Using (B.4) and (B.5), we find

$$\langle[\Delta_2(\tau)]^3\rangle = -6I_0^3(1+m^2)^3 \times \left\{1 - \left[1 - \frac{D_\chi(2\tau)}{2I_0^2(1+m^2)}\right]^2 + 2\frac{D_\chi(2\tau)}{I_0^2(1+m^2)} \times \left[1 - \frac{D_\chi(\tau)}{2I_0^2(1+m^2)}\right]^2\right\}. \quad \text{(B.11)}$$

For small $m^2 \ll 1$, we obtain the asymptotic relationship

$$\langle[\Delta_2(\tau)]^3\rangle \cong \frac{3}{2}D_I(2\tau)\frac{4D_I(\tau) - D_I(2\tau)}{I_0}. \quad \text{(B.12)}$$

Substituting this asymptotic relationship into (8)–(10) yields for $m^2 \ll 1$

$$\gamma_2(\tau) \cong \frac{3}{2}\frac{D_I(2\tau)}{[4D_I(\tau) - D_I(2\tau)]^{1/2}I_0}, \quad \text{(B.13)}$$

$$\gamma_{2,1}(\tau) \cong 3\left[D_I(2\tau)\right]^{1/2}/I_0, \quad \text{(B.14)}$$

$$A_{2,1}(\tau) \cong 3. \quad \text{(B.15)}$$

REFERENCES

1. V. I. Shishov, T. V. Smirnova, and S. A. Tyul'bashev, Astron. Zh. **82**, 281 (2005) [Astron. Rep. **49**, 250 (2005)].
2. G. Bourgois, Astron. Astrophys. **21**, 33 (1972).
3. G. Bourgois and G. Cheynet, Astron. Astrophys. **21**, 26 (1972).
4. J. W. Armstrong, W. A. Coles, and B. J. Rickett, J. Geophys. Res. **77**, 2739 (1972).
5. V. I. Shishov and A. V. Pynzar', Izv. Vyssh. Uchebn. Zaved., Radiofiz. **18**, 506 (1975).
6. M. Nakagami, in *Statistical Methods in Radio Wave Propagation,* Ed. by W. C. Hoffman (Pergamon, New York, 1960), p. 3.
7. R. P. Mercier, Proc. Cambridge Phylos. Soc. **58**, 382 (1962).
8. V. I. Tatarskii, *Wave Propagation in a Turbulent Medium* (Nauka, Moscow, 1967; McGraw-Hill, New York, 1961).
9. A. M. Prokhorov, V. F. Bunkin, K. S. Gochelashvily, and V. I. Shishov, Proc. IEEE **63**, 790 (1975).
10. R. J. Hill, R. G. Frehlich, and W. D. Otto, NOAA Tech. Memo. ETL-274, NOAA Environmental Research Laboratories (Boulder, Colorado, 1996).
11. R. J. Hill and R. G. Frehlich, J. Opt. Soc. Am. A **14**, 1530 (1997).
12. D. R. Stinebring, T. V. Smirnova, T. H. Hankins, *et al.*, Astrophys. J. **539**, 300 (2000).
13. V. I. Shishov, T. V. Smirnova, W. Sieber, *et al.*, Astron. Astrophys. **404**, 557 (2003).
14. T. V. Smirnova, V. I. Shishov, W. Sieber, *et al.*, Astron. Astrophys. (2005) (in press).
15. S. M. Rytov, *Introduction to Statistical Radio Physics* (Nauka, Moscow, 1976) [in Russian].

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 916–931. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 1025–1041.

Hard X-ray and Gamma-Ray Flares on the Sun: Stereoscopic Effects near the Limb from Observations on the 2001 *Mars Odyssey* Spacecraft and Near-Earth Spacecraft

M. A. Livshits[1], V. A. Chernetskiĭ[1], I. G. Mitrofanov[2], A. S. Kozyrev[2], M. L. Litvak[2], A. B. Sanin[2], V. I. Tret'yakov[2], W. Boynton[3], K. Shinohara[3], and D. Hamara[3]

[1] *Pushkov Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Moscow, Russia*

[2] *Space Research Institute, Moscow, Russia*

[3] *Arizona State University, Arizona, USA*

Received April 20, 2004; in final form, May 18, 2005

Abstract—We present the first results of data on solar flares detected by the HEND instrument onboard the 2001 *Mars Odyssey* interplanetary spacecraft during its flight to Mars and in orbit around Mars. The instruments carried by the spacecraft, which was developed at the Space Research Institute of the Russian Academy of Sciences, included a scintillation detector with two crystals, enabling the detection of photons with energies from tens of keV to 2.5 MeV with high time resolution. Several dozen flares were detected on both the sides visible from the Earth and back side of the Sun, supplementing other available data in a number of cases. A joint analysis of the HEND data together with data obtained in near-Earth orbits enabled the detection of variations in the integrated fluxes of photons with energies exceeding 80 keV during observations of flares near the limb from various directions. Two events were analyzed in great detail: the setting of a region displaying frequent very short flares on May 20, 2001, and the rising of the group 10486, which displayed numerous flare phenomena on the limb followed by extremely high activity in October–November 2003. These variations appear in simultaneous observations of limb M flares made at angles differing by only 8°–10°. Analyses of observations of rising sources obtained on two spacecraft lead to similar results. This indicates that the vast majority of emission at energies exceeding 80 keV arises at altitudes of no higher than seven to ten thousand kilometers. We briefly consider the powerful solar-disk gamma-ray flare of August 25, 2001. In this case, there are some differences in the behavior of the hard radiation in the decay phase for observations made at angles differing by 25°, which is most likely due to differences in the instrumental responses to radiation with this spectrum. The absence of hard radiation at great heights in the region of the "cusp" places some constraints on our picture of the physical processes occurring in powerful solar flares. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The high-energy radiation from solar flares contains information about the main energy-release processes and particle-acceleration mechanisms. A review of previous studies, beginning in 1980, using the SMM satellite, is given in [1]. Further, observations of powerful solar flares with $h\nu \geq 300$ keV were carried out using various instruments of the Compton Gamma-ray Observatory, the FEBUS instrument onboard the GRANAT observatory, the Yohkoh BGO detector, and instruments of the GAMMA-1 observatory. For the most powerful events, the gamma-ray emission is produced over 10 min, and some events are accompanied by weak flares occurring over a period of one to several hours. This was found to be the case for flares in June 1991 [2], but is most likely also true for other very powerful events.

This behavior of the gamma-ray emission is probably associated with the fact that the most efficient acceleration of particles on the Sun—apparently of protons as well as electrons—likewise occurs on time scales of several minutes, near spots in the flash phase of large flares; i.e., at the end of or just after impulsive events, at the onset of the formation of coronal flare loops. This conclusion was recently made based on a direct identification of a source of accelerated relativistic particles in the flare of July 14, 2000, based on multifaceted observations and a collection of indirect evidence for other powerful flares during the three last solar cycles [3]. This brought to the floor the question of whether data on the high-energy radiation of flares can be used to reach some conclusions about the sources of the main particle acceleration—their locations and the duration of the processes leading to the generation of high-energy particles in flares.

1063-7729/05/4911-0916$26.00

The 2001 *Mars Odyssey* satellite was launched on April 7, 2001. Onboard equipment included a Gamma-Ray Spectrometer (GRS) with a High Energy Neutron Detector (HEND). The main task of this satellite is to study Mars. However, the GRS and two scintillation HEND can also be used to detect hard X-ray and gamma-ray radiation from solar flares at various points in the solar system. Thus far, several dozen solar flares have been detected during the flight to Mars from April 7 to October 24, 2001, and during the satellite's orbiting of Mars. The HEND data supplement available data on the hard radiation of solar flares from observations in near-Earth orbits, helping to fill gaps in these data and providing the possibility of studying flares and the activity of groups on the side of the Sun that is not visible from the Earth. In some cases, the flare emission extends to 1.5–2.5 MeV, making it possible to study the physical conditions in high-energy flare sources on the Sun.

Indeed, these data provide a good indicator of particle acceleration on the Sun. This was facilitated by the fact that the energy range of the HEND detector enabled the detection of particles in the transition region from X-ray emission by the beams of electrons with energies of tens of keV that are often present in flares to gamma-rays that arise as a result of interactions between relativistic particles and the surrounding plasma. In the simplest case, the HEND detector can be used like two photometers that detect photons with energies higher than 80 and higher than 330 keV. While the former range includes radiation due to the large number of electrons accelerated to energies of about 100 keV, the second range includes only emission associated with the generation of solar cosmic rays. In addition, in many cases, scintillation and proportional counters enable the detection of not only gamma-rays, but also of solar high-energy particles. This also facilitates studies of the acceleration of particles on the Sun during their propagation in the heliosphere.

In addition, various regions of the solar surface have been observable by HEND at various times. After the *Mars Odyssey* launch, the angle between the Sun and the Earth and the Sun and the satellite changed. For example, in 2002, the satellite was located in a position opposite that of the Earth. The simultaneous use of the *Mars Odyssey* data and data obtained in near-Earth orbits enabled studies of sources of high-energy radiation from different directions. Such "stereoscopic" observations open new possibilities for investigations of flares. This topic is the main focus of the present study.

As is known, stereoscopic observations carried out in 1978–1979, using the Pioneer Venus Orbiter (PVO) and International Sun–Earth Explorer 3 (ISEE-3) spacecraft outside the Sun–Earth line elucidated that 95% of the radiation with energies exceeding 150 keV is produced at heights of less than 2500 km above the photosphere [4]. At the same time, analysis for three flares led to the conclusion that the corresponding radiation arose near the bases of flare loops. Subsequently, numerous images of flares at energies up to 50 keV were obtained, in which sources near the bases of loops were clearly distinguished. However, a third site near the loop apex is also sometimes visible at these energies, as was first identified in [5, 6]. In several impulsive events, the height of this site was close to 10 000 km. A high source in the region of the so-called "cusp" is more often observed in soft X-ray images (see below).

Unfortunately, the poor resolution of images of flares at hard energies has prevented investigation of the structure of the sources, in particular, their heights. The directivity of flare gamma-rays was widely discussed in the 1970s. The main argument for this was the concentration toward the limb of flares with appreciable fluxes of protons with energies exceeding 10 MeV [7, Fig. 9]. This was interpreted as indicating the presence of photons arising near the apices of loops during collisions between relativistic electrons and thermal ions, with the predominant emission of photons in the direction of propagation of the particle beam, i.e., parallel to the solar surface. A softening of the spectra in the transition from limb flares to central flares is sometimes observed at soft X-ray energies (tens of keV) [8], taken to provide confirmation of this view.

However, a subsequent statistical study of 28 flares detected by the SIGNE instrument onboard the *Venera-13* and *Venera-14* spacecraft and the HXRBS instrument onboard the SMM spacecraft demonstrated a near absence of directivity of the radiation at energies from 50 to 500 keV—the ratio of the fluxes did not exceed a factor of 2.5 (see [9] and references therein). This question has now again become topical in connection with the need to interpret the first results obtained by the RHESSI satellite [10].

In principle, the data analyzed here contain information about several large events for which photons with energies up to 2.5 MeV were detected, which can be used to investigate the directivity of this radiation. The high time resolution of these data is an important factor, enabling analysis of the directivity for individual impulsive events or post-eruptive episodes.

Observations of powerful solar flares on the disk have not revealed appreciable differences from the results obtained with HEND and in modern observations in near-Earth orbits. However, such comparisons are not trivial for events near the limb, where stereoscopic effects are most strongly manifested. In the first stage, we aimed to estimate the heights of sources with energies exceeding 80 keV. We then

focused our attention on analyzing the set and rise of two groups of spots, on May 20, 2001, and October 22–23, 2003. These groups were accompanied by numerous fairly hard flares. We briefly also discuss the "typical" powerful gamma-ray flare of August 25, 2001, for which HEND data were obtained, together with other numerous observational material. Interpretation of the results enables refinement of the conditions under which hard radiation arises in solar flares.

2. DESCRIPTION OF THE INSTRUMENT

The HEND instrument was developed and constructed at the Institute for Space Research of the Russian Academy of Sciences, and was intended for studies of neutron and gamma-ray emission from the Martian surface. It is one element of the GRS apparatus [11] of the NASA 2001 *Mars Odyssey* project, whose aim was a global study of the elemental composition of the Martian surface via gamma-ray and neutron spectroscopy. The GRS complex includes

—the GRS with a cooled Germanium detector operating at energies of 30–8000 keV;

—a neutron spectrometer (NS) for measuring the flux of thermal and epithermal neutrons;

—the HEND for measuring the flux of neutrons with energies from 0.4 eV to 15 MeV.

HEND has three ^{3}He proportional counters for the detection of epithermal neutrons and one scintillation block for the detection of energetic neutrons and gamma-rays. We will consider here data obtained using the scintillation block, whose main detecting element is a crystal of organic scintillating material.

It is necessary to distinguish counts from recoil protons from those made by charged cosmic-ray particles. For this, the scintillator is surrounded by an outer detector with anticoincidence protection based on a CsI crystal. When counts from particles in the outer detector are recorded, the signal is processed and used to reject the corresponding events detected in the inner scintillator.

In addition to the charged particles detected in the inner and outer scintillators, the instrument detects gamma-ray photons. These photons penetrate freely through the anticoincidence protection, so that their counts in the inner scintillator will be mixed with counts from recoil protons. An electronic analog division scheme is used to distinguish the counts from neutrons and photons, based on a comparison of the shapes of the pulses generated in the photomultiplier by the electron produced in the case of a detected photon or the proton produced by the detection of a neutron. Thus, the output of the inner scintillator block forms two independent sets of counts: protons from neutron detections and electrons from gamma-rays. Experimentation showed that the probability of erroneous identification of these counts is very small, $<10^{-3}$.

Thus, two HEND signals arise during the detection of gamma-rays: photons from the outer CsI crystal, which are determined as the counts from this crystal for which there is no simultaneous signal in the inner scintillator, and photons in the inner crystal, which are determined as "electron" counts in this crystal for which there is no simultaneous signal in the outer CsI crystal. The possible energy ranges for gamma-rays that can be registered in the inner and outer devices are 60 keV–2 MeV and 30 keV–1 MeV, respectively.

Two continuous profiles of gamma-ray measurements in the inner scintillator and outer CsI crystal are formed in the instrument, with time resolutions of 1.0 and 0.25 s, respectively. Simultaneously with the recording of these continuous time profiles, a spectrum of gamma-ray counts is formed every 20 s in 16 energy channels in each scintillator.

The data generated each 20 s are transferred to the GRS Central Electronic Board (CEB) and then to the onboard telemetry system, which transmits them to the Earth during tracking sessions. When the HEND data are written to the CEB, they are tied to an onboard timing system with accuracy to better than 1 ms. In turn, the *Mars Odyssey* onboard time system is tied to universal time (UT) with the same accuracy. In addition, positions of the spacecraft for both the flight from the Earth to Mars and in the subsequent orbits around Mars are derived with an accuracy of several tens of meters based on trajectory measurements.

Knowledge of the precise time during the detection of cosmic gamma-ray bursts and solar flares enables the use of the HEND data to localize the sources of the gamma-ray bursts and in stereoscopic observations of the Sun. Our analysis is concerned with the latter task, based primarily on time profiles obtained with the outer detector with a time resolution of 0.25 s at energies of 80 keV–1 MeV and with the inner detector with a resolution of 1 s at energies of 330 keV–2.5 MeV. For brevity, we will refer to these profiles as the signals from the X-ray and gamma-ray photometers.

3. METHODOLOGICAL COMMENTS ON STEREOSCOPIC OBSERVATIONS

Two approaches have been used when analyzing observations near the limb which are schematically illustrated in Fig. 1. First, if some formation is fully visible from one of the spacecraft at a given time and

is not visible from the other, its height h cannot exceed the height given by the relation

$$\cos\alpha = \frac{R_0}{R_0 + h},$$

where α is the angle between the two spacecraft (Fig. 1a) and R_0 is the radius of the Sun.

Second, in early X-ray studies of the Sun, the heights of soft X-ray sources were estimated based on observations of their rising and setting. This approach can be reformulated for the case of hard X-ray observations with two spacecraft as follows. Let us suppose that the rising of a flare is observed from a spacecraft near Mars at a time t_1 and from a spacecraft in a near-Earth orbit at a time t_2, after which a limb source is observed from the Earth. With knowledge of the angle

$$\beta = \Omega(t_2 - t_1),$$

where Ω is the angular velocity of rotation of the region at its known heliolatitude, determination of the angle α and estimation of the source height become more reliable. In the case of a very low source, the angle β turns out to be equal to the difference in the longitudes of the two spacecraft. High sources can begin to be observed when they are still behind the limb of the first spacecraft (Mars in Fig. 1b), and arrive at the limb for observations from the Earth after the expected time corresponding to the equality of β and the difference of the longitudes of the two spacecraft. In a number of cases, information from the second spacecraft has made it possible to determine the expected time for the appearance of a flare source behind the limb of a rising group (as was the case at the end of October 2003) which can then be used to estimate the height in the case of simultaneous observations (Fig. 1a). It is obvious that this reasoning operates well in the case of a nonvariable source or observations of similar events that repeat fairly often. Such series of homologous events are sometimes observed, one of which is discussed in the following section.

We calculated the longitudes of the spacecraft in the plane of the ecliptic on the path to Mars using the flight ephemerides; when the spacecraft was in orbit around Mars, we took the longitude of the spacecraft to be equal to that of Mars. As an example, Fig. 1c presents the positions of the Earth and Mars in their orbits on October 28, 2003, at 12^{h} Moscow time. In this case, when the ecliptic longitude of the Earth exceeded that of Mars, the *Mars Odyssey* spacecraft was able to observe regions on the Sun beyond the eastern limb for an Earth observer; flares in the eastern part of the disk were located closer to the disk center for the HEND observations compared to their position for an Earth observer. The cases discussed below refer to precisely this relative position of the spacecraft, except for the very beginning of the flight, when the situation was reversed.

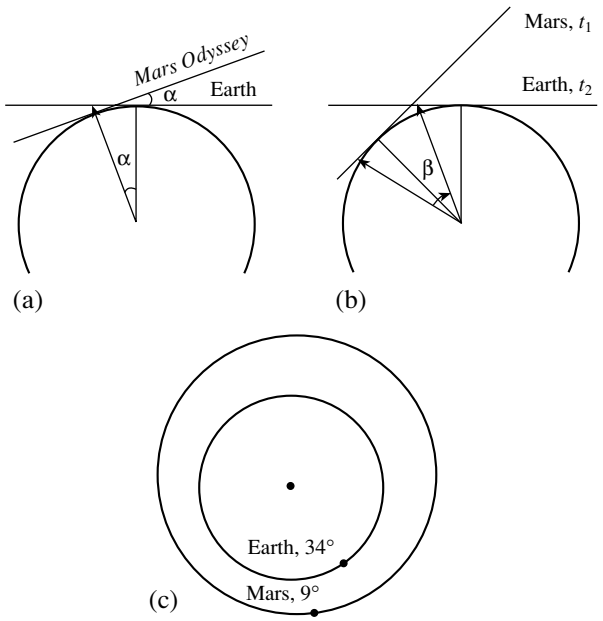

Fig. 1. Schematic depicting the estimation of the height of a source with the use of two instruments: (a) the spacecraft located in the plane of the ecliptic, separated in longitude by an angle α; (b) observations conducted at various times t_1, t_2 (the angle $\beta = \Omega(t_2 - t_1)$, where Ω is the rotational angular velocity of the active region during observations from the Earth); (c) an example of the mutual locations of the two spacecraft (end of October 2003).

4. SETTING OF THE FLARE ACTIVE REGION OF MAY 20, 2001

At the beginning of the flight, HEND detected a flare on May 20, 2001. The observations were carried out in approximately the same direction as those from near-Earth orbit, but, naturally, certain difficulties of subsequent observations of the Sun from near Mars had not yet arisen. Figure 2 shows the variations of the signal of the outer scintillator in the flare-detection regime (time resolution 0.25 s). A hard X-ray flare with its maximum at 6:03 UT (times everywhere have been recalculated to those for an observer in near-Earth orbit) was rapidly followed by the arrival of accelerated protons. We can see that the amplitude of the hard X-ray flare was only slightly smaller than the amplitude of the signal due to the accelerated particles—primarily protons with energies exceeding 60 MeV. In this case, the very rapid arrival of the accelerated particles to both the boundary of the magnetosphere and to the *Mars Odyssey* spacecraft is due to the favorable (western) position of the flare on the Sun and the very quiet conditions in the interplanetary medium on May 20, up until this event. Protons with energies of tens of MeV moved essentially along the force lines of the interplanetary magnetic field connecting the flare with the IMP-8,

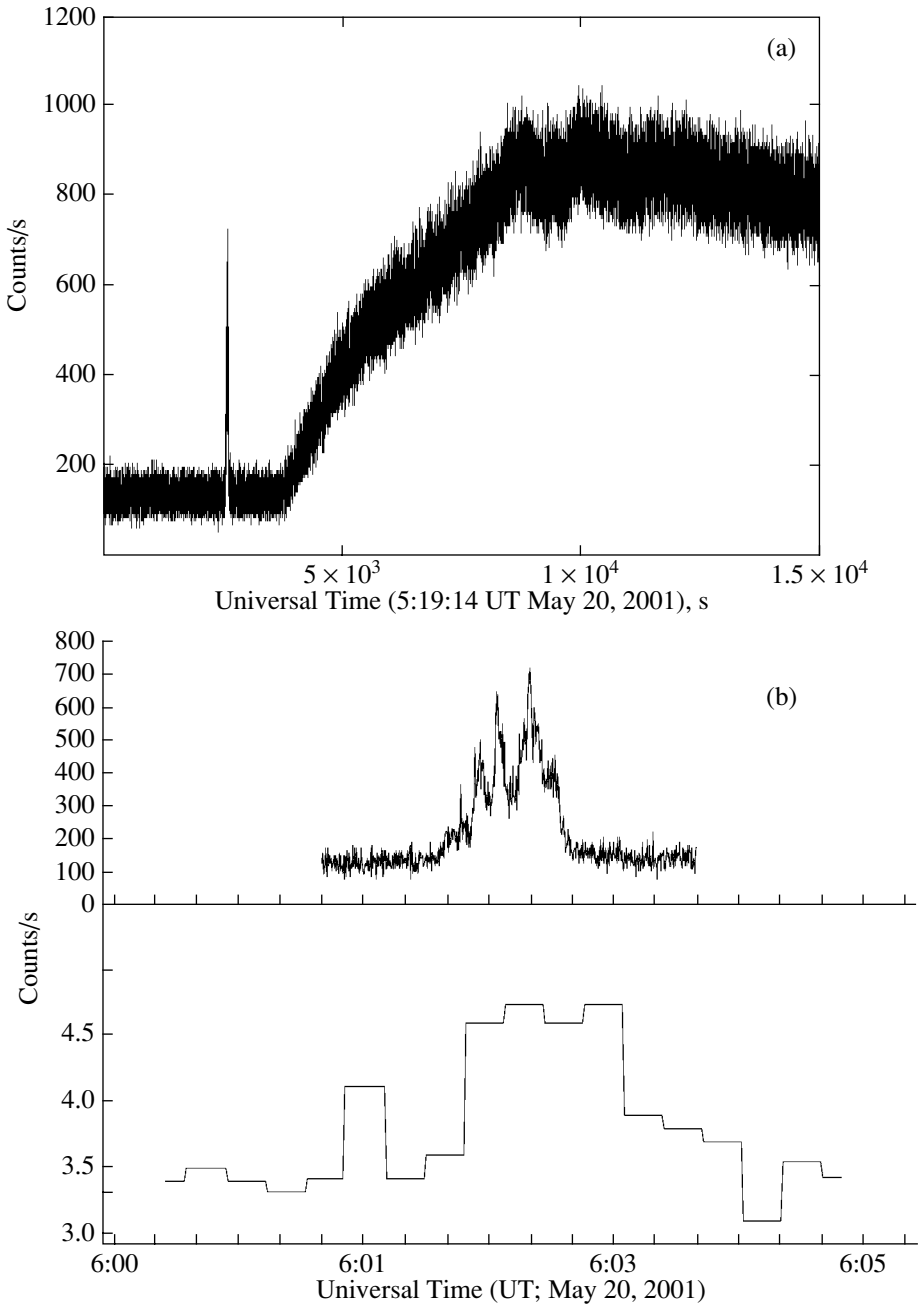

Fig. 2. (a) Count rate for the outer scintillator for the flare of May 20, 2001. The time resolution is 0.25 s. We can see that the amplitude of the hard X-ray flare does not exceed the instrumental response to solar protons with energies exceeding 60 MeV. (b) Count rates for the outer and inner scintillators for the flare of May 20, 2001, after 6^h UT. The resolution of the hard X-ray data is 0.25 s. The gamma-ray emission is presented as the sum of the counts in all the spectral channels; the time resolution here is 19 s.

GOES, and *Mars Odyssey* spacecraft, and were detected by instruments onboard these three spacecraft approximately 20 min after their acceleration. The corresponding time profiles of the accelerated protons turned out to be similar for all three spacecraft, and the fluxes of 10 MeV protons were close to 10 s.f.u., while the fluxes of protons with energies of tens of MeV approached 1 s.f.u. Analysis of the 1100 currently known proton enhancements observed in near-Earth orbit shows that the event of May 20 represents one of the most widespread cases. This is associated with the fact that precisely flares with X-ray strengths of about M5 are related to the weakest events that give rise to accelerated protons that arrive at the magnetosphere [12], and such sources of accelerated protons are the most numerous.

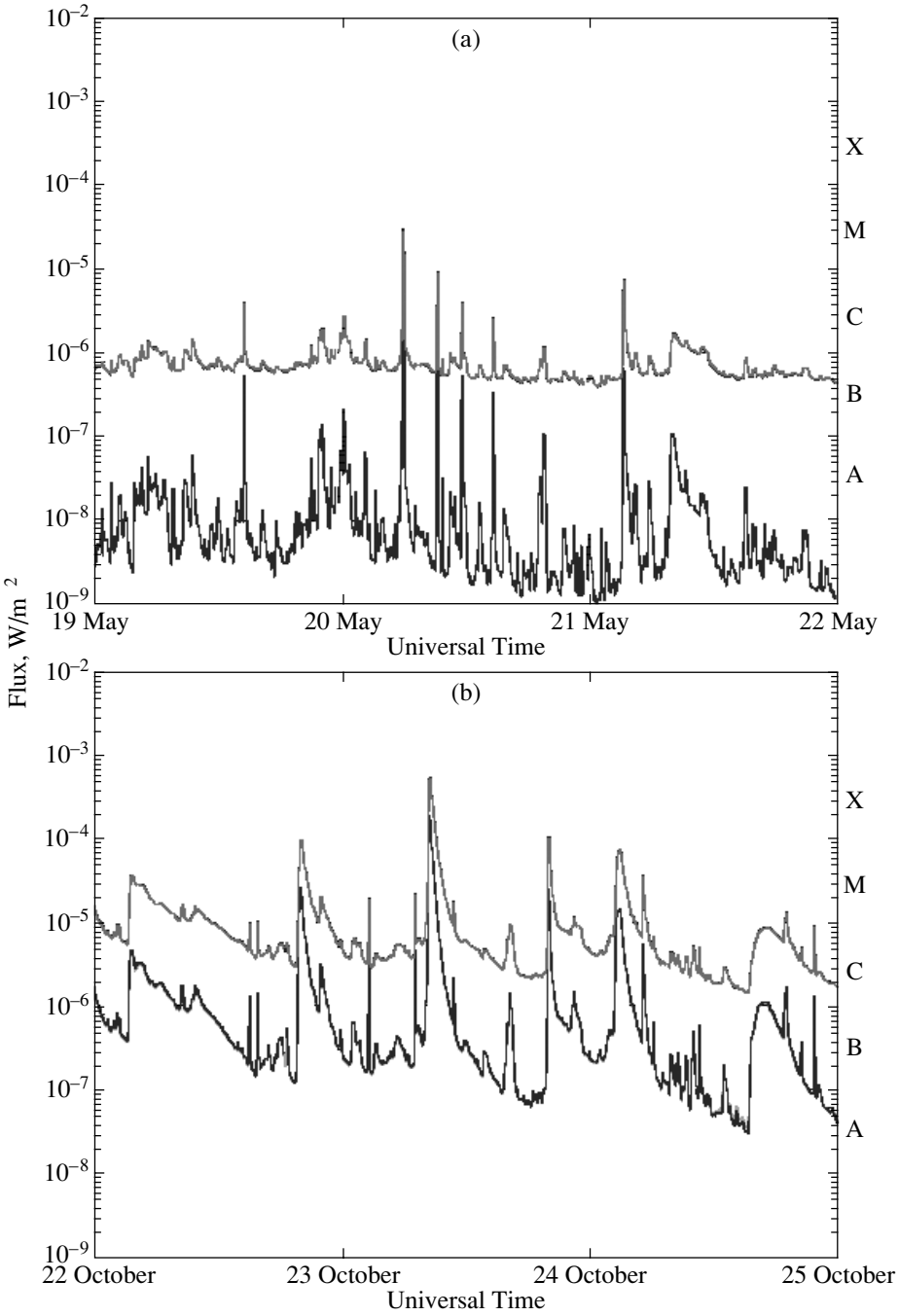

Fig. 3. (a) Setting of an activity center behind the limb on May 20, 2001, and (b) rising of another flare group on October 22, 2003, based on observations in the GOES 1–8 and 0.5–4 Å bands.

Figure 2b shows the responses of the outer and inner scintillators (i.e., the X-ray and gamma-ray photometers) in the flare-detection regime. The high time resolution of the X-ray flare enables the detection of structure: all the events are divided into four to five individual "elementary" events, each of which lasts about 10 s. Such fine structure is observed more often in events near the limb.

When analyzing the gamma-ray emission with a time resolution of 1 s, the useful signal was only weakly distinguished, but became more firmly detected in a 19 s integration regime, reaching approximately 80 photons in the entire flare. It is surprising that this relatively strong signal was not detected by Yohkoh, especially in the corresponding channels of the HXT X-ray telescope, which operated at the time of the flare and has a sensitivity that is somewhat

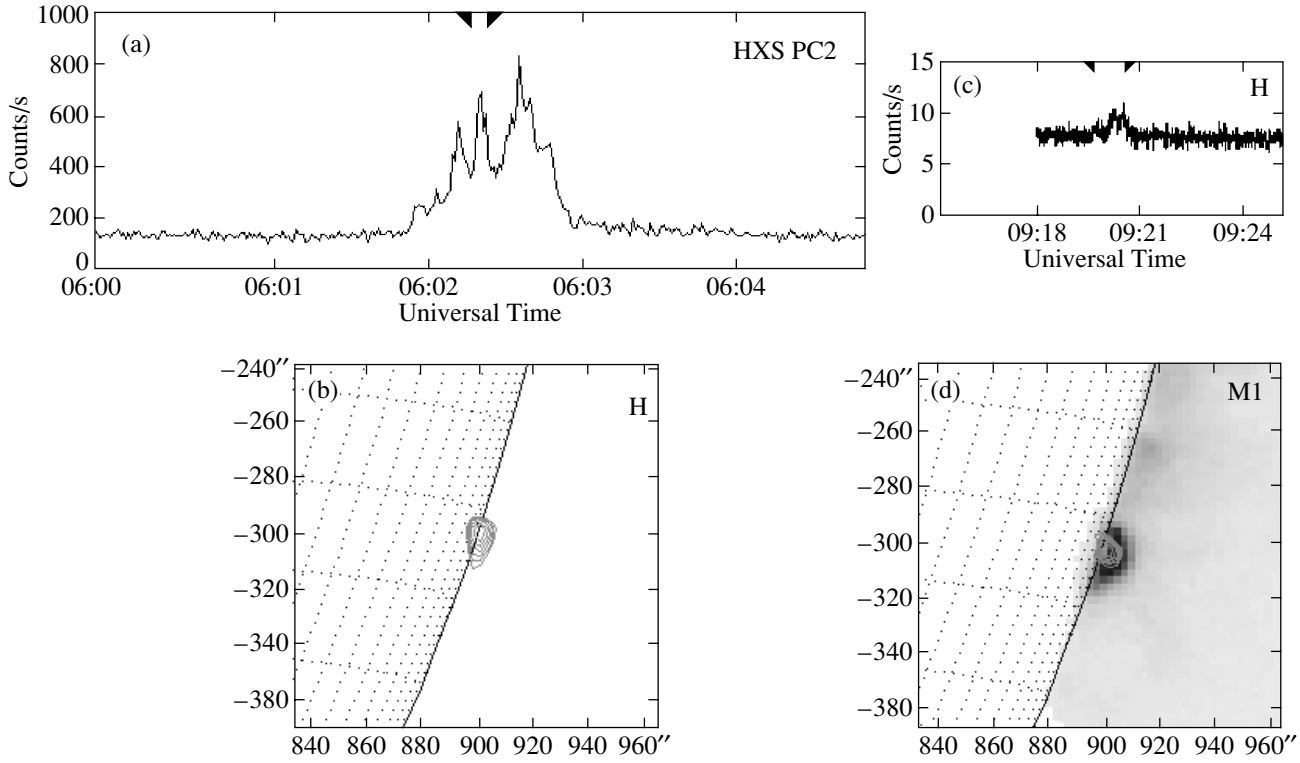

Fig. 4. Left: (a) time profile 75–872 keV in counts per second and (b) image in the Yohkoh XHT H channel (53–93 keV) for the flare of May 20, 2001, near 6^h UT. The time interval for the image is indicated on the upper axis of the flare time profile. Right: (c) time profile in the H channel (53–93 keV) in counts per second per channel and (d) image in the Yohkoh HXT M1 channel (23–33 keV) for the flare of May 20, 2001, near 9^h UT (it was not possible to construct an image in harder channels). The time interval for the image is indicated on the upper axis of the flare time profile.

higher than HEND (this flare was at the detection limit of the Yohkoh gamma-ray telescope).

When analyzing this discrepancy, it became clear that we were dealing with a very rare case of the setting beyond the limb of a group of spots in which rapid (or, as it has recently become popular to call them, "compact") flares arise with some regularity. The analyzed flare was the first, most powerful, flare of a series of events associated with the evolution of this group when it was on the side of the Sun hidden from the Earth. Instruments in near-Earth orbits operating in the soft X-ray and ultraviolet detected photons generated only at the apices of flare loops. This series is also clearly manifest in the regular data taken by the GOES satellite. All five flares visible in Fig. 3, which began on May 20, 2001, at 6:20 UT (M6.4), 9:12 UT (M1.5), 11:35 UT (C6.1), 14:32 UT (C2.9), and 19:30 UT (C1.3), occurred at the western limb during the setting of group 9455.

All these flares are characterized by certain peculiarities, which make them good candidates for gamma-ray flares. For example, the duration of the soft X-ray emission of the first flare of May 20, 2001, was only 6 min, while the growth phase was only 3 min (excluding a precursor at 5:46 UT, which is not shown in the figure). Note that several rapid flares occurred during the passage of this same group across the visible disk of the Sun, in particular, the M3.6 flare of May 13, 2001, beginning at 3:00 UT. This can be considered an example of the half of M flares that do not give rise to proton enhancements at near-Earth orbits and fluxes of hard radiation which are not accessible to measurements with HEND.

Usually, such powerful, bright flare systems of events as the flare beginning at about 6^h UT on May 20, 2001, last an hour or more. The simplest explanation for the brief duration of the flares in this series is that the primary energy release in these flares was so great that the magnetic field could not confine the forming system of loops. However, the idea that the system was completely disrupted is not always in agreement with the observations—often it is simply the conditions for prolonged existence that are disrupted. For example, at the end of the ejection of the flare of May 20, when a shock (type II radio flare) emerged into the middle corona from 6:06 to 6:26 UT, the SOHO data display a wide front with a cavity at its center and with strongly twisted formations at

the center and arcades; the large-scale streamer was disrupted. However, the post-eruptive development of this event that is usual for large flares did not occur.

Analogous phenomena with somewhat lower powers were observed in the second flare, which began at 9:12 UT and lasted only 11 min. Type-II radio emission was already detected at large distances at 9:28 UT. The variations of the soft X-ray flux for the last flares in the series and analysis of the radio data suggests that all the flares in the series were homologous (all the characteristic properties of the processes for the first two events were virtually identical).

The first two flares were observed by the Yohkoh HXT instrument (Fig. 4). The most striking thing about them is the extreme similarity of the main properties of the time profiles for the flare of May 20, 2001, obtained for the two instruments, which differed only in their time resolution (Figs. 2b, 4a). The similarity in the absolute values here is due to a chance coincidence in the count rates; in reality, the absolute flux corresponding to the HEND measurements is somewhat higher than the flux for the Yohkoh HXT data. However, after the HEND and HXT X-ray photometer signals were over, the gamma-ray photometer went on to observe another maximum. This delay in the hard emission by approximately 20 s is manifested in the data averaged to 19 s, as can be seen by comparing Figs. 2b (top) and 4a with Fig. 2b (bottom).

This difference between the HEND and HXT data can be expressed in another way. The flare of May 20, 2001, was already powerful at about 6:00, and involved a large number of accelerated particles. Nevertheless, in observations carried out in near-Earth orbit, the spectrum of photons with energies of 50–300 keV displayed a power-law index of −4.3 (according to the HXT spectrum). Such powerful flares usually have harder spectra, with indices from −3 to −2, especially in the case of limb events. This provides additional support for the reality of the inferred differences.

Note that this flare occurred precisely at the limb for an observer at the Earth. This follows, for example, from the SOHO 171 Å frames at 13^h UT, which show a large loop that is moving behind the limb in the setting part of the active region. In other words, the second base of the loop was probably already located on the side of the Sun that is hidden from an observer in near-Earth orbit; one of the flares with energy exceeding 330 keV was detected from the *Mars Odyssey*, but was not able to be observed from near-Earth orbit.

The second flare was observed on the Yohkoh HXT instrument to energies of about 60 keV, and could already not be distinguished very clearly in the standard H channel (53–93 keV). It was possible to construct images only in two of the four energy channels: L (14–23 keV) and M1 (23–33 keV); the latter image is presented in Fig. 4d. The decrease in the fluxes compared to the first flare is associated with both its lower power and its setting behind the limb over a time of only three hours. Note that, in this case, there is some evidence for a hardening of the spectra of flares near the limb at 20–100 keV, as can be noted in a statistical analysis of hard flares. The flares of May 20, 2001, at 9:12 UT (M1.5), 11:35 UT (C6.1), and probably 14:32 UT (C2.9) were also observed by the *Mars Odyssey*, but the detection of accelerated particles from the western flare at this time prevents a quantitative comparison of the time profiles for the X-ray emission detected by the two spacecraft.

Thus, even observations carried out using a single spacecraft contain some information about the location of the X-ray source. Applied to the case considered of a well defined setting source of rapid flares, these soft X-ray data (Fig. 3) enable us to estimate the height of ordinary loops associated with weak compact flares to be $h \leq 20\,000$ km. The Yohkoh HXT data can be used to analyze not only the "ideal setting" of the sources on May 20, 2001, but also some set of such cases. A preliminary analysis of this material suggests that the height of the sources of radiation with energies exceeding 50 keV is lower than the value of 20 000 km for the sources of softer photons. It is clear that the main emission in this range of energies arises near the loop bases, and the question is what contribution is made by high sources near the cusps of typical large prolonged flares.

The variety of geometries and other properties of flares hinders detailed studies of the setting of sources behind the solar limb. The presence of two spacecraft, preferably with identical instruments, fundamentally changes the situation, enabling simultaneous observations of a single event from different directions. The difference in the longitudes in the ecliptic plane for the *Mars Odyssey* spacecraft and the Earth was 8.3° at the time of the maximum of the May 20, 2001 flare. The flare, which was located precisely at the limb for the near-Earth observations, was observed by HEND on the visible disk 8.3° from the limb. In spite of the small difference in the spacecraft longitudes, when the source is located precisely at the limb for one of the instruments, the effect of setting behind the limb will already be manifest for the other. In our view, the most probable explanation is that the second base of a high loop associated with two parts of a bipolar group was not visible in the near-Earth observations. This testifies to a low height for the source, which can be estimated using the method depicted in Fig. 1a to be about 7300 km. Clearly, we are dealing here with a part of the source that is invisible to the near-Earth

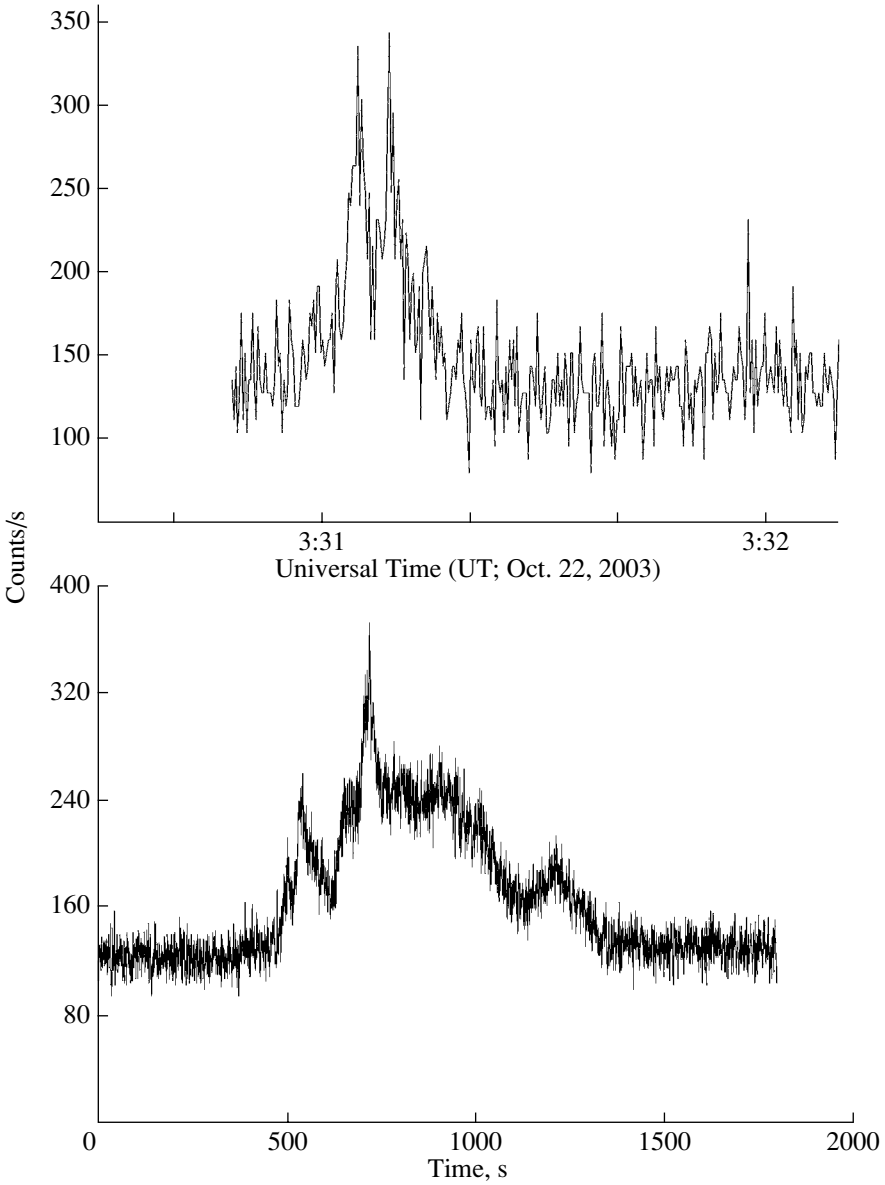

Fig. 5. (a) X-ray flare of October 22, 2003; the time resolution is 0.25 s. (b) The same for the October 24, 2003. Time in seconds, the beginning of the time scale is 2:36:41 UT.

observations and is localized near the base of the loop, rather than at the loop apex.

5. EVENTS DURING THE RISING OF GROUP 10486 IN OCTOBER 2003, BASED ON STEREOSCOPIC OBSERVATIONS

At the end of October 2003, in the decay phase of the 23rd solar cycle, a period of high flare activity began, which has been actively studied with both ground-based observatories and spacecraft in near-Earth orbits. At the beginning of this period, October 23, 2003, the *Mars Odyssey* spacecraft in its orbit around Mars detected nonstationary events on the Sun, which were located 24° to the west of their position for observations from the Earth. Thus, the rising of the groups, above all group 10486, was observed from a near-Mars orbit earlier than from the Earth.

This case is opposite to that discussed above in terms of both type of activity and observing conditions. While the setting of a source of a series of homologous, very short flares was observed in May 2001, in this case, we have the appearance from behind the limb of a source that was transformed over one to two days into a very complex site displaying powerful activity. A large number of weak, nonsta-

tionary processes were detected at the eastern limb south of the solar equator in the soft X-ray range by near-Earth instruments. Against the background of the two most powerful flares, which began at 3:28 UT (M3.7) and 19:47 UT (M9.9) on October 22, 2003, a number of weaker events were observed, such as those starting at 8:30 UT (M1.7), 9:37 UT (M1.7), 15:57 UT (M1.2), and 21:56 UT (Fig. 3b). Another difference from the case of May 2001 is that, the observations of these nonstationary events near the eastern limb did not reveal accelerated solar particles traveling in the direction of the spacecraft. Accordingly, there is no flux of particles hindering the detection of X-rays and gamma-rays, as was the case for the western flare of May 20, 2001.

At the end of October 2003, the activity on the disk was determined primarily by group 484, which preceded group 486 (here and below, we omit the 10 before the group number). The effect of a connection between a weak flare in group 484 and the beginning of activity in the following group, 486, can already be seen in the structure of a powerful coronal mass ejection on October 21, 2001 at 4:28 UT at the eastern limb 16° south of the equator. Our data indicate that the enhanced level of soft X-ray emission after 18^h UT on October 21, 2001 (shown by the GOES data) is clearly related to the activity of group 486. Indeed, a large flare with its maximum at 18:41 UT and lasting 12 min in the hard X-ray range was observed by the *Mars Odyssey* spacecraft. The count rate in the outer scintillator reached a maximum of 300 s^{-1}, and the radiation in the first peak remained appreciable at energies of 100–200 keV (see [13] for more detail). Traces of this flare, which realistically could have reached magnitude X but occurred far beyond the limb for an Earth observer, are visible on the 195 Å SOHO image obtained at 18:48 UT.

The following flare occurred at the end of the soft X-ray enhancement that began at about 18^h UT on October 21, 2003, at about 3:30 UT on October 22. According to the GOES data, its class reached M3.7, but in reality the flare was somewhat more powerful due to the invisibility of the dense bases of the flare loops from near-Earth orbit. This also follows from a 195 Å SOHO film, in which the flare loops in the southern part of the region move beyond the limb, in the direction of the main flare site. The time profile of the hard X-ray emission from the data obtained by the HEND outer scintillator is presented in Fig. 5a. There is no corresponding signal in the hard X-ray channels, beginning with energies of about 300 keV, or in the gamma-ray channel. In the data taken in near-Mars orbit (Fig. 5a), two peaks separated by 4–5 s, can be distinguished. This time structure is typical for high-energy phenomena that develop in low loops associated with impulsive events.

This source of hard X-ray emission was located beyond the limb for observations in near-Earth orbit and, thus, so was not visible. Several very weak events associated with phenomena in both groups were observed by HEND on October 22, 2003, but they were softer than the similar weak events on subsequent days. For comparison, the time profile of one of these is also presented in Fig. 5b. The flare of October 24 at 2:54 UT (Fig. 5b) was somewhat more powerful than the flare of October 22 (M7.6); the two X-ray flares were similar, but the hard radiation for the later flare continued for more than 10 min rather than only tens of seconds.

The X-ray emission from group 486 was first detected firmly in near-Earth orbits was on October 22, 2003 at 19:45 UT at energies near 40 keV. The CORONAS-F monitoring data testify that the detected fluxes were lower than expected for a M9.9 flare, suggesting that the bases of the loops may have been located beyond the solar limb (a more detailed comparison of CORONAS-F data for this flare and data taken in near-Mars orbit can be found in [13]).

Thus, the influence of group 486 began to be felt at soft X-ray energies at the end of October 21, 2003. The situation here is analogous to the setting of the soft source on May 20, 2001: the soft radiation is observed beyond the limb far longer than the hard radiation. In reality, the near-Earth spacecraft CORONAS-F and RHESSI detected at hard X-ray energies the final stage of a large flare that occurred on October 23.

The first large X5.4 flare of a series of powerful events from the end of October through November 4, 2003, occurred on October 23 at about 8:30 UT. Two main maxima are observed on the X-ray light curve (Fig. 6a), which correspond to two impulsive flare episodes separated by approximately 15 min. These two maxima were also observed on the radio at 10 cm (Fig. 6c), confirming the existence of two stages of the event. The fine time structure of the main X-ray impulse is also well defined in recordings with the maximum time resolution of 0.25 s when recordings of only this radiation are separated out.

This flare was detected in the inner HEND scintillator, as can be seen in Fig. 6b, which presents the gamma-ray count rate averaged over 8 s. A comparison of the X-ray and gamma-ray curves shows that radiation with energies exceeding 1 MeV is clearly distinguished not at the time of the first, most powerful X-ray impulse, but instead at the time of the weaker X-ray impulse that occurred 2 min later. It is most likely that this corresponds to the time when

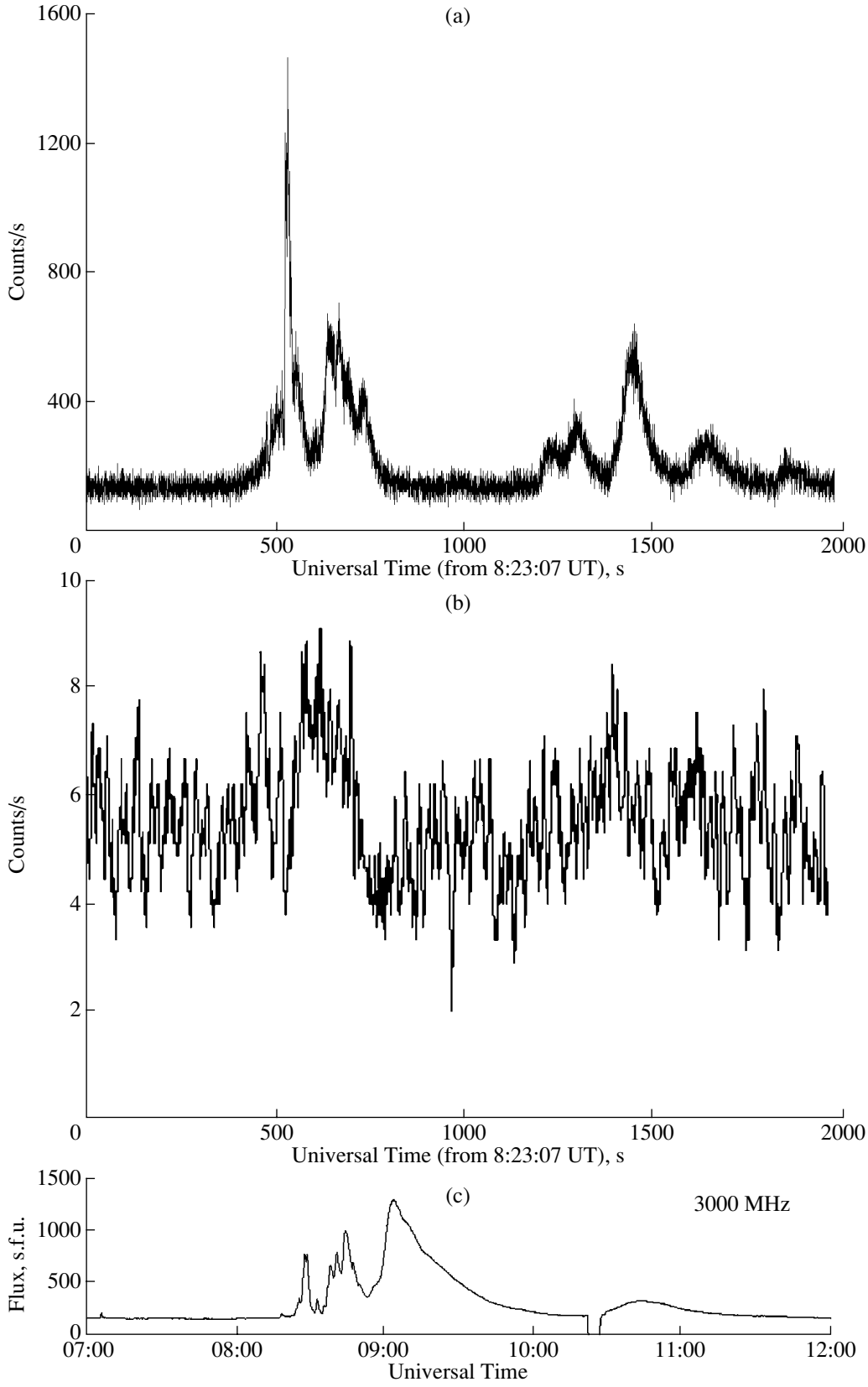

Fig. 6. (a) Hard X-ray emission of the flare of October 23, 2003; the time resolution is 0.25 s, time in seconds. (b) Gamma-ray emission for the same flare and the same time interval, averaged over 8 points. (c) Radio outburst on October 23, 2003, for IZMIRAN data.

the flare particles were accelerated to relativistic energies. The radiation becomes softer at later times, but there is a powerful energy release in both impulsive events. The quasi-stationary development of giant post-eruptive loops whose axis is directed toward another, northeastern, group began after 9^h UT. The development of this post-eruptive phase is clearly expressed and can be traced in both out-of-atmosphere and ground-based data from the ultraviolet to the radio ranges. We are considering this flare in somewhat more detail only because it displayed the main properties of phenomena that were developed in subsequent flares in this series. This event is worthy of a separate, more detailed discussion, all the moreso because other detailed observations from the Earth were hindered by the nearness of the flare to the limb and observations in near-Earth orbits were unsuccessful.

Returning to the main theme of the paper, we assert, based on an analysis of all the available observations, that the activity at the eastern limb of the Sun was first detected on the *Mars Odyssey* spacecraft at photon energies exceeding 80 keV one day earlier than the event was detected in near-Earth orbit. The boundaries of this time interval can be taken to be either the onset of visibility of the hard phenomena to the near-Mars and near-Earth instruments (18:40 UT October 21 and 19:50 UT October 22, 2003), or the movement behind the limb of the group for the two types of spacecraft. This occurred for the Martian spacecraft at 3:30 UT on October 22 (Fig. 5a), while activity at hard energies began to be detected at the beginning of October 23 by the RHESSI and CORONAS-F spacecraft.

Thus, assuming an approximately constant height for the sources of hard radiation associated with flares that are not very strong, we can estimate this height based on simple reasoning (Fig. 1b). The difference in the longitudes of the Earth and Mars at the time of the first flare was close to 22°. The active region shifts by 14° per day in the same direction as the motion of the two planets in their orbits (here, we use data for the siderial rotation of the Sun for an observer on Earth). This means that the base of the second flare was located 8° beyond the limb. Given that the time interval can somewhat exceed one day, it is reasonable to estimate the height of the hard source to be in the range 7000–10 000 km.

6. EVENT OF AUGUST 25, 2001, AS AN EXAMPLE OF A POWERFUL FLARE ON THE DISK

Hard radiation accompanied by particle acceleration on the Sun was detected by both HEND and instruments in near-Earth orbits. First and foremost, we carried out a qualitative comparison of the time profiles of flares occurring on the solar disk observed simultaneously from different directions. For about a dozen cases, such comparisons lead to a surprising coincidence in the shapes of the flares and, most often, fluxes of the radiation for M3–X2 flares. Several examples are presented above. Generally speaking, this supports the conclusion drawn earlier that the radiation of photons with energies exceeding 70 keV for events on the visible disk is not directed (see [9] and references therein).

However, this is not so obvious in the case of the largest events. As an example, let us consider a flare detected during the flight of the *Mars Odyssey* spacecraft to Mars. The powerful flare of August 25, 2001, with well-developed high-energy phenomena occurred on the solar disk at some distance from the southeast limb (class X5.3, onset at the Earth at 16:23 UT, coordinates S17 E34). No flux of accelerated protons was observed from this eastern flare, while a sporadic high-velocity flux of plasma arrived early, after approximately two days. Gamma-ray emission was detected by several spacecraft. In particular, the CORONAS-F and Yohkoh spacecraft detected appreciable fluxes of hard photons with energies up to 40 MeV during this flare [14]. Such events are fairly rare, as can be seen, for example, from an inspection of a list of events detected by FOBOS [15].

Figure 7 presents the results of gamma-ray observations with a time resolution of 1 s. In this case, a gamma-ray signal with a duration of about 100 s and with its maximum at about 16:31:44 UT was firmly detected. We can see that the entire flare represents a superposition of several impulses with durations of 10–30 s each. The time profile repeats the main properties of the data obtained in the HXT Yohkoh H channel. This suggests that we are dealing with the acceleration of individual beams of particles and their subsequent destruction in dense layers. The shapes of the profiles for higher-energy nonthermal radiation remain similar, reflecting the time dependence of the particle acceleration. However, a more detailed comparison shows that the decrease in the flux in time shown by the HEND data occurs somewhat more slowly than in the HXT Yohkoh H channel. A more detailed analysis is required in order to elucidate whether this is related to the effect of directivity of the radiation for observations at angles differing by 25°. Competition arises from the variations in band sensitivity for the different channel energies when incident radiation also becomes harder.

In many respects, the analyzed event can be considered to be a source of a "typical" powerful gamma-ray flare. Indeed, the flare arose at the point where a bright Hα floccule passed between two large spots in a group with opposite polarities (Fig. 8a). This

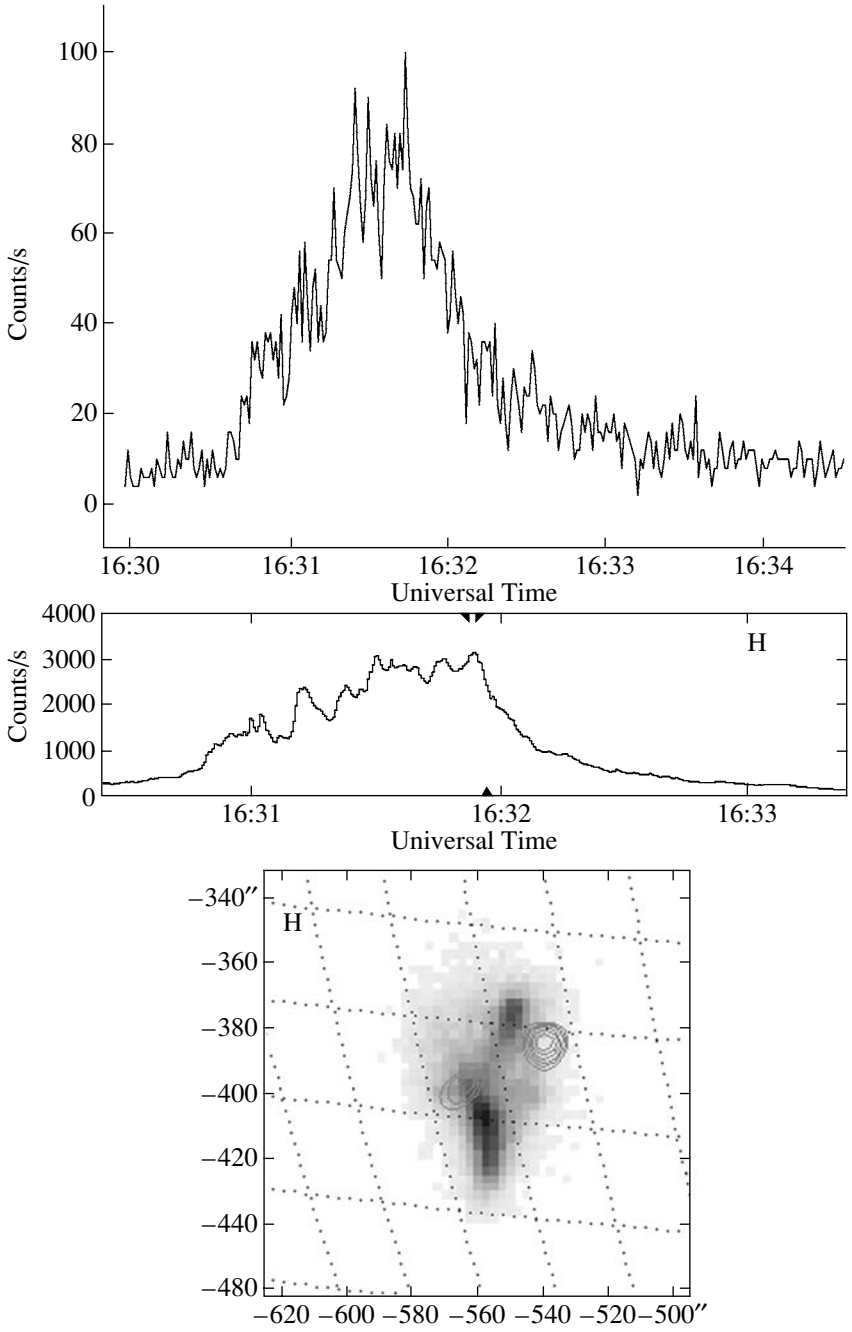

Fig. 7. Count rate for the inner HEND scintillator on August 25, 2001 (upper), time profile for the flare from Yohkoh data in counts/s per channel (middle), and image of the flare in the HXT Yohkoh H channel (53–93 keV) (lower). The map of contours of hard-radiation brightness is shown against the background of an image of the SXT Yohkoh soft radiation. The times when the images were obtained are indicated in the middle plot.

situation is the most favorable for the development of high-energy phenomena, most of which are produced in the penumbra of a single large spot during the rising of small islands with opposite magnetic polarities (i.e., in the presence of large differences in magnetic fluxes of opposite polarity). The extent of the 3B optical flare along the magnetic-field neutral line was modest compared to other flares with comparable powers (for a height of the main loop at the flare maximum of about 20 000–25 000 km; Fig. 7, lower). The compactness of the light sections of the flare in Figs. 8b and 8c is characteristic of gamma-ray flares. The difference images in four lines kindly presented to us by I.M. Chertok enable us to understand the

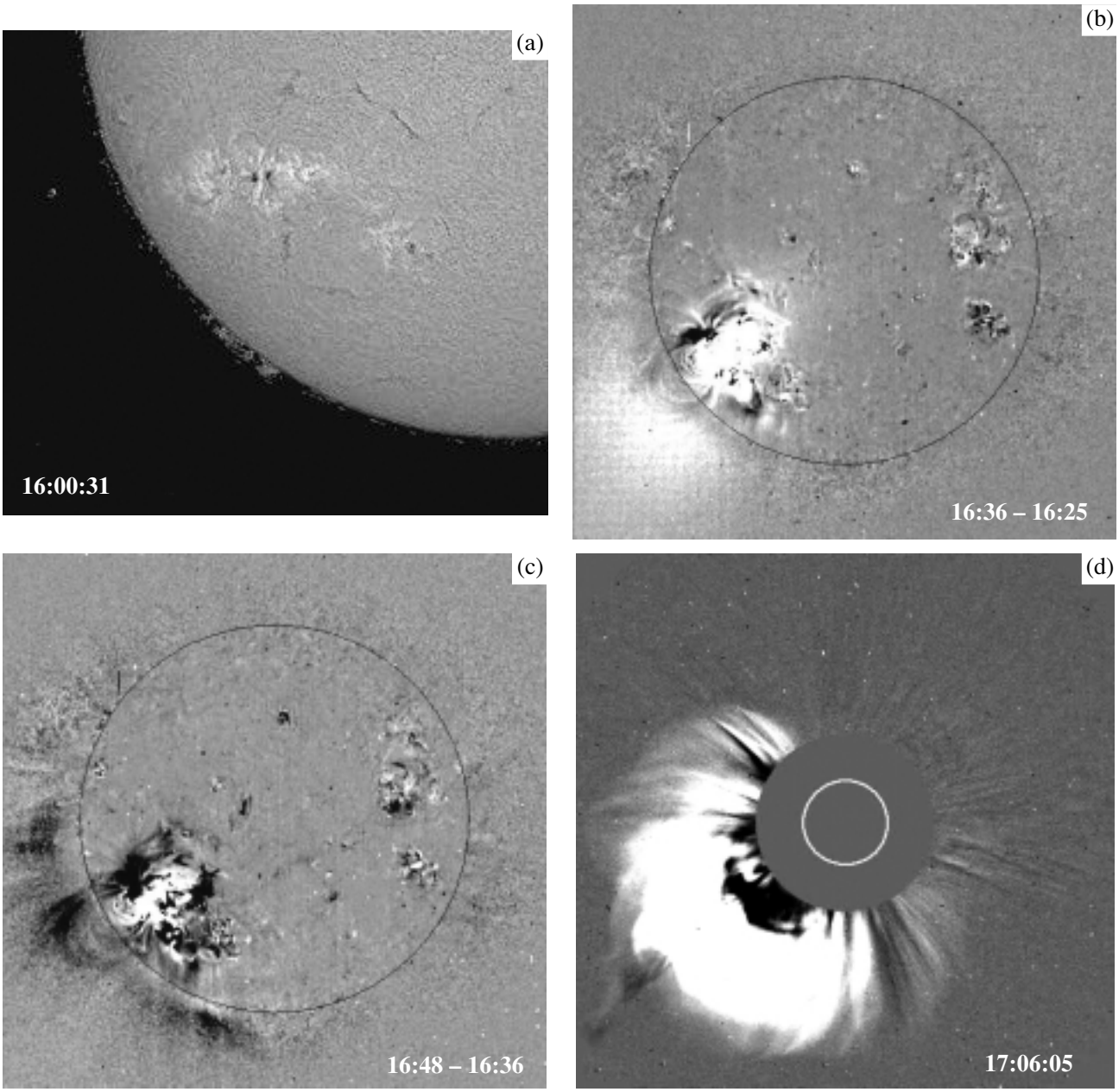

Fig. 8. Flare of August 25, 2001: (a) onset of the flare in Hα (16^h UT); (b) difference image at 195 Å from the SOHO data for the interval of the flare maximum; (c) same as in (b) for the next frame (the darkening/dimming can be noted to the North of the region in which the flare began, which makes a transition to a dark arc above the limb); (d) coronal mass ejection in the middle corona from the SOHO C2 data.

main properties of the plasma ejection, and possibly the outflow of accelerated particles (this analysis is analogous to that applied in [16]). Namely, there are grounds to suppose that a perturbation cone was directed to the Southeast, approximately 30°–40° from the radial direction passing through the flare.

Indeed, the development of the high-energy event of August 25, 2001, lasted over all several minutes. The rapid and dense coronal mass ejection that arose at that time (Fig. 8d) led to the development of the subsequent post-eruptive phase of the flare immediately after the high-energy event (Figs. 8b, 8c). This is quite common, and only in certain powerful events, such as the event of June 15, 1991, is there an interval of 10–20 min between these two phases. As a rule, after a powerful impulse, the system of new flare loops propagates in two directions along the neutral line of the longitudinal magnetic field. The structure of

the flare begins to resemble an S shape, referred to as a "sigmoid flare." One peculiarity of the event of August 25, 2001, is that the ejection was asymmetric: it was stronger toward the Northeast than toward the South. This was manifested via the substantial development of the system of giant arches, which was clearly visible on the eastern limb, as well as the stronger dimming to the Northeast. A corresponding darkening that arose due to the ejection of part of the high loops is clearly visible in Fig. 8c to the northeaster of the spots, in the location of the energy release.

Thus, powerful phenomena accompanying large fluxes of gamma-rays represent a combination of compact formations and the very unusual and prolonged subsequent development of nonstationary processes.

7. RESULTS AND DISCUSSION

We have presented the first results of data on solar flares obtained using the HEND instrument onboard the *Mars Odyssey* interplanetary space station during its flight to Mars and its orbits around Mars. The scientific instrument complex of the station includes the HEND, developed at the Space Research Institute of the Russian Academy of Sciences. The scintillation HEND with two independent crystals enabled the detection of photons with energies from tens of keV to 2.5 MeV. In the first stage, we restricted our use of these data to a regime with two "photometers" operating at more than 80 keV (hard X-ray) and more than 330 keV (gamma-ray), with maximum time resolutions of 0.25 and 1 s, respectively. The instrument also detects solar cosmic rays.

Consideration of these data together with the results of measurements made in near-Earth orbits enable the detection of some variations in the integrated fluxes of hard radiation from solar flares in the case of simultaneous observations from two different directions. The most interesting observations of this type are those near the solar limb. We have applied two approaches to the realization of such "stereoscopic" observations. The first consists of simultaneous observations of a source from two different directions, and the second, in determining the moments when a source appears at the limb at different spacecraft as a result of the corresponding rotation of the Sun.

We analyzed two cases as examples: the setting of a flare region on May 20, 2001, which displayed regular brief flares, and the rising of group 10486, which displayed numerous flare phenomena at the limb and extremely high activity in October–November 2003. In both cases, we found that the setting of a modest (class M) flare 8°–10° behind the limb led to an appreciable decrease in the nonthermal radiation with energies exceeding 80 keV. The corresponding height of formation of the bulk of this radiation is located below 7000–10 000 km. This supports the idea that this radiation is generated near the bases of flare loops, with the flux from the loop apices being negligible. We emphasize that this result was based on geometric reasoning, without serious assumptions being made about the structure of the source or the mechanism generating the radiation.

Essentially, the instrument we have used detects flares beginning with class M3, or more precisely, those of them in which a fairly large quantity of particles have been accelerated. If these flares are located in the western part of the disk, or even in the central part, accelerated solar cosmic-ray particles that are generated there can be detected in near-Earth orbits. The new information about these weak sources of proton perturbations is the following. Acceleration occurs in the impulsive or explosive phase of these flares, and continues no more than several minutes. If an ordinary system of high loops forms, it exists for only a very short time. An additional source of X-ray emission usually forms in the post-eruptive phase, high in the region of the cusp. RHESSI data clearly demonstrate that, in the powerful flare of July 23, 2002, this source is located at coronal heights, and moved upward beginning from a height of about 20 000 km [17]. Even in X flares, this radiation is concentrated near energies of tens of keV.

Analysis of HEND data leads us to conclude that the energies of photons arising in the source at high heights near the loop apices do not exceed 80 keV. This contradicts the picture in which the acceleration of relativistic particles occurs during the reconnection of magnetic-field lines in a vertical current sheet or at the front of a large-scale shock in the corona. Strictly speaking, this conclusion can be drawn only for M flares. One of the main problems, therefore, becomes the need to search for mechanisms for efficient particle acceleration at low heights in compact flares.

It is not possible to draw the same conclusion for powerful solar flares based on currently available data. Several examples only confirm the short durations of typical solar gamma-ray flares. We emphasize that comparisons of the HEND data with the results of radio observations indicate a coincidence between many interesting features, especially in the impulsive phases of flares. This is associated with the good time resolution of the HEND data. The meaning of these coincidences is that the radiation at energies exceeding 80 keV is nonthermal, and the appearance of a thermal component is not noticeable, even in such powerful gamma-ray flares as the event of August 25, 2001. Note also that, in a number of cases, HEND data occupy gaps in observations of the hard X-ray

radiation of several interesting events, and joint analyses could lead to new results for phenomena such as large solar flares. Our analysis also convinced us of the large diagnostic potential of data obtained with even modest instruments for which regions beyond the eastern limb of the Sun are accessible to observation.

ACKNOWLEDGMENTS

The authors thank the original investigators, who made data from Yohkoh (HXT), SOHO, GOES, and CORONAS-F instruments openly available over the Internet. We also thank I.M. Chertok for presenting us with the difference images and IZMIRAN data on radio flares, as well as A.V. Belov and I.N. Myagkova for discussions. This work was supported by the Russian Foundation for Basic Research (project no. 05-02-17105).

REFERENCES

1. H. Hudson and J. Ryan, Annu. Rev. Astron. Astrophys. **33**, 239 (1995).
2. G. Rank, J. Ryan, H. Debrunner, *et al.*, Astron. Astrophys. **378**, 1046 (2001).
3. M. A. Livshits and A. V. Belov, Astron. Zh. **81**, 732 (2004) [Astron. Rep. **48**, 665 (2004)].
4. S. R. Kane, Sol. Phys. **86**, 355 (1983).
5. S. Masuda, PhD Thesis (Univ. of Tokyo, Tokyo, 1994).
6. S. Masuda, T. Kosugi, H. Hara, *et al.*, Nature **371**, 495 (1994).
7. E. Rieger, Sol. Phys. **121**, 323 (1989).
8. S. V. Bogovalov, S. R. Kel'ner, and Yu. D. Kotov, Astron. Zh. **65**, 1275 (1988) [Sov. Astron. **32**, 664 (1988)].
9. P. Li, K. C. Hurley, C. R. Barat Talon, and Victoria Kurt, Astrophys. J. **426**, 758 (1994).
10. H. T. Gallagher, B. R. Dennis, S. Krucker, *et al.*, Sol. Phys. **210**, 341 (2002).
11. W. V. Boynton, W. C. Feldman, I. G. Mitrofanov, *et al.*, Space Sci. Rev. **110**, 37 (2004).
12. A. V. Belov, E. Garcia, V. G. Kurt, and E. Mavromichalaki, Kosm. Issled. (2005) (in press).
13. M. A. Livshits, V. A. Chernetskii, A. V. Bogovalov, *et al.*, Astron. Vestn., No. 6 (2005) (in press).
14. S. N. Kuznetsov, K. Kudela, I. N. Myagkova, and B. Yu. Yushkov, in *Proceedings of the 28th International Cosmic Ray Conference* (Universal Academy Press, 2003), p. 3183.
15. O. V. Terekhov, Doctoral Dissertation (Institute for Space Research, Russian Academy of Sciences, Moscow, 1996).
16. I. M. Chertok and V. V. Grechnev, Astron. Zh. **80**, 1013 (2003) [Astron. Rep. **47**, 934 (2003)].
17. S. Krucker, in *Proceedings of the 13th Cool Stars Cambridge Workshop, Hamburg* (2004), ESA SP-560, p. 101.

Translated by D. Gabuzda

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 932–938. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 1042–1048.

The Proportion of Hot and Cold Regions in the Polar Solar Corona during 1957–2002

V. I. Makarov[1] and A. G. Tlatov[2]

[1]*Main (Pulkovo) Astronomical Observatory, Pulkovskoe sh. 65, St. Petersburg, 196140 Russia*

[2]*Kislovodsk Mountain Astronomy Station, P.O. Box 1, Kislovodsk, 357741 Russia*

Received October 25, 2004; in final form, May 18, 2005

Abstract—Occultation observations of the intensity of the FeXIV 530.3 nm and FeX 637.4 nm forbidden lines detected at the Kislovodsk Mountain Station during 1957–2002, indicate long-term changes in the structure of the solar corona. The monthly average intensities of green ($KI_{530.3}$) and red ($KI_{637.4}$) spectral lines are calculated for all latitudes (0°–90°) and for a high-latitude zone (45°–90°). A strong correlation ($r = 0.91$) between the green $KI_{530.3}$ line intensity and the Wolf numbers is found and used to fill gaps in the observations. The ratio $KI_{637.4}/KI_{530.3}$ takes on its maximum value at the solar minimum. The $KI^{p}_{637.4}/KI^{p}_{530.3}$ ratio in the high-latitude solar zone (45°–90°) increased by more than a factor of two during 1957–2002. This means that the fraction of cool regions in the polar corona has more than doubled over these years. We suggest that this increase in the number of cool regions is related to an increase in the area of the polar solar zones occupied by magnetic field of a single polarity at the solar minimum, and possibly to an increase in the area occupied by polar coronal holes. This is associated with long-term variations in the internal structure of the Sun. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Studies of the high-temperature coronal plasma are concerned with both heating of the upper solar atmosphere and the dynamics of local magnetic fields. Variations in the intensities of forbidden coronal spectral lines are related to variations in the temperature and density, which, in turn, are related to long-term changes in the global structure of the corona, solar wind, and terrestrial climate. Studies of the long-term changes in the corona are important for our understanding of both internal solar processes and the physics of cosmic rays. It is obvious that long, uniform series of coronal intensities in several spectral lines are necessary for this purpose.

Observations of the solar corona during total solar eclipses are not sufficient for studies of long-term changes. Even if we consider every eclipse, this provides no more than one minute of observations per year, on average. The difficulty of occultation observations of the corona is that the intensity of the coronal luminosity is 10^{-5}–10^{-6} of the intensity at the center of the solar disk.

The green coronal line at 530.3 nm was discovered during the eclipse of 1869. Seventy years later, W. Grotrian [1] and then B. Edlēn [2] showed that the 530.3 nm line was emitted by FeXIV. B. Lyot [3] was the first to carry out occultation observations of the solar corona. His work initiated the practice of regular measurements at various observatories in Europe, America, and Japan: Pic du Midi (1947–1974), Arosa (1947–1978), Climax (1947–1966), Norikura (from 1950 to the present), Kanzelhöhe (1948–1961), Kislovodsk (from 1952 to the present), Lomnický Štit (from 1966 to the present), and Sacramento Peak (1948–1966 and from 1977 to the present).

These observations were analyzed in [4–16]. At present, there are three observational series of various durations for the intensity of the FeXIV 530.3 nm green coronal line [9, 11, 13, 16, 17] and one series (for Kislovodsk) for the FeX 637.4 nm red coronal line in the period 1957–2002.

We analyze here the intensities of the green FeXIV 530.3 nm ($KI_{530.3}$) and red FeX 637.4 nm ($KI_{637.4}$) coronal lines observed at the Kislovodsk Station of the Main Astronomical Observatory during the period of 1957–2002. We use these data to determine the proportions of hot and cold regions in the polar corona at an altitude of $1.04\,R_0$.

2. OBSERVATIONAL DATA

Difficulties of occultation observations of the corona are associated with the large amount of light scattered in the terrestrial atmosphere, at the edge of the objective, and inside the coronograph. As a rule, the total amount of scattered light exceeds the brightness of the corona by two orders of magnitude.

1063-7729/05/4911-0932$26.00

M.N. Gnevyshev pioneered the organization of the first occultation observations in Russia. In 1948, he chose a site on the Shat Zhad Mas Plateau in the central Caucasus at an altitude of 2070 m above sea level for a coronal station. The first photographs of coronal lines were obtained in 1950. Regular Lyot coronograph [3] observations of the corona in the FeXIV 530.3 nm and FeX 637.4 nm lines were started in 1952, and have been continued to the present [6–9, 16, 18].

When obtaining long series of coronal observations, it is important to ensure stability of the photometric system and accuracy of the measurements when the brightness of the coronal lines becomes equal to the brightness of the spectrum of the solar-disk center with a spectral width of 1 Å.

An international meeting for corona researchers proposed by M.N. Gnevyshev was held at the Pic du Midi Observatory in 1967. The meeting approved a research technique developed in Kislovodsk, which was later recommended to all corona researchers at the 12th IAU Congress in Prague. According to the IAU recommendations, the intensities of coronal lines should be expressed in absolute units—units of coronal intensity (uci), which are equal to 10^{-6} of the energy irradiated by the center of the solar disk within a continuum band with a width of 1 Å near the line. These regular observations of coronal lines must be carried out at 5° intervals around the Sun, with the intensity being detected at a height of 40″ above the limb.

A calibration system equipped with a nine-step photometric scale and dark filters is used to derive the absolute intensities $KI_{637.4}$ and $KI_{530.3}$. The Kislovodsk Station provides a large number of observing days and high stability of the photometric system, making it possible to obtain a continuous series of coronal brightnesses.

The coronal observations obtained at any observatory will naturally display some gaps. Therefore, the number of observing days is an important characteristic. The table presents the numbers of observing days per year for three main coronal stations. The Kislovodsk Station is among the leaders in terms of the number of observing days (120–150 days per year), being second only to the Sacramento Peak Observatory, whose data are beyond the conventional coronal observations recommended by the 12th IAU Congress.

Monthly average number of days of coronal observations per year

Month	Kis	Nor	LSt
1	10.0	5.4	6.1
2	8.4	4.0	5.5
3	8.5	6.3	4.3
4	6.7	7.2	4.0
5	7.0	6.7	4.0
6	8.2	4.7	4.2
7	10.1	4.8	5.2
8	11.8	7.3	5.8
9	12.3	7.1	6.0
10	13.5	11.1	7.3
11	10.5	9.8	5.2
12	10.1	6.8	4.8

Note: Kis: Kislovodsk Mountain Station of the Main Astronomical Observatory, data for 1952–2002; Nor: Norikura Observatory, data for 1950–1995; LSt: Lomnický Štit, data for 1966–2001.

3. KISLOVODSK DATA FOR THE GREEN AND RED CORONAL LINES

Figures 1 (upper plot) and 2 show well developed cyclic changes in the monthly average intensities of the $KI_{530.3}$ green coronal line (in erg cm^{-2} s^{-1} sr^{-1}) and their distribution in solar latitude during 1957–2002. The observations indicate that the maximum brightness of the green line in the inner corona is detected over active regions (sunspots, faculae, and flocculi). It is especially important that there is a high correlation between the maximum intensity of the $KI_{530.3}$ green line and corresponding indices of solar activity, such as the Wolf numbers $W(t)$, for all observed activity cycles. The maximum brightness of the FeXIV 530.3 nm line was observed during the 19th cycle, while the 20th and 23rd cycles were less active than others. Thus, there is a correspondence between the intensity of the green coronal line and the Wolf numbers, in contradiction with the conclusions of [13, 14].

Calculations show that the correlation coefficient between the monthly average intensity $KI_{530.3}$ averaged over all latitudes and the Wolf numbers W is 0.91 for the interval 1957–2002. The correlation function takes the form

$$KI_{530.3} = 0.228\,W + 12.7\,\text{uci}.$$

Figure 3 shows that the Kislovodsk observations of the intensity of the $KI_{530.3}$ green coronal line display a high correlation between $KI_{530.3}$ and $W(t)$. However, the coronal measurements in the FeXIV 530.3 nm line observed at the Norikura Station show that this correlation is not always so high [17].

The FeX 637.4 nm red coronal line displays a somewhat different relation to the Wolf

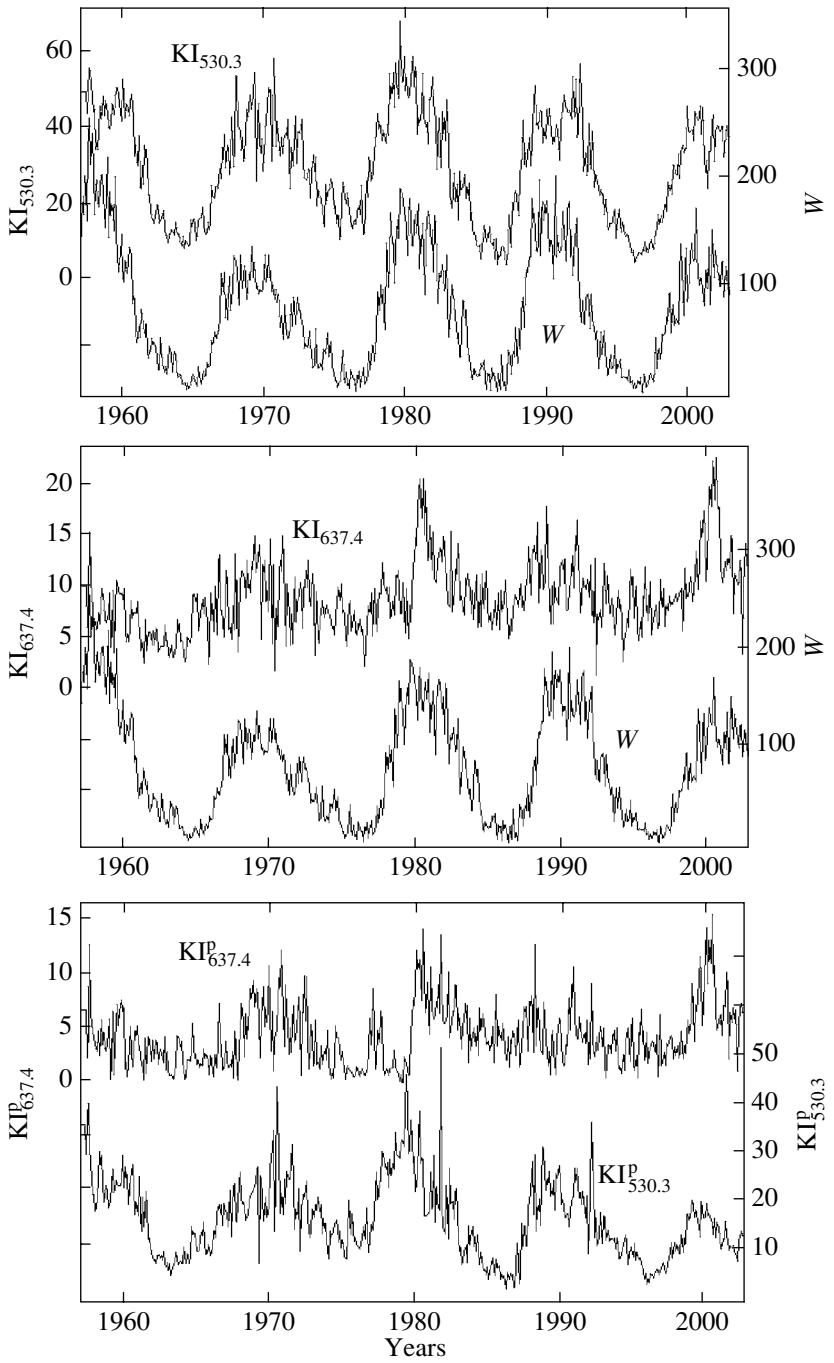

Fig. 1. Top: monthly average intensities of the $KI_{530.3}$ green coronal line (in units of 10^{-6} of the brightness at the center of the solar disk) for all solar latitudes (0°–90°) from the Kislovodsk observations for 1957–2002. The 11-year cycles of the Wolf numbers are also shown. Center: same as above for the $KI_{637.4}$ red coronal line. Bottom: monthly average intensities of the green ($KI^{p}_{530.3}$) and red ($KI^{p}_{637.4}$) coronal lines for latitudes 45°–90° during 1957–2002.

numbers $W(t)$. Figures 1 (middle plot) and 4 show the monthly average intensities $KI_{637.4}$ (in erg cm^{-2} s^{-1} sr^{-1}) and their distributions in solar latitude according to the Kislovodsk observations during 1957–2002. On average, the maximum intensity coincides with the maximum Wolf numbers, although the 11-year cyclicity is more weakly expressed than for the green coronal line. The correlation function takes the form

$$KI_{637.4} = 0.024\,W + 7\,\text{uci},$$

and the correlation coefficient for the relation between $KI_{637.4}$ and W is 0.5.

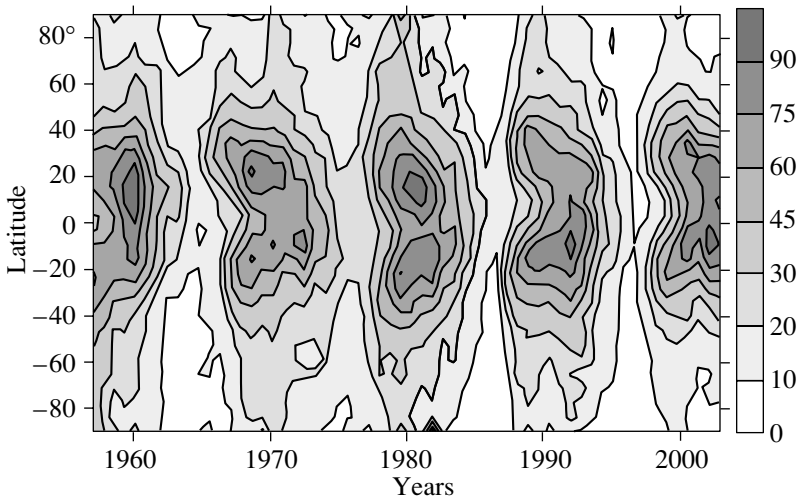

Fig. 2. Latitude–time distribution for the monthly average intensity of the $KI_{530.3}$ green coronal line (in units of 10^{-6} of the brightness at the center of the solar disk) detected at an altitude of $1.04\,R_0$ according to the Kislovodsk observations for 1957–2002.

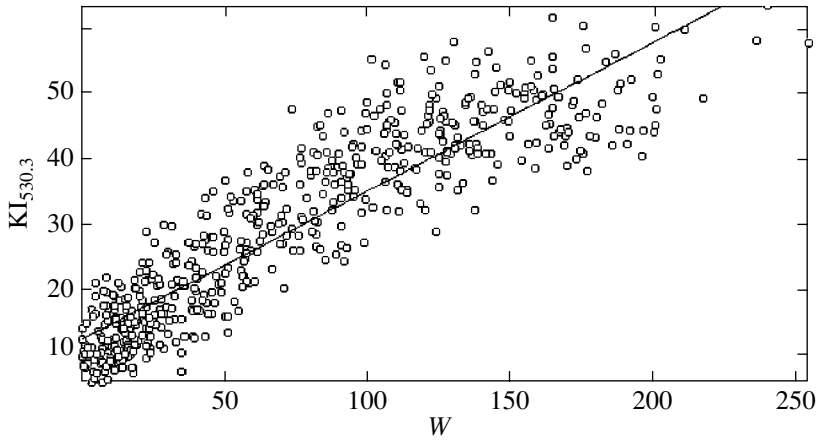

Fig. 3. Correspondence between the intensity of the $KI_{530.3}$ green coronal line (in units of 10^{-6} of the brightness at the center of the solar disk) averaged over all latitudes and the Wolf numbers W during 1957–2002.

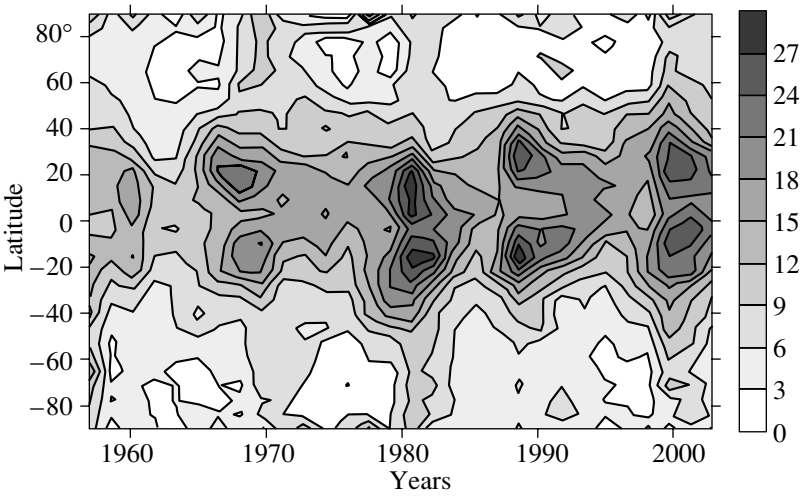

Fig. 4. Same as Fig. 2 for the intensity of the $KI_{637.4}$ red coronal line.

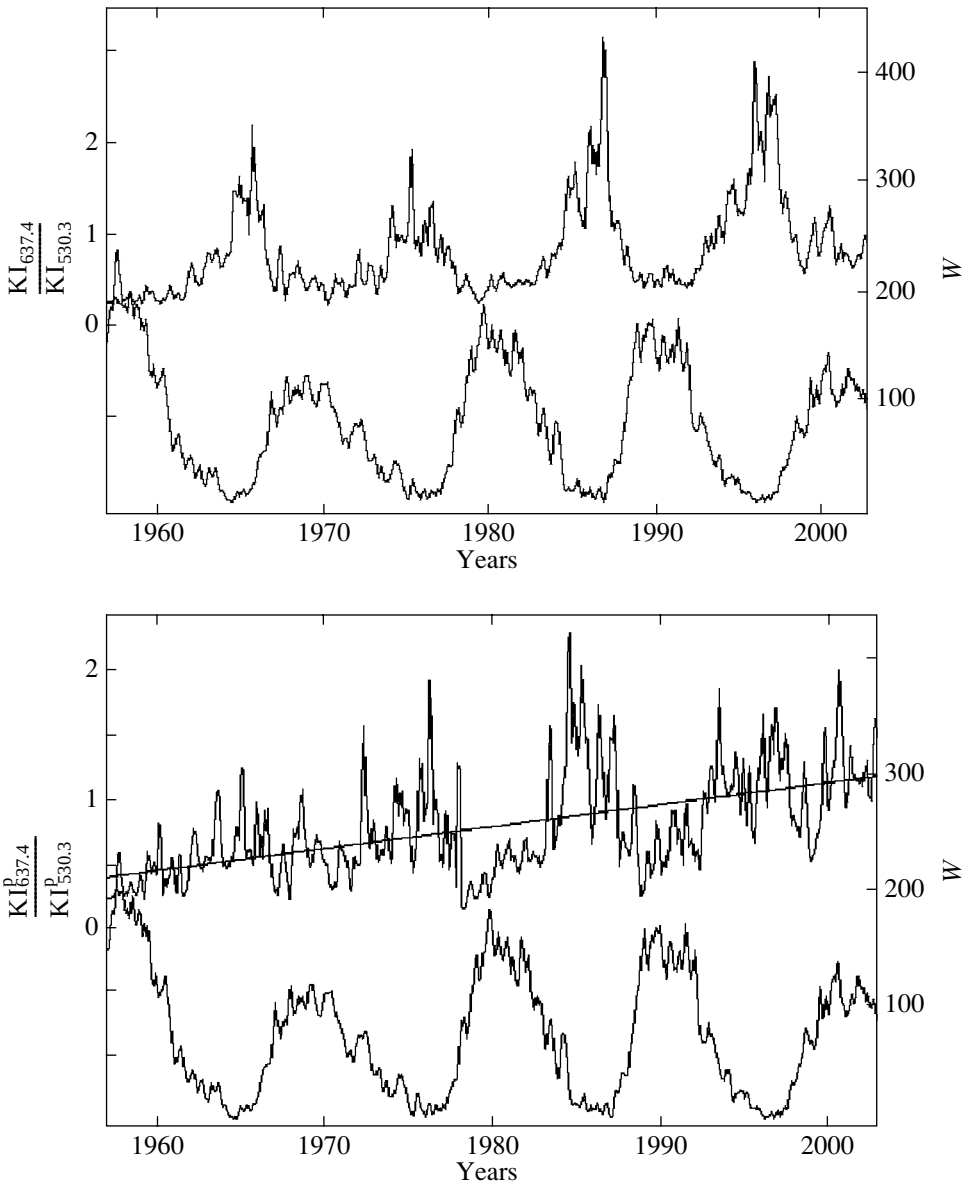

Fig. 5. Top: the ratio $KI_{637.4}/KI_{530.3}$ for latitudes 0°–90° of the Northern and Southern solar Hemispheres (upper curve) and Wolf-number cycles W for 1957–2002 (lower curve). Bottom: the ratio $KI^{p}_{637.4}/KI^{p}_{530.3}$ for latitudes 45°–90° of the Northern and Southern solar Hemispheres (upper curve and a linear approximation), and the Wolf-number W cycles during 1957–2002 (lower curve).

4. LONG-TERM CHANGES IN THE INTENSITIES OF THE FeXIV 530.3 nm AND FeX 637.4 nm LINES IN THE POLAR SOLAR CORONA

The green (FeXIV 530.3 nm) and red (FeX 637.4 nm) coronal lines are most intense and visible. The FeX and FeXIV ions originate in cool ($T \approx 0.9 \times 10^6$ K) and hot ($T \approx 1.3 \times 10^6$ K) coronal regions, respectively, and do not coincide in space. However, correlated increases in the brightnesses of the green and red lines are frequently observed over active regions. This means that the magnetic field prevents the conduction of heat, bringing about the prolonged coexistence of cool and hot regions.

It is well known that the temperature behavior of the ionization curves for FeXIV 530.3 nm and FeX 637.4 nm are very different [19]. This means that the red and green lines originate at different temperatures in different regions of the corona. Thus, the ratio of intensities of these lines $KI_{637.4}/KI_{530.3}$ describes the relative contribution of regions with different temperatures, rather than the temperature of the corona. Consequently, the proportions of cool and

hot coronal regions can be determined from the ratio FeX/FeXIV, since this quantity depends strongly on the temperature and only slightly on variations in density. Earlier calculations [4, 5, 20, 21] estimating the coronal temperature from the ratio $KI_{637.4}/KI_{530.3}$ do not take into account the fact that the red and green lines originate in different regions of the corona with different temperatures.

Figure 1 shows that the green and red coronal-line emission varies appreciably from solar minimum to solar maximum, both at all latitudes ($KI_{530.3}$ and $KI_{637.4}$ in the upper and middle plots) and in the polar zone ($KI^p_{530.3}$ and $KI^p_{637.4}$ in the lower plot). The ratios $KI_{637.4}/KI_{530.3}$ for all latitudes and $KI^p_{637.4}/KI^p_{530.3}$ for the polar zones likewise vary significantly (Fig. 5). The total set of data on the FeXIV 530.3 nm green coronal line for the combined observations at Arosa, Pic du Midi, Norikura, Kanzelhöhe, Wendelstein, and Lomnický Štit together with the Kislovodsk series on the FeX 637.4 nm red coronal line leads to the same conclusion [22, 23].

In the last decade, special attention has been paid to the polar solar activity [24, 25]. Figures 1, 2, and 4 present the cyclic variations and latitude distributions for $KI^p_{530.3}$ and $KI^p_{637.4}$ in the polar corona during 1957–2002. The $KI^p_{530.3}$ and $KI^p_{637.4}$ intensities at high latitudes are significantly weaker than those expected based on the latitude distribution of the sunspots. Some uncertainties in $KI^p_{637.4}/KI^p_{530.3}$ can lead to errors in estimates of the proportion of cool regions in the polar corona. However, the behavior of $KI^p_{637.4}/KI^p_{530.3}$ in Fig. 5 (lower plot) suggests that there is a long-term component in variations of the structure of the polar corona.

The upper plot in Fig. 5 shows the 11-year cycles of $KI_{637.4}/KI_{530.3}$ for latitudes 0°–90°. We can see that the maximum of this ratio corresponds to the solar minimum. According to the lower plot in Fig. 5, the maximum of $KI_{637.4}/KI_{530.3}$ at latitudes 45°–90° is also observed at the sunspot minimum.

It is well known that the most important phenomena in the high-latitude corona are related to coronal holes, which have reduced densities and temperatures. The ratio of the intensities of the red and green coronal lines, $KI^p_{637.4}/KI^p_{530.3}$, at latitudes 45°–90° can be used to derive long-term changes in the proportion of cool and hot regions in the high-latitude corona. The ratio $KI^p_{637.4}/KI^p_{530.3}$ has increased from 0.4 to 1.0 over the last 45 years. This corresponds to a decrease in the portion of hot regions in the high-latitude corona by approximately a factor of two. We believe that an increase in the area of the solar polar cap occupied by magnetic field of a single polarity, and thus an increase in the polar magnetic flux, results in an increase in the area of polar coronal holes [25, 26]. This phenomenon can explain the increase in $KI^p_{637.4}/KI^p_{530.3}$ at high latitudes, and, consequently, the increase in the number of cool regions in the polar corona during 1957–2002.

5. CONCLUSIONS

We have presented here long-term changes in the ratio $KI^p_{637.4}/KI^p_{530.3}$ detected at latitudes 45°–90°, and analyzed data on the green (FeXIV 530.3 nm, $KI_{530.3}$) and red (FeX 637.4 nm, $KI_{637.4}$) coronal lines observed during 1957–2002 at the Kislovodsk Station of the Main Astronomical Observatory. Figure 5 shows the cyclic changes in $KI_{637.4}/KI_{530.3}$ for all latitudes and the polar zones ($KI^p_{637.4}/KI^p_{530.3}$). The ratio $KI_{637.4}/KI_{530.3}$ behaves similarly with the cycle phase in both cases. Figure 5 shows that the monthly average $KI^p_{637.4}/KI^p_{530.3}$ has approximately doubled during 1957–2002. This corresponds to an approximately twofold increase in the proportion of cool regions, or a decrease in the proportion of hot regions in the polar corona during 1957–2002. The total set of data on the FeXIV 530.3 nm green coronal line for the combined observations at Arosa, Pic du Midi, Norikura, Kanzelhöhe, Wendelstein, and Lomnický Štit together with the Kislovodsk series on the red coronal line FeX 637.4 nm leads to only one conclusion [22, 23].

It is likely that the increase in the proportion of cool regions in the high-latitude solar zone is related to an increase in the area of polar zones occupied by magnetic field of a single polarity during the sunspot minimum [22]. This results in an increase in the area of coronal holes and, consequently, in an increase in the number of cool regions in the polar corona. These phenomena are associated with long-term changes in the internal structure of the Sun. They are related to the question of how solar activity can reach a state similar to the Maunder minimum, leaving this state within no more than 100 yrs. Such effects are also related to global warming on the Earth [22].

ACKNOWLEDGMENTS

The authors are grateful to M.A. Livshits for his helpful discussions and suggestions. This work was supported by the Russian Foundation for Basic Research (project nos. 05-02-16299 and 03-02-16091), the basic research program Solar Activity and Physical Processes in the Solar–Terrestrial System of the Presidium of the Russian Academy of Sciences, and the basic research program Plasma Processes in the Solar System of the Department of Physical Sciences of the Russian Academy of Sciences.

REFERENCES

1. W. Grotrian, Naturwissenschaften **34**, 87 (1939).
2. B. Edlēn, Z. Astrophys. **22**, 30 (1942).
3. B. Lyot, Mon. Not. R. Astron. Soc. **99**, 580 (1939).
4. M. Waldmeier, Z. Astrophys. **21**, 85 (1941).
5. M. Waldmeier, Z. Astrophys. **30**, 137 (1952).
6. M. N. Gnevyshev, Astron. Zh. **37**, 227 (1960) [Sov. Astron. **4**, 217 (1960)].
7. M. N. Gnevyshev, Astron. Zh. **42**, 488 (1965) [Sov. Astron. **9**, 387 (1965)].
8. M. N. Gnevyshev, Sol. Phys. **2**, 108 (1967).
9. M. N. Gnevyshev, Sol. Phys. **90**, 177 (1984).
10. J. Sýkora, Bull. Astron. Inst. Czech. **22**, 12 (1971).
11. J. Sýkora, Contrib. Astron. Obs. Skalnatė Pleso **22**, 55 (1992).
12. J.-L. Leroy and M. Trellis, Astron. Astrophys. **35**, 283 (1974).
13. V. Rušin and M. Rybanský, Sol. Phys. **207**, 47 (2002).
14. M. Rybanský and V. Rušin, Contrib. Astron. Obs. Skalnatė Pleso **22**, 229 (1992).
15. R. C. Altrock, *Synoptic Solar Physics*, Ed. by R. S. Balasubramanian, J. W. Harvey, and D. M. Rabin; Astron. Soc. Pac. Conf. Ser. **140**, 339 (1998).
16. V. I. Makarov and A. G. Tlatov, in *The Sun at an Epoch of Polarity Reversal of the Magnetic Field*, Ed. by V. I. Makarov and V. N. Obridko (Pulkovo, St. Petersburg, 2001), p. 259.
17. T. Sakurai, M. Irie, M. Imai, *et al.*, Publ. Nat. Astron. Obs. Jpn. **5**, 121 (1999).
18. A. G. Tlatov, S. A. Guseva, and Kim Gur der, *The Sun at an Epoch of Polarity Reversal of the Magnetic Field*, Ed. by V. I. Makarov and V. N. Obridko (Pulkovo, St. Petersburg, 2001), p. 385.
19. M. Arnaud and R. Rothenflug, Astron. Astrophys., Suppl. Ser. **60**, 425 (1985).
20. H. E. Mason, Mon. Not. R. Astron. Soc. **170**, 651 (1975).
21. M. Guhathakurta, R. R. Fisher, and R. C. Altrock, Astrophys. J. **414**, L145 (1993).
22. V. I. Makarov, A. G. Tlatov, and D. K. Callebaut, in *Solar Variability as an Input to the Earth's Environment*, Ed. by A. P. Wilson (ESA Publications Division, Noordwijk, 2003), ESA SP-535, p. 217.
23. V. I. Makarov, A. G. Tlatov, and D. K. Callebaut, in *SOLSPA: The Second Solar Cycle and Space Weather Euroconference,* Ed. by A. P. Wilson (ESA Publications Division, Noordwijk, 2002), ESA SP-477, p. 241.
24. V. I. Makarov and V. V. Makarova, Sol. Phys. **163**, 267 (1996).
25. V. I. Makarov, A. G. Tlatov, D. K. Callebaut, and V. N. Obridko, Sol. Phys. **206**, 383 (2002).
26. V. I. Makarov, A. G. Tlatov, D. K. Callebaut, and V. N. Obridko, in *The Sun at an Epoch of Polarity Reversal of the Magnetic Field*, Ed. by V. I. Makarov and V. N. Obridko (Pulkovo, St. Petersburg, 2001), p. 267.

Translated by V. Badin

Astronomy Reports, Vol. 49, No. 11, 2005, pp. 939–945. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 11, 2005, pp. 1049–1056.

The Radio Brightness Distribution along the Solar Limb Observed on the Siberian Solar Radio Telescope

B. B. Krissinel'

Institute of Solar–Terrestrial Physics, Siberian Division, Russian Academy of Sciences, P.O. Box 4026, Irkutsk, 664033 Russia

Received March 9, 2005; in final form, May 18, 2005

Abstract—The radio brightness distribution at 5.2 cm has been obtained along the entire solar limb down to a level of 0.01 of brightness temperature at the disk center T_0^c during the solar minimum. The measurements are based on strip scans from the Siberian Solar Radio Telescope. Data reduction included model fitting of an actual solar scan for each position angle. The maximum limb brightness, $1.37\,T_0^c$, in the derived distribution is at equatorial direction, where the radio radius exceeds the photosphere radius by 7%. Toward the poles, the brightness peak and radio radius gradually decrease to $1.01\,T_0^c$ and $1.005\,R_0$. The derived two-dimensional brightness distribution was used to calculate radio radius as a function of position angle for several wavelengths from 4 to 31.6 cm. These dependences are consistent with RATAN-600 observations at position angles 0° and 25°. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The distribution of radio brightness T_b over the solar disk, especially near the limb, is determined by the height dependence of the electron temperature and density, and is thus connected with a fundamental problem—heating of the corona. This explains the interest of numerous researchers in the radial distribution of T_b. The first summary of such studies at centimeter wavelengths, which are most suitable for studies of the corona and transition region, was made by Swanson [1]. Measurements of T_b at the limb in east–west or north–south directions [2–4] and of the radio radius along the limb [5] were carried out in a number of subsequent papers. These results yield a fairly complicated and often contradictory pattern for T_b distribution along the limb and its wavelength dependence. The characteristic radial distribution of T_b (uniform brightness over almost all the disk and a sharp increase in the brightness at the limb, especially in the equatorial direction) has not been thoroughly explained; it is not consistent with known theoretical and semiempirical models of the corona. Well-known attempts to explain this as an effect of spicules [5, 6] or small-scale, optically thick inhomogeneities in the transition region [7] have not yielded convincing results.

Studies with high angular resolution, which enable accurate measurement of the strength and width of the limb brightness peak, are of great importance in this connection, as well as for models of the coronal structure. However, Very Large Array (VLA) observations during the partial solar eclipse of October 12, 1997 [8], yielded an unexpectedly high peak at the limb at 4.9 GHz: 80% in position angle 27.8°. Observations at 6 cm on the Westerbork Synthesis Radio Telescope in 1979 [9], also led to unexpected results: the amounts of brightening at the western and southern limbs were virtually the same (40–50%); the radio radius was greater than the optical radius by 50″ and 40″, respectively. This differs from the Bonn results [5], which, in spite of poorer angular resolution, found a fairly strong dependence of the radio radius on the position angle at this same wavelength. The most comprehensive studies in this direction have been carried out at centimeter wavelengths on the RATAN-600 radio telescope [10]; however, these data refer only to position angles of 0° and 25°. A number of studies with high angular resolution were concerned with measurements of the mean solar radio radius and the brightness enhancement near the polar limb [11–14]. However, these works did not consider the shape of the brightness variations beyond the limb, which reflects the structure of the solar atmosphere.

The aim of this study was to analyze the distribution of T_b at 5.2 cm along the entire solar limb with high angular resolution. We wished to study the brightness distribution not only near the limb, but also at a considerable distance from it, as such is important for estimates of the contribution from loop structures in the corona. Our analysis was based on strip scans obtained with the arrays of the Siberian Solar Radio Telescope (SSRT) [15, 16] during the solar minimum (May–August 1995).

1063-7729/05/4911-0939$26.00

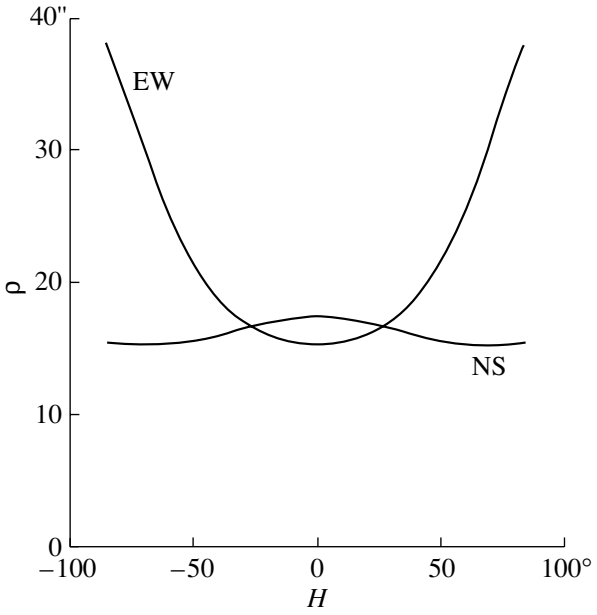

Fig. 1. Variation of the spatial resolution of the SSRT interferometers during the day in summertime.

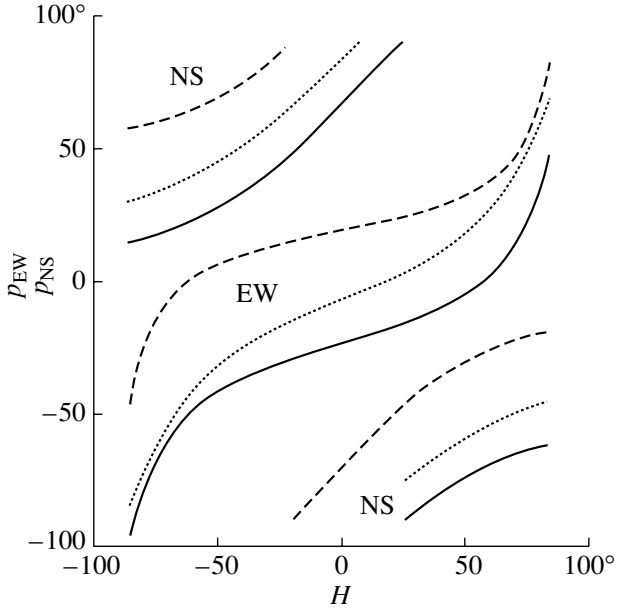

Fig. 2. Position angles of the beam peaks of the SSRT arrays on May 4 (solid), June 20 (dotted), and August 25 (dashed), 1995, with respect to the solar meridian. The vertical axis shows the position angle in degrees and the horizontal axis, the hour angle in degrees.

2. THE INSTRUMENT USED AND OBSERVATIONAL DATA

The SSRT consists of two orthogonal linear arrays operating at 5.2 cm, each of which includes 128 antennas with a diameter of 2.5 m separated by 4.9 m. The length of each array is 622.3 m. Two-dimensional maps can be obtained every 15 min by means of frequency scanning (using a 500-channel receiver). The angular resolution in this mode is $22'' \times 26''$.

Simultaneously with mapping, both arrays of instruments observe in a strip-scan mode. These data represent a sequence of responses to the Sun's transit across the interference fringes of the array's beam. The beam of the SSRT arrays is described by the formula

$$P(\theta) = \left(\frac{\sin\left(N\frac{\pi d}{\lambda}\cos\theta\right)}{N\sin\left(\frac{\pi d}{\lambda}\cos\theta\right)} \right)^2,$$

where $N = 128$ is the number of antennas, $d = 4.9$ m the antenna spacing, λ the wavelength, and $\cos\theta$ the direction cosine measured from the interferometer baseline. The solar-disk center is crossed by the beam peak at times given by the formula

$$H = \arcsin\left(\frac{k\lambda}{d\cos\delta}\right)$$

for the east–west (EW) array,

$$H = \arccos\left(\frac{\frac{k\lambda}{d} + \sin\delta\cos\phi}{\cos\delta\sin\phi}\right)$$

for the north–south (NS) array,

where k is the number of the interference maximum and H, δ, and ϕ are the solar hour angle, solar declination, and latitude, respectively. During the day, the scan duration, which is determined by the angular spacing of the beam fringes, changes from 150 s at culmination to 30 min at $H \sim 90°$ for the east–west array and from 40 min at culmination to 3 min at $H \sim 90°$ for the north–south array. In the summer, we obtain daily up to 160 scans on the EW array and up to 90 scans on the NS array.

Figure 1 shows the angular resolution ρ of the arrays as a function of the hour angle for the summer:

$$\rho = 0.885\frac{\lambda}{Nd\sqrt{1-(\sin\theta)^2}}.$$

We can see that it is not desirable to use the data obtained on the EW interferometer at large hour angles because of the degraded angular resolution.

To study the distribution of T_b beyond the solar limb, the instrument must have an appropriate field of view, which can be described by the quantity p_z: the ratio of the duration of the interference period

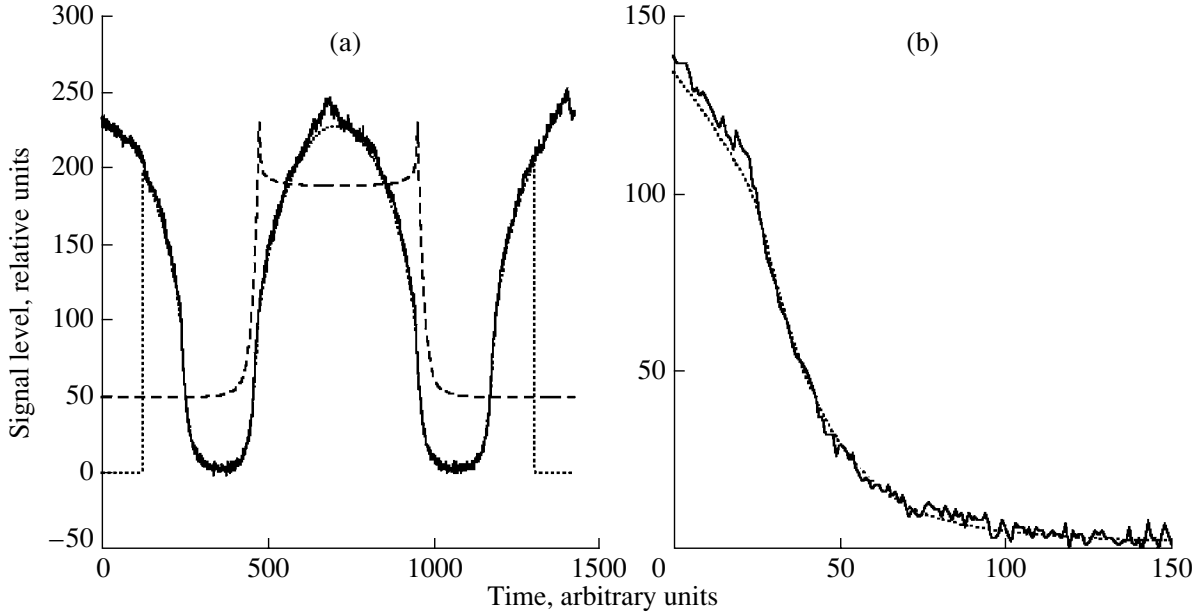

Fig. 3. (a) Example of a comparison of actual (solid) and calculated (dotted) scans. The dashed line shows the model radial distribution of T_b. (b) Enlargement of the right-hand part of the actual and calculated scans.

to the duration of the solar disk's transit across the interference fringe of the array beam. Analysis shows that observations on the EW interferometer are field-restricted in the range of hour angles near culmination and on the NS interferometer at large hour angles.

To study the distribution of T_b along the limb with the SSRT arrays, we used the fact that the angle between the direction of the interference maximum and the normal to the array baseline changes continuously during the day. The Sun's position angle p_0 also changes during the observation session. As a result, the position angle p of the kth interference maximum with respect to the solar meridian will be

$$p_{EW}(k) = \arctan(\sin\delta(k)\tan H(k)) + p_0,$$

$$p_{NS}(k) = \arctan\left(-\frac{\cos\phi\cos\delta(k)}{\sin\phi\sin H(k) - \sin\delta(k)\cot H(k)}\right) + p_0,$$

where $H(k)$ and $\delta(k)$ are the hour angle and declination of the kth interference maximum.

As an example, Fig. 2 presents the position angles p_{EW} and p_{NS} as functions of the hour angle for three days in spring–summer (May 4, June 20, and August 25). As is visible from the graphs, we can scan the solar disk in all directions if we use the May–August observations on the EW and NS arrays, when the Sun's declination is within 10°–23° and its own position angle changes from −24° to +20°. However, bearing in mind with the above restrictions on the spatial resolution and field of view, and in order to ensure statistical reliability, a detailed analysis of the brightness distribution along the entire limb requires the use of a fairly large amount of observational data.

We selected for the study observations on days when there were no active regions, filaments, or coronal holes near the limb (along the scan direction). The data were processed for observations obtained on May 4, 11, and 12, June 1, 7, and 20, July 4, and August 8, 10, 25, and 29, 1995.

3. DATA PROCESSING

Since we cannot estimate directly T_b for any part of the solar disk from the strip scan data, the distribution of T_b over the solar disk can be studied only by means of modeling. This means that we must choose a distribution of T_b over the solar disk whose calculated response would coincide with the envelope of the actual scan.

To model a scan whose envelope matches an actual scan, we must set the form of the radial brightness distribution over the Sun in the given direction. We used here the data of [10, 17] on the parameters of the radial distribution of T_b in the equatorial direction at several wavelengths as an input model. We created a large set of models with radial distances from 0 (disk center) to $2R_0$, with various values of emission peak, radio radius, and steepness of brightness decline beyond the limb.

The chosen model radial distribution must then be corrected for the effect of the beams of the SSRT paraboloids (HPBW = 80′). In the considered range

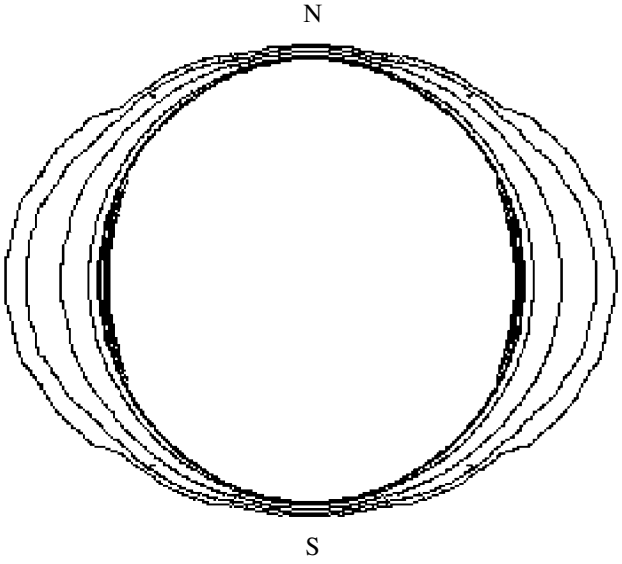

Fig. 4. Two-dimensional distribution of T_b at 5.2 cm (the contours range from 0.01 to 1.3 of the brightness temperature at the solar-disk center).

of radial distances, the beam is well described by the formula

$$F(\theta) = 2\frac{J_1(x)}{x}, \quad x = \frac{\pi D \sin\theta}{\lambda},$$

where D is the mirror diameter, J_1 the first-order Bessel function, and λ the wavelength.

We then form an axially symmetric model brightness distribution for the entire solar disk. For this purpose, we filled a 512×512 two-dimensional array with the brightness values for the one-dimensional model corresponding to the distance from the center to a given point in the array.

To take into account the smoothing effect of the beam, the obtained model distribution of T_b is convolved with the interferometer beam. The model scan yielded by this convolution is compared with the actual scan: the scans are superposed on the same screen and adjusted in level and time position. We then estimated the accuracy of the model based on the extent to which the calculated scan matched the envelope of the quiet Sun. If the model did not fit, we corrected it or used another model. We took into account some parameters of the radial distribution: the magnitude and position of the brightness peak at the limb, and the radial distance at the 0.5 and 0.1 levels. As an example, Fig. 3 presents a scan of the Sun obtained on the EW interferometer on June 1, 1995, together with a model scan ($k = 58$, $p_{EW} = 3.6°$, $p_z = 1.49$). This figure also shows the model radial distribution of T_b.

We chose scans for the analysis such that the step in position angle would not exceed, on average, 1°. The large set of models obtained was averaged over 3° intervals and then additionally smoothed over three points. Using this set of models, we created a two-dimensional brightness distribution model for all position angles, which was a first approximation, since the input radial models were selected under the condition of an axially symmetric brightness distribution over the Sun. In subsequent stages, we carried out a repeated gradual fitting of the two-dimensional model by comparing the calculated scans with the observational data for all position angles.

The majority of the actual scans turned out to be symmetric, and the model described equally well both slopes of the scan. However, in some cases pertaining to scans in the polar regions of the Sun, this symmetry was broken. This can be explained, on the one hand, by the presence of coronal holes in this region and, on the other hand, by the possible effect of inclination of the solar rotation axis to the plane of the ecliptic.

4. RESULTS

The resulting empirical two-dimensional distribution of T_b (Fig. 4) is elongated in the equatorial direction, and its brightness varies smoothly from the equator to the poles. The distribution of T_b over the disk is close to elliptical; the smallest uniform-brightness dimension is in the equatorial direction.

Figure 5 presents the magnitude of the radio brightness peak as a function of the position angle. The maximum values, T_{max}, which reach $1.37\,T_0^c$, correspond to the equatorial direction. The brightness peak monotonically decreases from the equator to the poles; with its magnitude in the polar zones being slightly less than $1.01\,T_0^c$. Note the good coincidence of T_{max} with the RATAN-600 results of [10], according to which the peak magnitude (at 5.2 cm at position angle 24°–26°) is $1.22\,T_0^c$, while the graph of Fig. 5 indicates $1.23\,T_0^c$.

Figure 6 shows radial distance R_r at various levels as functions of the position angle. At the 0.5 level, the maximum value of R_r exceeds the photospheric radius by 7%. With increasing position angle, the value of R_r gradually decreases, reaching $1.005\,R_0$ in the polar region. The radial distance in the equatorial direction is $1.2\,R_0$ at the 0.1 level, and $1.48\,R_0$ at the 0.01 level. In the polar zone, the radio brightnesses at the 0.1 and 0.01 levels are at heights of $1.027\,R_0$ and $1.065\,R_0$, respectively.

Figure 7a presents the radial brightness distribution along the equator and meridian, normalized to the brightness at the disk center. Figure 7b presents part of the equatorial distribution beyond the limb, as well as the corresponding part of the distribution

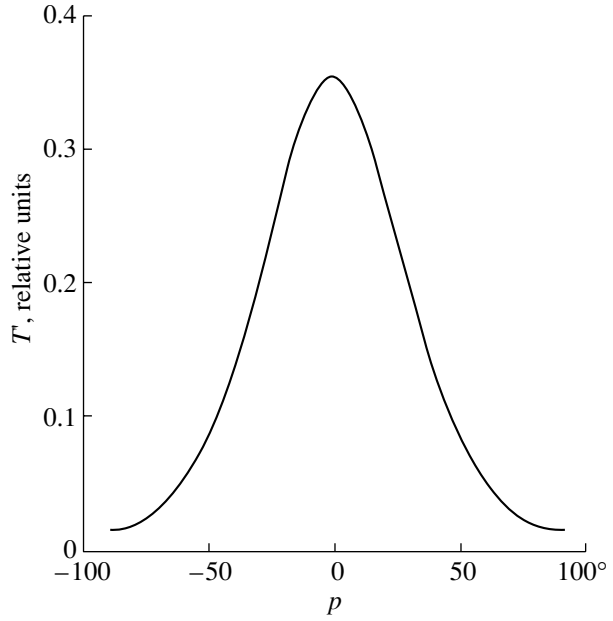

Fig. 5. Magnitude of the brightness temperature peak ($T' = T_{max}/T_0^c - 1$) as a function of the position angle.

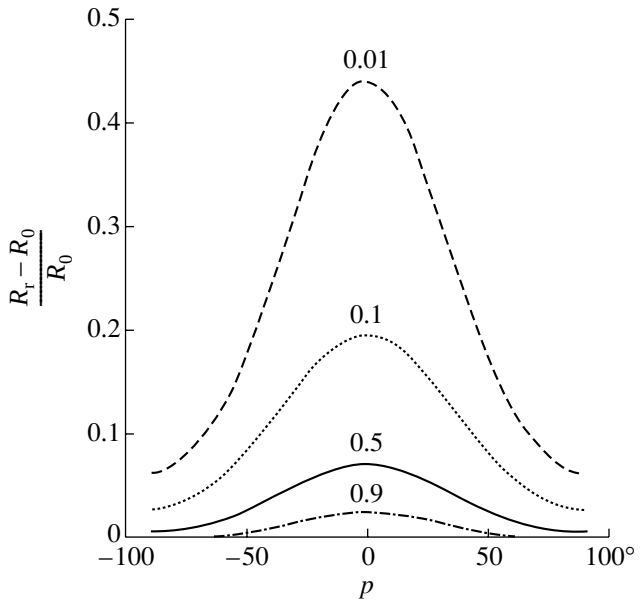

Fig. 6. Variation of the difference between the radial distance R_r and the photospheric radius R_0 at the 0.9, 0.5, 0.1, and 0.01 levels along the solar limb.

calculated in the empirical model for quiescent regions of the solar atmosphere [18]. The calculations of T_b for the solar-disk center in this model match the observational data well. However, like all the others, this model does not fit the T_b observations near and—as we can see from Fig. 7b—beyond the limb. The largest discrepancy with the calculated curve is near the 0.2 level; further out, this curve approaches the observed curve. The curves nearly coincide near the 0.01 level.

One possible origin of the discrepancy between the calculated radial distributions of T_b and the observational data is the loop structure of the corona [19]. In the above model, for the quiet regions of the Sun, the electron density begins to increase sharply toward the photosphere, beginning from heights of about 300 000 km. At radial distances corresponding to these heights, the brightness at 5.2 cm is 1% of the brightness at the solar-disk center. Thus, if we suppose that the sharp increase in electron density is associated with coronal loop structure, the 0.01 level corresponds to the boundary of the loop structure. The height of the loop structure gradually decreases from the equator to the poles, where it is $0.07\,R_0$. The value of R_r at the 0.5 level in the polar region, $1.005\,R_0$ (3500 km above the photosphere), testifies that the brightness temperature here is mainly determined by the transition zone, and that the number of coronal loops is small.

According to results obtained with the Nobeyama radioheliograph [12–14], the 17 GHz radio radius in the polar region is $1.0125\,R_0$; i.e., it exceeds the value of the radio radius we have measured at 5.2 cm. In the first instance, this can be explained by the effect of coronal holes, which are virtually always present in the polar regions. While coronal holes are visible at 5.2 cm as dips of the emission, they are usually either not manifest or display enhanced brightness at 1.7 cm [20]. On the other hand, Selhorst *et al.* [14] explain the increased radio radius in the polar zone at 17 GHz as a consequence of the effect of the radio telescope beamwidth on the response from the limb with enhanced brightness.

The magnitude of the T_b peak is $0.995\,R_0$. This is consistent with the results of [10], where this value for the equatorial zone is equal to or less than R_0 at all wavelengths from 2 to 31.6 cm. The shift of the brightness peak from the limb toward the disk center at centimeter wavelengths can be understood as an effect of refraction in the transition zone.

Since the radiation mechanism for the corona is known to be bremsstrahlung, the derived two-dimensional map of T_b can be extrapolated to other wavelengths for parts of the corona where the optical depth is small. Figure 8 presents R_r at the 0.5 level calculated as a function of position angle for 4.0, 6.0, 8.2, 11.7, 20.7, and 31.6 cm, together with the dependence observed at 5.2 cm. This figure also shows the observational data obtained by other researchers.

The value of R_r at 5.2 cm in the equatorial direction measured from the SSRT observations of a solar

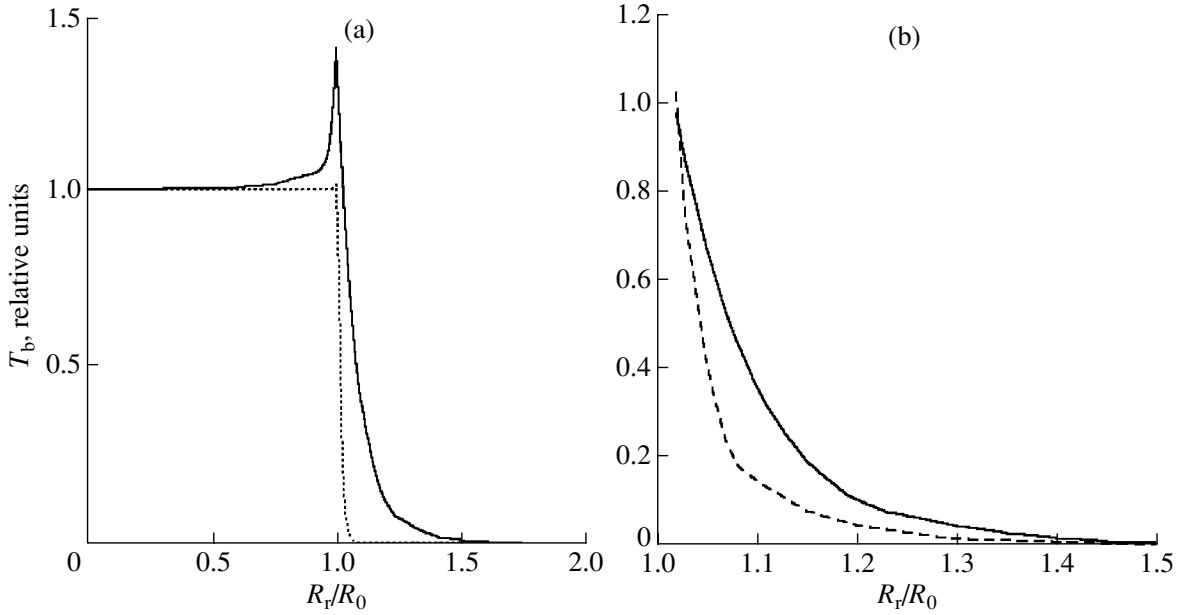

Fig. 7. (a) Radial distribution of T_b along the equator (solid) and meridian (dotted). (b) Part of the radial distribution of T_b along the equator beyond the limb (solid) and results of the calculations (dashed).

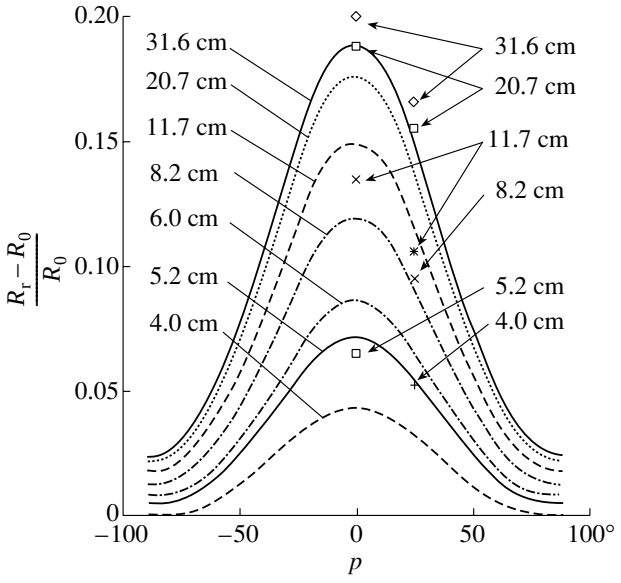

Fig. 8. Variation of the radio radius along the limb at 4.0, 5.2, 6.0, 8.2, 11.7, 20.7, and 31.6 cm. The indicated points show the observational data obtained in [10, 21].

eclipse [21] virtually coincides with the value obtained here ($1.065\,R_0$ and $1.07\,R_0$). The value $R_r = 1.032\,R_0$ obtained at 4.0 cm by extrapolating at $p = 25°$ is considerably lower than the RATAN-600 value [10]. At longer wavelengths, the calculation results fit the RATAN-600 data much better. The value obtained at 8.2 cm by extrapolating at $p = 25°$ is $R_r = 1.093\,R_0$, while the observed value is $1.095\,R_0$. The calculated 11.7 cm values at $p = 0°$ and $25°$ are $1.152\,R_0$ and $1.119\,R_0$, while the observed values are $1.135\,R_0$ and $1.106\,R_0$, respectively, indicating a fairly good agreement. The extrapolated values at 20.7 cm are $1.18\,R_0$ and $1.141\,R_0$, while the observed values are $1.188\,R_0$ and $1.155\,R_0$, respectively. Finally, at 31.6 cm, we have calculated values of $1.193\,R_0$ and $1.151\,R_0$ and observed values of $1.205\,R_0$ and $1.166\,R_0$.

Thus, the values of R_r calculated based on the map of T_b at 5.2 cm are consistent with the results of the RATAN-600 measurements at a number of wavelengths. This testifies to a sufficiently high accuracy of the derived distribution of T_b at 5.2 cm far off the limb.

5. CONCLUSION

We have obtained the brightness distribution over the solar disk and beyond the limb to a distance of two solar radii at wavelength of 5.2 cm at the solar minimum. We used observations obtained on the east–west and north–south arrays of the Siberian Solar Radio Telescope carried out from May to August 1995. The data processing consisted in fitting a radial brightness distribution model and matching the envelopes of actual solar scans for each position angle.

Taking into account the bremsstrahlung nature of the coronal radiation, we calculated the radio radius of the Sun at several wavelengths from 4 to 31.6 cm based on the obtained brightness map. The calculated radio radii agree well with the RATAN-600 measurements at position angles 0° and 25°, as well as with the SSRT data obtained during solar eclipse.

The brightness temperature distribution at wavelength of 5.2 cm displays a pronounced elongation in the equatorial direction. The maximum radio radius is 1.07 of the photospheric radius R_0. The limb brightness at the east and west points reaches 1.37 of the brightness temperature at the disk center T_0^c. Toward the poles, the limb brightness peak and radio radius gradually decrease to $1.01\,T_0^c$ and $1.005\,R_0$.

The derived brightness distribution as a whole defines the atmosphere corresponding to quiet solar regions, except for the polar regions, where the influence of the usually present coronal holes is manifest. For this reason, the radio radius in the polar regions is smaller than at 1.7 cm, where, according to [14], it is $1.0125\,R_0$.

If we assume that the nonuniform limb brightness distribution is due to coronal loop structure, we can consider the limiting size of this structure visible at 5.2 cm to correspond to the radial distances at the level $0.01\,T_0^c$. In the equatorial direction, this boundary corresponds to a height of about 300 000 km; according to empirical models of quiet regions [18], the electron density increases sharply beginning from such heights. The radial distances measured at the 0.01 level change along the limb from $1.44\,R_0$ to $1.07\,R_0$.

Our studies provide hope that, in the future, we will be able to use such results to derive statistical parameters regarding the loop structure of the solar corona.

ACKNOWLEDGMENTS

The author is grateful to the staff of the Radio Astrophysics Department of the Institute of Solar–Terrestrial Physics, Siberian Division of the Russian Academy of Sciences, for discussions of this work, as well as to V.I. Smirnov and V.V. Grechnev for valuable comments.

REFERENCES

1. P. N. Swanson, Sol. Phys. **32**, 77 (1973).
2. J. C. Ceballos and P. Lantos, Sol. Phys. **22**, 142 (1972).
3. E. Fürst, O. Hachenberg, and W. Hirth, Astron. Astrophys. **36**, 123 (1974).
4. F. Chiuderi-Drago, E. Fürst, W. Hirth, *et al.*, Astron. Astrophys. **39**, 429 (1975).
5. E. Fürst, P. Lantos, and W. Hirth, Sol. Phys. **63**, 257 (1979).
6. P. Lantos, E. Fürst, and W. Hirth, Sol. Phys. **63**, 271 (1979).
7. U. Feldman, Astrophys. J. **275**, 367 (1983).
8. K. A. Marsh, G. J. Hurford, and H. Zirin, Astron. Astrophys. **94**, 67 (1981).
9. M. R. Kundu, A. P. Rao, F. T. Erskine, *et al.*, Astrophys. J. **234**, 1122 (1979).
10. V. N. Borovik, Lect. Notes Phys. **432**, 185 (1994).
11. A. F. Bachurin, Izv. Krymsk. Astrofiz. Obs. **68**, 68 (1983).
12. A. Nindos, M. R. Kundu, S. M. White, *et al.*, Astrophys. J. **527**, 415 (1999).
13. C. L. Selhorst, A. V. R. Silva, J. E. R. Costa, *et al.*, Astron. Astrophys. **401**, 1143 (2003).
14. C. L. Selhorst, A. V. R. Silva, and J. E. R. Costa, Astron. Astrophys. **420**, 1117 (2004).
15. G. Ya. Smol'nik, B. B. Krissinel', T. A. Treskov, *et al.*, Issled. Geomagn. Aéron. Fiz. Solntsa **91**, 146 (1990).
16. V. V. Grechnev, S. V. Lesovoi, G. Ya. Smolkov, *et al.*, Sol. Phys. **216**, 239 (2003).
17. V. M. Bogod, A. S. Grebinskiĭ, and L. V. Opeĭkin, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **39**, 1450 (1996).
18. K. V. Getman and M. A. Livshits, Astron. Zh. **73**, 119 (1996) [Astron. Rep. **40**, 104 (1996)].
19. G. S. Vaiana and R. Rozner, Annu. Rev. Astron. Astrophys. **16**, 393 (1978).
20. B. B. Krissinel, S. M. Kuznetsova, V. P. Maksimov, *et al.*, Publ. Astron. Soc. Jpn. **52**, 909 (2000).
21. B. I. Lubyshev, G. Ya. Smol'nik, T. A. Treskov, *et al.*, in *Proceedings of the International Conference on Solar Physics, Troitsk, Russia, 1999* (Nauka, Moscow, 1999), p. 306.

Translated by G. Rudnitskiĭ