Astronomy Reports, Vol. 48, No. 8, 2004, pp. 611–621. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 675–686.
Original Russian Text Copyright © 2004 by Gusev, Kaĭsin.

Multicolor CCD Photometry of the Dusty, Giant, Late-Type Spiral Galaxy NGC 5351

A. S. Gusev[1] and S. S. Kaĭsin[2]

[1] *Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119992 Russia*

[2] *Special Astrophysical Observatory, Russian Academy of Sciences, Nizhniĭ Arkhyz, Zelenchuk, Karachaevo-Cherkesiya 369167, Russia*

Received April 21, 2003; in final form, March 15, 2004

Abstract—The paper reports the results of $BVRI$ surface photometry of the giant galaxy NGC 5351 based on CCD observations obtained on the 1-m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences. Analysis of the structure and radial brightness distribution in the galaxy shows that NGC 5351 has a complex and, in some places, asymmetric structure. The galaxy possesses a large quantity of dust. The average internal extinction due to dust is $A_V = 1.2^m \pm 0.4^m$. After correcting for the effect of this dust, the parameters of the galaxy are typical of late-type spirals. The compositions of the stellar population in various parts of the galaxy are estimated using two-color diagrams. Star-forming regions in NGC 5351 are identified and studied. Most of the star-forming regions are located in the ring of the galaxy. Evolutionary modeling is used to estimate the ages of regions of violent star formation. An elliptical companion galaxy to NGC 5351 was found. The rotation curve of the galaxy is modeled and its mass estimated. The disk of NGC 5351 is self-gravitating within its optical radius.
© *2004 MAIK "Nauka/Interperiodica".*

1. GENERAL INFORMATION

The giant and fairly distant barred spiral galaxy NGC 5351 (Fig. 1) remains a comparatively poorly studied object. Table 1 gives the principal parameters of the galaxy according to the RC3 catalog [1]. Various authors have estimated the position angle of the galaxy within a radius of $\sim 1'$ from the nucleus to be from 93° [2] to 104° [3].

Photometry of the galaxy has been performed in the V [4], R [3, 5, 6], and V and I [7] filters. However, so far, no detailed multicolor photometry has been carried out for NGC 5351. According to [3–6], the brightness of the galactic disk decreases exponentially with distance on a scale of about $20''$.

The galaxy has a regular, symmetrical shape; its two spiral arms wind up, forming an inner ring. Numerous long and continuous outer arms extend outside the ring [2, 8, 9]. NGC 5351 has a small companion—the galaxy NGC 5349—located $3.4'$ to the southeast [2].

21-cm HI observations revealed a powerful gaseous disk extending to 1.5 optical radii of the galaxy [10]. The gas motions in some local regions of the outer disk and in the spiral arms of the galaxy are strongly perturbed [2]. Given the accepted distance of 48.9 Mpc, the mass of the galaxy within $1'$ from the center is estimated to be $2.2 \times 10^{11}\ M_\odot$ [2, 10], and the mass of HI, to be $1.3 \times 10^{10}\ M_\odot$ [10]. The integrated mass-to-luminosity ratio is $M/L = 6.5(M/L)_\odot$ [2]. A rotation curve for NGC 5351 can be found in [2].

2. OBSERVATIONS AND DATA REDUCTION

We observed NGC 5351 on January 21–22 and 22–23, 1998, with the 1-m Zeiss-1000 telescope (focal distance 13.3 m) of the Special Astrophysical Observatory of the Russian Academy of Sciences equipped with a standard CCD photometer [11]. The K585 CCD camera combined with B, V, R, and I broadband filters realizes a photometric system that is close to the standard Johnson–Cousins $BVRI$ system. The 530×580 pixels CCD yields a $143'' \times 212''$ field of view for an image scale of $0.28'' \times 0.37''$ per pixel. Table 2 gives information about the observations.

We took five to six exposures in each filter, shifting the telescope by several arcseconds between each exposure to reduce the effect of CCD defects. The seeing was $2.5''–3.5''$.

The photometric calibration of the galaxy was based on standard stars (PG 0220, PG 1407, RU 149, and S 101429) from the list of Majewski *et al.* [12] observed on the same night. We also used the aperture photometry data for the galaxy from [13].

1063-7729/04/4808-0611$26.00 © 2004 MAIK "Nauka/Interperiodica"

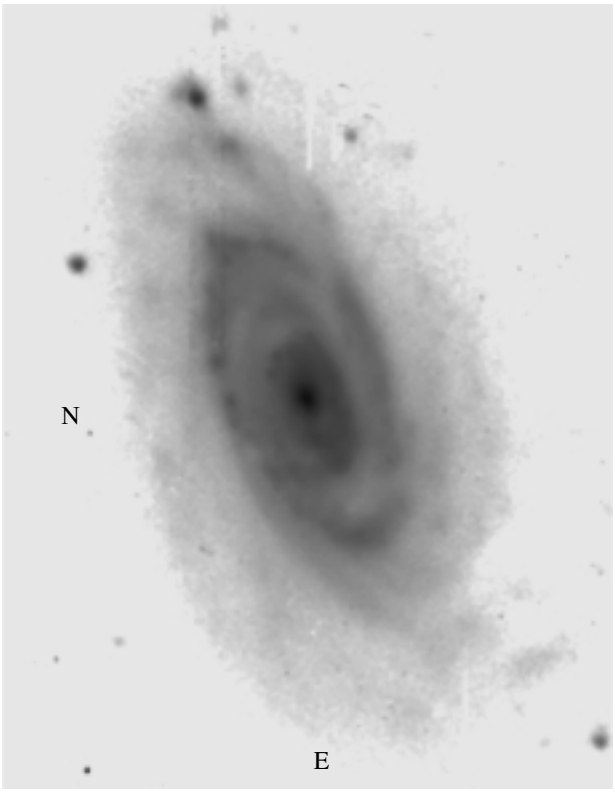

Fig. 1. CCD R image of NGC 5351. The size of the image is $2.4' \times 3.5'$.

The accuracy of the photometric calibration was 0.12^m, 0.08^m, 0.05^m, and 0.04^m in B, V, R, and I, respectively. The errors of the photometric measurements of the galaxy reported below in Section 3 do not include these calibration errors, which must be added to the errors listed in Section 3 to obtain the absolute photometric errors. This is not true, however, for the photometric data for the companion galaxy NGC 5351 (see Section 3.4 and Table 5 below); in this case, the corresponding errors include the photometric calibration errors.

We performed the preliminary reduction of the images at the Special Astrophysical Observatory of the Russian Academy of Sciences. To correct for electron bias and the effect of "hot pixels" and bad columns of the CCD array, we subtracted a dark frame that was the average of several exposures taken with a closed shutter and the same integration time as for the object. We used summed frames of the twilight sky with signal-to-noise ratios of no less than 60 to 80 taken in each of the filters as flat fields. We divided each image by its flat field to correct for the nonuniform sensitivity of the detector pixels.

We performed the subsequent data reduction at the Sternberg Astronomical Institute using a standard procedure incorporating the ESO MIDAS image processing procedures. The principal reduction stages included the following.

(a) Scaling of the images to $0.37''$/pixel and alignment of the galaxy images (with an accuracy of up to 0.1 pixel). The alignment of the images made it possible to efficiently eliminate cosmic-ray traces and the effects of individual hot pixels and bad columns of the CCD.

(b) Determining the sky background and subtracting it from each frame.

(c) Coadding the galaxy images taken in the same filters.

(d) Correcting for the air mass.

(e) Transforming the counts to a logarithmic scale (mag/arcsec^2) based on the results of the photometric calibration.

(f) Reducing the instrumental photometric system to the standard Johnson–Cousins photometric system (taking into account the derived color equations).

(g) Subtracting galaxy images taken in different filters in order to obtain color-index maps.

Table 1. Principal parameters of NGC 5351

Parameter	Value
Type	SBb
m_B, mag	13.00
$M_B^{0,i}$, mag	−21.19
V_{LG}, km/s	3667
R, Mpc ($H_0 = 75$ km s^{-1} Mpc^{-1})	48.9
D_{25}, arcmin	2.88
i, deg	59.2
PA, deg	100

Table 2. Log of observations

Date	Filter	Exposures, s
Jan. 21–22, 1998	B	3×600
	V	4×600
	R	4×600
	I	4×600
Jan. 22–23, 1998	B	2×600
	V	2×600
	R	1×600
	I	1×600

We also used images of NGC 5351 in the J, H, and K IR filters taken from the 2MASS catalog. We reduced these images using a procedure similar to that applied to the optical frames. The JHK images have a seeing of 1″ and a scale of 1.0″/pixel × 1.0″/pixel.

We corrected all the data (brightness and color indices) for Galactic extinction (based on the RC3 data). When constructing two-color diagrams and investigating the color characteristics of sites of star formation, we also corrected for extinction associated with the inclination of the disk of NGC 5351 (based on the RC3 data). In the latter case, the corrected quantities are marked by the subscripts "$0, i$." Table 3 lists the adopted corrections for the various filters. The image scale is 88 pc/pixel for the accepted distance of the galaxy.

Table 3. Galactic extinction and extinction due to the inclination of NGC 5351

Color index	E	$E(i)$
$B-V$	0.06	0.09
$V-R$	0.03	0.05
$R-I$	0.04	0.07

3. ANALYSIS OF THE RESULTS

3.1. Photometric Profiles and Morphology of the Galaxy

The galaxy has a diffuse, moderately bright nuclear region: the maximum brightness within several pixels of the nucleus exceeds $\mu_V = 19.6^m/\text{arcsec}^2$ (Fig. 2a). The diameter of the nucleus—3.2 kpc (14″; Figs. 2a and 2b)—is fairly large, but we should bear in mind that NGC 5351 is a large galaxy (according to [1], its diameter is 41 kpc). The isophotes are much less flattened in the nucleus than in the disk (especially in the IR), and their ellipticity reaches $e = 0.17$ in the I filter (Fig. 3a), indicating the presence of an extended bulge whose flux predominates in the central region (within 10″). The position angle of the isophotes varies as a function of the wavelength (from 136° in B to 101° in I, see Fig. 3b).

The 8.3 kpc × 5.2 kpc (35″ × 22″) central region of the disk of NGC 5351, whose isophotes are extended along the major axis of the galaxy, is of special interest. The isophotes in this region are nearly rectangular (see the $21.5^m/\text{arcsec}^2$ isophote in Fig. 2b), and the region itself is shifted to the east of the nucleus (see the asymmetry of the photometric profiles at nucleocentric distances of 7″–20″ in Fig. 2a). This region morphologically resembles an asymmetric bar. We show below that the asymmetry of the bar is due to the nonuniform distribution of dust in the galaxy. The B and V surface brightnesses of the bar to the west of the nucleus do not vary with distance and are equal to $20.8 \pm 0.2^m/\text{arcsec}^2$ in V (Fig. 2a). In the redder filters (R and, especially, I), the brightness decreases with galactocentric distance (e.g., from 19.2 to $19.7^m/\text{arcsec}^2$ in I), reflecting the known fact that the fraction of young stars increases toward the tip of the bar [14].

The observed structure of the bar east of the nucleus is complex: it has a semiring with a radius of 1.9 kpc (8″; Fig. 1) whose V brightness is 0.1^m–$0.15^m/\text{arcsec}^2$ higher than the brightness of the surrounding background. A short (in projection) spiral arm winding in the direction opposite to that of other arms emerges from the tip of the eastern part of the bar. However, in this case, we may be observing a disconnected ring (Figs. 1 and 2b). This is the brightest arm in the galaxy, and its surface brightness reaches $\mu_V = 21.3 \pm 0.1^m/\text{arcsec}^2$. Afanas'ev *et al.* [2] pointed out the presence of strongly noncircular gas motions in this region. The ellipticity of the bar isophotes is $e = 0.20 \pm 0.05$ in B, V, and R and somewhat higher in I (Fig. 3a). The position angle of the bar is 105.5° ± 0.5° in V, R, and I (Fig. 3b).

The inner spiral arms of the galaxy, which form a ring with a diameter of 17.5 kpc (74″), are barely visible against the disk. Only the western inner arm is conspicuous (Fig. 1). Figure 2a clearly shows the arrangement of the arms: the section made through the northwestern part of NGC 5351 ($r = -40''$ to $-20''$) passes nearly along the western spiral arm ($\mu_V = 21.65 \pm 0.15^m/\text{arcsec}^2$), while the section made through the southeastern part of the galaxy ($r = 20''$–$40''$) crosses the inner disk ($\mu_V = 22.1 \pm 0.2^m/\text{arcsec}^2$) and ring (the maximum $\mu_V = 21.2^m/\text{arcsec}^2$).

Two spiral arms can be seen in the outer disk of NGC 5351, in the western and eastern parts of the galaxy. A bright diffuse object is visible at the tip of the western arm (Figs. 1 and 2b). See Section 3.4 for a more detailed discussion of this object.

The position angle of the galaxy can be very accurately determined from the orientation of its outer isophotes: PA = 111.2° ± 0.6°. The inclination of the disk inferred from the flattening of the B, V, and R isophotes of the thick disk is $i = 62° \pm 1°$. I images of the galaxy yield a somewhat higher inclination, $i = 65°$. Figure 4 shows averaged photometric profiles of NGC 5351 computed for the derived PA and i. The region of the ring of the galaxy is prominent at $r = 37''$ and separates the galactic disk into its inner and outer parts.

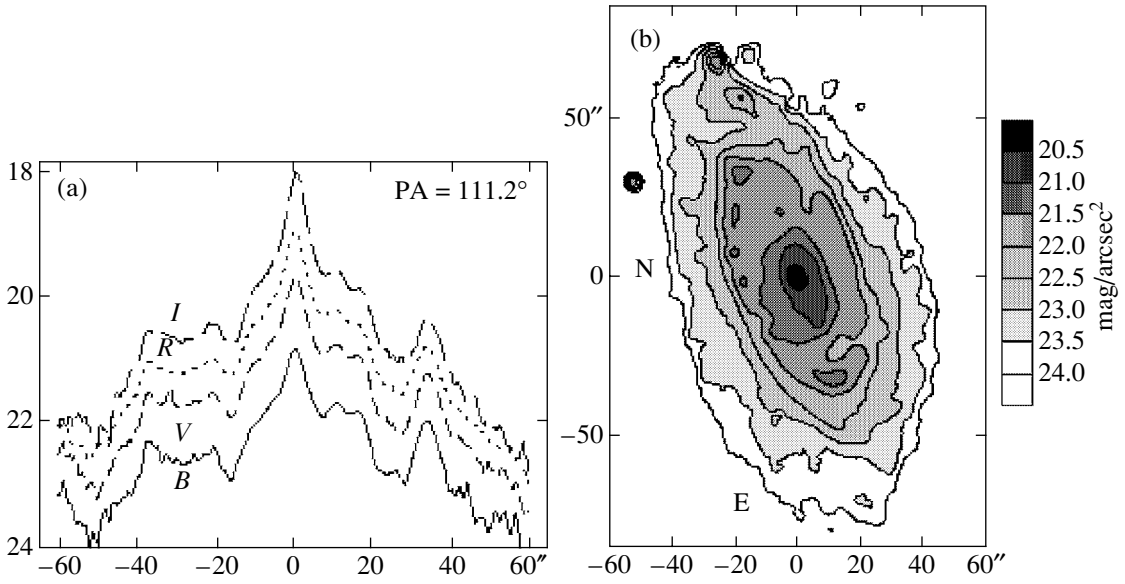

Fig. 2. (a) Photometric profiles along the major axis of NGC 5351 (in mag/arcsec^2) taken in the B (solid), V (short-dashed), R (dotted), and I (long-dashed) filters, and (b) the V image of the galaxy.

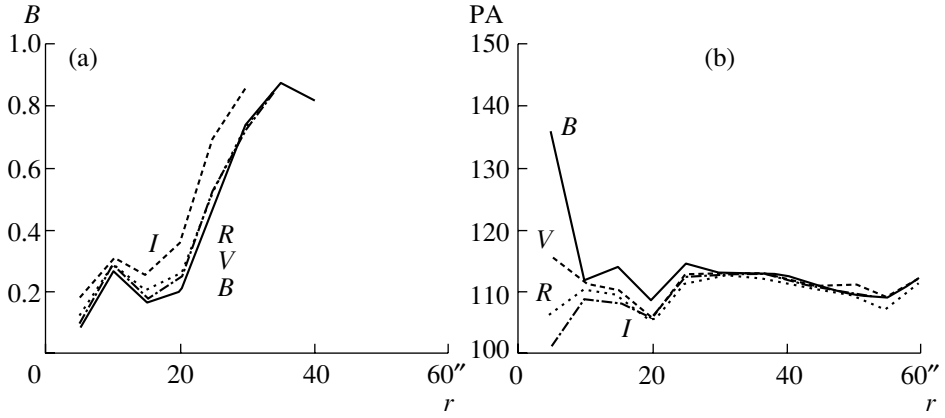

Fig. 3. (a) Isophote ellipticity $e = 1 - b/a$ and (b) position angle of the galaxy in B, V, R, and I as functions of distance r from the center of NGC 5351.

Analysis of the brightness profiles showed that the inner and outer parts of the disk have different parameters and that this difference becomes more pronounced at shorter wavelengths. The brightness of the outer disk ($r = 50''-70''$) decreases exponentially on a scale of 6.79 ± 0.17 kpc ($28.6'' \pm 0.7''$) in V and 5.25 ± 0.39 kpc ($22.2'' \pm 1.7''$) in K. The central brightnesses are $\mu_V = 20.78 \pm 0.20^m$/arcsec^2 and $\mu_K = 17.32 \pm 0.09^m$/arcsec^2 in V and K, respectively. The brightness in the region $r = 12''-28''$ occupied by the inner disk decreases exponentially in a somewhat steeper fashion, on a scale of 4.60 ± 0.24 kpc ($19.4'' \pm 1.0''$) in V and 4.60 ± 0.08 kpc ($19.4'' \pm 0.3''$) in K. The corresponding central brightnesses are $\mu_V = 20.23 \pm 0.27^m$/arcsec^2 and $\mu_K = 16.83 \pm 0.06^m$/arcsec^2 (without subtracting the parameters of the outer disk). The I and K brightnesses of the outer disk decrease faster than the brightness in bluer filters (B, V, and R), probably due to an increase in the number of hot, young stars toward the periphery of the galaxy. On the other hand, the scale length of the inner disk is virtually the same in all filters, which suggests the absence of a young stellar population in this region of the galaxy.

A decomposition of the central region of the galaxy into the inner disk and bulge shows that the brightness decreases in accordance with a de Vaucouleurs law at galactocentric distances $r = 2''-8''$. The bulge has effective radius $r_e(V) = 2.6 \pm 0.6$ kpc ($10.8'' \pm 2.6''$), $r_e(K) = 1.6 \pm 1.0$ kpc ($6.5'' \pm 4.2''$),

$\mu_V(r_e) = 26.65 \pm 1.15^m/\text{arcsec}^2$, and $\mu_K(r_e) = 22.91 \pm 1.48^m/\text{arcsec}^2$. We thus conclude that the spherical component (bulge) contributes negligibly to the total luminosity and mass of the galaxy: the luminosity of the bulge is $3.5 \pm 2.0\%$ of the total K luminosity of NGC 5351.

3.2. Distribution of Colors and Dust

On the whole, NGC 5351 is as a rather red galaxy—its integrated color index is $B-V = 0.8^m$ [1]. Such $B-V$ values are typical of S0 galaxies. However, we show below that the galaxy contains a large quantity of dust; the reddening of the galaxy due to selective extinction by dust is, on average, $E(B-V) \sim 0.4^m$. It follows that the integrated $B-V$ color index of NGC 5351 is 0.4^m, which is typical of Sc galaxies. The intrinsic color indices of various parts of the galaxy will also be smaller than their observed values (by $0.3^m-0.7^m$ for $B-V$).

The nucleus of NGC 5351 is redder than the galaxy as a whole; its color indices are $B-V = 1.10^m \pm 0.05^m$ and $V-R = 0.57^m \pm 0.03^m$ (Figs. 5a–5c). The nucleus is especially prominent in $R-I$ (Figs. 5a and 5d). The color of the galaxy becomes bluer with distance from the nucleus—all three color indices decrease with galactocentric distance ($B-V$ from $1.05^m \pm 0.10^m$ to $0.7^m \pm 0.1^m$, $V-R$ from $0.65^m \pm 0.05^m$ to $0.4^m \pm 0.1^m$, and $R-I$ from $0.9^m \pm 0.1^m$ to $0.3^m \pm 0.2^m$). Figure 5a shows the clear asymmetry of the color parameters of NGC 5351 to the west and east of the nucleus at galactocentric distances ranging from $5''$ to $35''$. This asymmetry can be explained by the complex morphology of the galaxy in this region: to the west of the nucleus ($r = -5''$ to $35''$), the section passes through the inner spiral arm, while the section to the east of the nucleus ($r = 5''-35''$) passes through the bar and inner disk.

As we noted in Section 3.1, the bar of the galaxy becomes bluer toward its periphery. Its color indices decrease from $0.96^m \pm 0.01^m$ to $0.86^m \pm 0.01^m$ in $B-V$, from $0.63^m \pm 0.01^m$ to $0.40^m \pm 0.01^m$ in $V-R$, and from $0.68^m \pm 0.02^m$ to $0.60^m \pm 0.02^m$ in $R-I$ (Fig. 5a). The color indices of the western inner spiral arm coincide with the corresponding mean color indices of the bar. As a result, the color-index contours in Fig. 5b ($B-V = 0.9^m$, $V-R = 0.50^m$, and $R-I = 0.55^m$) bound a region whose shape resembles that of a standard symmetrical bar. This indicates that the stellar population of the bar is the same as that of the western spiral arm (i.e., the region located within 5.6 kpc ($20''$) of the center).

The ring of NGC 5351 is bluer than the surrounding regions of the outer and inner disks. This is especially obvious in Fig. 5a at $r = \pm 35''$ (where the section intersects the ring). The color indices of the ring, $B-V = 0.75^m \pm 0.04^m$ and $V-R = 0.35^m \pm 0.03^m$, are $0.1^m-0.2^m$ smaller than the corresponding color indices of the surrounding regions. The brightest parts of the ring are clearly visible in Figs. 5b and 5c, but the ring cannot be distinguished in $R-I$ (Fig. 5a), suggesting a similarity between the old stellar populations of the disk and galactic ring.

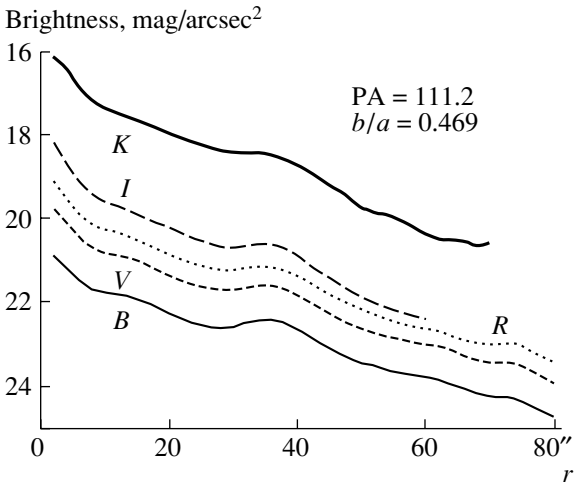

Fig. 4. Averaged B, V, R, I, and K photometric profiles of NGC 5351 (in mag/arcsec^2).

The $R-I$ color index decreases much faster along the minor axis than do $B-V$ and $V-R$ (Fig. 5d). The integrated IR color indices of the galaxy are $J-H = 0.68^m$ and $H-K = 0.25^m$.

We used the BVK images of the galaxy to estimate the distribution of dust and the extinction it produces in the galaxy. We applied the technique and model computations of Rhee and Albada [15], developed for the analysis of the distribution of internal extinction in a galaxy based on the $(B-V)_0-(V-K)_0$ two-color diagram. Figure 6 shows the distributions of dust and the associated extinction in NGC 5351. Note that the distribution of the internal extinction depends strongly on the adopted model for the composition of the stellar population, the star-formation history, and the distribution of the dust with respect to the stellar disk and is therefore only approximate. The derived extinctions A_V are accurate to $\pm 0.35^m$. For regions of ongoing star formation with an extremely young stellar population, A_V can be calculated unambiguously using the $(B-V)_0-(V-K)_0$ diagram. We analyze the extinction in star-forming regions in Section 3.4 and do not show this in Fig. 6.

Analysis of the distribution of dust yields some insights into the morphology of the galaxy. The nonuniform dust distribution in the central region of NGC 5351 explains the asymmetry of the bar (Fig. 2a). The numerous dust lanes in the disk distort its structure (in its northern and southeastern parts)

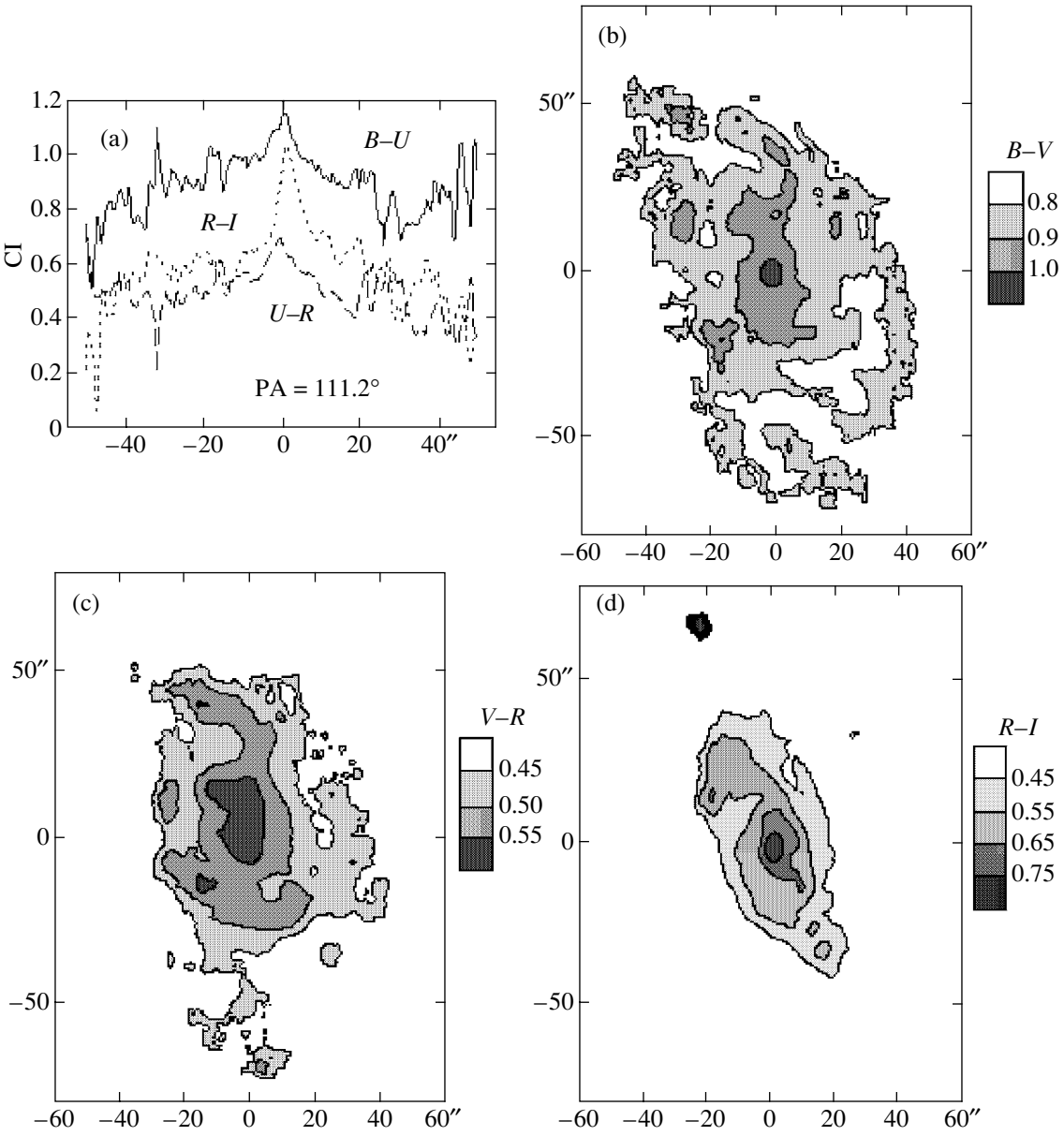

Fig. 5. (a) $B-V$ (solid), $V-R$ (dashed), and $R-I$ (dotted) color indices measured along the major axis of NGC 5351 and maps of the (b) $B-V$, (c) $V-R$, and (d) $R-I$ color indices.

and the spiral arms of the galaxy (in the northern, western, and southeastern parts of the galaxy). This all suggests that the observed "irregularity" of the structure of NGC 5351 can be explained by the nonuniform distribution of dust in the galaxy.

The B, V, and K photometric parameters of the galaxy can be used to estimate the internal extinction by dust, which proves to be fairly strong: $A_V = 1.2^m \pm 0.35^m$, on average, for the entire galaxy, with A_V increasing from 0.6^m in the central region to 2.1^m in the dust lanes of the outer disk and regions of ongoing star formation. These extinctions correspond to reddenings $E(B-V) = 0.2^m - 0.7^m$. We have $E(B-V) = 0.4^m \pm 0.1^m$, on average, for the entire galaxy. It follows that the intrinsic color of NGC 5351 is typical of late-type spiral galaxies. The strong internal extinction of the galaxy explains the discrepancy between its observed color parameters (typical of early-type galaxies) and the lack of an appreciable bulge (a characteristic feature of late-type galaxies).

3.3. Two-Color Diagrams

Figures 7a–7d show the $(B-V)_0^i - (V-R)_0^i$ and $(B-V)_0^i - (V-I)_0^i$ two-color diagrams for various regions of NGC 5351. The numbers in Figs. 7a and

7c indicate the colors of *1* the nucleus within 4″ of the center; *2* the bulge at galactocentric distances of 4″–7″; *3* the spiral arms; *4*, *5* the inner and outer disks, respectively; and *6*–*8* the bar at galactocentric distances of 1.3, 2.4, and 3.8 kpc (5.5″, 10″, and 16″), respectively.

The two-color diagrams have been corrected for extinction due to the inclination of NGC 5351, so that the color indices for different regions differ somewhat from those obtained in Section 3.2. The differences between the corresponding color indices—the reddening due to the inclination of NGC 5351—are given in Table 3.

The color indices are not corrected for the internal extinction by dust in the galaxy we have estimated here. Extinction by dust cannot explain the deviations of points in the diagrams from the normal integrated color sequence of galaxies (NCS); it shifts points upward and to the left along the NCS.

The points representing the color indices of different regions of the galaxy are aligned along a line that does not coincide with the NCS. The galactocentric distances of regions are correlated fairly well with their positions on the two-color diagrams: regions are shifted increasingly to the left and upward with increasing galactocentric distance (Figs. 7a and 7c). This behavior indicates an increase in the fraction of young stars toward the periphery of the galaxy. A number of regions are located to the left of the NCS, which cannot be explained by the effect of dust or the specific nature of the chemical composition of the galaxy, since these factors would shift points nearly along the NCS in the $(B-V)^i_0$–$(V-R)^i_0$ and $(B-V)^i_0$–$(V-I)^i_0$ diagrams. The regions in question are mostly in the outer parts of the bar. Their deviation from the NCS may be a consequence of a complex star-formation history in the disk of NGC 5351. The shift of points representing the color indices of the bar to the left of the NCS may be due to a deficit of intermediate-age stars in the outer parts of the bar [14]. The circumnuclear region of NGC 5351 occupies a position in the two-color diagrams that corresponds to systems with old stellar populations (with ages of about 10^{10} yr).

3.4. Star-Forming Regions and the Companion Galaxy

Nine small and mostly blue starlike and diffuse bright regions have been identified in NGC 5351. The color indices of a bright starlike object to the north of NGC 5351 (Figs. 1 and 2b) identified it as a field star, and we do not analyze this object below. Figures 7b and 7d show the positions of the remaining eight objects on the two-color diagrams. Table 4 lists the

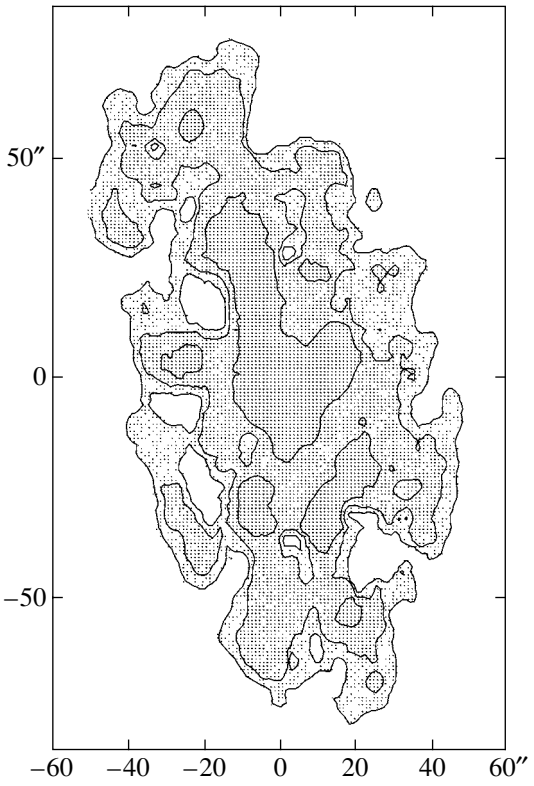

Fig. 6. Map of internal extinction by dust A_V in NGC 5351. Extinction contours are shown at 1.0^m, 1.5^m, and 2.0^m. Lighter regions correspond to stronger extinction.

derived parameters of these regions: (1) the number of the object; (2) coordinates of the region in arcsec relative to the center of the galaxy; (3–5) $(B-V)^i_0$, $(V-R)^i_0$, and $(V-I)^i_0$ color indices; (6) the diameter of the object in pc; and (7) the age of the object. We determined the diameter of each region from the best-quality V frame (with 2.1″ seeing, which corresponds to 500 pc). We determined the brightness of the regions studied in each filter by subtracting the brightness of the surrounding base of NGC 5351 from the brightness distribution of the region occupied by the object. We inferred the ages of the regions from their positions on the two-color diagrams using evolutionary tracks of aging stellar systems computed with the PEGASE2 program developed by Fioc and Rocca-Volmerange (Institut d'Astrophysique de Paris) [17]. The color indices of the stellar systems were modeled using the initial mass function of Kennicutt with upper and lower mass limits of 120 $M_\odot$ and 0.1 $M_\odot$. We considered systems with metallicities Z higher and lower than the solar value.

Five diffuse objects (2–4, 6, and 7) are located in the ring of NGC 5351. Four are located in the northern part of the ring (2–4, and 7), have characteristic sizes ranging from 650 to 1350 pc (so that these

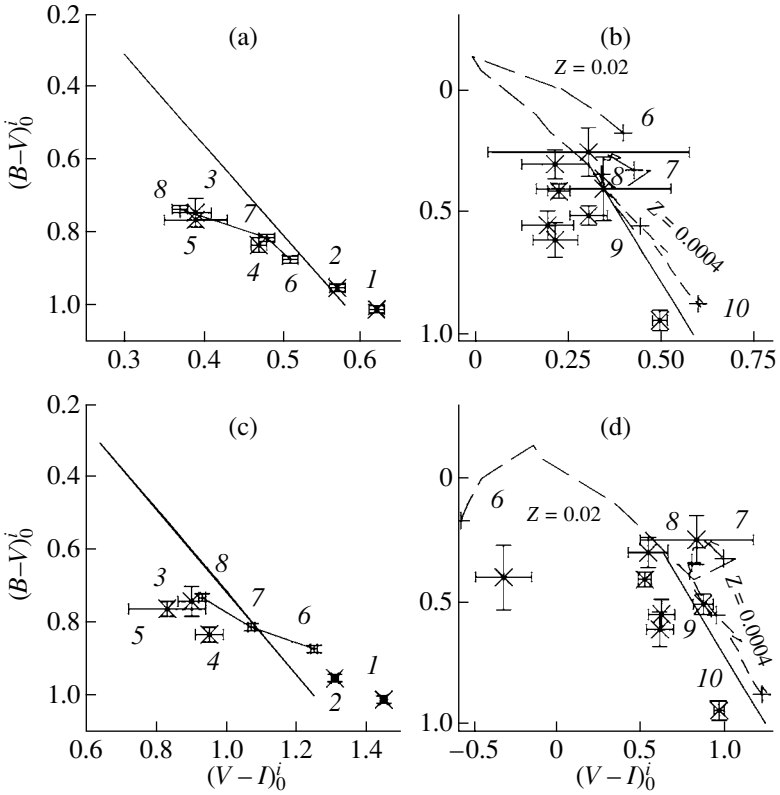

Fig. 7. (a, b) $(B-V)_0-(V-R)_0$ and (c, d) $(B-V)_0-(V-I)_0$ two-color diagrams for (a, c) NGC 5351 and (b, d) sites of star formation in the galaxy. The straight solid line shows the NCS of galaxies according to Buta and Williams [16]. The solid broken lines in diagrams (a, c) show the variation of the color indices along the major axis of the bar in NGC 5351. The dashed lines in diagrams (b, d) show the evolutionary tracks of an aging stellar system in which all its stars formed in a single burst of star formation during the first 10^6 yr at $Z = 0.02$ for an age of $10^6-5\times10^8$ yr (long-dashed line) and at $Z = 0.0004$ for an age of $5\times10^8-2\times10^{10}$ yr (short-dashed line). The numbers in diagrams (b, d) indicate the logarithm of the age of the system (in yr). Also shown are the errors. See text for other notation.

objects are stellar complexes in the classification of Efremov [18]), and are located somewhat to the left of the NCS and below the evolutionary tracks. Object *1*, which is in the outer western spiral arm, has similar parameters. Such strong deviations of the positions of these regions from the evolutionary tracks in the diagrams can be explained only by very strong selective extinction by dust (the reddening $E(B-V)$ inside the regions may reach $0.6^m - 0.7^m$ for objects 3 and 7). This ties in with the results we obtained in Section 3.2. Extinction by dust shifts points downward and to the right along the NCS [16] in the two-color diagrams. In this case, all the objects considered here are young star-forming regions with ages of the order of $(3-8)\times10^6$ yr and extinctions by dust of $\sim1^m-2^m$ in V (Figs. 7b and 7d).

Region 6, which is located in the southern part of the ring, has color indices that are typical of a very young stellar system with an age of $<3.2\times10^6$ yr (Figs. 7b and 7d). This region has a diameter of 1 kpc, implying that it is a stellar complex in the classification of Efremov.

Object *5*, which is located to the south of the nucleus in the region occupied by the bar, is of considerable interest. It is a compact object with a diameter of less than 500 pc. The large errors of the measured color indices prevent a confident determination of its age, but its position in the two-color diagrams indicates that this region should be populated by moderately young stars.

The position of the bright diffuse object at the western edge of the outer disk of NGC 5351 (at a deprojected galactocentric distance of 73″ (17.5 kpc)) in the two-color diagrams indicates an old composition for its stellar population. A bar emerging from the disk of the galaxy toward this region can be seen in Figs. 1 and 2b. The photometric parameters of the object are consistent with the colors of a globular cluster, but its size (3.4 kpc) exceeds the typical sizes of globular clusters by almost two orders of magnitude. If this object has a metal-poor stellar population (by analogy with globular clusters), its age is about 10^{10} yr for $Z = 0.0004 - 0.004$. The absolute magnitude of the object is $M_V^0 = -16.16^m \pm 0.09^m$ (for the

Table 4. Paramaters of star-forming regions in NGC 5351

No.	Coordinates, arcsec	$(B-V)_0^i$	$(V-R)_0^i$	$(V-I)_0^i$	d, pc	τ, 10^6 yr
1	17.6N, 57.0W	0.41	0.22	0.53	1490	4.7 ± 0.8
2	17.2N, 0.9E	0.30	0.21	0.55	670	6.6 ± 2.8
3	20.5N, 8.2W	0.55	0.19	0.63	980	4.5 ± 1.3?
4	18.7N, 31.7W	0.51	0.30	0.88	660	6.3 ± 0.7
5	5.5S, 0.6E	0.25	0.30	0.84	<500	$5 - 100$
6	16.1S, 0.9W	0.40	0.34	−0.32	1050	<3.2
7	11.7S, 31.4E	0.61	0.21	0.62	1350	3.8 ± 1.5?

accepted distance of NGC 5351). According to the results of model computations performed using the PEGASE2 program, the object has a mass of $(1.5 \pm 1.1) \times 10^9\ M_\odot$ and so must be an elliptical companion galaxy of the giant galaxy NGC 5351. In terms of its mass and photometric parameters, the object resembles the companions of M31—the elliptical galaxies NGC 147, NGC 185, and NGC 205. Table 5 gives the inferred parameters of the companion galaxy. The brightness distribution in the companion is fairly typical of small ellipticals (Fig. 8). The averaged photometric profiles of the companion galaxy can be qualitatively described by a de Vaucouleurs law with an effective radius of $2''-3''$ (0.5–0.8 kpc), which is smaller at longer wavelengths (Fig. 8). Unfortunately, a detailed analysis of the photometric parameters of this object was hindered by its small diameter, which exceeds the seeing of the images by only a modest factor. Thus, the deviation of the observed brightness distribution from a de Vaucouleurs law in the central region (Fig. 8) can be explained by "smearing" with radiation from the central region of the companion galaxy.

3.5. Rotation Curve of the Galaxy

We constructed the rotation curve of NGC 5351, $V(r)$, based on spectrophotometric observations obtained on the 6-m telescope of the Special Astrophysical Observatory [2]. Since the scatter of the data obtained by Afanas'ev *et al.* [2] is fairly large, we averaged the observed values over 2-kpc intervals (Fig. 9) before modeling the rotation curve. The last data point corresponds to the maximum rotational velocity implied by HI radio observations within a galactocentric radius of $90''$ (21.4 kpc), according to Rhee and Albada [19]. We reduced all the data to the inclination and position angle derived in Section 3.1.

The maximum of the rotation curve based on the optical data ($\sim$200 km/s) is located at a galactocentric distance of 9–15 kpc, which corresponds to the radius of the ring and the outer region immediately adjacent to the ring (Section 3.1).

We modeled the rotation curve using the GR4 program developed at the Sternberg Astronomical Institute. We considered a two-component model with an exponential disk and a dark halo, with the disk scale length set equal to 6.8 kpc. We ignored the contribution of the bulge to the integrated rotation curve due to the large uncertainty in the curve.

Good agreement with the observed rotation curve can be obtained for models with various disk and halo masses. However, multicomponent models with dark halos and the minimum permissible disk mass yield unrealistically low mass-to-luminosity ratios M/L (about unity for the B and V luminosities). The observed rotation curve is fit best by a simple single-component model with a thin exponential disk (Fig. 9). The contribution of the bulge to the total

Table 5. Parameters of the companion galaxy of NGC 5351

Parameter	Value
Coordinates relative to the center of NGC 5351, arcsec	25.3N, 68.0W
Type	E3
D, kpc	3.4 ± 0.3
m_V, mag	17.29 ± 0.09
M_V^0, mag	-16.16 ± 0.09
PA, deg	120 ± 5
$(B-V)_0$	0.94 ± 0.14
$(V-R)_0$	0.49 ± 0.10
$(V-I)_0$	0.97 ± 0.10
$M/M_\odot$	$(1.5 \pm 1.1) \times 10^9$
τ, yr	$\sim 10^{10}$

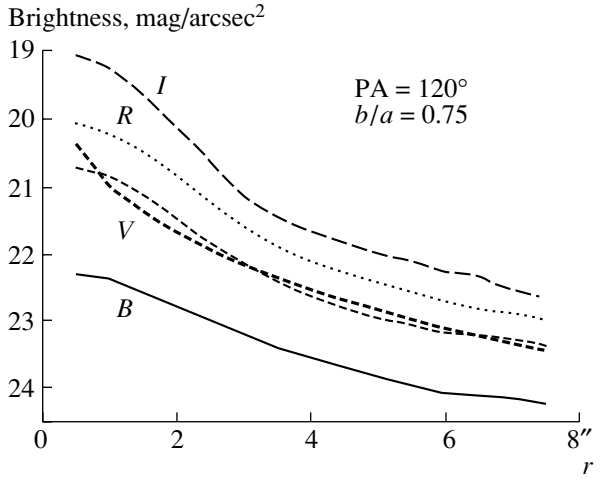

Fig. 8. Averaged B, V, R, and I photometric profiles of the companion galaxy (in mag/arcsec^2). The bold dashed curve shows the brightness distribution for a de Vaucouleurs law.

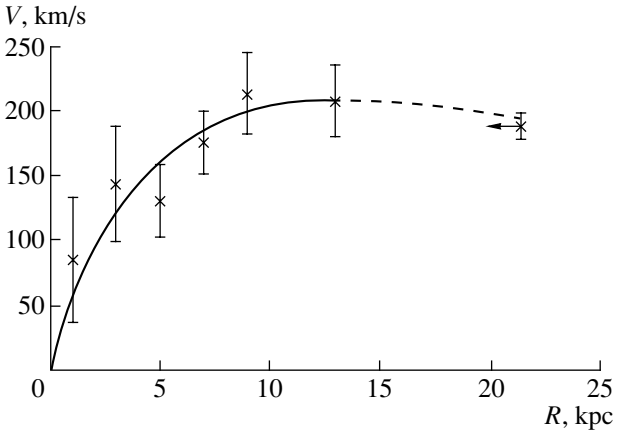

Fig. 9. Rotation curve of NGC 5351. The x's with error bars show the mean observed rotation curve based on the data of Afanas'ev *et al.* [2], with the maximum rotational velocity shown taken from Rhee and Albada [19] (the rightmost point). Also shown is the model rotation curve. All data are reduced to PA = 111.2° and inclination $i = 62°$.

mass of the galaxy within 90″ of its center can be assumed to be negligible in this case. The mass of the dark halo within the same radius does not exceed half of the total mass of the galaxy. It follows that the disk of NGC 5351 must be self-gravitating at these galactocentric distances.

The resulting model yielded disk masses M (which are equal to the mass of the entire galaxy) of $(1.8 \pm 0.2) \times 10^{11}\ M_\odot$ within 90″ (21.4 kpc) and $(1.1 \pm 0.2) \times 10^{11}\ M_\odot$ within 60″ (14.2 kpc) of the center. Given the internal extinction by dust, the luminosity of the disk within 60″ is $L_V = (5.4 \pm 1.4) \times 10^{10} L_\odot$ and $L_K = (1.1 \pm 0.1) \times 10^{11} L_\odot$, and the mass-to-luminosity ratio within the same radius is $M/L_V = (2.3 \pm 1.0)(M/L_V)_\odot$, $M/L_K = (1.0 \pm 0.3)(M/L_K)_\odot$, and $M/L_B = (2.0 \pm 0.8)(M/L_B)_\odot$, respectively. These M/L values are close to those expected for galaxies with similar color parameters (according to photometric evolutionary models [20]).

NGC 5351 has the properties of a late-type spiral galaxy. Its mass-to-luminosity ratios and the model computations of the dependence of M/L on color indices reported by Bell and De Jong [20] indicate that the star-formation history in NGC 5351 is similar to that expected for late-type spiral galaxies: the star-formation rate in the galaxy decreases exponentially on a time scale of $(9 \pm 4) \times 10^9$ yr, with no powerful global bursts of star formation in the past; the metallicity of the galaxy is approximately equal to the solar value. The small contribution of the bulge to the luminosity and kinematics of the galaxy and its relative HI content are also typical of the corresponding parameters of late-type spirals. The observed red color parameters of the galaxy are due to strong internal selective extinction by dust in NGC 5351.

The only region in NGC 5351 with ongoing active processes is its ring, which divides the disk into two parts with different photometric scale lengths (Section 3.1) and hosts most of the star-forming complexes (Section 3.4) discovered so far, as well as regions of noncircular gas motions [2].

4. CONCLUSIONS

(1) We have carried out $BVRI$ CCD photometry of the giant spiral galaxy NGC 5351. The galaxy has a complex structure consisting of a bulge without a conspicuous nucleus, an inner and outer ring, an asymmetric bar, a complex spiral pattern, a disk with different color parameters inside and outside the outer ring, and numerous dust lanes.

(2) A 37″-radius (8.8 kpc) ring divides the disk into outer and inner parts. The V brightness of the disk ($r = 12''-28''$) decreases exponentially on a scale of $19.4'' \pm 1.0''$ (4.60 ± 0.24 kpc), while that of the outer disk ($r = 50''-70''$) decreases exponentially on a scale of $28.6'' \pm 0.7''$ (6.79 ± 0.17 kpc).

(3) An asymmetrical bar $35'' \times 22''$ (8.3×5.2 kpc) in size can be seen. The B and V surface brightnesses of the bar to the east of the nucleus do not vary with galactocentric distance, $V = 20.8 \pm 0.2^m$/arcsec^2. The asymmetry of the bar is due to the nonuniform distribution of dust in the galaxy.

(4) The galaxy possesses large quantities of dust. The reddening of the galaxy due to selective extinction by dust is, on average, equal to $E(B-V) \sim 0.4^m$.

(5) We have identified seven starlike and diffuse star-forming regions (stellar complexes) in the galaxy. Most of these regions are located in the ring. The typical sizes and ages of the star-forming regions are 400–1350 pc and roughly $(3–8) \times 10^6$ yr.

(6) The giant elliptical galaxy NGC 5351 has a (type E3) elliptical companion galaxy to the west of its center at a distance of 73″ (17.5 kpc). The absolute magnitude and diameter of the companion are $M_V^0 = -16.16^m \pm 0.09^m$ and 3.4 kpc.

(7) The rotation curve of NGC 5351 within 90″ (21.4 kpc) of its center can be explained by a model with a single component—a stellar disk. The mass M of the disk in this case is $(1.8 \pm 0.2) \times 10^{11}\ M_\odot$. The mass and mass-to-luminosity ratios within 60″ (14.2 kpc) of the center are $M = (1.1 \pm 0.2) \times 10^{11}\ M_\odot$, $M/L_V = (2.3 \pm 1.0)(M/L_V)_\odot$, and $M/L_K = (1.0 \pm 0.3)(M/L_K)_\odot$, implying that the galactic disk must be self-gravitating at the galactocentric distances considered.

(8) NGC 5351 has the properties of a late-type spiral galaxy, as is demonstrated by the small contribution of the bulge to the luminosity and kinematics of the galaxy and the relative HI content of NGC 5351. The star-formation history in the galaxy is also typical of late-type spirals: the star-formation rate decreases exponentially on a time scale of $(9 \pm 4) \times 10^9$ yr without powerful bursts of star formation in the past. The red color of the galaxy can be explained as a consquence of internal selective extinction by dust.

ACKNOWLEDGMENTS

We are grateful to A.V. Zasov (Sternberg Astronomical Institute) for sharing a model of the rotation curve of the galaxy and fruitful discussions and to D.V. Bizyaev (Sternberg Astronomical Institute) for performing the preliminary reduction of the observational data. We thank the reviewer for a number of important comments and addenda. This work was supported by the Russian Foundation for Basic Research (project nos. 01-02-16800 and 01-02-17597).

REFERENCES

1. G. de Vaucouleurs, A. de Vaucouleurs, H. G. Corwin, *et al.*, *3rd Reference Catalogue of Bright Galaxies* (Springer, New York, 1991).
2. V. L. Afanas'ev, A. N. Burenkov, A. V. Zasov, and O. K. Sil'chenko, Astron. Zh. **69**, 19 (1992) [Sov. Astron. **36**, 10 (1992)].
3. P. J. Grosbol, Astron. Astrophys., Suppl. Ser. **60**, 261 (1985).
4. W. E. Baggett, S. M. Baggett, and K. S. J. Anderson, Astron. J. **116**, 1626 (1998).
5. S. M. Kent, Astrophys. J., Suppl. Ser. **59**, 115 (1985).
6. S. Courteau, Astrophys. J., Suppl. Ser. **103**, 363 (1996).
7. J. Roth, Astron. J. **108**, 862 (1994).
8. D. M. Elmegreen and B. G. Elmegreen, Astrophys. J. **314**, 3 (1987).
9. D. M. Elmegreen and B. G. Elmegreen, Astrophys. J. **445**, 591 (1995).
10. A. H. Broeils and M.-H. Rhee, Astron. Astrophys. **324**, 877 (1997).
11. I. I. Zin'kovskii, S. S. Kaisin, A. I. Kopylov, *et al.*, Tekh. Otchet SAO RAN **231** (1994).
12. S. R. Majewski, R. G. Kron, D. C. Koo, and M. A. Bershady, Publ. Astron. Soc. Pac. **106**, 1258 (1994).
13. P. Prugniel and P. Heraudeau, Astron. Astrophys., Suppl. Ser. **128**, 299 (1998).
14. A. S. Gusev, Astron. Zh. **77**, 654 (2000) [Astron. Rep. **44**, 579 (2000)].
15. S. A. Kassin, J. A. Frogel, R. W. Pogge, *et al.*, Astron. J. **126**, 1276 (2003).
16. R. Buta and K. L. Williams, Astron. J. **109**, 543 (1995).
17. M. Fioc and B. Rocca-Volmerange, Astron. Astrophys. **326**, 950 (1997).
18. Yu. N. Efremov, *Sites of Star Formation in Galaxies* (Nauka, Moscow, 1989) [in Russian].
19. M.-H. Rhee and T. S. Albada, Astron. Astrophys., Suppl. Ser. **115**, 407 (1996).
20. E. F. Bell and R. S. de Jong, Astrophys. J. **550**, 212 (2001).

Translated by A. Dambis

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 622–628. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 687–693.
Original Russian Text Copyright © 2004 by Denisyuk, Valiullin, Argyle, Gaĭsina.

Behavior of the Broad Hα Component in the Seyfert Galaxy NGC 4151 in 1976–2003

É. K. Denisyuk[1], R. R. Valiullin[1], R. W. Argyle[2], and V. N. Gaĭsina[1]

[1] *Fesenkov Astrophysical Institute, Almaty, Kazakhstan*

[2] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*

Received June 10, 2003; in final form, November 10, 2003

Abstract—Variations in the flux and profile of the broad Hα component in the Seyfert galaxy NGC 4151 are analyzed based on spectral observations acquired from 1976 to spring 2003. The procedure used to distinguish components is described. There is a strong correlation between the flux of the broad Hα component and the *U* flux of the galactic nucleus. There was an appreciable increase in the intensity of the broad Hα in 1990–2000, with the flux reaching a maximum in 1996. The intensity and shape of an emission feature that appeared in the red wing of the broad Hα component at the end of 2002 are estimated.
© *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

There has recently been considerable interest in studies of the bright, rapidly variable Seyfert galaxy NGC 4151, including the organization of multi-wavelength campaigns, when the object was observed over a comparatively short time but with a wide variety of telescopes and instruments. Studies of the variability of NGC 4151 on longer time scales are also of interest. A flare in the nucleus of the galaxy lasting 10–12 yr has been observed and monitored. Near the flare maximum in 1996, the *U* continuum emission was a factor of ten stronger than its minimum value. The behavior of the flare intensity as a function of time is fairly symmetrical and roughly Gaussian in shape.

Studies of this galaxy using the Hubble Space Telescope [1, 2] revealed the presence of roughly 20 individual clouds within 1″–2″ of the nucleus; these clouds have various radial velocities and radial-velocity dispersions and emit a large number of forbidden and permitted emission lines. The sizes of the clouds are comparable to or a factor of two to three larger than the resolution of the HST (0.1″). Ground-based spectral observations detect the sum of the emission lines of all the individual clouds whose light enters the slit or diaphragm of the spectrograph. The most detailed studies of the profiles of the narrow lines of NGC 4151 and a number of other Seyfert galaxies have been carried out by Veilleux [3, 4].

The first spectrograms of NGC 4151 obtained at the Fesenkov Astrophysical Institute of the Academy of Sciences of the Kazakh Republic were acquired in 1976. We present here an analysis of variations in the Hα broad component occurring over more than 25 years.

2. OBSERVATIONS

All the observations were obtained on the 70-cm telescope of the Observatory of the Fesenkov Astrophysical Institute in Almaty, which has a focal length of 11.2 m. The diffraction spectrograph described in [5] was used. Until 2000, the main light detector at the spectrograph output was a three-cascade UM-92 electrooptical device, and the light was recorded on either A-600 photographic film or another film with similar characteristics. Beginning in 2000, the light detector was an ST-7 CCD array, which was replaced by an ST-8 CCD at the end of 2002. The observations were carried out at green and red wavelengths, with the centers of the corresponding spectral intervals being near λ4900 and λ6700 Å. Depending on the diffraction grating and objective used, the spectrograms had spectral resolutions of 2.5–5.0 Å. The entrance-slit width was 4″–6″. The equipment used is descibed in detail in [5]. As a rule, to increase the accuracy of the results, the spectrum of the galaxy was taken from two to six times in a row with exposures of 2–20 min, depending on the dispersion.

3. REDUCTION OF THE SPECTROGRAMS

A large fraction of the spectrograms and negative spectrograms with calibration spectra were measured using the AMD-1 microdensitometer of the Fesenkov Astrophysical Institute. Roughly 20% of the

1063-7729/04/4808-0622$26.00 © 2004 MAIK "Nauka/Interperiodica"

negatives had optical densities that were too high for measurements with this microdensitometer, and, therefore, these spectrograms were processed on the PDS 1010A microdensitometer of the Royal Greenwich Observatory in England. We used our own software to correct for errors in the field of the electrooptical device and for the S-like distortion that is characteristic of amplifiers of this type. A night-sky spectrum was subtracted from the total spectrum, and the spectrum of the target was reduced to a strictly constant dispersion along the entire wavelength range.

The following stage in the reduction is distinguishing individual components in the complex profile formed by the Hα and [NII] narrow lines superposed on the broad Hα line. We consider this to have been done most successfully by Veilleux [3, 4], whose dates were obtained on a 3-m telescope using an echelle spectrograph and an 800×800 CCD array with high spectral resolution and high spectrophotometric accuracy. He used a $2.5'' \times 2.5''$ square diaphragm in place of the entrance slit of the spectrograph. The profiles of a number of narrow lines [3] in NGC 4151 and their relative intensities [4] are presented separately. The profiles of the [NII] and [SII] lines are very similar to those of other narrow emission lines presented in [3, 4], providing evidence that all of these lines formed under the same conditions and also testifying to the trustworthiness of the deblending performed by Veilleux.

Table 1 of [4] shows that the ratio of the flux in the narrow Hα line to the sum of the fluxes in the [NII] 6584, 6548 Å lines was $H(1) = 1.05$, while the ratio of the sum of the fluxes in the narrow Hα line and the two [NII] lines to the sum of the fluxes in the [SII] 6717 and 6731 Å lines was $W(1) = 3.62$.

Because our observations were carried with a wider entrance slit than those of Veilleux, our spectra include a contribution from regions of the galaxy located further from the nucleus. We used the results of slit spectral observations of NGC 4151 obtained using the Hubble Space Telescope with a slit width of $0.1''$ [2] to bring our data into consistency with the observations of Veilleux (in other words, to correct for the contribution of these outer regions). We used the data of Tables 2A–2D of [2] to calculate the coefficients H and W outside a central region $2.5'' \times 2.5''$ in size for 14 regions with sizes from $0.1'' \times 0.35''$ to $0.1'' \times 0.6''$, and derived the mean weighted values $H(2) = 0.866$ and $W(2) = 3.75$. The absolute fluxes in the [OIII] 5007 Å line for the corresponding regions in the galaxy were used as weighting factors. We estimated the total weight of the outer regions by comparing the fluxes measured in a $2.5'' \times 2.5''$ diaphragm (803×10^{-14} erg s^{-1} cm^{-2} [4]) and in a circular diaphragm with a diameter of $8.1''$ ($(1.20 + 0.06) \times 10^{-11}$ erg s^{-1} cm^{-2} [6]). We assumed that the outer regions of the galaxy made approximately the same contribution to our observations with a long

Characteristics of spectrograms used to construct Fig. 1

1	2	3	4	5	6
1	29 Mar 76	1	2/2	7/7	11.09
2	25 May 77	1	2/1	9/5	12.58
3	05 Jul 81	1	1/1	5/5	8.91
4	07 Feb 86	3	3/2	12.5/7.5	7.58
5	27 Feb 87	1	1/1	4/4	3.54
6	16 Mar 88	5	6/6	40/40	5.31
7	06 Apr 89	16	58/1	389.5/7	5.70
8	14 Feb 90	6	27/13	148/68	7.21
9	14 Jan 91	4	20/12	121/93	12.87
10	19 Apr 91	4	49/18	245/98	15.47
11	04 Feb 92	3	18/2	57/12	13.72
12	08 May 92	2	14/6	52/26	17.30
13	21 Feb 93	3	22/6	135/23	18.50
14	19 Jun 94	17	43/11	238/46	25.75
15	04 Mar 95	5	31/11	207/84	24.05
16	04 Apr 95	7	37/24	245/135	25.62
17	20 May 95	4	29/9	157/92	35.22
18	22 Apr 96	3	9/3	84/35	27.42
19	13 Feb 97	2	9/2	63/18	25.37
20	30 Dec 97	1	3/2	19/14	14.05
21	24 Feb 98	2	7/4	48/22	13.50
22	19 Dec 98	2	15/8	64/26	12.98
23	24 Jan 99	4	31/10	102/40	11.65
24	25 Mar 99	6	43/21	310/115	13.71
25	04 May 99	4	33/11	280/62	16.42
26	03 Jan 00	5	17/9	570/230	16.19
27	21 Feb 01	9	10/6	300/180	8.26
28	01 Dec 02	1	1/1	10/10	8.02
29	14 Feb 03	14	43/33	1290/990	7.41

Note: The columns present (1) the interval number, (2) the mean epoch for the given group of spectra, (3) the number of observing nights for the given group, (4) the number of obtained/reduced spectrograms, (5) the total exposure time for the obtained/reduced spectrograms in minutes, and (6) the integrated intensity of the Hα broad component in units of 10^{-13} erg s^{-1} cm^{-2}.

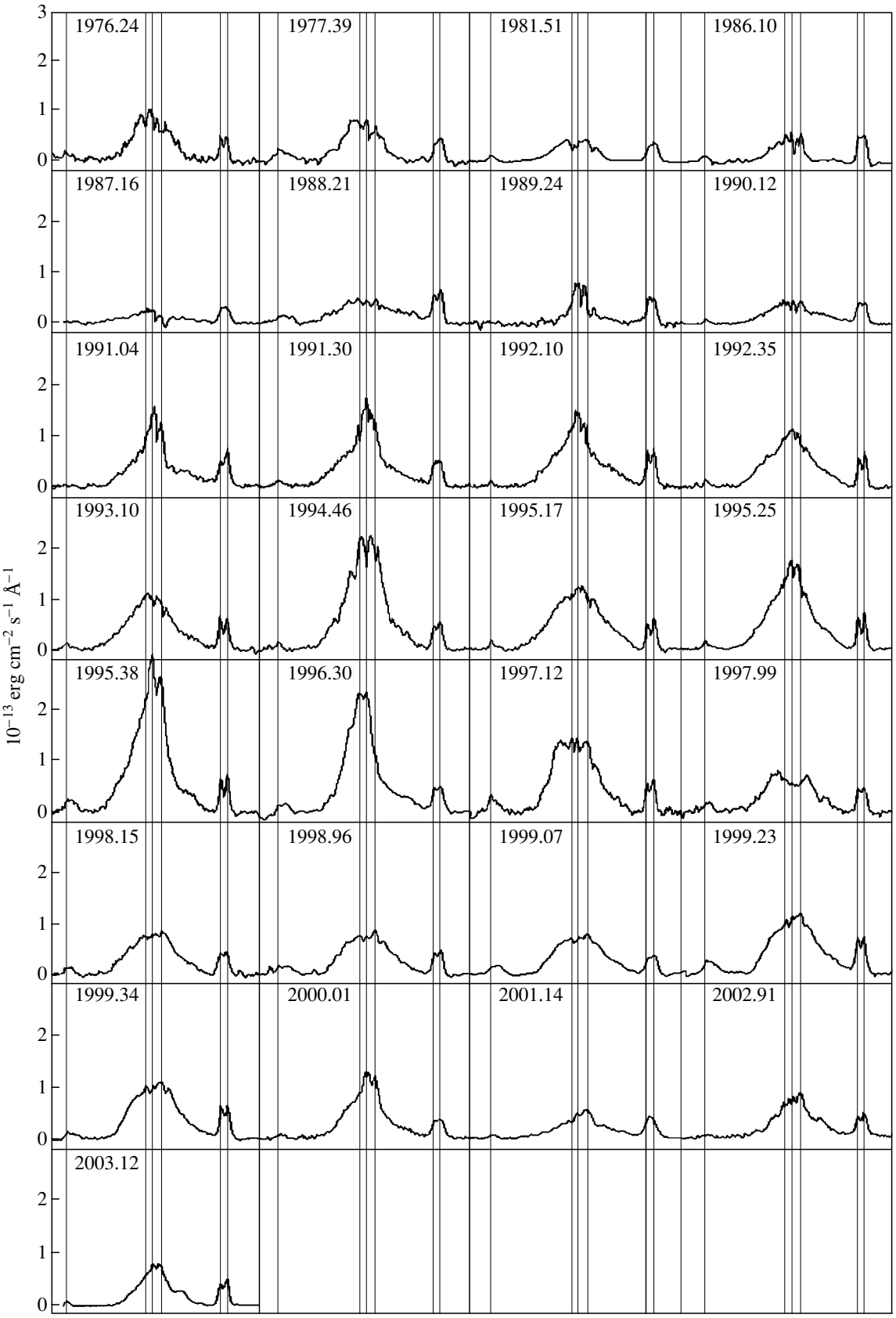

Fig. 1. Spectrum of NGC 4151 after subtracting the narrow Hα and [NII] lines. The spectral interval in each panel is 6320–6800 Å. See text for details.

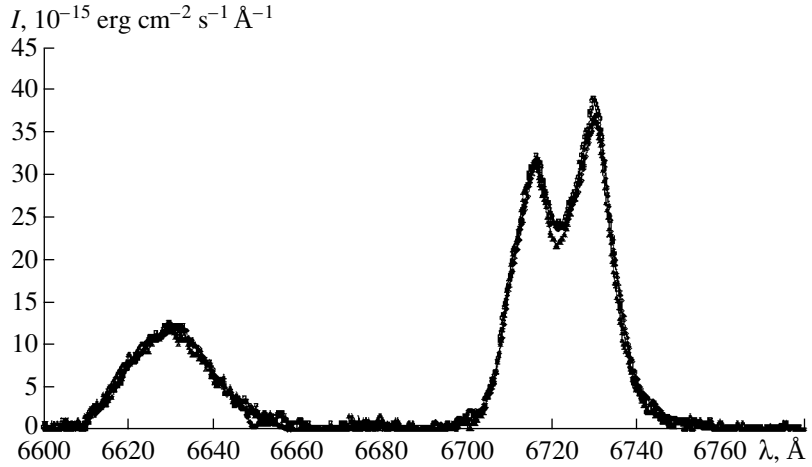

Fig. 2. New feature in the spectrum of NGC 4151 after subtraction of the continuum and the red wing of the broad Hα component. The [SII] line has been left for comparison.

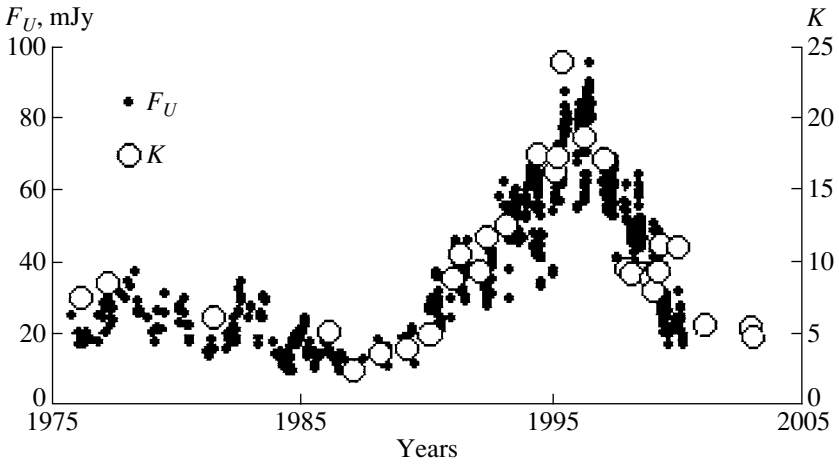

Fig. 3. Comparison of time variations of the parameter K (the ratio of the flux of the broad Hα component and the total flux in the [SII] lines) and of the flux of the nucleus of NGC 4151 in the U filter. The arrows indicate the direction of increasing time.

slit with width $4''$ as they did in the observations with a circular diaphragm with a diameter of $8''$. This assumption is justified by estimates of the [OIII] flux (114×10^{-13} erg s^{-1} cm^{-2}) obtained using a long slit with width $2.3''$ [7]. Thus, the [OIII] 5007 Å flux in outer regions is 387×10^{-14} erg s^{-1} cm^{-2}. Using this value as a weight for the outer regions, we obtained the final values $H = 0.99$ and $W = 3.66$. Estimates of the accuracies of the various data used to derive these coefficients are far from always published; we believe that the errors are no more than 5%.

We used the following method to distinguish blended lines in the broad Hα profile. We first measured the integrated flux of the sulfur doublet without its base in relative units, using a polynomial approximation for the wing of the broad Hα component. Further, an artificial profile was constructed from the three blended lines—[NII] 6548 Å, Hα, and [NII] 6584 Å—using the coefficients presented above. The instrumental profiles for the spectrograms obtained with various spectral resolutions were determined from the corresponding profile of the [OI] λ6300 Å line. The scale for the three-component artificial profile was always chosen so that the total flux of the components was equal to the flux of the [SII] line multiplied by the factor $W = 3.66$. When we used the data of [4, 5] applied to the size of our entrance slit and our calibration for 2003 obtained using standard stars, the derived flux for the three-component profile was 5.4×10^{-12} erg s^{-1} cm^{-2}. Data for the absolute flux in the [SII] 6717 and 6731 Å lines were obtained in January to March 2003; observations of the galaxy and standard stars were carried out over nine nights with good seeing. The resulting absolute flux for the

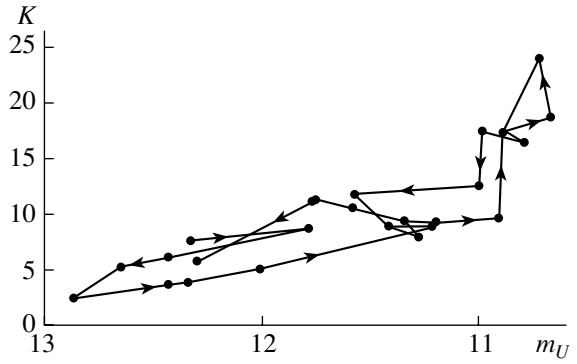

Fig. 4. Relationship between K (the ratio of the flux of the broad Hα component and the total flux in the [SII] lines) and the flux of the nucleus of NGC 4151 in the U filter.

sum of the [SII] 6717 and 6731 Å lines is $(1.48 \pm 0.06) \times 10^{-12}$ erg s^{-1} cm^{-2}.

The constructed three-component profile was subtracted from the total observed profile (Hα + [NII]) to obtain the profile only of the broad Hα component.

4. RESULTS

We obtained 584 spectrograms of NGC 4151 at red wavelengths from 1976 through 2003. The observations were carried out irregularly in time due to both varying observing conditions and the seasonal visibility of the object. In addition, the spectral variability of Seyfert galaxies on short time scales was not firmly established in the 1970s, so that it was not thought to be necessary to conduct regular observations. Only two spectrograms of NGC 4151 were obtained in each of 1976 and 1977, and one, in 1987. After analyzing the time distribution of our data, we decided to present our results in the form of average values for 28 individual time intervals. A list of spectrograms included in this study is given in the table, which presents (1) the interval number, (2) mean epoch for the given group of spectra, (3) number of observing nights for the given group, (4) number of obtained/reduced spectrograms, (5) total exposure time for the obtained/reduced spectrograms in minutes, and (6) integrated intensity of the Hα broad component in units of 10^{-13} erg s^{-1} cm^{-2}. The rms errors for the early observations are rather high ($\leq 40\%$), while the accuracy of the data obtained with the CCD array is appreciably better (5–7%).

Figure 1 presents the mean profile of the broad component for the epochs given in the table. The X axis plots a linear wavelength scale from 6320–6800 Å in the system of the galaxy, and the Y axis, the spectral flux density in absolute units. The thin vertical lines indicate the positions of the [OI] 6364 Å, [NII] 6548, 6584 Å, Hα 6563 Å, and [SII] 6717, 6731 Å emission lines.

At the beginning of the observing season for NGC 4151 in December 2002, a new emission feature appeared in the red wing of the broad Hα component. This feature is clearly visible on all subsequent spectrograms taken up until the end of April 2003, and its profile, intensity, and radial velocity remained nearly constant within the measurement errors over these four months. Figure 2 presents a superposition of spectrograms of this feature obtained in January, the end of March, and the middle of April 2003 (eight 30-min exposures). The red wing of the Hα broad component was approximated using a seventh- to ninth-order polynomial and subtracted. The [SII] 6717, 6731 Å lines have been left in Fig. 2 for comparison. The radial velocity of the feature relative to the central peak of the Hα line was $+3090 \pm 10$ km/s, and its integrated intensity was $(4.9 \pm 0.2) \times 10^{-13}$ erg s^{-1} cm^{-2} with a velocity dispersion of 425 ± 10 km/s. There is no doubt that this feature is a hydrogen emission line, since it is also visible near the Hβ line with the same relative radial velocity.

The appearance of this feature may be analogous to the events described in [8, 9], but the object radiating these hydrogen lines must be appreciably further from the center of the galaxy.

5. DISCUSSION

Figure 3 presents the values of K (the ratio of the flux of the broad Hα component from the table and the total flux in the [SII] lines) and of the flux of the nucleus of NGC 4151 in the U filter (in mJy) from [10] as functions of time. A good correlation between these two parameters is clearly visible. For example, in 1987, when the flux in the U filter was minimum, K was $\sim$3, while, in 1995–1996, when the U flux was close to its maximum (about 85 mJy [10]), the value of K reached 25. Since K had decreased appreciably in March 2001 (to $\sim$4), we expect a corresponding drop in the U flux to 5–10 mJy in the spring of 2003.

Figure 4 plots the data for NGC 4151 in the K – nuclear U flux plane. Points that are neighboring in time are connected by line segments with arrows indicating the direction of increasing time. Both figures display an event that occurred between 1990 and 2000 and has been noted in a number of studies (see, for example, [10]): a substantial increase in the intensity of ionizing radiation of the central nucleus and, as a result, a nearly simultaneous increase in the brightness of the broad component. We believe that this event corresponds to the partial or complete disruption of an object (gas cloud or star) in the immediate

vicinity of the central nucleus, which was manifest as an increase in the luminosity of the nucleus in the region of high-frequency ionizing radiation. This object either could be completely destroyed or could pass by the nucleus along an approximately parabolic orbit, with some fraction of its mass being reprocessed into short-wavelength radiation. Although this process may be quite short-lived in the coordinate system of the object, general relativistic time dilation near the Schwartzschild radius could give rise to a prolonged enhancement of the short-wavelength brightness of the nucleus lasting, for example, ten years.

With time, the large enhancement of the nuclear flux of ionizing radiation should also lead to appreciable variability of the intensities and intensity ratios of emission lines radiated in regions that are more distant from the nucleus. In this way, the ten-year flare of the nuclear flux can be used as a tool to study the spatial distribution of clouds in the circumnuclear region and the physical properties of the material in these clouds. The distance to NGC 4151 is 13 Mpc for a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1}. At this distance, a cloud with a diameter of $0.2''$ should have a physical size of about 13 pc. If the radiation front propagates at the speed of light, the thickness of the ionization layer in space will be approximately 3 pc, so that the maximum of the radiation front crosses an average cloud over about four years, acting on roughly one-quarter of the cloud volume at any given time and substantially changing the conditions for the formation of emission lines within this volume. However, the variations in the spectrum extend in time over tens of years for an observer on Earth due to the difference of the distances of the near and far edges of the clouds.

It is obvious that the variations of the emission lines in distant clouds should be observed with some delay relative to the variations of the short-wavelength radiation of the nucleus. This delay is equal to the difference between the core–cloud–observer and core–observer paths, converting this difference from distance into time using the speed of light. If the time delay for the reaction of some cloud Δt and the cloud's distance from the nucleus in the plane of the sky L are known and the speed of propagation of the front is taken to be C, the obvious formula

$$h = \pm C\frac{L^2 - \Delta t^2}{2\Delta t}$$

can be used to find the heights of the cloud above $(-)$ or below $(+)$ the plane of the sky; i.e., we can reconstruct its full spatial position relative to the nucleus. This approach does not work only in the case of clouds that are located directly in the line of sight from the observer to the nucleus.

Similar flares could also have occurred earlier, and the result of their action on circumnuclear clouds detected at some later time. It is possible that this is occurring now as well, since the delay of such manifestations comprises tens or hundreds of years. It is possible that the new emission feature described above is an example of this. Comparison of profiles of the most intense narrow emission line in NGC 4151 in the visible, [OIII] 5007 Å , observed in 1976 [11] and 1987 [3] shows appreciable variations, associated with variability of the intensities of the emission lines in individual clouds, as well as possible variability of their radial velocities. Therefore, it is important to conduct regular observations such as those of [1, 2]. Note that monitoring of the behavior of the individual clouds giving rise to the narrow lines can be carried out using ground-based telescopes with entrance-slit widths of several arcseconds. However, a detailed discussion of this topic extends beyond the framework of the current study.

6. CONCLUSIONS

The results presented above were obtained using a single telescope and spectrograph over a time interval spanning more than a quarter of a century. On average, the quality of the spectrograms is appreciably lower than the quality that can be obtained using modern CCD arrays and substantially larger telescopes. However, the value of these early observations is that they are unique measurements that cannot be repeated and so can play an important role in variability studies. This is particularly true with regard to slow variations of the hydrogen emission radiated in the immediate vicinity of the nucleus. If the ten-year increase in the power of the ionizating radiation of the nucleus referred to above is associated with the passage and partial disruption of an object moving along a closed, fairly elongated elliptical orbit, this process may be repeated after some time, analogous to the passage of periodic comets in the solar system.

ACKNOWLEDGMENTS

This work was partially supported by an INTAS grant (96-0328). The authors thank É. Vil'kovisskiĭ for help with the organization of the reduction of some of the material using the English microdensitometer, and also the authors of [10] V. Doroshenko and V. Lyutyĭ for making available additional photometric data for NGC 4151 obtained in 1999–2000.

REFERENCES

1. M. E. Kaiser, L. D. Bradley, II, J. B. Hutchings, *et al.*, Astrophys. J. **528**, 260 (2000).
2. C. H. Nelson, D. Weistrop, J. B. Hutchings, *et al.*, Astrophys. J. **531**, 257 (2000).
3. S. Veilleux, Astrophys. J., Suppl. Ser. **75**, 357 (1991).
4. S. Veilleux, Astrophys. J. **369**, 331 (1991).
5. E. K. Denisyuk, Astron. Astrophys. Trans. **22**, 175 (2003).
6. R. R. J. Antonucci and R. D. Cohen, Astrophys. J. **271**, 564 (1983).
7. Yu. F. Malkov, V. I. Pronik, and S. G. Sergeev, Astron. Astrophys. **324**, 904 (1997).
8. E. K. Denisyuk, Astron. Astrophys. Trans. **10**, 279 (1996).
9. B. I. Demchenko and E. Denisyuk, Astrofizika **44**, 667 (2001).
10. V. M. Lyutyi and V. T. Doroshenko, Pis'ma Astron. Zh. **25**, 403 (1999) [Astron. Lett. **25**, 341 (1999)].
11. D. Pelat and D. Alloin, Astron. Astrophys. **105**, 335 (1982).

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 629–637. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 694–702.
Original Russian Text Copyright © 2004 by Shchekinov, Zinchenko.

The Origin of Giant Molecular Clouds

Yu. A. Shchekinov[1] and I. I. Zinchenko[2,3]

[1] *Research Department of Physics, Rostov State University, and Isaac Newton Institute of Chile, Rostov-on-Don Branch, pr. Stachki 194, Rostov-on-Don, 344090 Russia*

[2] *Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhni Novgorod, 603600 Russia*

[3] *Helsinki University Observatory, Tähtitorninmäki, P.O. Box 14, 00014, University of Helsinki, Finland*

Received May 26, 2003; in final form, January 9, 2004

Abstract—The parameters of molecular clouds formed via the Parker instability with dominant radiative losses are estimated. In this scenario, the cloud parameters (such as their mean densities and masses) should depend on galactocentric radius. This dependence is determined mainly by radial variations of the gas metallicity and the flux of heating radiation in the Galaxy. Due to the development of the interchange mode of the Parker instability, the angular momentum of the clouds will not necessarily be parallel to the galactic rotation axis. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In spite of the considerable progress in analyses of the physical conditions in giant molecular clouds (GMCs), their origin remains unclear (see review [1]). Three scenarios are usually considered: (a) collisional mergers of smaller clouds [2–4], (b) the compression of material by supernova shocks or galactic spiral shocks [5, 6], and (c) the compression of material during the development of the Parker instability with subsequent gravitational instability [7–10]. Though there are no direct observations that confirm or reject particular scenarios, there has recently appeared indirect evidence that the last scenario is the most likely. This includes observational evidence for short lifetimes of GMCs, $2 \times 10^7 < \tau < 10^8$ years [1]; observations of molecular clouds at high Galactic latitudes ($b \sim 5°$) [11] with masses $\sim 3 \times 10^4 M_\odot$ [12]; and observations of molecular gas in galaxies with poorly developed spiral structure [13], where, according to standard concepts, cloud–cloud collisions should not be efficient [3]. In addition, the dispersion of molecular-cloud velocities does not depend on their masses, which is difficult to explain in a scenario with collisional mergers of small clouds.

If we assume that molecular clouds are formed from diffuse gas, the accumulation of the required mass, $M \sim 10^5 M_\odot$, requires that gas be collected from scales of $\sim$150 pc. Since the Jeans length in diffuse interstellar gas is considerably larger than this, the most natural mechanism under these conditions is the Parker instability. Another circumstance testifying to an important role for the Parker instability is the fact that the specific angular momentum of the clouds ($\sim$60 pc km s^{-1}) is considerably lower than the specific angular momentum of the differential rotation of the interstellar gas ($\sim$250 pc km s^{-1}). During the development of the Parker instability, the gas accumulates in "valleys" aligned with the magnetic field, i.e., almost parallel to the local linear velocity of the rotational motion; consequently, its specific momentum should be less than the galactic value.

We note also an interesting fact that may represent an additional indirect argument in favor of the Parker instability (if it is not a coincidence). The total kinetic energy of the proper motions of a molecular cloud is $K \simeq 5.6 \times 10^{53}$ erg [14], while the gravitational energy of the same mass of gas ($M \sim 10^9\ M_\odot$) in a diffuse state occupying a height scale $\sim$100 pc in a gravitational field with acceleration $g \sim 10^{-9}$ cm/s^2 is $G \sim 6 \times 10^{53}$ erg. It is also important that convective motions during the development of the Parker instability (more specifically, the development of its interchange mode perpendicular to the magnetic field) can result in the appearance of condensations whose total momentum is directed at some (generally speaking, arbitrary) angle to the plane of the Galaxy; therefore, in this case, the momentum of the forming molecular clouds can have an arbitrary orientation. In this connection, we note that the observed orientations of the angular rotational velocities display a two-peaked distribution with maxima near position angles

1063-7729/04/4808-0629$26.00 © 2004 MAIK "Nauka/Interperiodica"

$\Theta \simeq 90°$ and $\simeq 270°$, approximately in the direction of the north and south Galactic poles [15]. However, the dispersion of these distributions is very large: the peak near $\Theta \simeq 90°$ occupies an angle interval $\Delta\Theta \simeq 45°-120°$, and the peak near $\Theta \simeq 270°$, an angle interval $\Delta\Theta \simeq 240°-330°$; in addition, an appreciable fraction of the clouds have their rotation oriented along the Galactic plane.

In this paper, we present additional arguments providing indirect evidence that the Parker instability is responsible for the formation of GMCs. In the next section, we briefly describe the Parker-instability mechanism and its dynamical features. Section 3 discusses the effect of radiative losses on the development of the Parker instability. In Section 4, we consider the physical characteristics of clouds born via this instability and the consequences of the proposed scenario, as well as probable observational confirmations. Section 5 presents a general discussion of our results and our main conclusions.

2. DESCRIPTION OF THE PARKER INSTABILITY

The Parker instability [16] develops in a magnetized, vertically stratified plasma in an external gravitational field with a horizontally directed magnetic field. In the simplest case of an isothermal gas in a uniform gravitational field, $\mathbf{g} = (0, 0, g)$ and $g =$ const, a constant ratio of the gas to magnetic pressures, $\beta = 8\pi P/B^2 =$ const, and a constant ratio of the cosmic-ray and gas pressures, $\alpha = P_{CR}/P =$ const, the distributions of the gas density and magnetic field are exponential:

$$\rho(z) = \rho_0 \exp(-z/H), \qquad (1)$$
$$B(z) = B_0 \exp(-z/2H),$$

where $H = (1 + \beta^{-1} + \alpha)c_0^2/g$ is the gas scale height and c_0 is the isothermal sound speed in the $z = 0$ plane. If such a distribution is derived from an equilibrium state by a perturbation whose wave vector is aligned with the magnetic field and which has a wavelength greater than some critical value, the gravititational force of the gas flowing along the lines of force is so large that the magnetic pressure cannot restore the former equilibrium—an instability develops. The corresponding critical wavelength is [16–18]

$$\lambda_c = 4\pi H \qquad (2)$$
$$\times \left[\frac{\beta\gamma}{2[1+\beta+\beta(\alpha-\gamma)][1+\beta(1+\alpha)]-\beta\gamma}\right]^{1/2},$$

where γ is the gas adiabatic index. With subsequent increase of the wavelength, the instability increment reaches its maximum and then decreases. The wavelength corresponding to the maximum increment at $\alpha \sim \beta \sim \gamma \sim 1$ is $\lambda_m \simeq 1.8\lambda_c$ [16], so that $\lambda_c \sim 4H$. For a scale height $H \sim 200-300$ pc, the critical wavelength is $\sim 0.8-1.2$ kpc, and the wavelength corresponding to the maximum increment is $\lambda_m \sim 1.4-2.2$ kpc. This size is close to that of the large-scale cells bordered by dust lanes observed in some spiral galaxies, such as NGC 253, NGC 1232, and M 83.

Note that λ_c and λ_m are more sensitive to variations of the cosmic-ray pressure than to the magnetic-field pressure: for example, in a close-to-isothermal mode (i.e., $\gamma \gtrsim 1$) and for $\beta \simeq 1$, the ratios λ_c/H and λ_m/H vary as α^{-1}, whereas they vary as $\beta^{1/2}$ for $\alpha \simeq 1$. Essentially, this could lead to corresponding variations in the sizes and masses of structures forming due to the Parker instability. However, if there is equipartition between the various kinds of energy in the interstellar medium (ISM), as can be considered to be the case for our Galaxy, the values of α and β do not deviate significantly from unity (see Section 4).

One of the characteristics of the Parker instability that prevents us from giving absolute preference to this as the mechanism for the formation of GMCs is that the density enhancement in valleys (sites of gas accumulation) is only a factor of two over the density of the unperturbed gas density in the plane, whereas the required density increase should be orders of magnitude (see the discussion in [9, 19], as well as the numerical model [20]). However, all numerical calculations of the nonlinear dynamics of the Parker instability carried out thus far have been restricted to a gas equation of state with a fixed adiabatic index (as a rule, $1.05 \leq \gamma \leq 1.6$); in this case, since the mass of the gas column above a valley increases by only a modest factor (approximately a factor of three—the ratio of the characteristic horizontal size of a perturbation to the characteristic size of the valley), the density increase at the bottom of a valley cannot exceed this value. At the same time, it is fairly obvious that radiative losses of the gas on a timescale appreciably shorter than the timescale for the development of the Parker instability (see below) will result in a loss of the thermal energy of the gas and, therefore, to a decrease in pressure; in turn, this will lead to a compression of the gas as it sinks to the valley bottom that is even greater than the compression in an isothermal process. In this paper, we estimate the molecular-cloud parameters expected for the development of the Parker instability in a radiatively cooled gas. A numerical calculation of the Parker instability proper in such a medium will be presented in a future paper.

3. THERMAL REGIME OF THE GAS

One property of the thermodynamics of the interstellar gas is its ability to undergo phase transitions

under the influence of thermal instability. Diffuse HI gas with a temperature $T \sim 10^4$ K makes a transition to a cooler phase with a temperature $T \sim 10^2$ K if the pressure in the gas exceeds some critical value, determined by its metallicity Z and the intensity of external heating radiation Γ (by soft X rays, cosmic rays, the ultraviolet radiation of stars). The gas temperature at the critical point depends very weakly on metallicity and the sources of heating (logarithmically), whereas the density and pressure depend on Γ and Z in a power-law fashion.

As a rule, the characteristic thermal timescales (for radiative cooling and heating by external sources, $\tau_c \sim 10^5$–10^6 years) are considerably shorter than the timescale for the development of the Parker instability, which is comparable to the free-fall timescale, $\tau_p \sim 10^8$ years. This allows us to assume that the gas is in a quasi-equilibrium state, in which its thermal properties are, on average, described by an equilibrium between radiative losses and external radiation sources; for temperatures $T < 2 \times 10^4$ K, this equilibrium is characterized by a two-phase equation of state [21, 22]. For the ionization and heating of gas by cosmic rays and soft X rays, the gas density and temperature at the critical point are [23]

$$n_{cr} \simeq 1.5 \times 10^{14} \frac{\xi}{Z}\ \mathrm{cm}^{-3}, \quad T_{cr} \simeq 5900\ \mathrm{K}, \qquad (3)$$

where ξ is the primary ionization rate, which is $\xi_{cr} \simeq 3 \times 10^{-17}$ s^{-1} for cosmic rays and $\xi_{xr} \simeq 10^{-15}$ s^{-1} for soft X rays [24]; the value of Z is given in units of the solar metallicity. It is currently accepted that the main mechanism for heating of the interstellar gas is the photoemission of electrons from the surfaces of grains [24]. The gas parameters at the critical point are described by the approximation formulas

$$n_{cr} = 0.46 \frac{G_0^{0.6}}{Z^{0.55}}\ \mathrm{cm}^{-1}, \quad T_{cr} \simeq 6300\ \mathrm{K}, \qquad (4)$$

$$P_{cr} \simeq 2520 \frac{G_0^{0.6}}{Z^{0.55}}\ \mathrm{cm}^{-3}\ \mathrm{K},$$

where G_0 is the flux of far-ultraviolet radiation in units of the Habing value, 1.7×10^{-3} erg cm^{-2} s^{-1} [25]. It is assumed in [26] that the flux G_0 changes in the vertical (z) direction in the same way as in the galaxy M81, namely,

$$G_0 = 1.1(1 + 0.014z^2)^{-1}, \qquad (5)$$

where z is the distance from the Galactic plane in kiloparsecs. Thus, the formula for the critical pressure becomes

$$P_{cr} = 2670 g_0^{0.6} (1 + 0.014z^2)^{-0.6} Z^{-0.55}\ \mathrm{cm}^{-3}\ \mathrm{K}, \qquad (6)$$

where we have retained the parameter g_0 (equal to unity in the solar neighborhood), which can characterize radial variations of the ultraviolet flux.

The actual vertical pressure profile is set by the condition of hydrostatic equilibrium:

$$\frac{1}{\rho}\frac{dP}{dz} = -\frac{d}{dz}\phi(R, z), \qquad (7)$$

where the potential $\phi(R, z)$ is determined by the distribution of the gravitating mass in the Galaxy. Neglecting the contribution of the gas component in ϕ and assuming the gas to be isothermal and to have a constant sound speed in the diffuse component, $c_s = 0.12T^{1/2}$ km/s, Wolfire *et al.* [26] obtained the following pressure distribution in the hot (X-ray) component of the halo:

$$P(R, z) \qquad (8)$$
$$= 2250 T_6^{1/2} p(R)(1 + 0.05z^2)^{-1.35/T_6}\ \mathrm{cm}^{-3}\ \mathrm{K},$$

where $p(R)$ describes the radial dependence of the pressure, $p(R) = 1$ in the solar neighborhood, and $T_6 = T/10^6$ K. It is easy to see that the resulting dependence $P(z)$ is quite sensitive to the temperature of the hot (X-ray) component of the halo. In this case, $P_{cr} > P(R, z)$ everywhere when the temperature of the X-ray gas is $T_6 \leq 2$ [24], so that the warm gas can be transformed into cool clouds only under the action of sufficiently strong perturbations, such as a spiral shock or supernova shocks. The total pressure, including the partial pressures of all the components of the halo gas (neutral and ionized hydrogen, as well as the cloud component), can far exceed the pressure of the X-ray component. Kalberla and Kerp [27] presented the profile of the total pressure in the interstellar gas, which can be approximated in the solar neighborhood as

$$P_\odot(z) = 1.8 \times 10^4 e^{-0.85z + 0.065z^2}; \qquad (9)$$

this could mean that, at heights smaller than some value, the total pressure exceeds the critical pressure, and thermal instabilities can develop even from weak perturbations. To find the radial dependence of the ratio of the critical and actual gas pressures, P_c and P, we will take the radial scale of the gas distribution to be 15 kpc, as in [27], i.e.,

$$P(R, z) = 1.8 \times 10^4 e^{-0.85z + 0.065z^2 - 0.067(R - R_\odot)}; \qquad (10)$$

the radial scale of the interstellar radiation field to be 5 kpc (see the discussion in [28]), i.e.,

$$G_0(R, z) = 1.1[1 + 0.014z^2]^{-1} e^{-0.2(R - R_\odot)}; \qquad (11)$$

and the metallicity to vary with the radius as

$$Z = e^{-0.18(R - R_\odot)}. \qquad (12)$$

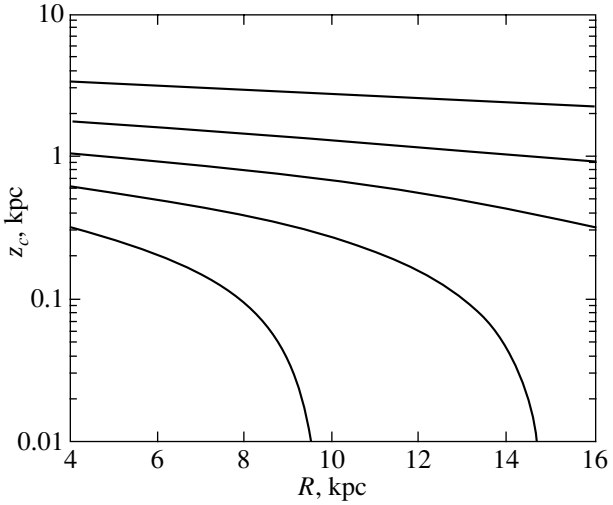

Fig. 1. Critical height above the Galactic plane. Below this height, the pressure of the surrounding diffuse gas is higher than the critical pressure at which the diffuse gas can make a transition to a cool phase due to the thermal instability. Curves are plotted for the ratios P^0_{cr}/P^0 =0.15, 0.35, 0.55, 0.75, 0.95 (from top to bottom).

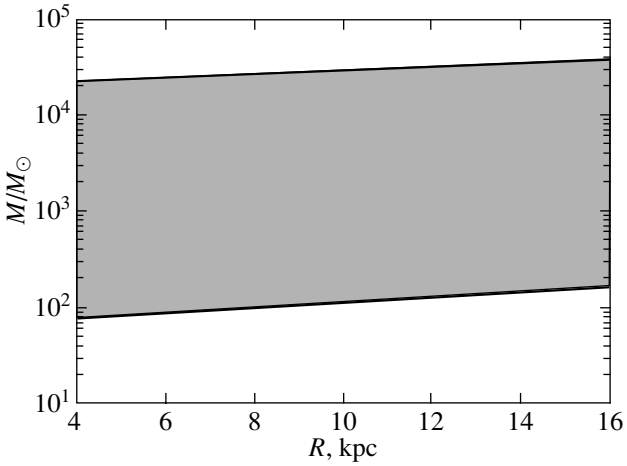

Fig. 2. Radial dependences of the upper (M^u_m) and lower (M^l_m) boundaries of possible molecular-cloud masses (top and bottom curves, respectively). The shaded area is the allowed interval of masses for clouds produced via the Parker instability. When estimating M^l_m, we assumed that the cooling rate is inversely proportional to the concentration of metals: $\Lambda \propto Z^{-1}$.

This corresponds to a decrease in the metallicity by 0.078 dex/kpc, which is close to the values for oxygen (0.07 dex/kpc), nitrogen (0.09 dex/kpc) [29], and iron (0.099 dex/kpc) [30]. Under these assumptions, the critical height z_c below which the thermal instability can develop is

$$z_c = 6.54 \qquad (13)$$
$$- \sqrt{42.7 + 15.4[\ln(P^0_{cr}/P^0) + 0.046(R - R_\odot)]},$$

where P^0_{cr} and P^0 are the critical and actual pressures of the gas at $z = 0$ and $R = R_\odot$. Figure 1 shows this dependence for a set of values of (P^0_{cr}/P^0). We can easily see that, if the actual total pressure is equal to the pressure obtained in [27], thermal instability can develop at all heights $z < 2$ kpc.

4. CLOUD MASSES AND DENSITIES

The formula

$$M_m \sim 2\rho_0(R)\lambda_m H \lambda_\perp \qquad (14)$$

can serve as a rough estimate of the mass of a cloud formed in a magnetic well under the action of instability and radiation cooling; here, $\rho_0(R)$ is the gas density in the Galactic plane, λ_m is the wavelength corresponding to the maximum increment of the Parker instability, and $\lambda_\perp$ is the characteristic length limiting the instability region perpendicular to the magnetic field and the gravitational acceleration. Generally speaking, it is necessary to substitute $\min[H, z_c]$ in place of H in (14); however, since $z_c > H$ in the interval $R \le 15$ kpc when $P^0_{cr}/P^0 < 0.7$, formula (14) gives an exact estimate of the mass. Formally, magnetohydrodynamic perturbations with a wave vector $\mathbf{k} = (0, k, 0)$, corresponding to the interchange instability, are characterized by an increase in the increment with increasing wave number; however, this increase is fairly gradual, and asymptotically (for $kH \gg 1$), the interchange-instability increment does not exceed the reciprocal free-fall timescale [17, 31]. Under these conditions, it is natural to assume that the minimum transverse dimension of the instability region is limited by the radiative length, i.e., the distance traveled by an acoustic perturbation over the cooling timescale, $\lambda_\perp \sim \lambda_c = c_s\tau_c$, where $\tau_c = kT/\Lambda(T)n$ is the scale on which all the motions are synchronized by the rapid, isobaric compression. Thus, we can write a lower limit for M_m as

$$M^l_m \sim 2\lambda_m m_h \qquad (15)$$
$$\times \int_0^H n(z)\frac{kTc_s}{\Lambda(T)n(z)}dz = 2\lambda_m c_s m_h \frac{kT}{\Lambda(T)}H.$$

For $T \sim (1\text{–}3) \times 10^4$ K (corresponding to $\Lambda(T) \sim 10^{-23}$–10^{-22} erg cm^3 s^{-1}), this yields

$$M^l_m \simeq 1.6 \times 10^4 g_9^{-2} M_\odot, \qquad (16)$$

where $g_9 = g/10^{-9}$. Adopting for the characteristic value of the gravitational acceleration g the approx-

imation

$$g_9 \simeq 12e^{-0.12(R-R_\odot)} \quad (17)$$

obtained from the model of [27], we find $M_m^l \simeq 10^2 \times e^{0.24(R-R_\odot)} M_\odot$. If we allow for the fact that the radiative loss rate is inversely proportional to the concentration of metals, $\Lambda \propto Z^{-1}$, this dependence becomes $M_m^l \simeq 10^2 e^{0.06(R-R_\odot)} M_\odot$ (Fig. 2).

The upper limit on the cloud masses is determined by the characteristic size of the coherence of the perturbations in the perpendicular direction, equal to the scale of the inhomogeneity of the magnetic field in the Galaxy, $\lambda_\perp \sim \lambda_B \sim 100$ pc. This yields for the total mass of gas that has been accumulated in magnetic wells as a result of the Parker instability

$$M_m^u \sim \frac{3.9 \times 10^6}{g_9^2} e^{-0.2(R-R_\odot)} M_\odot \quad (18)$$
$$= 2.7 \times 10^4 e^{0.04(R-R_\odot)} M_\odot,$$

with the parameters adopted above; this dependence is given by the upper curve in Fig. 2. Thus, if we assume that this gas will produce molecular clouds upon subsequent cooling, the masses of these clouds should be in the interval bounded by the curves M_m^l and M_m^u.

Observations demonstrate that the spectrum of interstellar-cloud masses is essentially continuous from $\sim 1\ M_\odot$ to $\sim 10^6\ M_\odot$ and is close to $dN(M)/dM \propto M^{-1.5}$ [32, 33]. According to the generally accepted classification, structures usually classified as GMCs have masses $M \gtrsim 10^4\ M_\odot$ [34]. Note that the dependence of GMC properties on Galactocentric radius has been little studied [34].

We should underscore the qualitative character of the $M_m^l(R)$ and $M_m^u(R)$ dependences given above. Under real conditions, the radial variations of the diffuse-gas temperature and characteristic scale of magnetic-field inhomogeneity, which determine M_m^l and M_m^u, could considerably change the radial dependences of these masses, although the approximate relations should be correct in order of magnitude. Thus, the meaning of these dependences is that they give the correct orders of magnitude for the interval of probable molecular-cloud masses, in accordance with specific physical mechanisms.

As the clouds are cooled to temperatures at which a heat balance is established, the density in the clouds becomes approximately $n_c \sim 10 n_{cr}$ [24] if the external pressure is equal to the critical pressure, $P \simeq P_{cr}$. For the parameters adopted above,

$$n_c \simeq 4.6 G_0^{0.6} Z^{-0.55} \quad (19)$$
$$\simeq 4.6 e^{-0.076(R-R_\odot)}\ \mathrm{cm}^{-3}.$$

If the external pressure exceeds the critical value, $P > P_{cr}$, the steady-state density in the clouds varies as $n_c \sim P^{1.4}$ [24]; therefore, the radial dependence of n_r will be considerably steeper than is given by (19). In Fig. 3, these dependences are shown by solid and dashed curves, respectively. It is easy to see that the interval of expected densities n_c is close to the average density observed in molecular clouds, $\langle n \rangle \sim 50\ \mathrm{cm}^{-3}$ [1]. If we continue with the hypothesis that cooling gas accumulating in magnetic wells produces molecular clouds and assume that the density n_c we have found characterizes the mean density of molecular clouds, then we expect large-scale radial variations of the molecular-cloud density in our model with a characteristic scale $R_d \lesssim 13$ kpc, as follows from (19). In this connection, we note the observations of large-scale radial variations of the intensity ratios of the CO $J = 2-1/J = 1-0$ lines [35] in the interval of Galactocentric radii $R = 4$–8 kpc, which may testify to an increase in the mean gas density in molecular clouds. It is also interesting that the densities of the dense cores of molecular clouds display a rather strong dependence on Galactocentric radius [36]; although we cannot in general indicate a unique dependence between the density of a molecular cloud as a whole and the density of its core, it is nevertheless natural to suppose that such a relationship does, on average, exist, in which case the observed large-scale variations of the molecular-cloud core densities may reflect similar variations in the cloud densities. In [35], the observed large-scale density gradient in molecular clouds was attributed to the direct action of a Galactic shock. However, the fact that the molecular-cloud lifetime corresponds to a short timescale that is close to the dynamic timescale (i.e., the time for crossing of a cloud by a sound wave; see the discussion in [37]) leads us to consider this interpretation to be unconvincing. The Galactic shock could play a role in stimulating the Parker instability.

In the above analysis, we have implicitly assumed that the partial pressures of the magnetic field and cosmic rays are comparable to the gas pressure, $\alpha \sim \beta \sim 1$. This commonly used assumption corresponds to the hypothesis of equipartition between the various kinds of energy in the interstellar medium, which has essentially been confirmed by observations, if by the gas pressure we mean the total pressure, including the turbulent pressure [38]. Since the critical wavelength of the Parker instability considerably (by an order of magnitude) exceeds the characteristic scale for turbulence in the ISM (~ 100 pc), including the turbulent pressure in the overall pressure balance is justified. At the same time, deviations from equipartition are substantial in a number of cases. A typical example is the galaxy NGC 6946, where the magnetic-field energy density exceeds the energy density of

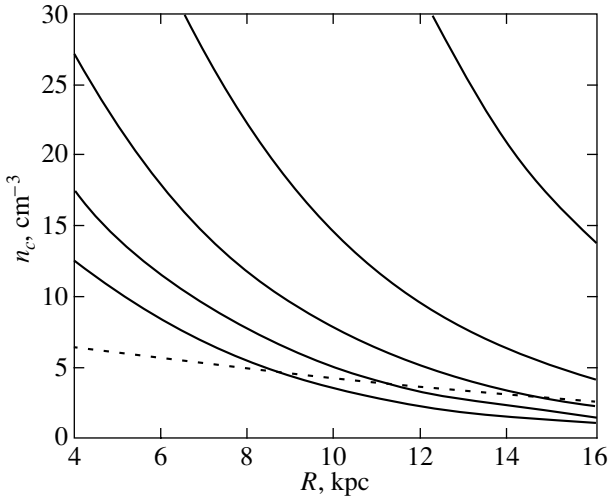

Fig. 3. Gas density in the clouds as a function of Galactocentric radius. The dotted curve shows the dependence described by (19) with $P = P_{cr}$. The solid curve shows the dependences $n_c(R)$ for the case when $P(R)$ and $P_{cr}(R)$ are characterized by radial relations similar to (10) and (6), respectively, but with $P^0_{cr}/P^0 = 0.15$, 0.35, 0.55, 0.75, 0.95 (from top to bottom).

turbulent motions by an appreciable factor outside the central 1.5 kpc [39]. From the point of view of the formation of molecular clouds via the Parker instability, one important consequence of a lack of equipartition could be the suppression of instability. Indeed, the maximum Parker-instability increment decreases when the pressures of the magnetic field and cosmic rays deviate from their equipartition values. When $\alpha, \beta^{-1} \ll 1$, the instability increment is of order $n \sim O(\alpha, \beta^{-1})$, while it decreases as $n \sim O(\alpha^{-1}, \beta)$ when $\alpha, \beta^{-1} \gg 1$ (see, e.g., the discussion in [18]).

The hypothesis of equipartition is probably valid in our Galaxy. This is supported by the analysis of the large-scale distributions of the 21-cm line emission, soft X-ray radiation, gamma-ray radiation with energies $E > 100$ MeV, and 408-MHz radio continuum in the Galaxy carried out in [27], as well as by recent data on the distribution of the magnetic field [40]. We therefore expect that the Parker instability develops with an increment that is close to the maximum increment and with wavelengths within the intervals considered above.

5. DISCUSSION AND CONCLUSIONS

We have analyzed probable variations of the parameters of molecular clouds on Galactic scales, hypothesizing that the main mechanism for the accumulation of diffuse gas with the subsequent formation of molecular clouds is the Parker instability with radiative losses. Due to the large-scale radial variations of the metallicity and diffuse ultraviolet radiation in the Galaxy, clouds born when warm diffuse gas makes a transition to a cool phase will have a higher density in central regions of the Galaxy. This is consistent with the observed increase of the CO $R_{2-1/1-0}$ line intensity ratio [35]. Moreover, such an increase in the cloud densities toward the Galactic center could explain the similar behavior of the densities of molecular-cloud cores [36] if the properties of the cores reflect to some extent the properties of the parent clouds. In the proposed model, the total number of molecular clouds in the Galaxy can be estimated to be

$$N_{mc} \sim 2\pi \frac{\tau_{mc}}{\tau_p} \int\limits_0^{15\ \mathrm{kpc}} \frac{r}{\lambda_m} \frac{dr}{\lambda_\perp}, \qquad (20)$$

where λ_m and $\lambda_\perp$ are the Parker-instability scales along and perpendicular to the magnetic field, $\tau_{mc} \sim$ 20–100 billion years is the lifetime of molecular clouds, and $\tau_p \sim 100$ billion years is the timescale for the development of the Parker instability. This estimate yields

$$N_{mc} \sim (1200\text{–}6000) \left(\frac{\lambda_m}{1\ \mathrm{kpc}}\right)^{-1} \left(\frac{\lambda_\perp}{0.1\ \mathrm{kpc}}\right)^{-1}, \qquad (21)$$

which is equal in order of magnitude to the total number of molecular clouds in the Galaxy ($\sim$2000).

One of the basic difficulties associated with the formation of molecular clouds via the Parker instability is that, as was noted above, perturbations grow faster perpendicular to the magnetic field than along the field. This must tangle adjacent flux tubes and destroy the coherence of the motion; such coherence is necessary to form a sufficiently massive, dense clump of gas (for a detailed discussion, see [41]). In numerical experiments that include a spectrum of initial perturbations perpendicular to the field extending to extremely short wavelengths, the interchange mode of the Parker instability splits the flux into a multitude of field-aligned filaments, probably turbularizing it in the process [42]. However, as was noted above, radiative losses and the associated thermal instability synchronize the perturbations at short wavelengths corresponding to the radiative scale. In this case, the combined operation of the thermal instability and interchange mode of the Parker instability could lead to convective motions of the gas fragments involved in the instability, i.e., to rotation of gas cells with their angular-velocity vectors parallel or antiparallel to the magnetic field. We expect that, during the subsequent falling of such a convective cell along the increasingly strong magnetic field into a magnetic well, its

rotational motion will become aligned with the magnetic lines (transferring to them its excess angular momentum). Thus, clouds with angular momentum antiparallel to the total Galactic angular momentum could form at the bottom of the well, as is found in the distribution of rotation directions for molecular clouds [15]. Another physical effect that can diminish the tangling of the flux tubes during the development of the interchange mode and desynchronization of their motion is the reconnection of tangling magnetic lines belonging to different flux tubes and their unification into a more coherent flow. However, this possibility requires a more careful estimation of the efficiency of reconnection in such a flow.

One of the basic paradoxes of the physics of molecular clouds that may be associated with their formation is that, on the one hand, the internal structure of the clouds can be fractal (the zone of fractality encloses a fairly large volume outside the densest central region [43–45]), while, on the other hand, the clouds should be much more homogeneous and spherical if self-gravitation is important (which should be the case if stars are born in such clouds). This seeming inconsistency can be explained by the fact that the Jeans timescale for a molecular cloud with a mean density of $\langle n\rangle \sim 50$ cm^{-3} is $t_j \sim$ 10 billion years, which is comparable to (or even exceeds) the timescale for variations in the cloud radius, $t_r \sim R/v_c$, for $R \sim 30$ pc and a velocity of infalling material $v_c \sim 10$ km/s, if the growth of the cloud is associated with material infalling during the nonlinear stages of the development of the Parker instability. This is consistent with the results of recent three-dimensional numerical modeling [46], in which molecular clouds can be treated as turbulent density fluctuations with lifetimes determined by either the dynamical timescale or the timescale for the destruction of such fluctuations by young massive stars, if these are born in the cloud interiors. In the formation of molecular clouds via the Parker instability, the Parker instability ensures an inflow of material into a restricted volume where a region of enhanced density is formed. The upper limits on the lifetimes of such regions depend on how long the material falls into the magnetic well; this timescale is from one to several hundreds of millions of years. The actual lifetime of a molecular cloud is determined by the time required for stars born in the cloud to ionize the infalling gas and eject it from the central region.

An important mechanism regulating the formation of molecular clouds is the dynamical coupling between the interstellar gas and the magnetic field. The magnetic fields observed in molecular clouds demonstrate the correlation $B \propto \rho^{2/3}$ [47], which is expected for an isotropic compression of the gas. This may provide evidence against the formation of molecular clouds via the compression of gas by supernova shocks (which should yield $B \propto \rho$) or to mergers of smaller clouds (when we might expect the annihilation of randomly oriented fields and an absence of a significant correlation between B and ρ). At the same time, in the formation of molecular clouds via the Parker instability, the process of accumulation of the gas is fairly isotropic when the gas contracts perpendicular to the field from $\lambda_\perp \sim 100$ pc and $H \sim$ 200 pc to the radius of a molecular cloud, $R \sim 20$ pc.

The formation of molecular clouds and cloud complexes is closely related to star formation; therefore, we expect that studies of the properties of young stars in galaxies can shed light on features of molecular-cloud formation.

In this connection, it is of interest that spiral galaxies demonstrate at most an extremely weak correlation between the star-formation rate and the surface or volume density of gas [48, 49, 50]. On the other hand, numerous observations demonstrate that the star-formation rate decreases sharply in the outer regions of galaxies, without being accompanied by an equally sharp decrease of the gas density [51]. This has usually been interpreted [52–54] as indicating the existence of a critical surface density of gas σ_c related to the Toomre gravitational-stability parameter $Q = \sigma/\sigma_c$, where $\sigma_c = \kappa c_s/\pi G$, κ is the epicyclic frequency, c_s is the sound speed in the gas, and G is the gravitational constant. If $Q < 1$, gravitational instability develops in the system, leading to the birth of stars. However, this interpretation was criticized in [55]; if the Toomre parameter played such an important role in star formation, a system should be self-regulating: star formation should increase the dispersion of the sound speed c_s, keeping the Toomre parameter (and, accordingly, the star-formation rate) at a level close to $Q \sim 1$. An alternative mechanism that sharply decreases the star-formation rate [55–58] is associated with the possible existence of a stable, cool phase in the equation of state of the interstellar gas, i.e., in essence, with thermal instability. Thus, our arguments concerning the possible formation of molecular clouds via the radiative Parker instability concretize this idea, which describes the transition between the two-phase (but predominantly atomic) halo and the formation of stars.

Recent numerical calculations [20] have shown that an important role in the accumulation of diffuse gas may be played by the magneto-Jeans instability. The timescale for the development of this instability exceeds that for the development of the Parker instability, since it is determined by the self-gravitation of the gas component alone. Therefore, the radiative Parker instability develops first, and only later can it be modulated by the longer-wavelength

magneto-Jeans instability. Moreover, as is shown in [20], the magneto-Jeans instability develops more readily in spiral arms, so that it can stimulate mergers of already formed (or forming) molecular clouds in spiral arms into larger structures, such as giant molecular complexes with masses appreciably exceeding the limit M_m^u found above (probably, $M \lesssim 3 \times 10^7 M_\odot$ [20]). Based on these considerations, we suggest that the molecular gas observed in galaxies with poorly developed spirals (e.g., NGC 5055) cannot form giant molecular complexes and is instead concentrated in clouds of intermediate mass, $M_m^l < M < M_m^u$, whereas, in our Galaxy, precisely the magneto-Jeans instability may be responsible for the formation of the most massive molecular clouds.

ACKNOWLEDGMENTS

This program was supported by the Russian Foundation for Basic Research (project nos. 99-02-16938, 02-02-17642, and 03-02-16307), INTAS (project 99-1667), and the Federal Science and Technology Program in Astronomy (project 1.3.1.3). This research made use of the NASA Astrophysics Data System Abstract Service.

REFERENCES

1. L. Blitz and J. P. Williams, in *The Origin of Stars and Planetary Systems*, Ed. by C. J. Lada and N. D. Kylafis (Kluwer, 1999), p. 3.
2. J. Kwan, Astrophys. J. **229**, 567 (1979).
3. L. L. Cowie, Astrophys. J. **236**, 868 (1980).
4. J. Kwan and F. Valdez, Astrophys. J. **271**, 604 (1983).
5. W. Herbst and G. E. Assousa, Astrophys. J. **217**, 473 (1977).
6. P. R. Woodward, Astrophys. J. **207**, 466 (1976).
7. S. B. Pikelner, Astrophys. Space Sci. **7**, 489 (1970).
8. S. B. Pikel'ner, Astron. Zh. **47**, 254 (1970) [Sov. Astron. **14**, 208 (1970)].
9. T. Mouschovias, F. Shu, and P. Woodward, Astron. Astrophys. **33**, 73 (1974).
10. B. G. Elmegreen, Astrophys. J. **253**, 655 (1982).
11. M. H. Heyer and S. Terebey, Astrophys. J. **502**, 265 (1998).
12. R. J. Oliver, M. R. W. Masheder, and P. Thaddeus, Astron. Astrophys. **315**, 578 (1996).
13. M. D. Thornley, Astrophys. J. **469**, L45 (1996).
14. L. Blitz, in *The Evolution of the Interstellar Medium* (Astron. Soc. Pac., San Francisco, 1990), p. 273.
15. J. P. Phillips, Astron. Astrophys., Suppl. Ser. **134**, 241 (1999).
16. E. N. Parker, Astrophys. J. **145**, 811 (1966).
17. E. N. Parker, *Cosmical Magnetic Fields* (Clarendon, Oxford, 1979).
18. T. Mouschovias, in *Solar and Astrophysical Magnetohydrodynamics Flows* (Kluwer, Dordrecht, 1996), p. 475.
19. A. Santillán, J. Kim, J. Franco, *et al.*, Astrophys. J. **545**, 353 (2000).
20. W.-T. Kim, E. C. Ostriker, and J. M. Stone, Astrophys. J. **581**, 1080 (2002).
21. Ya. B. Zel'dovich and S. B. Pikel'ner, Zh. Éksp. Teor. Fiz. **56**, 310 (1969) [Sov. Phys. JETP **29**, 170 (1969)].
22. M. C. Begelman and C. F. McKee, Astrophys. J. **358**, 375 (1990).
23. A. A. Suchkov, Al. A. Suchkov, and Yu. A. Shchekinov, Pis'ma Astron. Zh. **7**, 617 (1981) [Sov. Astron. Lett. **7**, 342 (1981)].
24. M. G. Wolfire, D. Hollenbach, C. F. McKee, *et al.*, Astrophys. J. **443**, 152 (1995).
25. H. J. Habing, Bull. Astron. Inst. Netherlands **19**, 421 (1968).
26. M. G. Wolfire, C. F. McKee, D. Hollenbach, *et al.*, Astrophys. J. **453**, 673 (1995).
27. P. M. W. Kalberla and J. Kerp, Astron. Astrophys. **339**, 745 (1998).
28. A. Ferrara, Astrophys. J. **407**, 157 (1993).
29. R. A. Shaver, R. X. McGee, L. M. Newton, *et al.*, Mon. Not. R. Astron. Soc. **204**, 53 (1983).
30. J.-L. Hou, R.-X. Chang, and L. Chen, Chinese J. Astron. Astrophys. **2**, 17 (2002).
31. H. Kamaya, T. Horiuchi, R. Matsumoto, *et al.*, Astrophys. J. **486**, 307 (1997).
32. P. M. Solomon, A. R. Rivolo, J. Barrett, *et al.*, Astrophys. J. **319**, 730 (1987).
33. J. M. Dickey and R. W. Garwood, Astrophys. J. **341**, 201 (1989).
34. L. Blitz, in *The Physics of Star Formation and Early Stellar Evolution*, Ed. by C. J. Lada and N. D. Kylafis (Kluwer, 1991), p. 3.
35. S. Sakamoto, T. Hasegawa, T. Handa, *et al.*, Astrophys. J. **486**, 276 (1997).
36. I. Zinchenko, L. Pirogov, and M. Toriseva, Astron. Astrophys., Suppl. Ser. **133**, 337 (1998).
37. J. E. Pringle, R. J. Allen, and S. H. Lubow, Mon. Not. R. Astron. Soc. **327**, 663 (2001).
38. R. Beck, A. Brandenburg, D. Moss, *et al.*, Annu. Rev. Astron. Astrophys. **34**, 155 (1996).
39. R. Beck, astro-ph/0310287 (2003).
40. R. Beck, Space Sci. Rev. **99**, 243 (2001).
41. B. G. Elmegreen, in *The Evolution of the Interstellar Medium*, Ed. by L. Blitz (Astron. Soc. Pac., San Francisco, 1990), p. 247.
42. J. Kim, D. Ryu, and T. W. Jones, Astrophys. J. **557**, 464 (2001).
43. E. Falgarone, J.-L. Puget, and M. Perault, Astron. Astrophys. **257**, 715 (1992).
44. E. Falgarone and T. G. Phillps, Astrophys. J. **359**, 344 (1990).
45. E. Falgarone, T. G. Phillips, and C. K. Walker, Astrophys. J. **378**, 186 (1991).
46. J. Ballesteros-Paredes, E. Vazqyez-Semadeni, and J. Scalo, Astrophys. J. **515**, 286 (1999).
47. A. E. Dudorov, Astron. Zh. **68**, 695 (1991) [Sov. Astron. **35**, 342 (1991)].
48. R. C. Kennicutt, Astrophys. J. **344**, 685 (1989).
49. V. Buat, Astron. Astrophys. **264**, 444 (1992).

50. J. S. Young and N. Scoville, Astrophys. J. **258**, 467 (1982).
51. P. C. van der Kruit and L. Searle, Astron. Astrophys. **95**, 105 (1981); **95**, 116 (1981).
52. J. S. Gallagher and D. A. Hunter, Annu. Rev. Astron. Astrophys. **22**, 37 (1984).
53. E. D. Skillman and G. D. Bothun, Astron. Astrophys. **165**, 45 (1986).
54. B. Guiderdoni, Astron. Astrophys. **172**, 27 (1987).
55. J. P. Gerritsen and V. Icke, Astron. Astrophys. **325**, 972 (1997).
56. B. G. Elmegreen and A. Parravano, Astrophys. J. **435**, L121 (1994).
57. B. G. Elmegreen, Astrophys. J. **577**, 206 (2002).
58. J. Schaye, astro-ph/0205125 (2002).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 638–648. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 703–713.
Original Russian Text Copyright © 2004 by Bychkov, Sitnik.

Observed Features of a Two-Phase Stellar Wind from WR 136 in the Vicinity of NGC 6888

K. V. Bychkov and T. G. Sitnik

Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119992 Russia

Received December 25, 2003; in final form, March 15, 2004

Abstract—A number of features are detected outside the nebula NGC 6888, within 1.2° (30 pc) of the star WR 136, which can be explained in a two-phase stellar-wind model. These include regions with fine filamentary gas structure that do not contain sources of stellar wind, extended radial "streams," ultra-compact HII regions with high-velocity gas motions, and high-velocity gas motions outside the envelope of NGC 6888. The two-phase wind consists of a rarefied component and dense compact condensations, or "bullets." The bullets generate cylindrical shocks in the interstellar gas, resulting in the presence of high-velocity gas up 20–30 pc from the star, outside the cavity formed by the rarified component of the wind.
© 2004 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

This paper discusses the two-phase structure of the wind from hot stars based on observations of a region 1.4° in radius around the nebula NGC 6888. We assume that the stellar wind consists of both rarefied gas and dense, compact condensations, or "bullets," moving at a speed comparable to the outflow velocity of the rarefied gas. In this two-phase model, the conventional understanding of stellar wind corresponds to the rarefied component. One distinguishing feature of the model is that the bullets penetrate the interstellar gas much more deeply than the rarefied wind, forming perturbed structures there [1]. The bullets also take part in complex gas-dynamical interactions outside the cavity occupied by the wind's rarefied component.

The earliest stellar-wind models considered only the rarefied phase [2, 3], but it later became necessary to include a dense component as well. For example, a model with a two-phase stellar wind from a presupernova, presumably a Wolf–Rayet star, was suggested in 1979 to explain the fine filamentary structure of supernova remnants [1]. In 1984, Cherepashchuk *et al.* [4] suggested that the absorption of radio waves by a two-phase stellar wind should be lower than by a single-phase wind, because the clumpy structure of the wind makes it more "transparent." In 1990, Cherepashchuk [5] applied the idea of a clumpy stellar wind to reduce the discrepancy between the theoretical X-ray luminosity of the WN5 component of the binary system V444 Cyg and the luminosity observed by the Einstein observatory. Evidence for a dense stellar wind is also provided by estimates of the star's mass-loss rate determined independently in different wavebands: radio observations give an $\dot{M}$ that exceeds $\dot{M}$ derived from optical observations by an order of magnitude [6]. In addition, a cloudy structure for the wind from the WR star can explain the rapid spectroscopic and photometric variability of the V444 Cyg system [4, 7, 8]. Applying the two-phase stellar wind model to another binary, WR 140, made it possible to explain the parameters of its X-ray, infrared, and nonthermal radio emission [9, 10].

However, all the evidence for clumpy structure of stellar winds is indirect. It is difficult to detect bullets near a star directly due to their small expected size (of the order of the solar radius). However, fast bullets can be detected from their interactions with the interstellar gas outside the cavity, in regions as large as several tens of parsecs. We consider below the parameters of bullets (Section 2), their interactions with the interstellar medium, and their evolution (Section 3). In Sections 4 and 5, we discuss manifestations of such interactions in the vicinity of the ring nebula NGC 6888 formed by the wind from the star WR 136.

2. PARAMETERS OF THE BULLET COMPONENT OF THE WIND

Let us consider the behavior of the bullets and try to estimate their parameters. The bullets move with the rarified gas in the cavity formed by this gas until they reach the boundary between the rarefied wind and the interstellar gas. In the case of NGC 6888, the radius of the cavity is $r_{\rm rar} = 5$ pc (see Section 4).

Because of its high density, a bullet is not slowed significantly either at the shock front or at the boundary between the wind and the perturbed interstellar

1063-7729/04/4808-0638$26.00 © 2004 MAIK "Nauka/Interperiodica"

gas. We can also neglect a bullet's deceleration during its exit to the unperturbed interstellar medium. During its stay in a high-density region, the bullet experiences additional compression and becomes smaller in size.

Let us estimate the bullet's size when it exits to the interstellar medium (at a distance of $r_{\rm rar}$). We will assume that half of the stellar mass flows as the rarefied wind and half as bullets. Then,

$$\dot{M}_{\rm rar} = \dot{M}_{\rm b} = \frac{1}{2}\dot{M} = 3.15 \times 10^{-5}\, M_\odot\ {\rm yr}^{-1}. \quad (2.1)$$

Here and below, the subscript "b" is used for parameters of the bullets, and the subscript "rar," for those of the rarefied gas. We take the parameters of the star WR 136 (except for the parameters of the bullets) from [11]: the outflow rate $\dot{M} = 6.3 \times 10^{-5} M_\odot$/yr, the flow velocity $V_{\rm flow} = 1800$ km/s, and the mass of the star $M_* = 19\ M_\odot$. The lifetime of WR 136 in the Wolf–Rayet stage,

$$\tau_{\rm WR} = 2 \times 10^5 \text{ years}, \quad (2.2)$$

is taken from [12].

We estimate the parameters of the bullets using the results of studies of the binary WR 140, which has been well studied in various wavebands. Based on the interpretation of its X-ray, infrared, and nonthermal radio emission from [9, 10], the following ranges seem reasonable for the bullet's density, $\rho_{\rm b}$; size, $l_{\rm b}$; and mass, $m_{\rm b}$, at a distance from the star of $r_0 \approx 3 \times 10^{14}$ cm:

$$\rho_{\rm b}(r_0) \sim 10^{-11}\text{–}10^{-8}\ {\rm g/cm^3}, \quad (2.3)$$
$$l_{\rm b}(r_0) \sim 10^{10}\text{–}10^{12}\ {\rm cm},$$
$$m_{\rm b} \sim 10^{22}\text{–}10^{25}\ {\rm g}.$$

We will take the remaining wind parameters to be the same as for the WR 136 system. According to (2.1), (2.2), and (2.3), an order-of-magnitude estimate of the number of bullets ejected by the star during the active outflow stage is

$$N_{\rm b} = \frac{\dot{M}_{\rm b}\tau_{\rm WR}}{m_{\rm b}} \approx 6 \times 10^{10}.$$

The size of a bullet when it arrives at the edge of the cavity depends on the wind's thermal regime, and we must consider several possibilities.

I. First, let us assume that the pressures inside the bullet and in the rarefied gas of the wind are equal:

$$p_{\rm b} = p_{\rm rar}. \quad (2.4)$$

This is true if the bullet is effectively compressed by the ambient wind. If the outflow rate is steady, the density of the rarefied gas is inversely proportional to the square of the distance to the star:

$$n_{\rm rar} \sim r^{-2}. \quad (2.5)$$

The temperature of the rarefied gas is determined by its interaction with the star's radiation field. Let us consider two types of such interaction.

I.1. As the rarefied gas expands, its optical depth to the star's ionizing radiation decreases. In a transparent medium, the rates of photoionization, recombination, and electron collisions have similar dependences on the distance from the star, r. Thus, it is reasonable to assume that the star's temperature remains approximately constant as long as its radiation heats the gas efficiently. It then follows from (2.5) that

$$p_{\rm rar} \sim n_{\rm rar} \sim r^{-2}. \quad (2.6)$$

On the contrary, if the radiation does not efficiently heat the dense gas of the bullet, and its expansion is fairly adiabatic, then

$$p_{\rm b} \sim \rho_{\rm b}^{5/3}. \quad (2.7)$$

Formulas (2.4), (2.6), and (2.7) lead to the following relation between the bullet's density and its distance from the star:

$$\rho_{\rm b} \sim r^{-6/5}. \quad (2.8)$$

For such an expansion rate, the bullet's density when it exits to the unperturbed interstellar gas, $\rho_{\rm out}$, decreases by a factor of $(r_{\rm rar}/r_0)^{6/5} \approx 10^6$ and, according to the parameters (2.3), it will be in the range

$$\rho_{\rm out}^{\rm (T)} \sim 10^{-17}\text{–}10^{-14}\ {\rm g/cm^3}.$$

The superscript "T" here denotes the bullet's isothermal expansion. Accordingly, the bullet's size increases by approximately (100) of magnitude and becomes

$$l_{\rm out}^{\rm (T)} \sim 10^{12}\text{–}10^{14}\ {\rm cm}.$$

I.2. In the case of a weak interaction between the rarefied part of the wind and the star's radiation, the expansion of the rarefied component can be considered to be nearly adiabatic. We then obtain instead of (2.6)

$$p_{\rm rar} \sim n_{\rm rar}^{2/3} \sim r^{-10/3}. \quad (2.9)$$

The expansion rate of the bullet is much higher in this case:

$$\rho_{\rm b} \sim r^{-2}. \quad (2.10)$$

Thus, it becomes less dense and larger in size:

$$\rho_{\rm out}^{\rm (ad)} \sim 10^{-20}\text{–}10^{-17}\ {\rm g/cm^3},$$
$$l_{\rm out}^{\rm (ad)} \sim 10^{13}\text{–}10^{15}\ {\rm cm}.$$

Here, the superscript "ad" denotes the adiabatic expansion of the bullet.

II. Now, consider the case when the pressure inside the bullet is not completely compensated by the

pressure of the rarefied gas, so that condition (2.4) is not satisfied. In this case, the expansion of the bullet is virtually free. The expansion velocity, V_{exp}, at a distance r_0 from the star is determined by the bullet's temperature, T_0. According to the interpretation of the infrared radiation from the WR 140 system, T_0 can be taken to be 1000 K, so that $V_{exp} \approx 3 \times 10^5$ cm/s. The bullet's time of flight to the boundary of the rarefied gas can be estimated as $r_{rar}/V_{flow} \approx 10^{11}$ s. During this time, the bullet expands to a size of the order of $l_{out} \approx 3 \times 10^{16}$ cm, and its density becomes lower than that of the interstellar gas. However, the bullet's free expansion leads to a rapid decrease of its dynamical pressure, and, after some time, its expansion is slowed due to the pressure of the wind's rarefied component. A wide range of sizes and densities is possible for bullets. For example, a bullet size of $l_{out} \approx 5 \times 10^{15}$ cm and mass of about 10^{25} g corresponds to a density of $\rho_{out} \approx 10^{-22}$ g/cm^3.

Thus, it is reasonable to suppose that bullets entering the interstellar gas through the boundary of the rarefied wind of a Wolf–Rayet star can have sizes and densities covering a wide range:

$$l_{out} \sim 10^{13}\text{–}10^{16} \text{ cm}, \qquad (2.11)$$
$$\rho_{out} \sim 10^{-23}\text{–}10^{-14} \text{ g/cm}^3.$$

Below, we consider the interaction of a bullet with the interstellar medium.

3. EXPANSION OF A BULLET AND OF THE GAS IT PERTURBS

The most interesting manifestations of the two-phase model begin when a bullet exits to the interstellar medium, which has not yet been perturbed by the wind's rarefied component. A bullet's evolution for the case of supernova remnants was considered in [1], where it was demonstrated, in particular, that the formation of fine filaments could be due to perturbations of the interstellar gas by a bullet. Here, we consider the case of a stellar wind.

3.1. Interstellar Medium Perturbed by a Bullet

Like a supersonic aircraft, a dense bullet generates a shock in the interstellar gas. However, the Mach number is very large in this case, so that the shock can be modeled as being infinitely strong, and the Mach cone is essentially a cylinder. The perturbed gas compresses the bullet and hinders its expansion until the pressure behind the shock front formed by the bullet drops due to the bullet's deceleration. The shock first propagates adiabatically, then radiative losses begin.

During the adiabatic stage, the velocity of the shock, V_{sh}, and the cylinder's radius, r_{sh}, are related as

$$r_{sh} V_{sh} = \text{const}, \qquad (3.1)$$

leading to the approximate formula describing the evolution of the front

$$t_{sh}(r_{sh}) = \frac{r_{sh}^2}{V_{flow} l_{out}}. \qquad (3.2)$$

Here, $t_{sh}(r_{sh})$ is the time taken by the shock to cover the distance r_{sh}. The adiabatic stage corresponds to high velocities of the front. For this reason, the gas at the front is heated to millions of degrees and radiates in the X-ray.

As the cylindrical shock decelerates, the radiative stage begins, when the kinetic energy of the gas is carried away in the form of radiation. Under the conditions analyzed, the transition between the stages is determined primarily by the ionization of hydrogen in the unperturbed gas. The radiation cooling time [13],

$$t_{cool} = \frac{kT}{(1-x) n_e \varphi(T)}, \qquad (3.3)$$

depends on the gas temperature T, electron density n_e, degree of ionization of the hydrogen x, and radiation coefficient $\varphi(T)$. The value of n_e in the right side of (3.3) is a factor of four higher than the number density of particles in the unperturbed interstellar gas, n_0. Below, we use both the density, ρ, and the number density, n: $\rho = \mu m_{amu} n$, where μ is the molecular weight of the gas and m_{amu} is the atomic mass unit. It is reasonable to assume that n_0 in the region around WR 136 is in the range

$$n_0 \sim 0.1\text{–}0.5 \text{ cm}^{-3}. \qquad (3.4)$$

Here, the principal cooling mechanism behind the shock front is excitation of the second level of the hydrogen atom, which we can estimate [13] as

$$\varphi(T) \simeq 1.40 \times 10^{-24} \frac{e^{-\beta}}{\sqrt{kT}} \text{ erg cm}^3 \text{ s}^{-1}, \qquad (3.5)$$

where k is Boltzmann's constant. It follows from (3.3) and (3.5) that

$$t_{cool} \approx 7.12 \times 10^{23} \frac{(kT)^{3/2} e^{\beta}}{(1-x) n_e} \text{ s}, \qquad (3.6)$$

where $\beta = E_{12}/kT$ and $E_{12} = 10.2$ eV. The temperature near the front is expressed in terms of the velocity by the relation

$$kT = \frac{3}{16} \mu m_{amu} V_{sh}^2.$$

Using this last formula, we can rewrite $t_{\rm cool}$ as a function of $V_{\rm sh}$:

$$t_{\rm cool} \approx 1.24 \times 10^{-13} \frac{\mu^{3/2} V_{\rm sh}^3 e^{\beta}}{(1-x) n_e} \text{ s}. \quad (3.7)$$

The radiative stage begins when $t_{\rm cool}$ becomes comparable to the dynamical time scale:

$$t_D \approx 0.1 t_{\rm sh}(r_{\rm sh}). \quad (3.8)$$

The factor of 0.1 takes into account the steep density gradient, due to which the radiative cooling occurs only in a narrow region behind the front. Using (3.2) and (3.7), we derive the ratio

$$\gamma = \frac{t_{\rm cool}}{t_D} = 5 \times 10^{-12} \frac{\mu^{3/2} V_{\rm sh}^5 e^{\beta}}{(1-x) n_e V_{\rm flow} l_{\rm out}}. \quad (3.9)$$

The condition $\gamma = 1$ approximately describes the beginning of the radiative stage. When it is fulfilled, the parameters $\mu = 0.7$–1.0, $x = 0.3$–0.8, $n_e = 0.5$–1.5 cm^{-3}, $V_{\rm flow} = 1800$ km/s, and $l_{\rm out} = 10^{15}$–10^{16} cm correspond to $V_{\rm sh} \approx 60$–100 km/s. Thus, local velocity variations up to 60–100 km/s can be observed outside the cavity filled with the rarefied gas at distances of the order of 0.01–0.5 pc. The radius of the cylindrical front increases in proportion to the decrease in its velocity and reaches the value $r_{\rm sh} = V_{\rm flow} l_{\rm out} / V_{\rm sh} = 3 \times 10^{16}$ cm by the onset of the radiative stage. As long as the velocity of the front exceeds 30 km/s, the perturbed gas radiates a considerable fraction of its energy in optical lines [13], resembling a compact HII region.

3.2. Evolution of a Bullet

After it leaves the cavity, a bullet moves at a nearly constant velocity, experiencing comparatively little deceleration due to sweeping up of the interstellar gas. We provisionally call this stage of its motion the "inertial" stage. The bullet gradually slows as the mass of gas it has perturbed increases. We will consider the time when the mass of swept-up gas, $m_{\rm sw}$, becomes equal to the bullet's mass to be the end of the inertial stage.

Let us estimate the distance the bullet can cover during the inertial stage, $r_{\rm iner}$. The condition of mass equality, $m_{\rm sw} = m_{\rm b}$, leads to the proportion

$$\frac{r_{\rm iner}}{l_{\rm b}(r_{\rm rar})} \approx \frac{\rho_{\rm out}}{\rho_0}, \quad (3.10)$$

which contains the bullet's density and size as it exits into the interstellar medium at a distance $r_{\rm rar}$ from the star. By the end of the inertial stage, the bullet will reach a distance $r_{\rm rar} + r_{\rm iner}$ from the star. The value of $r_{\rm iner}$ is determined by the bullet's mass and size, as well as by the density of the interstellar gas:

$$r_{\rm iner} \approx \frac{m_b}{\rho_0 l_{\rm out}^2}. \quad (3.11)$$

For example, a bullet with a mass of 10^{25} g and a size of $l_{\rm out} \approx 10^{15}$ cm will cover a distance of about 10–15 pc from the boundary of the rarefied-wind region before beginning to appreciably decelerate.

According to these estimates, before the onset of the phase of strong deceleration, a bullet can cover a distance that exceeds its size by a factor of several million. Consequently, $r_{\rm iner}$ can be as large as tens of parsecs. Of course, only sufficiently dense and compact bullets are capable of covering distances this great, so that the cross-sectional area of the regions they perturb are not large. More appreciable perturbations are brought about by larger bullets, which travel correspondingly shorter distances.

A bullet's deceleration is accompanied by its expansion due to the decreasing external pressure of the surrounding gas, which, in turn, enhances the deceleration. At the same time, the bullet disintegrates into individual fragments due to various types of developing instabilities, such as the tangential instability. The deceleration distance is larger for the fragments than for a large bullet. The rate of development of the instabilities is determined by the hydrodynamical time scale, which is quite short compared to the bullet's age. Consequently, the bullet's fragmentation leads to its rapid deceleration. During the deceleration stage, the bullet will cover a distance equal to some 20–30% of the distance covered during the inertial phase, so that the total distance covered can be about 15–20 pc.

The kinetic energy accumulated in the bullet, $m_{\rm b} V_{\rm flow}^2/2$, is sufficient to heat a region of gas and dust about 10^{18} cm in size with a particle density of 0.2 cm^{-3} to a temperature of about 1000 K.

Thus, a bullet forms small (fractions of a parsec) multilayer structures in the interstellar medium with steep temperature and density gradients. The various layers of these structures are manifest as X-ray sources, optical HII regions, and heated regions of dust emitting in the infrared. Such structures for supernova remnants are described in [1].

4. THE INTERSTELLAR MEDIUM SURROUNDING NGC 6888

The well-known nebula NGC 6888 surrounding a close binary system (WN6 + a compact source) is located in the Cygnus arm (Fig. 1). It has the appearance of an obvious shell in the optical. The size of the shell depends on the spectral line observed [12, 14];

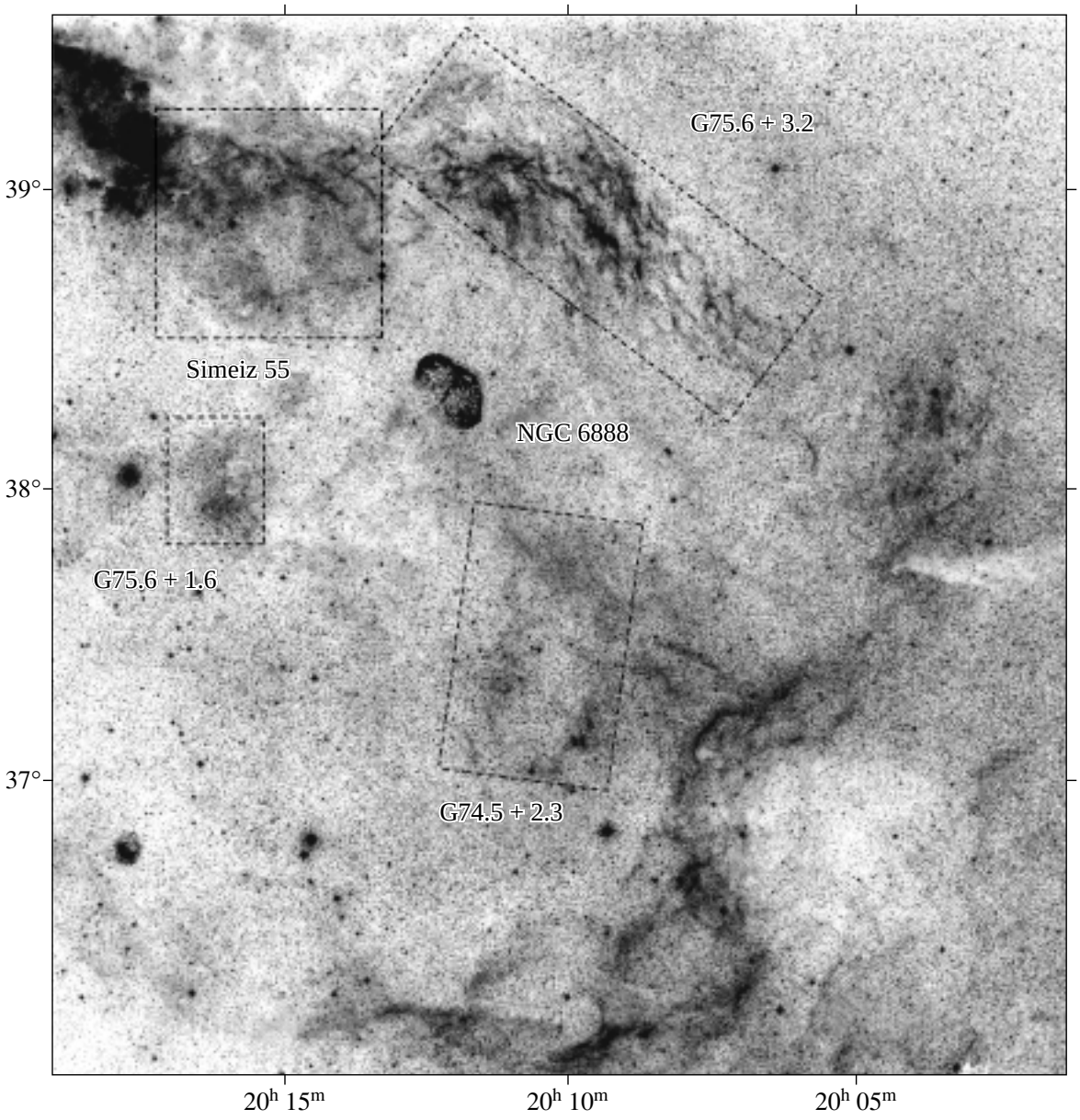

Fig. 1. Vicinity of NGC 6888 (red Palomar image). The boundaries of emission nebulas are shown as rectangles. The molecular cloud G75.6+2 is within $\alpha_{2000} = 20^h12^m - 20^h15^m$, $\delta_{2000} = 37^\circ45' - 38^\circ30'$.

it is somewhat larger in [OIII] lines than in Hα+[NII] lines. This shell structure of NGC 6888 provides evidence that it was formed by the stellar wind from the central star, WR 136. Radial streams and ejections of gas outside the nebula are observed along the minor axis. Bright compact features are observed in the X-ray both inside and outside the optical nebula [15].

WR 136 has been identified as a member of the association Cyg OB1 [16], whose distance is usually taken to be 1.5 kpc. This is the mean of the distance estimates made in [17–19], and this same distance to NGC 6888 is given in [20].

In the radiative stage, the stellar-wind material is confined to a dense shell [2, 3]. Therefore, in the case of NGC 6888, the size of the cavity occupied by the rarefied wind, r_{rar}, is equal to the nebula's maximum radius, determined from its size in [OIII] lines: $r_{rar} \approx 5$ pc.

In our search for signs of interaction between bullets and the interstellar gas, we will consider observations of a region with a radius of 1.4° around NGC 6888 (Fig. 1) in various spectral bands. This analysis demonstrates that the ionizing radiation and bullets from WR 136 penetrate to distances appreciably larger than the size of NGC 6888.

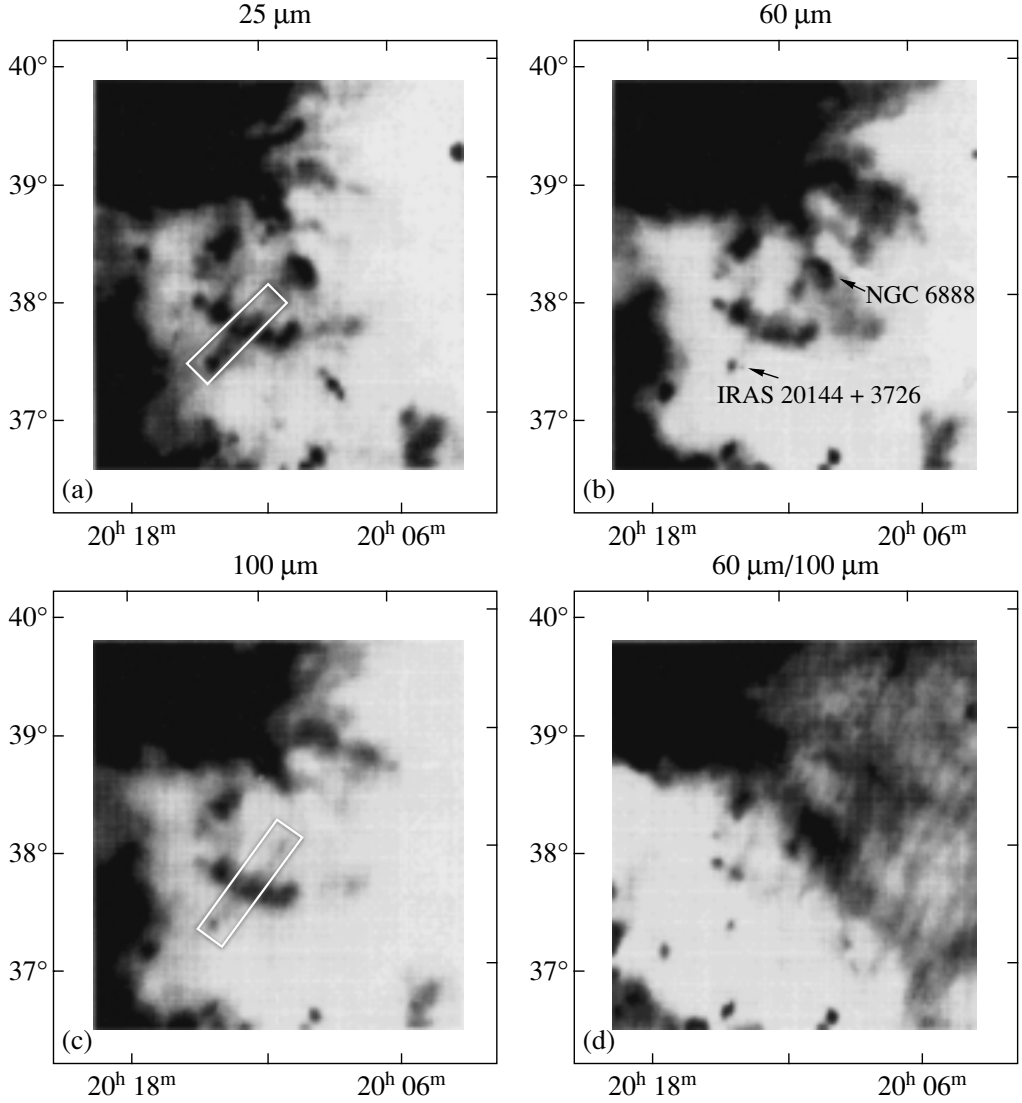

Fig. 2. IRAS infrared images of the vicinity of NGC 6888 in the (a) 25 μm, (b) 60 μm, and (c) 100 μm bands and (d) an image based on the $I(60\ \mu\text{m})/I(100\ \mu\text{m})$ intensity ratio [21]. The rectangle marks the southeastern stream (a, c), and the arrows point to NGC 6888 and the IR source at the end of the southeastern stream (b). The equatorial coordinates are given for epoch 1950.

4.1. IR Emission: Shell, Streams, and Point Sources

The radiation of the interstellar medium in the vicinity of NGC 6888 was first associated with the star WR 136 in the IR [21, 22]. A clumpy IR shell $1.5° \times 1.8°$ in size was identified in the IRAS 25, 60, and 100 μm bands, corresponding to 40 pc × 50 pc at a distance of 1.5 kpc (Fig. 2). NGC 6888, which emits brightly in the IR, is at the center of the shell.

The IR radiation from the region displays the following two characteristic features (Fig. 2).

1. "Streams"—comparatively narrow lanes of emission directed toward the northeast, southwest, and southeast from NGC 6888 in the plane of the sky. The first two streams are visible at 60 μm, while the southeastern stream can be traced in each of the bands. This latter stream resembles a narrow radial lane about 1° long with a brightening towards its edge.

2. Four point IR sources identified with ultracompact HII regions (Fig. 3) [23]. These sources were identified with HII regions based on their spectra; the ratios of the fluxes in different bands are not typical of stars or galaxies. Note especially the bright point source IRAS 20144+3726—an ultracompact HII region located where the southeastern stream ends in the plane of the sky (Figs. 2b, 3). Two other ultra-compact HII regions, IRAS 20138+3751 and IRAS 20148+3757, are observed within 1.2° in the same sector. The HII region IRAS 20101+3815 is

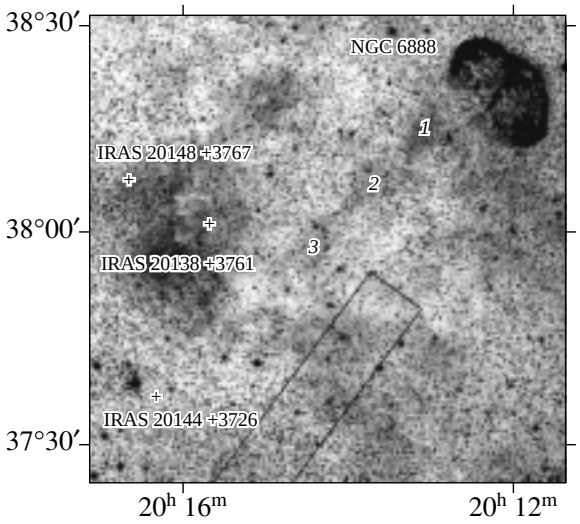

Fig. 3. Southeastern sector of the vicinity of NGC 6888. The molecular cloud G75.6+2, which is manifest via the radiation of its dust, is in the center and is crossed by the southeastern stream (marked by the numbers 1–2–3). The crosses show ultra-compact HII regions and the rectangle is the "kinematic stream."

projected onto the northern part of the NGC 6888 shell.

The possible connection of the point sources and IR streams with WR 136 will be discussed below. We note here only that, if the point sources are at the distance of WR 136, they are within 30 pc of the star.

4.2. Filamentary Nebulas and Shells

The structure of the optical and IR radiation is determined largely by the cloudy component of the interstellar medium. While fine filamentary nebulae and shells dominate in the northern half of the region (Fig. 1), the molecular cloud G75.6+2 dominates in the southeast [24, 25]. We will consider the southeastern sector separately.

The emission nebulae Simeiz 55 [26] and G75.6+3.2 are in the northern part of the studied region at distances of 0.5°–1.2° (13–30 pc) from WR 136 in the plane of the sky (Fig. 1). The nebula G74.5+2.3 is observed not more than 1.4° (36 pc) to the southwest. Let us consider several pieces of evidence indicating that these nebulae, together with NGC 6888, form a single complex associated with the central star, WR 136.

1. There are no Cyg OB1 stars (or any other association of the Cygnus arm) hotter than B0 in the direction toward the nebulae. However, the nebulae show bright emission not only in the Hα+[NII] and [SII] lines, but in [OIII] lines as well [14]. For these latter lines to be bright, strong ionizing radiation is needed. Therefore, all three nebulae were probably formed by WR 136.

2. The northern nebula, G75.6+3.2, is no closer than 0.5° to NGC 6888 and has a filamentary structure, with the filaments elongated predominantly in the same direction as the filaments inside NGC 6888. Either a stellar wind or a supernova explosion is needed to create such a structure. However, G75.6+3.2 cannot be a supernova remnant, since there is no evidence for nonthermal radio emission here [27].

3. Simeiz 55 and G74.5+2.3 are located in opposite directions relative to WR 136 and are bright shells, though the nebulae themselves contain no stars with strong winds. A stream observed in the Hα+[NII] and [OIII] lines extends from NGC 6888 to the southwestern shell, G74.5+2.3 [14].

4. Simeiz 55 has different morphologies in different optical lines: it has a horseshoe shape (open towards NGC 6888) in the Hα + [NII] lines, whereas only the northwestern part of the horseshoe, the part closest to WR 136 in the plane of the sky, remains bright in [OIII] lines [14].

5. The nebulae form a multilayer shell structure around NGC 6888 in the plane of the sky in both the Hα+[NII] and [OIII] lines [14]. In general, the shells in different lines coincide, though some inner details of the shell are more pronounced in [OIII], while the outer details are more clearly visible in Hα+[NII]. The described features of the shell's multilayer structure (with the inner regions being hotter) also testify to a thermal mechanism for its formation.

Thus, we have reason to believe that these nebulae are not isolated objects and form a single gas and dust complex together with NGC 6888 with a radius of 30–36 pc (see also Section 5.1). The optical line emission of the nebulae is due to ionizing radiation from the central star, WR 136, and the fine filamentary structure of G75.6+3.2 could have formed via the interaction of bullets from WR 136 with interstellar clouds, as described in [1].

4.3. Population of the Northeastern Sector

Within 1.2° to the southeast is the molecular cloud G75.6+2 ($V_{LSR} \sim 5$–12.5 km/s) (Figs. 1, 3). Its dust emission makes it easy to find in red Palomar images (Fig. 1).

Among the optical structures in this sector, we note especially a narrow radial stream whose weak emission can be traced at least 0.5° (13 pc) to the southeast of NGC 6888 (Fig. 3). This stream is the optical counterpart of the southeastern IR stream that ends in a bright point IR source, an ultra-compact HII region, at a distance of 1° (26 pc) (Fig. 2). The IR stream is nearly twice as long as the optical stream.

In addition, the stream is projected against a region of reduced CO emission, which is radially elongated and observed at velocities of $V_{LSR} \sim 10-12.5$ km/s [24]. Thus, this same feature—a radial stream ending in an ultra-compact HII region—is observed in three wavebands. Another ultra-compact HII region in the southeastern sector, IRAS 20138+3751, is observed 0.7° (18 pc) from WR 136 in the plane of the sky and is also associated with optical features, being surrounded by a fine dust envelope with characteristically curved small filaments (Fig. 3).

Thus, a radial stream and ultra-compact HII regions that are probably related to the star WR 136 (their locations are discussed in more detail below, Section 5.3) are observed in the southeastern cloud. It was suggested in [23] that the ultra-compact HII regions surrounded a stellar cocoon. However, the estimates in Section 3.2 demonstrate that the bullets' kinetic energy is sufficient to form such HII regions that are up to 10^{18} cm in size (see also Section 5.3).

Let us summarize the results of Section 4, where we have considered the population of a region 30–36 pc in radius around WR 136. Clear evidence for gas and dust perturbations that are undoubtly related to the central star is observed outside NGC 6888. It is possible that bullets tunneling through the nearby southeastern molecular cloud are observed as the optical and IR streams. The ultra-compact HII regions could also be tracers of the bullets' interactions with dense clouds, whereas the numerous northwestern filaments result from bullets "fired" at the molecular clouds.

5. KINEMATICS OF THE INTERSTELLAR GAS NEAR NGC 6888

The kinematics of the ionized hydrogen in the vicinity of NGC 6888 was considered in detail in a series of studies of the gas and dust complex around the Cyg OB1 association [25, 28, 29]. Let us consider the radial-velocity field of the region in order to locate the populations and study the kinematics of the bullets.

5.1. Line-of-Sight Velocities of the Ionized Hydrogen near NGC 6888

The line-of-sight velocities of the most typical motions of ionized hydrogen around NGC 6888 are $V_{LSR}^{main} \sim -5$ to 25 km/s and are characteristic of the gas in the Cygnus arm [28, 29]. The observed radial-velocity range is due primarily to large-scale gas motions: the Galactic rotation and streaming motions related to the spiral-density waves [30]. The radial velocities in NGC 6888 and in its immediate vicinity, within 0.2°, as well as those in Simeiz 55, are $V_{LSR}^{main} \sim 10 \pm 4$ km/s. The radial velocities of the molecular cloud G75.6+2 ($V_{LSR} \sim 5-12.5$ km/s) testify that it is spatially close to NGC 6888.

Ionized hydrogen with large negative (from $V_{LSR}^{high} \sim -100$ km/s) or positive (to $V_{LSR}^{high} \sim 80$ km/s) velocities is also observed in the neighborhood of NGC 6888. The high-velocity motions with $V_{LSR}^{high} > 25$ km/s should correspond to peculiar gas motions in the Cygnus arm, whereas the gas motions with large negative velocities could be either peculiar motions in the Cygnus arm or large-scale motions in the distant Perseus arm [25]. Peculiar motions related to the expansion of the shell of NGC 6888 itself are observed, with the expansion velocity being dependent on the direction and equal to 110 km/s along the major axis and 55 km/s along the minor axis [31].

5.2. Characteristics of High-Velocity Motions near NGC 6888

To detect possible manifestations of the bullet component of the stellar wind from WR 136 outside NGC 6888, we studied variations of the observed V_{LSR} of the ionized hydrogen with distance r from the central star. Since the structure of the studied region is nonuniform, we considered $V_{LSR}(r)$ for the directions toward each of the four objects in the region: the nebulae Simeiz 55, G75.6+3.2, G74.5+2.3, and the molecular cloud G75.6+2 (Fig. 4). The diagrams present V_{LSR} as a function of the distance from WR 136.

Figure 4 shows the following.

(1) High ionized-hydrogen velocities are observed outside the shell of NGC 6888. The velocity of the receding gas reaches 60–80 km/s, and that of the approaching gas, (−100 to −90) km/s.

(2) The speed for the high-velocity motions, V_{LSR}^{high}, decreases towards the nebulae Simeiz 55, G75.6+3.2, and G74.5+2.3 and, within 1.5° of WR 136, becomes equal to V_{LSR}^{main}, the velocity of the ambient interstellar medium (Figs. 4a, 4b, 4d). This provides evidence that the wind from WR 136 is the source of the high-velocity motion in these sectors. Note that the velocity of the high-velocity component of the ionized hydrogen in the southeastern sector does not vary in the region considered. This may be due to the superposition of the winds from the stars WR 137 and WR 138, which also belong to the Cyg OB1 association.

Thus, we find additional evidence that the wind from WR 136 influences the gas outside NGC 6888. We associate the high-velocity motions outside the

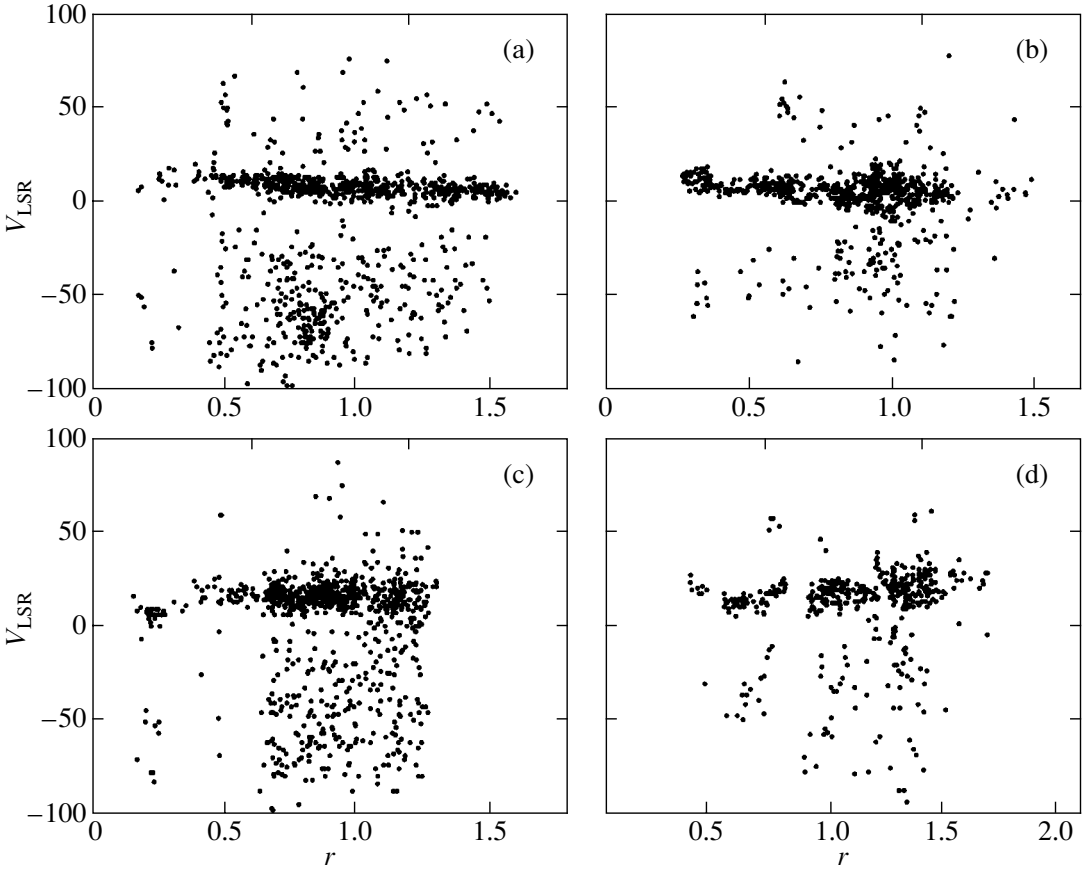

Fig. 4. Distribution of radial velocities with distance (in degrees) from WR 136 in the direction of (a) Simeiz 55, (b) G75.6+3.2, (c) the molecular cloud G75.6+2, and (d) G74.5+2.3. The nebula NGC 6888 is within 0.2° of the star.

cavity occupied by the rarefied wind with bullets that interact with the interstellar medium and accelerate it to high velocities (see Section 3).

5.3. Kinematics of Streams and Ultracompact HII Regions

Let us consider the kinematics of gas in the direction of the streams. In the southeastern stream (Fig. 3), the ionized hydrogen moves with $V_{LSR}^{main} = 8 \pm 3$ km/s (the number of measurements is $N = 34$) within 0.5°, then at a speed of $V_{LSR}^{main} = 17 \pm 2$ km/s ($N = 12$). The velocities V_{LSR}^{main} at the beginning and end of the stream coincide with the velocities of ionized hydrogen and CO emission observed near the stream. This stream does not stand out kinematically against the interstellar medium.

At a distance from NGC 6888 not exceeding 0.5°, we find a "kinematic stream": a radially elongated region about 0.5° in length (Fig. 3). In this stream, motions of ionized hydrogen at velocities $V_{LSR}^{main} = 16 \pm 3$ km/s and $V_{LSR}^{main} = (-80$ to $-40)$ km/s stand out [25]. This kinematic stream does not display increased optical or IR emission.

Let us see what kind of motion is observed in the two ultra-compact HII regions (Fig. 3). The ultra-compact HII region IRAS 20144+3726 is interesting because it is located at the end of the southeastern stream. Ionized hydrogen moving at $V_{LSR}^{main} = 12 \pm 1$ km/s ($N = 2$) and $V_{LSR}^{main} = -52 \pm 20$ km/s ($N = 3$) is observed within 1′ of this HII region. The same velocity ($V_{LSR} = -57$ km/s) was found from observations of this source in CS emission [23]. High-velocity ionized hydrogen is present in the vicinity of NGC 6888, and it cannot be ruled out that this gas is likewise in the Cygnus arm rather than in the Perseus arm. It could be the trace of a bullet ejected by WR 136. In this case, the size of the ultra-compact HII region ($\sim 2'$) does not exceed 0.5 pc.

The ultra-compact HII region IRAS 20138+3751 in the vicinity of NGC 6888 is remarkable in that it is projected against the outer edge of the southeastern molecular cloud and is surrounded in the plane of the sky by a dust envelope about 8′ in size, which is clearly visible in red Palomar images (Fig 3). The

envelope consists of tiny filaments that are curved as if they are flying from WR 136. The radial velocities of the ionized hydrogen are $V_{\rm LSR}^{\rm main} = 15 \pm 3$ km/s ($N = 11$) in the direction of the dust filaments and up to $V_{\rm LSR}^{\rm high} = -42 \pm 25$ km/s ($N = 4$) inside the dust envelope. Thus, we observe an HII region with high-velocity motions surrounded with a dust envelope where the radial velocities are typical of the vicinity of NGC 6888. This region could also be associated with the interaction of a bullet from WR 136 and the ambient interstellar medium. The size of the dust envelope is then 3 pc.

Thus, in the neighborhood of WR 136 (18–30 pc), outside the nebula NGC 6888, there are ultracompact HII regions where no exciting stars are found, but where high-velocity gas motions are present.

Thus, we can summarize the main result of Section 5 as follows. High-velocity gas motions related to the star WR 136 are observed outside the nebula NGC 6888. We associate such perturbations of the gas with cylindrical shocks created by the passage of bullets in the interstellar medium.

6. CONCLUSIONS

Let us summarize the observations that we believe provide evidence for interactions between the bullet component of the wind from WR 136 and the interstellar medium. These observations cover an extended area with a radius of about 30 pc, whereas the radius of the shell NGC 6888 is about 5 pc. The shell NGC 6888 delineates the region occupied by the rarefied wind.

1. High-velocity motions of gas are detected outside the nebula NGC 6888. These are related to the star WR 136, as is indicated by the decrease in the magnitude of the radial velocity with distance from the star that is observed along most directions.

2. Radial streams from WR 136 as long as 25 pc are observed. Among them, we especially note the southeastern stream extending to an ultra-compact HII region. This stream is observed as a narrow lane of optical and IR emission and is projected against a region of reduced CO emission.

3. Several IR sources are identified with ultra-compact HII regions without exciting stars. High-velocity motions are observed in two of these HII regions.

4. The nebula G75.6+3.2, which does not contain any stellar-wind sources, has a fine filamentary structure. G75.6+3.2 is outside the zone of the rarefied wind from WR 136 at a distance of 13–30 pc.

All these facts can be explained in a model with a two-phase stellar wind consisting of rarefied gas and bullets. The bullets penetrate the interstellar gas much more deeply than the rarefied wind, forming perturbed structures there. Some massive bullets can get as far as 15–20 pc beyond the cavity and heat the ambient gas to 1000 K. This may be how the ultra-compact HII regions observed within 30 pc from WR 136 formed. A dense bullet generates a shock in the interstellar gas, which can explain the presence of high-velocity motions at distances of up to 30 pc from WR 136, as well as the high velocities in the ultra-compact HII regions. The velocity of the cylindrical shocks generated by the bullets is in the range 60–100 km/s. Precisely such velocities are observed in the interstellar medium surrounding NGC 6888.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 01-02-16118 and 02-04-16042), the Federal Scientific and Technology Program "Astronomy" (contract 40.022.1.1.1102), and the Program of Support for Leading Scientific Schools of Russia (grant NSh-388.2003.2).

REFERENCES

1. K. V. Buchkov, Astron. Zh. **56**, 781 (1979) [Sov. Astron. **23**, 438 (1979)].
2. J. Castor, R. McCray, and R. Weaver, Astrophys. J. **200**, L107 (1975).
3. R. Weaver, J. Castor, R. McCray, *et al.*, Astrophys. J. **218**, 377 (1977).
4. A. M. Cherepashchuk, J. A. Eaton, and Kh. F. Khaliullin, Astrophys. J. **281**, 774 (1984).
5. A. M. Cherepashchuk, Astron. Zh. **67**, 955 (1990) [Sov. Astron. **34**, 481 (1990)].
6. A. M. Cherepashchuk, in *IAU Symp. No. 163: Wolf-Raet Stars: Binaries, Colliding Winds, Evolution* (1995).
7. A. F. J. Moffat, L. Drissen, R. Lamontagne, and C. Robert, Astrophys. J. **334**, 1038 (1988).
8. I. I. Antokhin, T. Nugis, and A. M. Cherepashchuk, Astron. Zh. **69**, 516 (1992) [Sov. Astron. **36**, 260 (1992)].
9. O. V. Aleksandrova and K. V. Bychkov, Astron. Zh. **77**, 883 (2000) [Astron. Rep. **44**, 781 (2000)].
10. O. V. Aleksandrova and K. V. Bychkov, Astron. Zh. **78**, 327 (2001) [Astron. Rep. **45**, 281 (2001)].
11. T. Nugis and H. J. G. L. M. Lamers, Astron. Astrophys. **360**, 227 (2000).
12. B. D. Moore, J. J. Hester, and P. A. Scowen, Astron. J. **119**, 2991 (2000).
13. S. A. Kaplan and S. B. Pikel'ner, *The Interstellar Medium* (Fizmatgiz, Moscow, 1963) [in Russian].
14. R. A. R. Parker, T. R. Gull, and R. P. Kirschner, *An Emission-Line Survey of the Milky Way*, NASA SP-434 (1979), p. 51.
15. M. Wrigge and H. J. Wendker, Astron. Astrophys. **391**, 287 (2002).

16. I. Lundstrom and B. Stenholm, Astron. Astrophys., Suppl. Ser. **58**, 163 (1984).
17. C. Blaha and R. Humphreys, Astron. J. **98**, 1598 (1989).
18. C. D. Garmany and R. E. Stencel, Astron. Astrophys., Suppl. Ser. **94**, 211 (1992).
19. T. G. Sitnik and A. M. Mel'nik, Pis'ma Astron. Zh. **22**, 471 (1996) [Astron. Lett. **22**, 422 (1996)].
20. H. J. Wendker, L. F. Smith, F. P. Israel, *et al.*, Astron. Astrophys. **42**, 173 (1975).
21. J. Nichols-Bohlin and R. A. Fesen, Astron. J. **105**, 672 (1993).
22. A. P. Marston, Astron. J. **109**, 2257 (1995).
23. L. Bronfman, L.-A. Nyman, and J. May, Astron. Astrophys., Suppl. Ser. **115**, 81 (1996).
24. H. O. Leung and P. Thaddeus, Astrophys. J., Suppl. Ser. **81**, 267 (1992).
25. T. A. Lozinskaya, V. V. Pravdikova, T. G. Sitnik, *et al.*, Astron. Zh. **75**, 514 (1998) [Astron. Rep. **42**, 453 (1998)].
26. V. F. Esipov, T. A. Lozinskaya, V. V. Mel'nikov, *et al.*, Pis'ma Astron. Zh. **22**, 571 (1996) [Astron. Lett. **22**, 509 (1996)].
27. T. A. Lozinskaya, V. V. Pravdikova, I. V. Gosachinskiĭ, and S. A. Trushkin, Astron. Zh. **74**, 376 (1997) [Astron. Rep. **41**, 327 (1997)].
28. T. A. Lozinskaya, V. V. Pravdikova, T. G. Sitnik, *et al.*, Pis'ma Astron. Zh. **23**, 515 (1997) [Astron. Lett. **23**, 450 (1997)].
29. T. G. Sitnik, Pis'ma Astron. Zh. **23**, 896 (1997) [Astron. Lett. **23**, 779 (1997)].
30. T. G. Sitnik, A. G. Mel'nik, and V. V. Pravdikova, Astron. Zh. **78**, 40 (2001) [Astron. Rep. **45**, 34 (2001)].
31. T. A. Lozinskaya, Astron. Zh. **47**, 122 (1970) [Sov. Astron. **14**, 98 (1970)].

Translated by N. Samus'

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 649–658. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 714–725.
Original Russian Text Copyright © 2004 by Abubekerov.

The Mass of the Compact Object in the X-ray Binary 4U 1700-37

M. K. Abubekerov

Moscow State University, Moscow, Russia

Received November 17, 2003; in final form, January 9, 2004

Abstract—The results of a systematic analysis of master radial-velocity curves for the X-ray binary 4U 1700-37 are presented. The dependence of the mass of the X-ray component on the mass of the optical component is derived in a Roche model based on a fit of the master radial-velocity curve. The parameters of the optical star are used to estimate the mass of the compact object in three ways. The masses derived based on information about the surface gravity of the optical companion and various observational data are $2.25^{+0.23}_{-0.24}M_\odot$ and $2.14^{+0.50}_{-0.43}M_\odot$. The masses based on the radius of the optical star, $21.9R_\odot$, are $1.76^{+0.20}_{-0.21}M_\odot$ and $1.65^{+0.78}_{-0.56}M_\odot$. The mass of the optical component derived from the mass–luminosity relation for X-ray binaries, $27.4M_\odot$, yields masses for the compact object of $1.41^{+0.08}_{-0.08}M_\odot$ and $1.35^{+0.18}_{-0.18}M_\odot$. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The X-ray binary 4U 1700-37 was discovered by the UHURU satellite in December 1970 [1]. Subsequent observations revealed eclipses of the X-ray source with a period of 3.412 d. The system was identified with the O6.5Iaf supergiant HD 153919 [2, 3], which displays brightness variations with the same period. The system is 1.8 kpc from the Earth and is made up of a compact object accreting material supplied by an optical companion.

The X-ray radiation of 4U 1700-37 has a hard spectrum that resembles the spectrum of an accreting neutron star [4, 5]. However, no regular pulses associated with the lighthouse effect, which arises during accretion onto a neutron star, are observed. This hinders the construction of a radial-velocity curve for the compact object and, hence, determination of the masses of the binary components.

In the early study [6], a least-squares solution for the observed radial-velocity curve was obtained using a model with two point masses. If the mass of the optical component is $35M_\odot$ and the inclination of the orbit is 90°, the mass of the compact object yielded by this model is $2.4M_\odot$.

Heap and Corcoran [7] estimated the masses of the binary components using the algorithm of [8] together with information about the duration of the eclipses and the radius of the optical component. Based on a mass for the optical component of $52 \pm 2M_\odot$, they obtained a mass for the compact object of $1.8 \pm 0.4M_\odot$. An earlier analysis carried out using the same method [9] yielded a mass for the relativistic component of $1.3M_\odot$ for a mass of the optical component of $27M_\odot$.

The Monte Carlo method used by Jones [10] yields a mass for the compact object of $2.6^{+2.3}_{-1.4}M_\odot$ for a mass of the optical component of $30^{+11}_{-7}M_\odot$.

Clark [11], who also used a Monte Carlo method, derived a mass for the compact object of $2.44 \pm 0.27M_\odot$ for a mass of the optical companion of $58 \pm 11M_\odot$. The modeling in this particular study was not carried out entirely correctly. In place of the half-amplitude of the radial-velocity curve for the optical component, $K_v = 18.7 \pm 1.0$ km/s, corresponding to a circular orbit, Clark [11] used the value $K_v = 20.6 \pm 1.0$ km/s, leading to a slight overestimation of the mass of the compact object.

The uncertainty in the estimated mass of the X-ray component of the binary 4U 1700-37 is obvious. Let us consider the factors hindering accurate estimation of the mass of the relativistic component in this binary system.

First, uncertainty in the mass of the relativistic component in 4U 1700-37 is due to the unknown mass of the optical component, which has only been determined from its spectral type and luminosity class with comparatively large uncertainties.

Second, the Monte Carlo method used for the mass estimates, as well as the method of [7, 9], relied on relations obtained for a point-mass formalism. Models approximating the O supergiant as a point mass of electromagnetic radiation do not take into account a wide range of phenomena associated with the interaction between the components. Therefore, a correct estimation of the mass of the relativistic component is not possible using this type of model

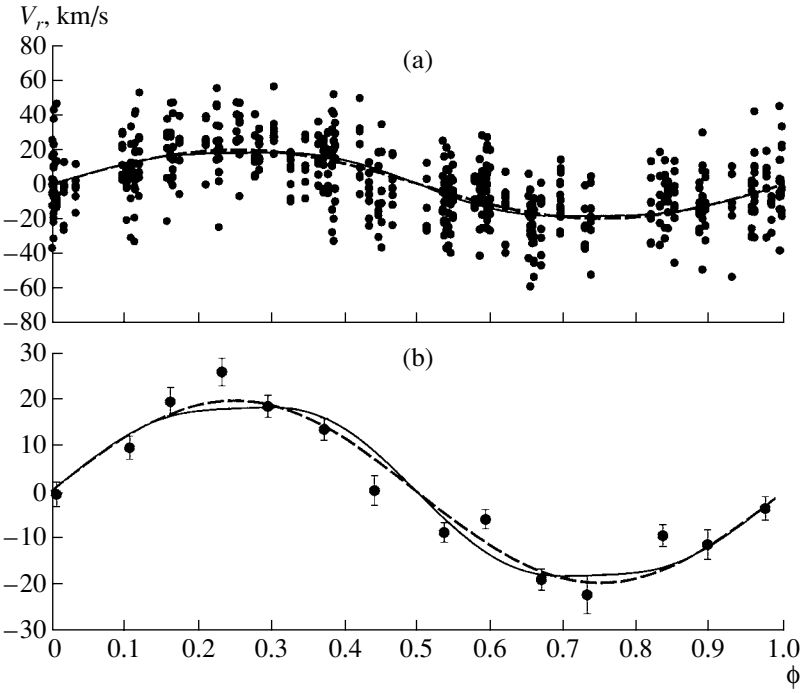

Fig. 1. (a) Master observed radial-velocity curve for the close X-ray binary 4U 1700-37. The filled circles show the radial velocities derived from hydrogen absorption lines [12]. For comparison, we also show theoretical radial-velocity curves for a Roche model (solid) and point-mass model (dashed) for $m_x = 2.11M_\odot$, corresponding to the minimum residual in the Roche model with $m_v = 50M_\odot$ computed using method 2 (excluding the mean radial velocities at phases 0.4–0.6). The parameters of the Roche model are presented in Table 1. (b) Radial velocities averaged within phase intervals (filled circles show the mean radial velocities within given phase bins). For comparison, we also show theoretical radial-velocity curves for the Roche model (solid) and point-mass model (dashed) for $m_x = 2.11M_\odot$ and $m_v = 50M_\odot$.

even if the mass of the optical component is precisely known.

Third, the Monte Carlo method uses the half-amplitude of the radial-velocity curve of the optical component (which is known only to within ∼5–10%), ignoring information contained in the shape of this curve.

It is clear that the collected spectroscopic observations must be interpreted using a more realistic model. A correct estimation of the masses of the components in 4U 1700-37 will represent another step forward in our understanding of the nature of the X-ray source.

2. OBSERVATIONAL MATERIAL

We used the data of [12, 13] to construct a master radial-velocity curve, adopting the center of the eclipse of the X-ray source by the optical component as the zero phase in both cases.

The spectral data of [12] were obtained in 1973–1976 on the 152-cm telescope of the European Southern Obervatory. The radial velocities were determined from the Hβ, Hγ, Hδ, H8, H9, H10, H11, and H12 hydrogen Balmer absorption lines. The radial velocities were measured in terms of the absolute shift of the absorption-line core relative to its laboratory wavelength. Therefore, we needed to correct the spectral data for the systemic velocity before using them in the master radial-velocity curve.

Note that the systemic velocity of the stellar system determined from the hydrogen Balmer absorption lines grows with decreasing line number in the series—the so-called Balmer progression [14, 15]. For example, the systemic radial velocities for the Hδ, Hγ, and Hβ lines are −82.7, −110.3, and −152.2 km/s, respectively. The increase in the systemic radial velocity with decreasing line number in the series of absorption lines is associated with the fact that the cores of the lines at the beginning of the series are formed in higher layers of the stellar atmosphere, where a radial outflow has already developed in the form of a stellar wind. The systemic velocities obtained using the higher-order Balmer H10, H11, and H12 absorption lines are −72.7, −83.3, and −78.4 km/s, respectively. Since these lines form at the base of the photosphere of the optical star and do not experience strong perturbation by the stellar wind, their mean value, −78.1 km/s, can be considered to be the γ velocity of the 4U 1700-37 system.

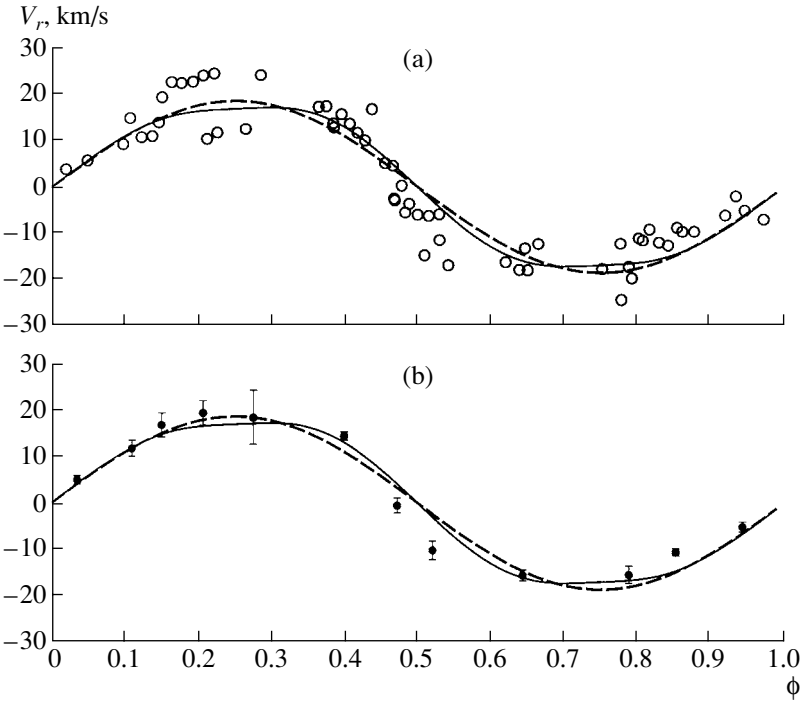

Fig. 2. (a) Same as Fig. 1a but with open circles showing radial velocities determined from the IUE spectral data of [13] and theoretical radial-velocity curves shown for $m_x = 2.01M_\odot$ and $m_v = 50M_\odot$. (b) Same as Fig. 1b but with theoretical radial-velocity curves shown for $m_x = 2.01M_\odot$ and $m_v = 50M_\odot$.

The data of [12] gave us 570 radial-velocity values distributed approximately uniformly in phase. The master radial-velocity curve is presented in Fig. 1a.

The IUE satellite has taken more than 60 spectra of 4U 1700-37. The results of these high-accuracy spectral observations are presented in [13]. The radial-velocity measurements in this case were obtained by cross correlating intervals of the spectrum in the range 1300–1850 Å. The data of [13] gave us 61 radial velocities; the corresponding radial-velocity curve is presented in Fig. 2a.

The complex processes occurring at the surface of the optical component and in various gaseous structures in the binary hinder determination of the intrinsic radial velocity of the optical component. The observed radial velocities are unavoidably distorted by random errors. We averaged the radial velocities within phase intervals in order to decrease the influence of such random errors. These mean observed radial-velocity curves are presented in Figs. 1b and 2b. Note that, in light of the recent study [16], the errors introduced into the observed radial velocity of the optical star by tidal–gravitational waves should also be random and can be suppressed by averaging over many nights of observations.

3. INTERPRETATION OF THE RADIAL-VELOCITY CURVES

The optical component in the close binary 4U 1700-37 is close to filling its critical Roche lobe [17]. Due to the tidal action of the relativistic component, the shape of the optical star is not spherical. The side facing the relativistic component is heated by X-ray radiation from this star. These effects associated with the interaction of the components must be taken into consideration when interpreting the observed radial-velocity curve of 4U 1700-37. Therefore, the mean radial velocities should be interpreted using a Roche model, thereby making it possible in a first approximation to include these effects. A detailed description of the Roche model is given in [18, 19], and we do not present this information again here. The numerical values of the Roche-model parameters for 4U 1700-37 are presented in Table 1.

The masses of both components were treated as unknown parameters. Since the process of finding a best fit is fairly cumbersome in this case, we adopted an approach in which we carried out a series of exhaustive parameter searches. For each mass of the optical component from the discrete set $m_v = 20, 30, 40, 50, 58,$ and $70M_\odot$, we carried out an exhaustive search for the optimum mass of the compact object, m_x. This process yielded a

Table 1. Numerical values of parameters used to synthesize radial-velocity curves for 4U 1700-37 in the Roche model

P, days	3.411581	Period
e	0.0	Eccentricity
ω, deg	0.0	Longitude of periastron of the optical component
i, deg	67*	Orbital inclination
μ	0.93*	Roche-lobe filling factor for the optical component at periastron
f	0.9	Rotation-asynchronization coefficient
$T_{\rm eff}$, K	36 000**	Effective temperature of the optical component
β	0.25	Gravitational-darkening coefficient
k_x	0.0005	Ratio of the X-ray luminosity of the relativistic component to the bolometric luminosity of the optical component, L_x/L_v
A	0.5	Coefficient for reprocessing of the X-ray radiation
u	0.3***	Limb-darkening coefficient

* Data taken from [17].
** Data taken from [11].
*** Data taken from [20].

relationship between the masses of the compact and optical components.

The residual between the mean observed radial-velocity curve and the theoretical curve was calculated using the formula

$$\Delta(m_x) = \frac{\sum\limits_{j=1}^{M}(n_j-1)}{M}\,\frac{\sum\limits_{j=1}^{M} n_j(V_j^{teor}-\bar{V}_j^{obs})^2}{\sum\limits_{j=1}^{M} n_j(n_j-1)\sigma_j^2}, \quad (1)$$

where $\bar{V}_j^{obs}$ is the observed mean radial velocity in a phase interval centered on phase $\bar{\phi}_j$, V_j^{teor} is the theoretical radial velocity at this phase, σ_j is the rms deviation of $\bar{V}_j^{obs}$ from the observed radial velocities in the phase interval centered on phase $\bar{\phi}_j$, M is the number of phase intervals, and n_j is the number of averaged observed radial velocities in this phase interval.

The quantity $\Delta(m_x)$ is distributed according to a Fisher law, $F_{M,\sum\limits_{j=1}^{M}(n_j-1),\alpha}$ [21]. We can therefore find the confidence set for the parameter m_x for a given value of m_v corresponding to a specified significance level α. This consists of the three values of m_x for which

$$\Delta(m_x) \le F_{M,\sum\limits_{j=1}^{M}(n_j-1),\alpha}.$$

We obtained fits for both a Roche model and a point-mass model. The latter model served to identify divergences of the two model results. The fitting of the mean radial-velocity curve was carried out separately for the spectral data of [12] and [13].

We should say a few words about the stellar wind of the optical component in the 4U 1700-37 system. The optical star is an O supergiant. Nonuniformity in the gravitational forces at its surface and heating of the surface facing the relativistic companion disrupts the isotropy of the stellar-wind outflow. The wind speed near the Lagrange point L_1 grows, manifest as an excess negative radial velocity near phase 0.5 (when the X-ray source is in front of the O supergiant; Figs. 1b and 2b). The anisotropy of the stellar wind introduces systematic errors in the observed radial-velocity curve. Therefore, we obtained fits to the mean observed radial-velocity curves in two ways:

Method 1. Using all the mean observed radial velocities.

Method 2. Excluding the mean observed radial velocities at phases 0.4–0.6, where the distorting effect of the anisotropic stellar wind is strongest.

We chose the 5% significance level as the critical level for our work. The fits obtained for the spectral data of [12, 13] using method 1 were not acceptable at this level for either the Roche model or the point-mass model. The behavior of the residuals for the Roche-model fit to the data of [12, 13] are presented in Figs. 3a and 3b. It is not possible to construct a relation between the masses of the components in this case.

Fitting the mean observed radial velocities while indirectly taking into account the anisotropy of the

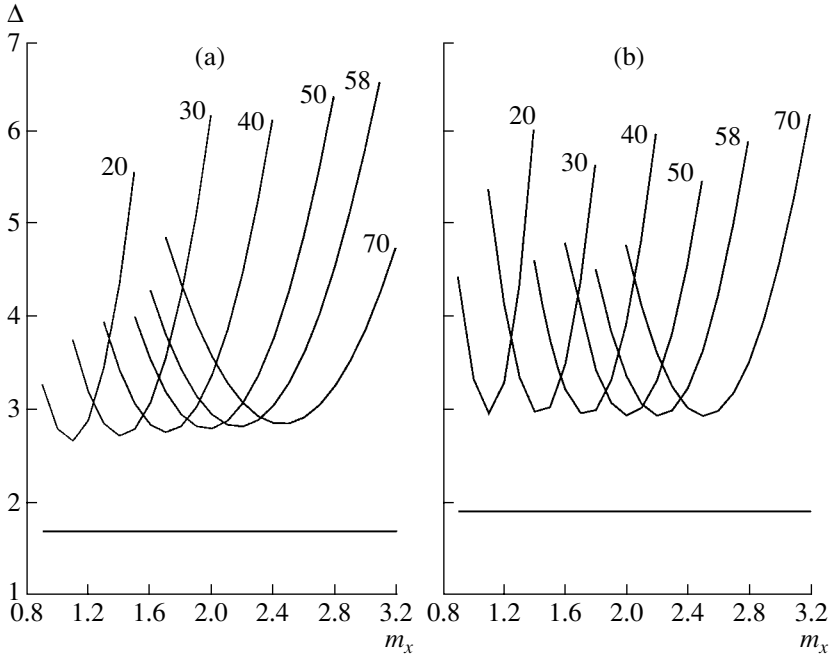

Fig. 3. (a) Residuals between the mean observed radial-velocity curve for 4U 1700-37 for the data of [12] and the curve synthesized in the Roche model, obtained using method 1 (i.e., using all of the mean radial velocities). The horizontal line corresponds to the critical Fisher criterion $\Delta_{14,570} = 1.69$ for the 5% significance level. The masses of the optical component in solar masses for which the residuals were obtained are indicated near the curves. (b) Same for the data of [13]. The horizontal line corresponds to the critical Fisher criterion $\Delta_{12,61} = 1.916$ for the 5% level.

wind (method 2) enabled us to obtain an acceptable model at the 5% significance level. This indicates the importance of including the anisotropy of the stellar wind when interpreting the radial-velocity curves of OB stars in close binary systems [22]. Figures 4a and 4b show the residuals obtained for the Roche-model fits to the mean observed radial-velocity curves. We used the results of these fits to construct the dependence of the mass of the X-ray component on the mass of the optical component (Figs. 5a and 5b). For a mass of the optical component of $69M_\odot$, the minimum residual in the analysis using the data of [12], achieved for $m_x = 2.61M_\odot$, is equal to the quantile. Therefore, the error corridor for the component-mass dependence in Fig. 5a is cut off at $m_v = 69M_\odot$. The numerical results of our fitting of the data of [12] and [13] using point-mass and Roche models are given in Tables 2 and 3.

We can see from these tables that the masses of the compact object obtained in the Roche and point-mass models are similar. This can be explained in two ways. First, due to the small gravitational force near the Lagrange point L_1, the temperature of the "nose" of the optical star is lower than that of its remaining surface (gravitational darkening). Second, the heating of the part of the optical star facing the X-ray source is very low ($k_x = 0.0005$). Therefore, the contribution of the nose, which introduces the largest perturbations into the observed radial-velocity curve, to the integrated radiation of the optical component is small. The similarity of the results for the Roche and point-mass models comes about because radiation of the optical component is dominated by the spherically symmetric part of its surface. The influence of the heating coefficient on the shape of the radial-velocity curve is discussed in more detail in [18].

The relations between the masses of the optical and X-ray components obtained using the data of [12] and [13] are in good agreement with each other (Figs. 5a and 5b, Tables 2 and 3).

Because the orbital inclination is not accurately known, we also obtained fits applying method 2 to the spectral data of [13] for inclinations of 62° and 72°, with the remaining parameters for the Roche model taken to be the same as before (Table 1). The numerical results are presented in column 1 in Tables 4 and 5, and the results are shown graphically in Figs. 6a and 6b.

The dependence of the mass of the compact object on the orbital inclination in the point-mass model is given by the relation $m_v \sim \sin^{-3} i$. We carried out a test similar to that of [23] to verify how accurately this relation was obeyed for the Roche model.

The mass of the compact object derived in the Roche model for an orbital inclination $i = 67°$ was

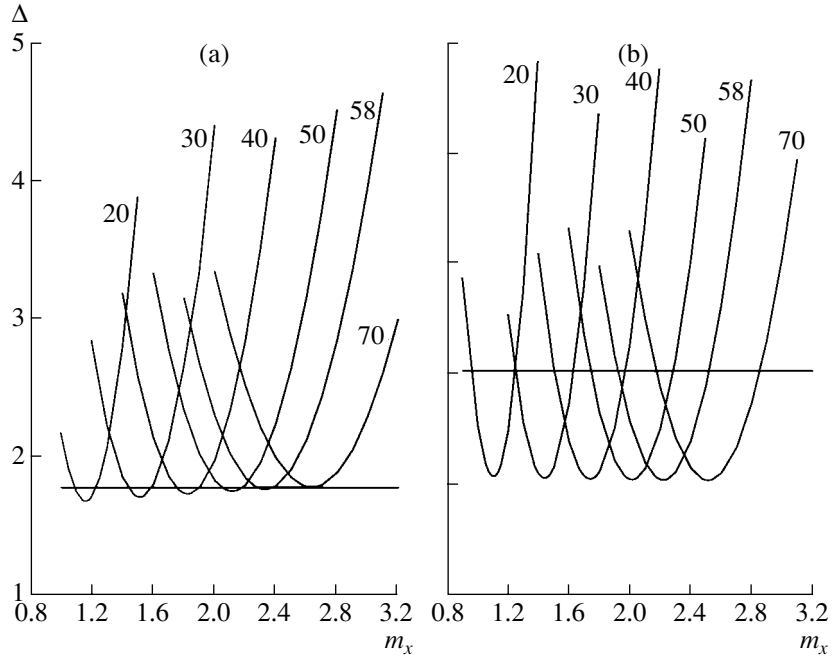

Fig. 4. (a) Residuals between the mean observed radial-velocity curve for 4U 1700-37 for the data of [12] and the curve synthesized in the Roche model obtained using method 2 (i.e., excluding the mean radial velocities at phases 0.4–0.6). The horizontal line corresponds to the critical Fisher criterion $\Delta_{11,428} = 1.79$ for the 5% significance level. The masses of the optical component in solar masses for which the residuals were obtained are indicated near the curves. (b) Same for the data of [13]. The horizontal line corresponds to the critical Fisher criterion $\Delta_{10,48} = 2.03$ for the 5% level.

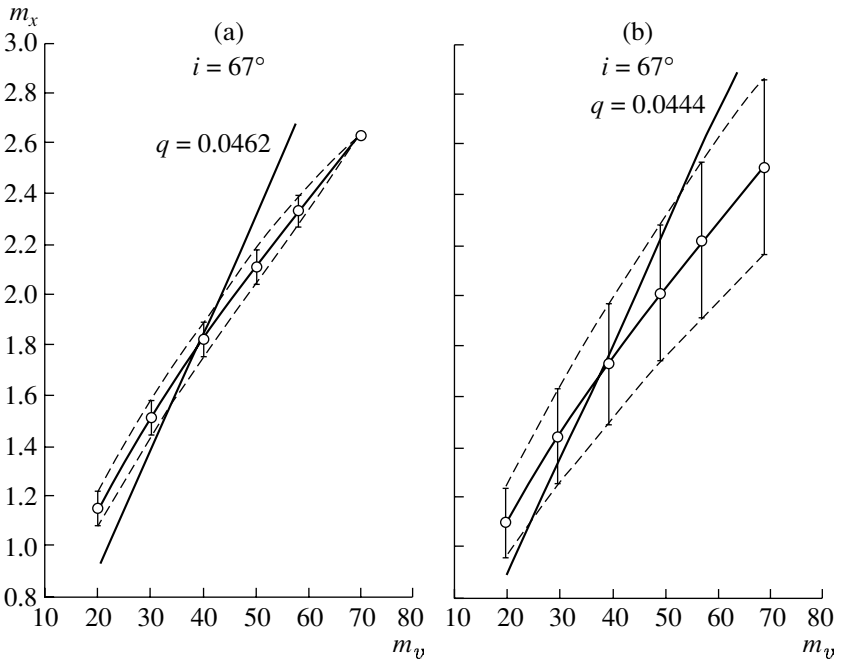

Fig. 5. (a) Dependence of the mass of the compact object in 4U 1700-37 on the mass of the optical star for an orbital inclination of 67°. The fit was obtained in a Roche model for the radial velocities determined from hydrogen Balmer absorption lines in [12] using method 2 (i.e., excluding the mean radial velocities at phases 0.4–0.6). The line corresponds to the relation between m_x and m_v for the component-mass ratio $q = 0.0462$ found from (4). (b) Same for the IUE data of [13]. The line corresponds to the relation between m_x and m_v for the component-mass ratio $q = 0.0444$ found from (4).

Table 2. Mass of the relativistic component of 4U 1700-37 obtained from the fit of the data of [12] in Roche ("R") and point-mass ("P") models

m_v, $M_\odot$	m_x(P), $M_\odot$	m_x(R), $M_\odot$
20	$1.15^{+0.13}_{-0.12}$	$1.15^{+0.08}_{-0.06}$
30	$1.51^{+0.16}_{-0.16}$	$1.51^{+0.08}_{-0.07}$
40	$1.82^{+0.20}_{-0.20}$	$1.82^{+0.08}_{-0.07}$
50	$2.11^{+0.23}_{-0.23}$	$2.11^{+0.07}_{-0.06}$
58	$2.32^{+0.26}_{-0.25}$	$2.33^{+0.05}_{-0.06}$
70	$2.64^{+0.28}_{-0.30}$	2.64*

* Confidence interval not indicated, since the model is rejected at the 5% significance level.

Table 3. Mass of the relativistic component of 4U 1700-37 obtained from the fit of the data of [13] in Roche ("R") and point-mass ("P") models

m_v, $M_\odot$	m_x(P), $M_\odot$	m_x(R), $M_\odot$
20	$1.09^{+0.10}_{-0.10}$	$1.10^{+0.15}_{-0.14}$
30	$1.42^{+0.13}_{-0.13}$	$1.44^{+0.19}_{-0.19}$
40	$1.72^{+0.16}_{-0.16}$	$1.73^{+0.25}_{-0.23}$
50	$1.99^{+0.18}_{-0.19}$	$2.01^{+0.28}_{-0.27}$
58	$2.19^{+0.21}_{-0.20}$	$2.22^{+0.31}_{-0.31}$
70	$2.49^{+0.23}_{-0.24}$	$2.51^{+0.35}_{-0.34}$

recalculated for $i = 62°$ and $i = 72°$ using formulas (2) and (3):

$$m_x(62°) = m_x(67°)\frac{\sin^3 67°}{\sin^3 62°}, \qquad (2)$$

$$m_x(72°) = m_x(67°)\frac{\sin^3 67°}{\sin^3 72°}. \qquad (3)$$

Here, $m_x(62°)$, $m_x(67°)$, and $m_x(72°)$ are the masses of the relativistic component for inclinations of 62°, 67°, and 72°. The results of recalculating the mass of the compact object using (2) and (3) are presented in the second columns of Tables 4 and 5.

A comparison of the first and second columns in Tables 4 and 5 shows that the relation $m_v \sim \sin^{-3} i$ yields results that are close to, but not strictly coincident with, those for the Roche model. Nevertheless, the masses agree within the errors (Figs. 6a and 6b). Therefore, if the inclination i is refined, the mass of the compact object m_x can be approximately recalculated using the relation $m_x \sim \sin^{-3} i$. A more precise determination of the component masses for specified values of i can be carried out by interpolating the results of Tables 2–5 (see also Figs. 5 and 6).

4. DETERMINATION OF THE COMPONENT MASSES

Based on the Gravitational Acceleration of the Optical Component

A detailed non-LTE analysis of spectra of the optical star in the 4U 1700-37 system was carried out in [11]. The gravitational acceleration $\log g$ derived from the wings of Balmer and HeI absorption lines in the spectrum of the optical component, HD 153919, lies in the range 3.45–3.55 [11]. The bolometric luminosity and effective temperature of the optical component were found to be $\log(L/L_\odot) = 5.82$ and $35\,000 \pm 1000$ K, yielding a radius of $21.9R_\odot$. Using the values $\log g = 3.45$–3.55 and $R_v = 21.9R_\odot$, we obtain a lower limit for the mass of the optical star of $55^{+7}_{-6}M_\odot$. The mass of the optical star derived from the observed free-fall acceleration will be a lower limit, since the true acceleration will be greater due to the action of centrifugal forces. Clark [11] suggested that the lower limit for the mass of the optical star is close to $50M_\odot$, while the upper limit, derived from the position of HD 153919 in the Hertzsprung–Russell diagram, is no more than $60M_\odot$. The mass of the compact object corresponding to a mass for the optical star of $55^{+7}_{-6}M_\odot$ and an inclination of 67° based on the relation between the component masses constructed using the spectral data of [12] is $2.25^{+0.23}_{-0.24}M_\odot$ (Fig. 5a). The relation between the component masses constructed for the IUE spectral data of [13] yields a mass for the relativistic component of $2.14^{+0.50}_{-0.43}M_\odot$ (Fig. 5b).

The masses of the relativistic component obtained from the component-mass relation based on the IUE spectral data of [13] are $2.26^{+0.49}_{-0.44}M_\odot$ and $2.06^{+0.47}_{-0.39}M_\odot$ for inclinations of 62° and 72°, respectively (Fig. 6b).

Based on the Radius of the Optical Component

The radius of the optical component R_v, Roche-lobe filling factor μ, eccentricity e, inclination i, orbital period P, mass function for the optical component $f_v(m)$, and component-mass ratio $q = m_x/m_v$ are related by (4) (see, for example, [17]):

$$\sin i = \frac{0.38\mu}{R_v}\left(\frac{GP^2 f_v(m)}{4\pi^2}\right)^{1/3}\frac{1+q}{q^{1.208}}, \qquad (4)$$

where $f_v(m)$ is defined to be

$$f_v(m) = \frac{P(1-e^2)^{3/2}}{2\pi G}K_v^3.$$

Table 4. Column 1: Mass of the relativistic component obtained for the Roche model, method 2, the data of [13], and orbital inclination $i = 62°$. Column 2: Result of using (2) to recalculate the mass of the compact object obtained for the Roche model and $i = 67°$ (see text for details)

$m_v, M_\odot$	$m_x, M_\odot$	
	1	2
20	$1.16^{+0.15}_{-0.15}$	1.25
30	$1.51^{+0.20}_{-0.20}$	1.63
40	$1.82^{+0.24}_{-0.24}$	1.96
50	$2.11^{+0.28}_{-0.28}$	2.28
58	$2.33^{+0.30}_{-0.31}$	2.52
70	$2.63^{+0.35}_{-0.34}$	2.84

Table 5. Column 1: Mass of the relativistic component obtained for the Roche model, method 2, the data of [13], and orbital inclination $i = 72°$. Column 2: Result of using (3) to recalculate the mass of the compact object obtained for the Roche model and $i = 67°$ (see text for details)

$m_v, M_\odot$	$m_x, M_\odot$	
	1	2
20	$1.06^{+0.15}_{-0.14}$	1.00
30	$1.39^{+0.19}_{-0.20}$	1.31
40	$1.67^{+0.24}_{-0.23}$	1.57
50	$1.94^{+0.27}_{-0.27}$	1.82
58	$2.13^{+0.30}_{-0.29}$	2.01
70	$2.51^{+0.33}_{-0.33}$	2.28

Consequently, if we have information about the radius of the optical component and the remaining parameters of the binary system in (4), we can determine the component-mass ratio q.

The values for μ, e, P, and i were taken from Table 1. We carried out a search for the velocity of the center of mass of the optical component K_v in a point-mass model. We constructed a radial-velocity curve for a point-mass model for each pair of component masses determined in the Roche model (Tables 2 and 3). Further, we determined the mean value of K_v. The results for the spectral data of [12] and [13] were $K_v = 19.72 \pm 0.02$ km/s and $K_v = 18.81 \pm 0.01$ km/s, respectively.

We can obtain an equation for q by substituting the radius of the optical star $R_v = 21.9R_\odot$ [11] into (4). Solution of (4) using the value $K_v = 19.72$ km/s based on the data of [12] yields the component-mass ratio $q = 0.0462$; using the value $K_v = 18.81$ km/s based on the data of [13] yields $q = 0.0444$.

We show the corresponding curves on the component-mass relations in Figs. 5a and 5b. The intersection of a line with the region of allowed masses identifies the corresponding mass of the compact object. The resulting masses for the relativistic component based on the data of [12] and [13] are $1.76^{+0.20}_{-0.21}M_\odot$ (Fig. 5a) and $1.65^{+0.78}_{-0.56}M_\odot$ (Fig. 5b).

Based on the Mass–Luminosity Relation

The mass of the optical component determined from the mean linear mass–luminosity relation for main-sequence stars in noninteracting binaries is $49.5M_\odot$ [24]. Using the relations between the component masses based on the data of [12] and [13], this value corresponds to masses for the compact object of $2.08^{+0.07}_{-0.07}M_\odot$ (Fig. 5a) and $1.98^{+0.27}_{-0.26}M_\odot$ (Fig. 5b).

However, the mass–luminosity relation for the optical components of X-ray binaries differs from the corresponding relation for single stars [25]. When the Roche lobe is filled or there is an intense stellar wind, the optical star in a close binary loses the upper layers of its atmosphere, so that the temperature of its surface and its luminosity are higher than for an isolated star of the same mass. The observed overabundances of carbon and nitrogen on the surface of the optical star in the 4U 1700-37 system [11] confirm that this star has lost the upper layers of its atmosphere. Figure 7 presents a mass–luminosity relation for OB stars in X-ray binaries based on the data of [26] and [27]. We can see that the luminosity of the optical component, $\log(L/L_\odot) = 5.82$ [11], corresponds to a mass of $27.4M_\odot$. For this mass, the masses of the relativistic component that are implied by the component-mass relations constructed using the data of [12] and [13] are $1.41^{+0.08}_{-0.08}M_\odot$ (Fig. 5a) and $1.35^{+0.18}_{-0.18}M_\odot$ (Fig. 5b).

Thus, based on the mass–luminosity relation for OB stars in X-ray binaries, the mass of the relativistic object in the 4U 1700-37 system agrees within the errors with the mean mass of neutron stars, $1.35^{+0.04}_{-0.04}M_\odot$ [28].

5. CONCLUSIONS

Our main result is the relationships between the masses of the optical and relativistic components in the X-ray binary system 4U 1700-37 obtained for orbital inclinations $i = 62°$, $67°$, and $72°$ (Figs. 5 and 6). We have also estimated the mass of the compact object in 4U 1700-37 based on the known parameters of the optical component. It has not been conclusively demonstrated whether the compact object in

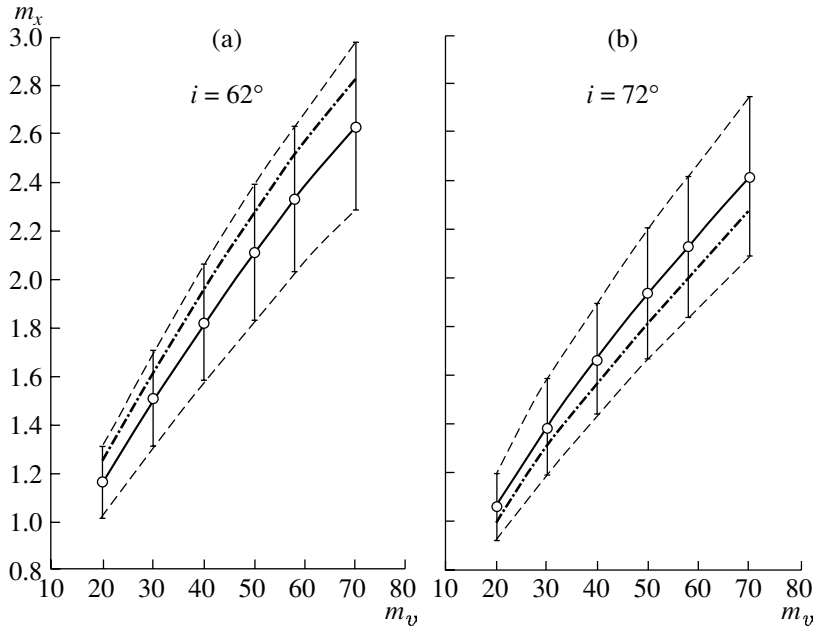

Fig. 6. (a) Dependence of the mass of the compact object in 4U 1700-37 on the mass of the optical star for $i = 62°$. The fit was obtained for the Roche model for the data of [13] using method 2 (i.e., excluding observed radial velocities at phases 0.4–0.6). The dot-dashed line corresponds to the masses of the compact object obtained by using (2) to recalculate the masses obtained for the Roche model and $i = 67°$. (b) Same for $i = 72°$. The dot-dashed line corresponds to the masses of the compact object obtained by using (3) to recalculate the masses obtained for the Roche model and $i = 67°$.

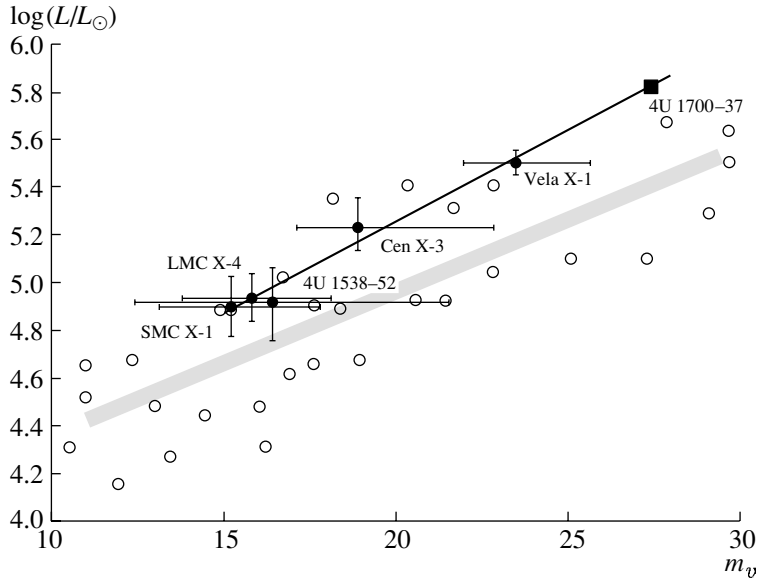

Fig. 7. Mass–luminosity relation for OB supergiants in X-ray binary systems (thin, dark line). The filled circles mark the positions of OB giants according to the data of [26, 27], with the names of the X-ray binary systems indicated nearby. The filled square marks the position of the optical component of the 4U 1700-37 system based on the luminosity $\log(L/L_\odot) = 5.82$ [11]. The hollow circles mark the positions of OB supergiants in noninteracting binaries on the mass–luminosity relation according to the data of [24]; the wide, gray line shows a least-squares linear approximation to this relation.

the 4U 1700-37 system is a low-mass black hole or a neutron star. Our results support the idea that it is a neutron star.

The masses for the relativistic component obtained using the mass–luminosity relation for OB stars in X-ray binaries are $1.41^{+0.08}_{-0.08} M_\odot$ and $1.35^{+0.18}_{-0.18} M_\odot$, which agrees within the errors with

the mean mass of neutron stars, $1.35^{+0.04}_{-0.04}M_\odot$ [28]. The masses of the compact object obtained based on the surface gravitational acceleration of the optical component are $2.25^{+0.23}_{-0.24}M_\odot$ and $2.14^{+0.50}_{-0.43}M_\odot$, while those based on the radius of the optical component are $1.76^{+0.20}_{-0.21}M_\odot$ and $1.65^{+0.78}_{-0.56}M_\odot$. These estimates for the mass of the compact object in 4U 1700-37 coincide within the errors with the mass of the Vela X-1 X-ray pulsar, $1.93^{+0.19}_{-0.21}M_\odot$ [23]. It may be that, similar to the situation with Vela X-1, the compact object in 4U 1700-37 is a massive neutron star. The hard X-ray spectrum of 4U 1700-37, which resembles the spectra of X-ray pulsars [4, 5], also provides evidence that the compact object is a neutron star. The absence of periodic X-ray pulsar from 4U 1700-37 associated with the rotation of an accreting neutron star can be understood if the magnetic-dipole axis and rotational axis are coincident.

It is not currently possible to give an unambiguous estimate of the mass of the relativistic component in 4U 1700-37. Additional information about the optical component is required to refine the mass of the relativistic component. For example, the space astrometric project GAIA planned by the European Southern Observatory will measure the trigonometric parallaxes and distances to millions of stars in the Galaxy. Knowledge of the distance to the 4U 1700-37 system will make it possible to obtain a direct estimate of the radius of the optical star, which can, in turn, be used to derive a trustworthy estimate of the mass of the compact object. Further careful searches for phenomena associated with the possible presence of an X-ray pulsar in 4U 1700-37 can also play an important role in elucidating the nature of this system.

ACKNOWLEDGMENTS

This work was supported by a grant from the program "Leading Scientific Schools of Russia" (HSh-388.2003.2). The author thanks A.M. Cherepashchuk, É.A. Antokhina, and V.M. Lipunov for useful discussions and advice, as well as Dr. G. Hammerschlag-Hensberge, who kindly presented his spectral data for this study.

REFERENCES

1. C. Jones *et al.*, Astrophys. J. **781**, 64 (1973).
2. C. Jones *et al.*, Bull. Am. Astron. Soc. **5**, 313 (1972).
3. C. Jones and W. Lillei, IAU Circ., No. 2503 (1973).
4. A. P. Reynolds *et al.*, Astron. Astrophys. **349**, 873 (1999).
5. L. Kaper and A. Cherepashchuk, in *Black Holes in Binaries and Galactic Nuclear: Diagnostic, Demography and Formation,* Ed. by L. Kaper, E. P. J. van Heuvel, and P. A. Noudt (Springer, Berlin, 2001), p. 289.
6. J. B. Hutchings, A. D. Thackeray, B. L. Webster, and P. J. Andrews, Mon. Not. R. Astron. Soc. **163**, 13 (1973).
7. S. R. Heap and M. F. Corcoran, Astrophys. J. **387**, 340 (1992).
8. R. G. Aitken, *The Binary Stars* (Dover, New York, 1964).
9. P. S. Conti and A. P. Cowley, Astrophys. J. **200**, 133 (1975).
10. C. Jones *et al.*, Astrophys. J. **459**, 259 (1996).
11. J. S. Clark *et al.*, Astron. Astrophys. **392**, 909 (2002).
12. G. Hammerschlag-Hensberge, C. De Loore, and van den Heuvel, Astron. Astrophys., Suppl. Ser. **32**, 375 (1978).
13. G. Hammerschlag-Hensberge, M. H. van Kerkwijk, and L. Kaper, Astron. Astrophys. **407**, 685 (2003).
14. J. B. Hutchings, Astrophys. J. **235**, 413 (1980).
15. D. Crampton, J. B. Hutchings, and A. P. Cowley, Astrophys. J. **299**, 839 (1985).
16. H. Quantrell, A. J. Norton, T. D. C. Ash, *et al.,* Astron. Astrophys. **401**, 313 (2003).
17. A. V. Goncharskiĭ, S. Yu. Romanov, and A. M. Cherepashchuk, *Finite-Parametric Inverse Problems* (Mosk. Gos. Univ., Moscow, 1991), pp. 99, 107 [in Russian].
18. E. A. Antokhina and A. M. Cherepashchuk, Astron. Zh. **71**, 420 (1994) [Astron. Rep. **38**, 367 (1994)].
19. E. A. Antokhina, Astron. Zh. **73**, 532 (1996) [Astron. Rep. **40**, 483 (1996)].
20. A. A. Rubashevskiĭ, Astron. Zh. **68**, 799 (1991) [Sov. Astron. **35**, 626 (1991)].
21. D. Hudson, *Statistics. Lectures on Elementary Statistics and Probability* (Geneva, 1964; Mir, Moscow, 1970).
22. M. Milgrom, Astron. Astrophys. **70**, 763 (1978).
23. M. K. Abubekerov, E. A. Antokhina, and A. M. Cherepashchuk, Astron. Zh. **81**, 108 (2004) [Astron. Rep. **48**, 89 (2004)].
24. A. Herrero, *IAU Symposium No. 212: A Massive Star Odyssei: From Main Sequence to Supernova*, Ed. by K. van der Hucht, A. Herrero, and C. Esteban (Publ. Astron. Soc. Pac., 2003), p. 3.
25. J. Ziolkowski, *Nonstationary Evolution of Close Binaries*, Ed. by A. N. Zitkov (PWN, Warsaw, 1978), p. 29.
26. A. M. Cherepashchuk, N. A. Katysheva, T. S. Khruzina, and C. Yu. Shugarov, *Highly Evolved Close Binary Stars: Catalog* (Netherland Gordon and Breach Sci., 1996), Vol. 1, Part 1, p. 82.
27. M. H. van Kerkwijk, J. van Paradijs, and E. J. Zuiderwijk, Astron. Astrophys. **303**, 497 (1995).
28. S. E. Thoresett and D. Chakrabarty, Astrophys. J. **512**, 288 (1999).

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 659–664. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 726–731.
Original Russian Text Copyright © 2004 by Matveyenko.

Profile of the H_2O Supermaser Emission: Structure of the Ejector Region in Orion KL

L. I. Matveyenko

Space Research Institute, Russian Academy of Sciences, Profsoyuznaya ul. 84/32, Moscow, 117810 Russia

Received November 20, 2003; in final form, March 15, 2004

Abstract—The structure of the ejector region in the active star-forming region Orion KL has been studied over a broad dynamic range with a high angular resolution of 0.1 milliarcsec, or 0.05 AU. The line profile of the H_2O supermaser emission has broad wings and can be represented as a superposition of two Gaussians with frequency widths $\Delta f_1 = 31$ kHz and $\Delta f_2 = 163$ kHz. The line intensities are $I_1 \approx 3 \times 10^5$ Jy/beam and $I_2 \approx 400$ Jy/beam, and the brightness temperatures, $T_{b1} \approx 5 \times 10^{16}$ K and $T_{b2} \approx 6 \times 10^{14}$ K. The broadband ejector emission is determined by a rotating bipolar outflow with a rotational period of 5 months. The ejector emission in the 31-kHz component at a velocity of 7.64 km/s is amplified by more than two orders of magnitude by the surrounding envelope. The maser amplification regime is partially saturated.
© 2004 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

Powerful water-vapor maser emission accompanies processes associated with the formation of stars and planetary systems and is a sensitive tracer of such ongoing processes. The emission of a maser source is determined by the physical conditions, geometry, pumping level, and kinetic temperature. In contrast to laboratory masers, cosmic masers have no reflecting mirrors and the emission propagates in only one direction. Of special interest are flares of supermaser H_2O emission. Flares were first observed in the object W49 [1], then in Orion KL [2]. Certain difficulties are associated with our understanding of their nature and, in particular, their triggering mechanism. Some progress in this direction has been achieved due to measurements of the hyperfine spatial structure of the flares. The H_2O supermaser emission region in Orion KL has been studied in detail over a broad dynamic range using the global VLBI network and VLBA [3, 7–9]. A compact source—the ejector of the bipolar outflow—has been detected. The brightness temperatures of the components of the structure reached $T_b = 10^{17}$ K. Below, we consider the results of measurements of the spectral line profile and of the ejector structure in the active star-forming region in Orion KL with very high angular resolution.

2. FLARE OF THE H_2O MASER EMISSION

The high level of activity in one of the star-forming regions in Orion KL in 1979–1987 was accompanied by flares of powerful water-vapor maser emission at a velocity of $v \approx 8$ km/s. The flux density of some flares reached 8 MJy [2, 3]. In the subsequent period of quiescence the flux density did not exceed 1 kJy. In 1997, a second period of activity started: the emission began to increase, reaching $F \approx 14$ kJy by the end of the year. Subsequently, the flux density of the flare emission grew exponentially, reaching 4.3 MJy in October 1998. The line profile was basically Gaussian with width $\Delta V \approx 0.53$ km/s. Figure 1 shows the line profile of the flare with a resolution of 0.053 km/s per channel (epoch 1998.7). In November, a general exponential decay began, and the emission returned to its former level in May and June 1999 [4–6]. The supermaser emission considerably exceeded the instrumental noise, enabling us to observe with very high angular resolution and to analyze in detail the line profile over a broad dynamic range. We have studied the hyperfine structure of the flare region throughout the activity period with an angular resolution of 0.1 mas, i.e., 0.05 AU [7]. We have found that the development of the activity was accompanied by substantial changes of the structure of the emission region.

3. STRUCTURE OF THE FLARE REGION

The structure of the region of the supermaser emission flare is complex and lies within an area several AU in size. Figure 2 shows maps of the region at epochs corresponding to various activity levels. The date is indicated at the top, and the size of the synthesized beam is shown in the lower left corners of the maps. In the stage of increasing flare emission in March to October 1998, as well as in the period of

1063-7729/04/4808-0659$26.00 © 2004 MAIK "Nauka/Interperiodica"

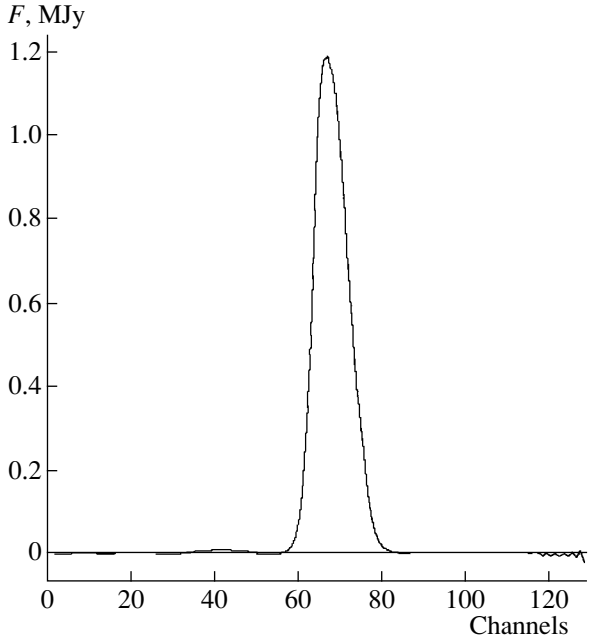

Fig. 1. Line profile of the H_2O maser emission in Orion KL during its high activity (epoch 1998.7). The emission flux density is in MJy and the channel spacing is 0.053 km/s.

quiescence in May 1995, the structure of the radiating region represented a compact core and two-sided jet—bipolar outflow. The visible size of the jet was $\sim 8 \times 0.15$ mas, while that of the core was ~ 0.1 mas. However, the brightness of the components increased considerably compared to their values in the period of quiescence. The brightness temperature of the jet increased by three orders of magnitude, reaching $T_{jet} \approx 2 \times 10^{15}$ K, and the brightness temperature of the central compact core–ejector—increased by four orders of magnitude ($T_b \approx 5 \times 10^{16}$ K). The radial velocities of the southeast and northwest parts of the jet with respect to the ejector differ by $\Delta v \approx \pm 0.3$ km/s. Bright, compact components were observed in the jet itself, whose projected velocity in the plane of the sky reached $v \approx 6$ km/s [8]. The structure of the flare region during the decrease of the activity between January 23, 1999, and April 24, 1999, underwent substantial changes. In January 1999, two components separated by 0.8 milliarcseconds (mas) in position angle $X \approx 4^\circ$ and located symmetrically about the core appeared in the central region. In August, a toroidal structure emerged around the ejector (Fig. 2). The brightness temperature of the central compact source–ejector—is $T_b \approx 2 \times 10^{16}$ K, and its radial velocity in the emission line is $v_o \approx 7.65$ km/s.

Thus, the line profile of the H_2O maser emission flare (Fig. 1) was determined by the complex spatial structure of the star-forming region and its physical conditions. The center of the structure is a bright compact source–the ejector of the bipolar outflow. The core size does not exceed 0.1 mas, or 0.045 AU, and its brightness temperature is $T_b \approx 10^{17}$ K. The thin, extended components-ejected outflows—are 0.05×0.2 AU in size at the nozzle exit, and their brightness temperature is $T_b \approx 10^{17}$ K. The broad dynamic range of the measurements has enabled us to isolate the emission of this region, to analyze specific features of the spectral profile, and to link these to the kinematics of the system.

4. LINE PROFILE OF THE EJECTOR EMISSION

The emission of the ejector region within 0.15 mas (0.07 AU) is concentrated in a narrow spectral line with an approximately Gaussian shape (Fig. 3). The measurements of the line profile were carried out over a broad dynamic range with a velocity resolution of 0.053 km/s. To improve the accuracy of the measurements for low signal levels (in the wings), the adopted frequency resolution was three channels, or 0.159 km/s. The shape of the line profile remained virtually unchanged throughout the active period. Figure 3 shows the emission-line profile on April 24, 1999. The full width at half maximum (FWHM) is $\Delta v = 0.42 \pm 0.02$ km/s, or $\Delta f \approx 31$ kHz. The line velocity is $v = 7.64 \pm 0.05$ km/s. For comparison, the dashed curve in the figure shows a Gaussian distribution with the same width. We can see that the line top is slightly flatter than the Gaussian curve, and the emission level in the wings is higher. The wings are more easily distinguished when the profile is plotted on a logarithmic scale. Figure 4a presents the line profile at the onset of the flare (observations of May 10, 1998, and September 9, 1998), and Figure 4b, the line profile in the final stage of the flare (observations of January 23, March 29, and April 24, 1999). The line intensities have been reduced to epoch 1999.3. These line profiles can be represented by two Gaussian components, shown by the dashed curves. Their intensities are $I_1 \approx 3 \times 10^5$ Jy/beam and $I_2 \approx 400$ Jy/beam, corresponding to brightness temperatures of $T_{b1} \approx 5 \times 10^{16}$ K and $T_{b2} \approx 6 \times 10^{14}$ K. The FWHM of the profiles are $\Delta v_1 = 0.42$ km/s ($\Delta f_1 = 31$ kHz) and $\Delta v_2 = 2.2$ km/s ($\Delta f_2 = 163$ kHz).

5. INTERPRETATION OF THE OBSERVATIONAL DATA

The observed line profile with its broad wings could have a number of origins of a macro- or micronature: kinematics or scattering. Let us consider some of these.

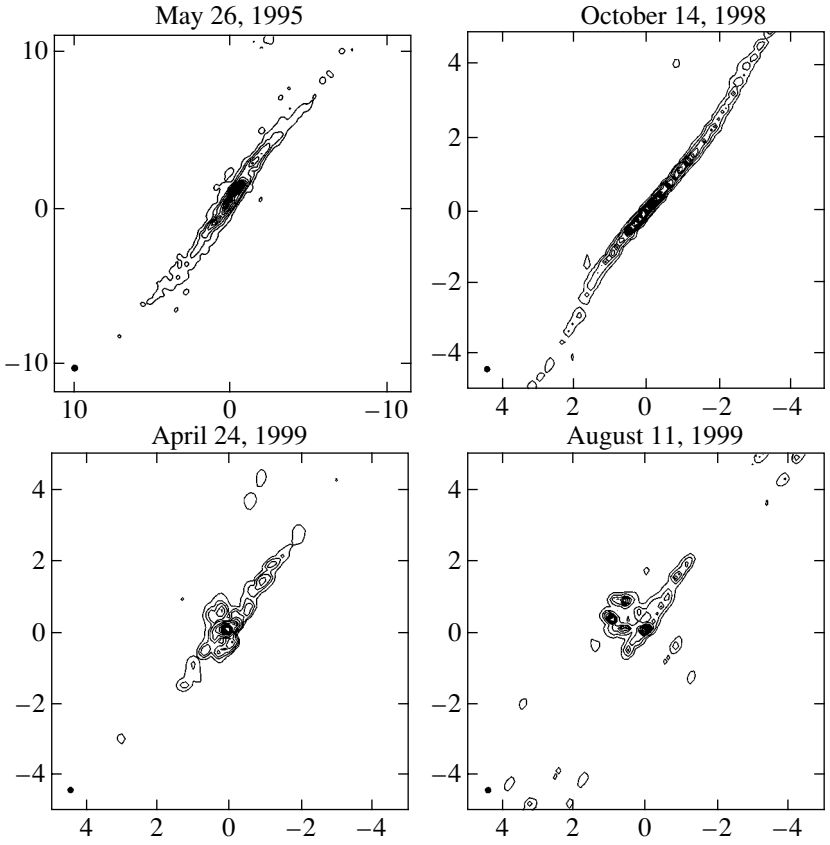

Fig. 2. Fine structure of the active star-forming region in Orion KL in the H_2O maser emission line at various epochs (indicated at the top of each map). The axes scale is in milliarcseconds (mas). The size of the synthesized beam is 0.15 mas (shown in the lower left corner of the map).

a. Doppler Linewidth

The Doppler broadening of a spectral line is due to the motion of the radiating particles. In the case of thermodynamic equilibrium or random turbulent motion (a Maxwellian velocity distribution of the particles), the line profile has a Gaussian shape. The line FWHM is

$$\Delta f_D = (2f_{mn}/c)[\ln 2(2kT_k/\mathrm{M} + \langle v^2\rangle)]^{0.5}\ \mathrm{Hz},$$

where f_{mn} is the frequency of the transition between levels m–n (for the H_2O maser line, $f_{mn} = 22235.08$ MHz), $M = 29.8975 \times 10^{-27}$ kg is the mass of the water molecule, $k = 1.38 \times 10^{-23}$ erg/K is Boltzmann's constant, $c = 3 \times 10^8$ m/s is the speed of light, T_k is the kinetic temperature of the gas in Kelvin, and $\langle v^2\rangle^{0.5}$ is the rms microturbulence velocity in km/s.

The Gaussian line profile corresponds to an ideal Maxwellian velocity distribution. In reality, the line shape is affected by collisions of particles, scattering of the emission, and the pumping level. This can result in large, extended wings (a Lorentz profile) and flattening of the line top (a Voigt profile). In the ejector region, the density of the particles in the outflow and the turbulence can be significant, and this will affect the line shape.

We can see from our data that the spectral line with its extended wings can be represented as two Gaussian components. The broad Gaussian with intensity $I_2 \approx 400$ Jy/beam, $T_b \approx 10^{14}$ K, and $v_2 \approx 7.64$ km/s corresponds to the maser emission of the bipolar outflow right at the ejector output. For a linewidth $\Delta v_2 = 2.2$ km/s ($\Delta f_2 = 163$ kHz), the effective kinetic temperature of the water vapor molecules is $T_k = 435$ K.

The kinetic temperature of the H_2O maser is determined by the excitation temperature, which usually does not exceed $T_{ex} = 50$–100 K. In this case, the linewidth should be $\Delta f_D = 26.6$–37.5 kHz, or $\Delta v_D = 0.36$–0.5 km/s, which is smaller than the considered value. The broadening of the line may be due to the microturbulence or kinematics of the outflow.

The ejector emission, $I_2 \exp\{-[(v - v_o)/\sigma_1]^2\}$, is amplified by the surrounding envelope to the level

$$I_1 = I_2 \exp[-\{(v - v_o)/\sigma_1\}]^2$$
$$\times \{1 + G_o \exp[-\{(v - v_o - \Delta v)/\sigma_2\}]^2\},$$

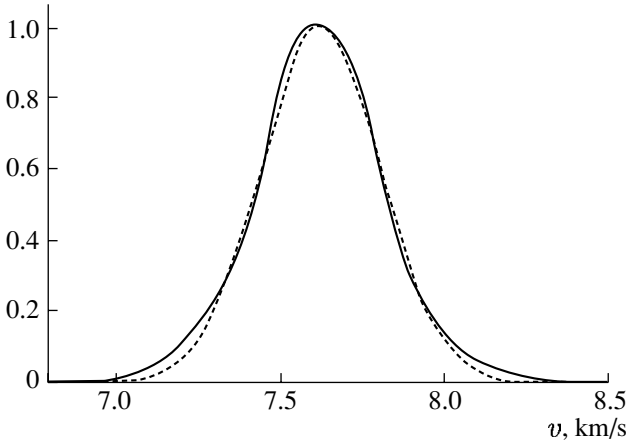

Fig. 3. Line profile of the ejector emission in relative units (epoch 1999.4). Dashed curve: Gaussian profile with the same FWHM.

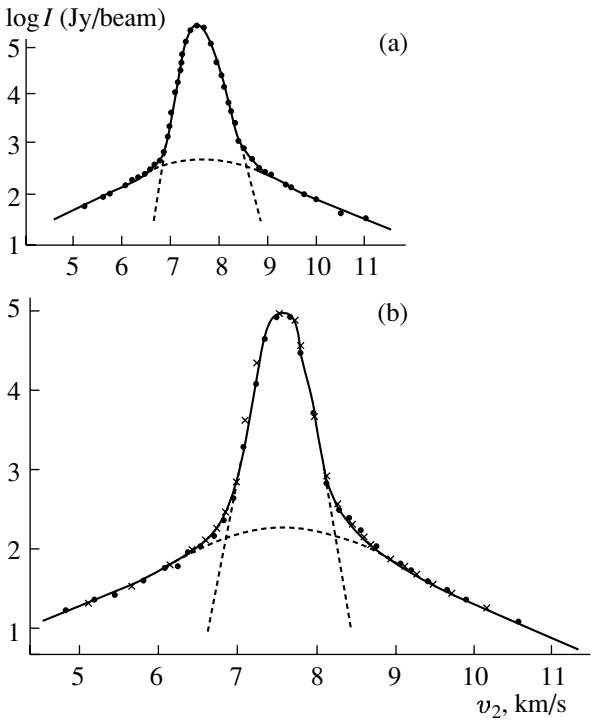

Fig. 4. Line profiles of the ejector-region emission on a logarithmic scale: (a) epochs 1998.42 and 1998.75, (b) epochs 1999.15, 1999.33, and 1999.41. The line intensities have been reduced to epoch 1999.36.

where $\sigma_1 = 1.32$ and $\sigma_2 = 0.25$ are determined by the widths of the Gaussian profiles of the signal, $\Delta v_1 = 2.2$ km/s, and by the amplification bandwidth, $\Delta v_2 = 0.42$ km/s; Δv is the difference between the profile velocities of the signal and amplifier. The radial velocity of the envelope is $v_o = 7.64$ km/s.

The intensity of the observed (amplified) emission is $I_1 \approx 3 \times 10^5$. It follows from the intensity ratio of the Gaussian profiles I_1/I_2 that the maser amplification in the envelope at a velocity of $v_o = 7.64$ km/s is $G_o \approx 750$. The maser linewidth (amplifier bandwidth) depends on the kinetic temperature and pumping level. For an excitation temperature of $T_{ex} = 100$ K, the linewidth is $\Delta f_D = 37.5$ kHz or $\Delta v_D \approx 0.5$ km/s. The high brightness temperature, $T_b \approx 10^{17}$ K, and Gaussian profile shape imply an unsaturated or partially saturated maser amplification regime. In the unsaturated regime, the gain G is proportional to e^τ and the linewidth is $\Delta f \approx \Delta f_D \tau^{-0.5}$; i.e., it decreases with increasing optical depth τ. The optical depth of the envelope for a gain of $G_o \approx 750$ in an unsaturated regime is $\tau \approx 6.6$. This should narrow the amplification bandwidth by approximately a factor of 2.6, to $\Delta f \approx 14$ kHz or $\Delta v \approx 0.2$ km/s. In reality, the observed line is more than twice as broad: $\Delta f_1 = 31$ kHz ($\Delta v_1 = 0.42$ km/s). This could indicate partial saturation of the amplification. Thus, the observed line profile with its extended wings can be explained by amplification of the ejector emission in a partly saturated regime.

The maser pumping in the envelope could be collisional. In this case, the velocity of the envelope expansion should be ≥ 10 km/s with respect to the surrounding medium. The radial velocity of the motion of the envelope will result in a relative shift of the emission and amplification line profiles. We observed virtually coincident profiles; this implies coincidence of the velocities of the envelope and central body, with their velocity in the Local Standard of Rest being $v_o = 7.64$ km/s. However, it is possible that the ejector profile has a more complex appearance. In turn, the pumping could be determined by the infrared emission of the central object, which may be the source IRc 4.

b. Bipolar Outflow Emission

Models for star formation involve the infall of material onto an accretion disk rotating around a central massive body, with the subsequent ejection of matter along the disk rotation axis. As a result, a highly collimated, rotating, bipolar outflow ejected in two opposite directions is formed. The radiating collimated outflow with embedded water vapor molecules resembles a cylinder whose emission is concentrated in both directions along its axis. The emission directivity is determined by the cylinder cross section $\oslash$ and length L; $\Omega = [\oslash/L]^2$.

The possible velocity gradient along the outflow dv/dl imposes certain limitations on the effective length of the cylinder. The outflow velocity difference should not exceed the maser window width, $\Delta v \approx$

0.5 km/s. Accordingly, the effective maser cylinder length is $L \approx 0.5/dv/dl$, and $\Omega = (2\oslash dv/dl)^2$.

The emission profiles of the two lobes of the outflow are identical, and their velocities are determined by the outflow velocity v at the cylinder input. In the presence of an outflow velocity gradient, the line velocities for the forward and backward flows will differ. The line velocity of the approaching lobe is $v = -v_{ej}$, while that of the receding one is, accordingly, $v = (v_{ej} + dv/dlL)$. Their relative shift is $\Delta v = 2v_{ej} + dv/dlL$. This shift results in a flattening and broadening of the total line profile of the bipolar outflow:

$$I = \{1 + G_o \exp(-[(v - v_o)/\sigma]^2)\} \times \{\exp(-[(v - v_{ej})/\sigma]^2) + \exp(-[(v + v_{ej} + dv/dlL)/\sigma]^2)\}.$$

The radial component of the line velocity, $v_{rad} = v\cos\chi$, depends on the viewing angle χ. The observed radial velocities of the outflows are $\sim$0.3 km/s. At the same time, the projected outflow velocities of the bipolar structure in the plane of the sky reach $\sim$6 km/s. The observed velocity projections testify that the ejector axis is close to the plane of the sky. Thus, we observe the emission of the bipolar outflow in a direction that is nearly orthogonal to the direction of its motion. The emission intensity is fairly high, implying low directivity.

c. Emission of the Rotating Bipolar Outflow

The observed maser line profile may be due to the particulars of the emission of the rotating outflow. The rotation of the outflow broadens the line in the direction orthogonal to the outflow motion. The optical depth of the outflow near the ejector is large. Let us consider the emission of the outflow—cylinder—at some angle χ. The radial velocity of an elementary radiating area in the cross section is determined by the velocity of the outer wall of the outflow and is $v_{||} = v_{rot}\sin\chi\sin\varphi$. The angle φ in the cross section of the cylinder is measured from the point nearest to the observer and varies in the range $-\pi/2 \le \varphi \le \pi/2$.

In the case of a constant outflow velocity, the linewidth is determined by the outflow rotation rate: $\Delta v_{||} \sim 2v_{rot}\sin\chi$. According to the observational data, the linewidth is $\Delta v_{||} \sim 2.2$ km/s, and, for a viewing angle $\chi \approx 40°$, the rotation rate is $v_{rot} \approx 1.4$ km/s. The radius of the outflow cross section is $R \approx 0.02$ AU, which corresponds to an outflow rotation period of $T \approx 5$ months.

d. Scattering of the Maser Emission on Free Electrons

The extended wings in the line profile could arise due to scattering of the maser emission on free electrons; in our case, this corresponds to the component $I_2 \approx 400$ Jy/beam. The scattered emission I_2 is a small fraction ($\sim 10^{-3}$) of the maser emission intensity, $I_1 \approx 3 \times 10^5$ Jy/beam. The fraction of the emission scattered on free electrons is determined by the optical depth and is equal to $\tau_e \sim \int \sigma_T N_e dl$, where $\sigma_T \approx 6.56 \times 10^{-25}$ cm^{-2} is the Thomson cross section. For low values of $\tau_e < 0.5$, the fraction of energy scattered in broad lines is of the order of $\tau_e/(1 - \tau_e)$. In our case, this corresponds to $\tau_e \sim 10^{-3}$, and the column density of free electrons is $N_e l \approx 1.5 \times 10^{21}$ cm^{-2}.

The molecular cloud OMC-1 is screened by an HII region; the electron density in the HII region is $N_e \approx 10^3$ cm^{-3}, and its size is $l \approx 0.6$ pc. The electron column density is $N_e l \approx 2 \times 10^{21}$ cm^{-2}, which is consistent with the above estimate.

The linewidth of the scattered emission is determined by the velocity of the scattering particles, i.e., their effective temperature. However, in contrast to the Doppler broadening considered above, M is the electron mass, $m_e = 9.1 \times 10^{-31}$ kg. The scattered linewidth $\Delta f_D = 163$ kHz at frequency $f_{mn} = 22235.08$ MHz corresponds to an effective electron temperature of $T_D \approx 0.5$ K, which is incompatible with the parameters of the HII region. Thus, the linewidth is determined solely by the radiating bipolar outflow.

6. CONCLUSIONS

Our study of the supermaser emission region in Orion KL over a broad dynamic range with high angular resolution (0.1 mas) has shown the following:

(i) In the central part of the structure, there is a compact bright source of H_2O maser emission—the ejector. Its size is ≤ 0.05 AU and its brightness temperature is $T_b \approx 10^{17}$ K.

(ii) The profile of the H_2O maser emission line of the ejector region is Gaussian with broad wings. The profile FWHM is $\Delta v = 0.42 \pm 0.02$ km/s ($\Delta f \approx 31$ kHz) and its velocity is $v = 7.64 \pm 0.05$ km/s. The shape of the line profile is essentially preserved during the entire interval of activity of 1998–1999.

(iii) The line profile can be represented as two Gaussian components with intensities $I_1 \approx 3 \times 10^5$ Jy/beam and $I_2 \approx 400$ Jy/beam with FWHM $\Delta v_1 = 0.42$ km/s ($\Delta f_1 = 31$ kHz) and $\Delta v_2 = 2.2$ km/s ($\Delta f_2 = 163$ kHz). The radial velocities of both Gaussians nearly coincide, $v = 7.64$ km/s.

(iv) The emission line profile is determined by a rotating bipolar outflow. The outflow rotation period is about 5 months. The rotation axis lies almost in the plane of the sky.

(v) The ejector maser emission is amplified in the surrounding medium (envelope) by more than two orders of magnitude at the velocity $v = 7.64$ km/s in the band $\Delta f_1 = 31$ kHz. The amplification is partially saturated.

(vi) The coincidence of the radial velocities of the emission profiles of the ejector and the envelope amplification implies that this velocity corresponds to the velocity of the Local Standard of Rest, $v_{LSR} = 7.64$ km/s.

ACKNOWLEDGMENTS

The author thanks K.M. Zakharin for help with the data processing and R.D. Dagkesamanskiĭ for useful comments. This work was supported by the Russian Foundation for Basic Research (project no. 02-0.2-16179) and the Program on Nonstationary Phenomena in Astronomy of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. B. F. Burke, K. J. Johnston, V. A. Efanov, *et al.*, Astron. Zh. **49**, 465 (1972) [Sov. Astron. **16**, 379 (1972)].
2. L. I. Matveenko, Pis'ma Astron. Zh. **7**, 100 (1981) [Sov. Astron. Lett. **7**, 54 (1981)].
3. L. I. Matveyenko, D. A. Graham, and P. J. Diamond, Pis'ma Astron. Zh. **14**, 101 (1988) [Sov. Astron. Lett. **14**, 468 (1988)].
4. A. M. Tolmachev, Pis'ma Astron. Zh. **26**, 41 (2000) [Astron. Lett. **26**, 34 (2000)].
5. H. Kobayashi, T. Shimoikura, T. Omodaka, and P. J. Diamond, in *Proceedings of the VSOP Symposium*, *Institute of Space and Astronautical Science*, *2000*, p. 109.
6. S. Horiuchi and O. Kameya, Preprint No. 4 (Mizusawa Astrogeodynamics Observatory, 2000).
7. L. I. Matveyenko, P. J. Diamond, and D. A. Graham, Pis'ma Astron. Zh. **24**, 723 (1998) [Astron. Lett. **24**, 623 (1998)].
8. L. I. Matveyenko, K. M. Zakharin, P. J. Diamond, and D. A. Graham, Pis'ma Astron. Zh. **29**, 45 (2003) [Astron. Lett. **29**, 641 (2003)].
9. L. I. Matveyenko, K. M. Zakharin, P. J. Diamond, and D. A. Graham, Pis'ma Astron. Zh. **31**, 1 (2004) [Astron. Lett. **30**, 100 (2004)].

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 665–677. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 732–745.
Original Russian Text Copyright © 2004 by Livshits, Belov.

When and Where are Solar Cosmic Rays Accelerated Most Efficiently?

M. A. Livshits and A. V. Belov

Institute of Terrestrial Magnetism, Ionosphere, and Radiowave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia

Received September 25, 2003; in final form, March 15, 2004

Abstract—The acceleration of particles by solar flares with extremely large proton fluxes whose energies exceed 100 MeV is considered. Most importantly, the location of the source of such acceleration in the flare of July 14, 2000, is determined assuming that the acceleration time coincides with the observed burst of hard line and continuous gamma-ray emission. The onset of this event corresponds to 10:19 UT, when data taken by the TRACE space observatory show that one of the flare ribbons reached a large sunspot in a group. The time interval for the development of the flare, 10:20–10:28 UT, is associated with the beginning of an increasing proton flux at the Earth. The region of efficient acceleration is estimated to be approximately two to three times higher than the height where the hard X-ray pulse usually originates (about 7000 km). The results are generalized for 28 powerful flares with extremely efficient acceleration of relativistic particles—in particular, for the well-studied events of June 15, 1991, and May 24, 1990—and are compared with the results of a statistical analysis of over 1100 increasing-proton-flux events. Efficient particle acceleration seems to be associated with the powerful impulsive episodes of the large flares analyzed. The results suggest that, along with sources of local (as in impulsive flares) and post-eruptive acceleration, there is an additional, very efficient, moderate-scale "accelerator" in tenuous regions with fairly strong magnetic fields and magnetic-field gradients. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of particle acceleration is one of the most difficult in the physics of solar flares. The presence of accelerated electrons is revealed in impulsive phenomena by their nonthermal spectra from 10–15 to 100 keV. Extensive data on ions accelerated during solar flares have been obtained over the past 50 years via the direct detection of the particle fluxes at the Earth and in interplanetary space. Catalogs of proton events include all increases in particle fluxes in the above energy range above a few to 10 MeV. The low-energy detection threshold ranges from 0.01–1 particle per square centimeter per second per steradian in periods of deep minimum and maximum solar activity, respectively. The highest proton activity during the last 40 years for which all the necessary data are available was observed in the XXII solar cycle [1]. Despite the availability of multiple records of proton events, the mechanism and sources of the particle acceleration remain unclear.

As follows from studies conducted over the last several decades, particles with energies of 10–80 MeV are probably accelerated by shocks formed by coronal mass ejections (CMEs). A substantial contribution to solving this problem has been made by D. Reams, S. Kahler, and O.C. St.Syr. The subject of the present paper will be more energetic particles, namely, protons with energies above 100 MeV (and electrons with energies above 50 keV). For brevity, we will call such particles relativistic, since an analysis of their motion must be carried out in a relativistic treatment.

A few years ago, shock acceleration was commonly assumed to take place for relativistic particles as well, particularly for protons with energies up to 1 GeV and more, i.e., for solar cosmic rays (CRs) observed at ground-based detectors. However, this view is now in contradiction with new results that follow both directly from observations of high-energy phenomena and from their theoretical interpretation. For example, in some cases, the main shock is formed well after the time when the particles with the highest energies have left the Sun, reaching detectors on the Earth within the following 15–20 min. In addition, it is difficult to explain the acceleration of particles to energies above 1 GeV by shock fronts with the physical parameters of those that are implied by observations of type-II radio bursts in the corona and interplanetary space.

As is known, ground-level enhancements (GLEs) of the cosmic-ray intensity are usually associated with the most powerful flares, which often occur in large spot groups or activity complexes. As a result, several successive GLEs are sometimes observed when such a group moves over the disk near the solar-activity maximum. The particles usually come to the Earth from powerful flares in the western part of the disk at longitudes from +20° to +75°,

1063-7729/04/4808-0665$26.00 © 2004 MAIK "Nauka/Interperiodica"

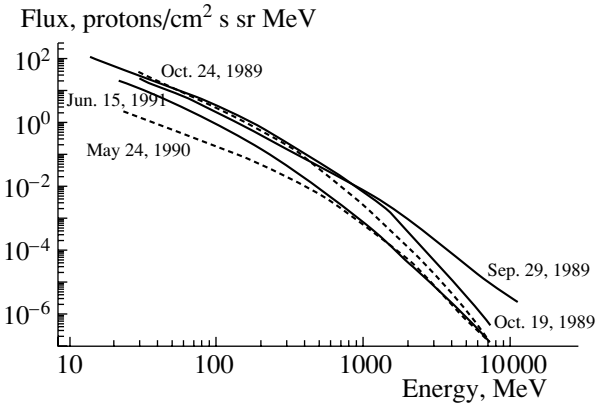

Fig. 1. Proton spectra observed at the Earth and in near-Earth orbit at the times of maxima of the most prominent enhancements of cosmic rays in the XXII cycle. The hardest solar cosmic-ray spectrum was detected in a flare behind the limb on September 29, 1989.

although a number of such events are associated with central flares, as well as processes near or behind the western limb. The amplitude of a GLE, the shape of its temporal profile, and its energy spectrum are sensitive to the longitude of the solar event resulting in the GLE. There are sometimes two maxima in the ground-based observations of solar CRs. The first is characterized by high anisotropy: the number of particles leaving the Sun along certain magnetic field lines is considerably greater than the fluxes from other directions. The spectrum of particles arriving directly from the Sun was found to be harder than the spectra of particles arriving later from various directions. The enhancement detected by a particular cosmic-ray station depends on its geomagnetic latitude and can vary from a fraction of a percent to to several thousand percent.

Proton spectra detected at the Earth and in near-Earth orbits are shown in Fig. 1 for the maxima of the largest enhancements in the intensity of solar CRs in the XXII cycle [2]. These spectra are thought to be close to the particle-emission spectra of the source of acceleration. The hardest solar CR spectrum was detected in a flare behind the limb on September 29, 1989. Note that the proton spectrum at energies of about 1 GeV in a short, powerful flare on May 24, 1990, almost coincided with the spectra of prolonged powerful flares in activity complexes. The proton energies in some of these events exceeded 10 GeV. The number of protons with energies over 30 MeV accelerated over the entire duration of the flares presented in Fig. 1 is greater than 2×10^{33}.

The temporal profile of the solar CR intensity can give us some information about the particles leaving the Sun. Starting from the 1960s, such information was derived using models for the diffusive propagation of high-energy protons. In recent years, it has become possible to utilize additional information about the acceleration process, such as the proton-acceleration time, which is derived from the temporal profile of the intensity of γ-ray radiation associated with π^0-meson decays. In [3], the temporal behavior of the intensity of solar protons with $E \leq 100$ MeV was approximated for several flares in the XXII cycle using a diffusive model for the propagation of solar CRs for a chosen proton-injection function on the Sun. In the case of disk flares, the main acceleration takes place during the hard γ-ray burst (about 10–20 min). The characteristics of the temporal profile of solar CRs in [3] are associated with another one or two phases of acceleration of the relativistic particles. Support for this possibility is provided by the presence of prolonged γ-ray radiation from powerful flares [4]. Unfortunately, it is not possible to choose a unique model for the acceleration of relativistic protons by elaborating the simplest diffusive models.

In general, we can assume that the relativistic protons are accelerated several times by various mechanisms during a single flare. This could include pulsed acceleration either near spots or in a vertical current sheet in the posteruptive phase and, sometimes, during the interaction of coronal loops. In principle, one can develop a detailed model describing the observations, but there is no guarantee that such a model will correspond to the real situation. There has been little progress in understanding the mechanism for accelerating relativistic particles in the Sun since the 1980s, i.e., during the period of operation of the SMM satellite (see review [5]). This is not due purely to the lack of a sufficient number of observations for the most energetic particles and hard γ-ray radiation by large flares.

The aim of the present article is to consider the acceleration of these particles, as far as possible, without using any *a priori* assumptions about the acceleration mechanisms or the nature of the source. It is now reliably established that electrons are accelerated to energies exceeding a few tens of keV in numerous flares that are sources of hard X-ray bursts.

This acceleration takes place in one or several episodes of impulsive and flash phases, with the source usually being localized at a small height within a very small region near spots. We will call this case "local acceleration." In other cases, especially in the gradual phase of the flare, when the Kopp–Pneuman model is applicable, the particles seem to be accelerated in a vertical current sheet. We will call this "posteruptive acceleration." The available evidence seems to suggest that these two possibilities—local and posteruptive acceleration for proton energies above 100 MeV (where shock acceleration can be

ignored)—are not able to explain the number and spectra of particles in observed phenomena with the maximum fluxes of relativistic particles. Indeed, it is difficult to accelerate a considerable number of particles with energies of several hundreds of MeV with a very small source of pulsed acceleration. In addition, it is known that such sources are localized near spots, in regions of strong closed magnetic fields, making it very difficult for particles to escape. The posteruptive stage is characterized by acceleration in a region of weak magnetic fields that slowly vary in space; this is unfavorable for the efficient acceleration of relativistic particles. In addition, there is direct observational evidence in several cases that considerable fluxes of energetic particles arrived at the Earth before most of the posteruptive phenomena on the Sun had been observed. These facts stimulated us to study a selected group of phenomena with extremely large relativistic-particle fluxes. Physically, we must answer the question of whether there is an additional source of very efficient particle acceleration in these phenomena and, if so, what its properties are.

This formulation of the problem determined our choice of material and the analysis method. We first refine the list of flares in which the acceleration of particles up to relativistic energies is most efficient, using both data on the direct detection of particles and data on radio- and γ-ray emission. Next, we shall discuss in detail the problem of identifying the source of relativistic particles in the well-studied flare of July 14, 2000. A separate section will be devoted to a retrospective view of all events since 1956 with the largest fluxes of protons with $E \geq 100$ MeV, using the information provided by modern observations of well-studied powerful phenomena. In conclusion, we shall summarize the results of our morphological study of flare particle acceleration in the most powerful events.

2. FLARES WITH THE MOST EFFICIENT ACCELERATION OF RELATIVISTIC PARTICLES

Analysis of the accumulated data on the most powerful solar flares suggests that there exist phenomena possessing an anomalously high acceleration efficiency. For example, direct detections of charged particles and neutrons sometimes indicate a total or maximum number of particles that is an order of magnitude greater than the average value for a given class of events. Such events are usually simply identified as GLEs. Nearly 70 GLEs have been recorded since the first measurements in 1942. It has just recently become clear that it is not entirely accurate to identify phenomena with high acceleration efficiencies with GLEs for the following two reasons. First, the probability of detecting a GLE depends strongly on the position of the flare with respect to the Earth. Therefore, even flares that are not very powerful give rise to an appreciable effect if they are localized in a region of the solar surface that is most favorable for observations from the Earth; the opposite is also true: the effect of a powerful flare that is far to the east of the central meridian can be negligible. Second, neutron monitors are sensitive to the hardest particles with $E \geq 1$ GeV. As a result, events with hard spectra are more likely to be detected than other events with the same number of relativistic particles. Therefore, the list of GLE includes both flares with the highest efficiency of the generation of relativistic particles and approximately the same number of ordinary powerful events, whose average fluxes of protons with $E \geq 100$ MeV are not exceptional.

We use here a database of X-ray flares supplemented by uniform material on the fluxes of protons with energies $E \geq 10$, 60, and 100 MeV. Events with fluxes of protons with $E \geq 100$ MeV exceeding 100 $\mathrm{cm}^{-2}\ \mathrm{s}^{-1}\ \mathrm{sr}^{-1}$ near the Earth were selected from the above database, which includes data on over 1000 proton flares in 1976–2003. There are only 12 such events. In these powerful flares, the condition of the relativistic particles' escape from the solar corona enabled them to reach near-Earth space, causing the flux of protons with $E \geq 100$ MeV to exceed the above limit.

The fluxes of relativistic particles reaching near-Earth space are often attenuated because the flare was beyond the optimal longitude interval from $-20°$ to $+90°$. In such cases, we deduced the acceleration efficiency using indirect data, primarily the radio flux at 15.4 GHz and the intensity of the hard γ-ray burst. These data were compiled in the catalog of proton events [1] and in [4], as well as in some other works. In particular, we analyzed all events with maximum fluxes at 15.4 GHz exceeding 2×10^4 s.f.u. This approach enabled us to estimate in a first approximation the fluxes of relativistic particles at the Sun for eastern flares, where direct measurements suffer from propagation effects. The extensive observational material for flares in the last two (XXII and XXIII) cycles enabled us to identify events of interest with reasonable reliability—the flares of June 1991 and others, including some cases when the direct detection of the flux of relativistic protons with energies of 100 MeV was problematic, as well as the event with the maximum number of detected neutrons, which occurred on May 24, 1990. A few powerful flares of the previous (XXI) cycle were also included in our final dataset, although the efficiency of their acceleration of relativistic particles was slightly lower than in the most powerful flares of the XXII cycle.

The attenuation of the flux of relativistic protons due to the distance of the source from the optimal

longitude zone can also be estimated using the statistical results presented, for example, for GLEs in [6]. In this case, the statistics are based on approximately 50 GLEs and a few hundred proton flares, together with simultaneous measurements of the same events by several spacecraft located at remote points in interplanetary space (see review [7]). Using this approach for small separations from the optimal longitude zone leads to estimates that are close to those derived from indirect data; however, using statistical results is not effective when the flare is considerably shifted to the east from the central meridian.

Our consideration also demonstrates that our estimate of the overall efficiency of the acceleration of relativistic particles from the flux of protons with $E \geq 100$ MeV is much less dependent on the accelerated-particle spectrum than the corresponding estimate derived from the more energetic particles producing GLEs.

Our final results for 28 flares are presented in the table, which lists the date, time of the onset of the flare, X ray magnitude, total duration dt, time for the increase of soft X-rays $dt1$ (both times were taken directly from the database of the GOES satellites), optical magnitude, and latitude and longitude of the flare (positive values refer to north of the equator and west of the central meridian). The particle fluxes in near-Earth space are presented for energies $E \geq 10$, 60, and 100 MeV. the presence of GLEs is indicated in the notes.

If the above limit on the dates of the flares were not imposed, the total list of events would include several flares in the 1940s, the well-known flare of February 23, 1956, and the powerful flares of August 1972. A few powerful flares that occurred at the eastern limb (for example, on June 1, 1991) were not included in the final table, because there was not sufficient information about the acceleration processes, especially at low heights. Our analysis indicates that, over approximately the last 30 years, the number of events with extremely efficient relativistic-particle acceleration is about equal to the number of GLEs. However, our sample includes the most powerful flares resulting in GLEs, which is about half our list; the second half represents flares with almost the same efficiency of relativistic-particle acceleration occurring in the eastern part of the solar disk.

Before presenting the main contents of our study, let us make two general comments about the phenomena listed in the table. In many moderate-power flares, the particle acceleration takes place during or just after one burst of the impulsive phase. This is close to the time when the derivative of the flux of soft X-ray radiation is maximum. This is also true for some of the powerful phenomena listed in the table. On the other hand, the most powerful events often represent several successive episodes, in some of which particles could be accelerated. The episodes with the most efficient acceleration in each event will be discussed below.

In addition, detailed data on the hard gamma-ray emission have been obtained in recent years for some flares. These data show that continuum radiation by electrons at 1–10 MeV appears simultaneously with continuum radiation at 60–100 MeV due to π^0-meson decays and line radiation by nuclei. This was found to be the case when the observations enabled us to carry out such an analysis for large flares: the flares of March 26, 1991 (GAMMA-1) [8], May 24, 1991 (GRANAT, FEBUS apparatus) [9], November 24, 2000 ($\approx$15 UT, Yohkoh) [10], and in some other cases when the observations were not complete. The observed and calculated spectra of high-energy photons for the flare of March 26, 1991, are presented in Fig. 2 [11, 12]. The decay of π mesons formed by nuclear reactions between the accelerated protons and background particles produces the observed high-energy radiation at $h\nu \geq 100$ MeV. The onset of the radiation by the electron and proton components is coincident to within an accuracy of a few to ten seconds. In the particular case of the flare of March 26, 1991, the acceleration of the electrons and protons with energies up to 120 and 300 MeV, respectively, took place simultaneously. Therefore, it is reasonable to suppose that the relativistic electrons and ions are accelerated simultaneously, at least in the main episodes discussed here. A weighty argument in favor of the possibility that the relativistic electrons and ions are accelerated by a single process is the tight correlation between the continuum radiation at $\geq$0.5 MeV and the nuclear line radiation at 4–8 MeV [5, Fig. 4].

3. LOCALIZATION OF THE SOURCE OF THE MAIN ACCELERATION OF RELATIVISTIC PARTICLES IN THE FLARE OF JULY 14, 2000

Modern observational data make it possible in some cases to localize the source of acceleration of the relativistic protons ($E \geq$ 100 MeV). The present section will be devoted to an analysis of the flare that occurred on Bastille day, 2000. It was observed for several hours by the TRACE satellite during the time when hard X-ray and γ-ray radiation was detected by Yohkoh. X-ray observations at energies above $\approx$25 keV were carried out at the beginning of the flare, from 10:05 to 10:14 UT by the hard X-ray spectrograph (HXRS) onboard the Multispectral Thermal Imager satellite, and after 10:19 UT by Yohkoh. Although the time of the maximum derivative of the soft X-ray emission (1–8 Å), 10:18 UT, which

Data on the flares and particle fluxes

Flare Date	Time, UT	X-ray magnitude	dt, min	$dt1$, min	Optical magnitude	Latitude, deg	Longitude, deg	Proton flux, $cm^{-2}\ s^{-1}\ sr^{-1}$			Notes
								$E >$ 10 MeV	$E >$ 60 MeV	$E >$ 100 MeV	
Sept. 23, 1978	9:47	X1.0	43	42	3B	35	50	1000	159	48.3	GLE
Oct. 12, 1981	6:22	X3.1	55	14	2B	−18	−30	590	84.8	26.27	GLE
June 3, 1982	11:41	X8.0	105	7	2B	−9	−72	30	0.89	0.283	
June 6, 1982	16:30	X12.0	162	24	3B	−9	−25	15	1.76	0.332	
July 9, 1982	7:31	X9.8	53	7	3B	17	−73	38	2.2	0.2	
Dec. 7, 1982	23:36	X2.8	71	18	1B	−19	79	900	111	56.1	GLE
Apr. 24, 1984	23:56	X13.0	104	5	3B	−12	−43	1110	224	5.51	
Aug. 12, 1989	13:57	X2.6	134	27	2B	−16	37	6000	875	237	GLE ?
Aug. 16, 1989	1:08	X20.0	80	9	2N	−18	84	1500	171	83.7	GLE
Sept. 29, 1989	10:47	X9.8	228	46		−30	90	4500	696	559	GLE
Oct. 19, 1989	12:29	X13.0	464	26	4B	−27	−10	2500	690	400	GLE
Oct. 22, 1989	17:08	X2.9	240	49	2B	−27	31	4450	811	380	GLE
Oct. 24, 1989	17:36	X5.7	528	55	3B	−30	57	5000	398	272	GLE
May 21, 1990	22:12	X5.5	87	5	2B	35	36	276	67.1	30.9	GLE
May 24, 1990	20:46	X9.3	59	3	1B	33	78	96	25	20	GLE
Mar. 22, 1991	22:43	X9.4	34	2	3B	−26	−28	10000	866	257	
June 4, 1991	3:37	X12.0	233	15	3B	30	−70	43	4.96	3.522	
June 6, 1991	0:54	X12.0	81	18	4B	33	−44	150	25.2	4.78	
June 9, 1991	1:37	X10.0	167	3	3B	34	−4	30	5.1	1.94	
June 11, 1991	2:09	X12.0	71	20	3B	31	17	2300	317	95.2	GLE
June 15, 1991	6:33	X12.0	284	118	3B	33	69	950	124	116	GLE
Nov. 2, 1992	2:31	X9.0	57	37	2B	−23	90	630	108	24.1	GLE
Nov. 6, 1997	11:49	X9.4	12	6	2B	−18	63	490	109	78.1	GLE
July 14, 2000	10:03	X5.7	40	21	3B	22	7	7000	1200	623	GLE
Nov. 8, 2000	22:42	M7.4	83	46	3F	10	77	14800	1580	451	halo NW
Apr. 15, 2001	13:19	X14.4	36	31	2B	−20	85	951	223	250	GLE
Sept. 24, 2001	9:32	X2.6	97	66	2B	−16	−23	12900	100	20	
Nov. 4, 2001	16:03	X1.0	54	17	3B	6	18	31700	200	220	GLE

often coincides with the impulsive phase of the flare and particle acceleration, was missed in this case, all the indirect evidence discussed in [13] indicates that there was no efficient acceleration of particles before the Yohkoh observations. This satellite detected two maxima at hard X-ray energies, 100–300 keV, in the delayed and prompt γ-ray lines. The lines associated with electron–positron annihilation and the capture of a neutron by a proton, as well as the rapid unresolved lines due to the excitation of oxygen and nitrogen nuclei, reached their maximum intensity on July 14, 2000, at 10:21:05 UT. A second, more powerful maximum in these lines was reached at ≈10:27 UT [14, Fig. 1].

The observations unambiguously fixed the time of powerful particle acceleration in the flare of July 14,

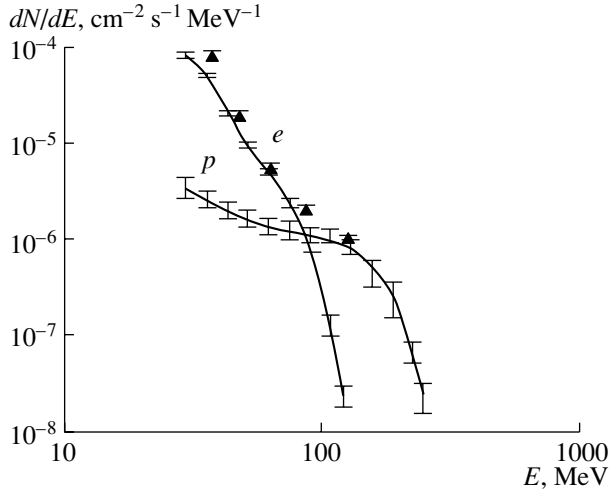

Fig. 2. Differential photon spectrum emitted by the flare of March 26, 1991 (near the disk center), at an angle below 25° from the radial direction. The calculated contributions from electrons and protons are plotted by the solid curve [11, 12]. The observed values are presented by triangles.

2000. Electron bremsstrahlung, hard radiation due to π^0-meson decays, and the onset of emission in the 0.511 keV and 2.22 MeV lines were synchronous in the second act of acceleration. In this case, the simultaneous acceleration of relativistic electrons and protons was confirmed directly by observations.

Thus, the most efficient acceleration of relativistic particles in this flare took place in a time interval that coincided with the hard γ-ray burst, from 10:19 to 10:30 UT, and the process seems to be separated into two stages. This heliophysical information can be compared with other observational data on the flare and the arrival time of the accelerated particles at the Earth.

In their analysis of GLEs of solar cosmic rays, Belov *et al.* [15] concluded that all or most of the energetic particles arriving at the Earth were accelerated between 10:19 and 10:29 UT, probably before 10:27 UT. Based on ground-based observations, Bieber *et al.* [16] calculated the characteristics of the solar source of accelerated particles using two different models for the propagation of the charged particles in interplanetary space. They obtained good agreement between the model calculations and the fluxes observed near the Earth under the assumption that the main emission by the charged particles was relatively brief (< 5 min) and possessed a maximum at 10:26–10:27 UT.

Therefore, we know the time for efficient particle acceleration on the Sun and can analyze the processes occurring there during this period.

The flare began at approximately 10:03 UT in the western part of an active region. The ordinary impulsive phase at 10:04–10:06 UT was weak, and it was impossible to predict that it would grow into a powerful event. Next, an ordinary flare with somewhat complex ribbons began to develop in the central part of the group. This is illustrated in the 195 Å TRACE satellite picture obtained at 10:14:24 (Fig. 3a). The situation changed dramatically at approximately 10:19 UT, when one of the ribbons of the flare, which formed at 10:12 UT and was immediately adjacent to the large spot, began to develop, and luminous plasma crossed the spot from west to east. This process was observed most clearly by the TRACE satellite in the continuum near 1600 Å (Fig. 4a) and in the line emission near 171 Å, i.e., in the emission by plasma with a lower temperature than the emission at 195 Å.

Along with the first bright point, which corresponded to the first maximum at 10:19 UT, another region to the east of the main spot is visible in subsequent pictures at 195 Å (Figs. 3b and 3c). Starting from 10:25 UT, the emission at 195 Å began to propagate to the northeast part of the group.

Note that, during 10:24–10:26 UT, the brightening at 171 Å propagated along a broken line that seems to correspond to a boundary of cells of the chromospheric network in the active region; further, it continued along a straight line, and emission in other feet of the loops of the flare arcade appeared on the opposite side of the neutral line. Next, this arcade covered the region of the former two-ribbon event; the top part of the arcade was heated over several hours, followed by the gradual ascent of the loops. The development of the first and then the second, more powerful, type-II radio outburst [17] began just at 10:19 UT; these are probably associated with several plasma ejections.

Thus, the flare process on July 14, 2000, reached the large spot at 10:19 UT; after 10:24 UT, the emission propagated toward the northeast; and the hard X-ray and γ-ray radiation reached their maxima at 10:27 UT (Fig. 3d). Although the flare continued to develop after 10:27 UT, the rate of acceleration seemed to drop sharply, and nonthermal processes gradually diminished. This can be seen in the rapid decay of the gamma-ray burst. Posteruptive heating of the plasma began near the top of the helmet structure just near this time (10:27 UT), and only at 10:38 UT did this structure decay into individual high loops. This marked the end of the formation of the arcade, which encompassed the eastern and western flare ribbons (Fig. 4c, 10:48 UT).

Let us now consider in more detail the specific features of this flare as a source of photons and energetic

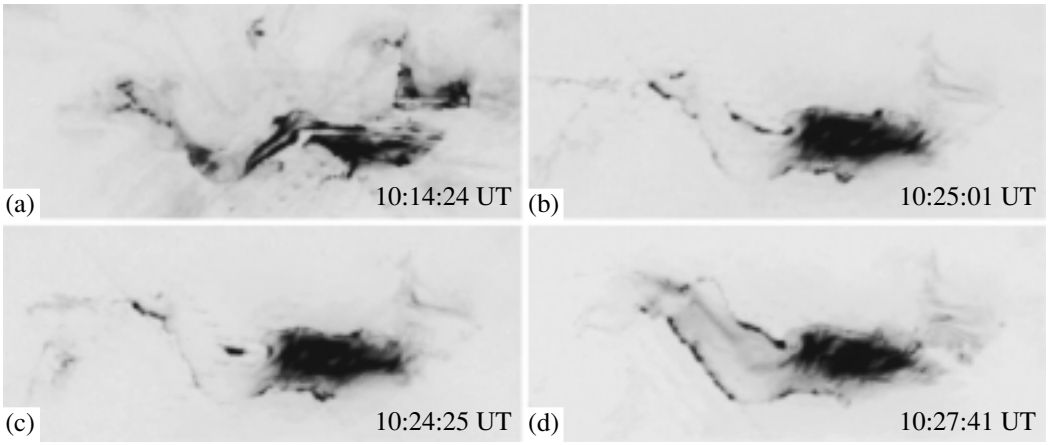

Fig. 3. Pictures of the flare of July 14, 2000, obtained by the TRACE satellite at 195 Å.

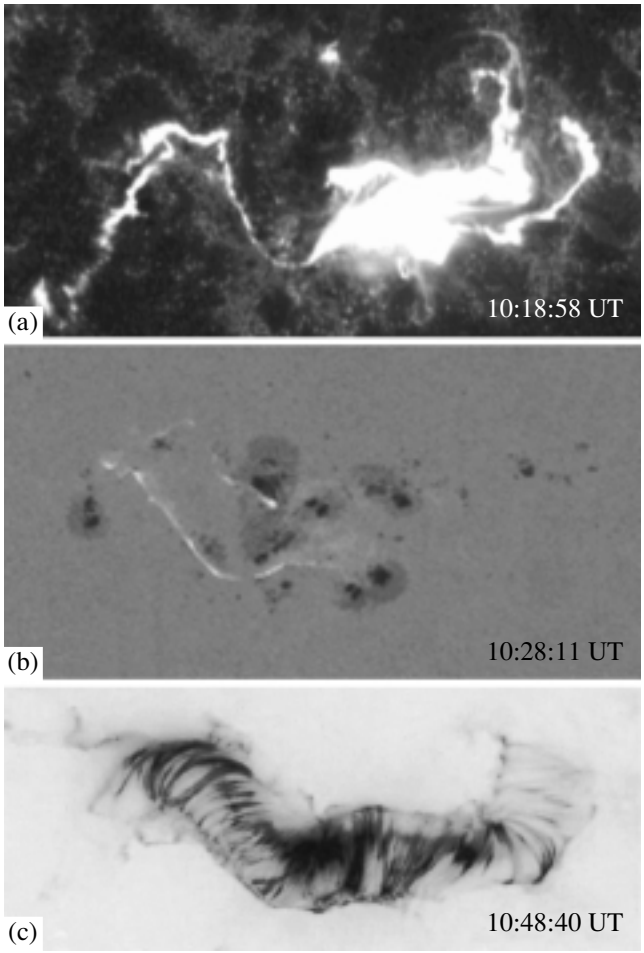

Fig. 4. Pictures of the flare of July 14, 2000, obtained by the TRACE satellite (a) at 1600 Å (with a short exposure of 30.4 s), where one can see an arm approaching the large spot from the west; (b) in white light; and (c) at 195 Å. The scales for pictures and the field of view were not accurately matched to each other.

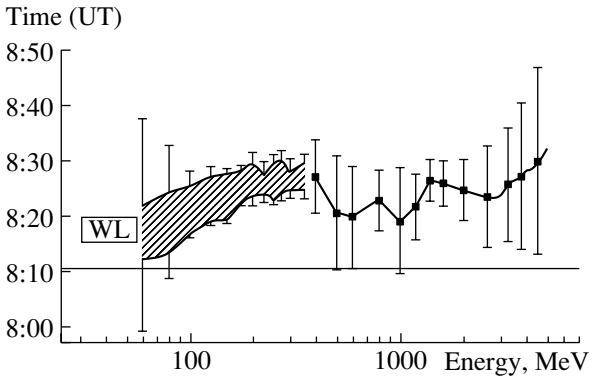

Fig. 5. Time interval in which particles of various energies left the solar corona during the flare of June 15, 1991 [21]. Only one time of escape was determined for the particles with energies above 350 MeV. The rms error in these times is plotted. The rectangle shows the lifetime of the white-light (WL) flare derived from ground-based observations.

particles. In most flares, nonthermal processes are characteristic of impulsive or explosive phases, which are close to the onset of the entire phenomenon. This phase is characterized by the maximum increase in the soft X-ray emission (at 1–8 Å). In some flares (especially, powerful ones), high-energy phenomena are manifest after a delay of up to tens of minutes rather than at the very beginning of the event. As a rule, this occurs when the flare process approaches large spots. The importance of this factor was discussed, in particular, by Ishkov [18] (this paper also contains references to earlier works by this author). The flare of July 14, 2000, is an example of a flare in which the "spot component" of the overall nonstationary process is clearly manifest beginning from 10:19 UT, 15 min after the onset of the flare and 10 min after the time of the maximum rate of increase in the soft X-ray radiation.

After 10:19 UT, the phenomenon develops, as do powerful impulsive events, in the immediate vicinity of strong magnetic fields and field gradients, but it is not absolutely identical to such impulsive events. Usually, the flare ribbons arise from a single center in a spot penumbra and then ignite bright points along a neutral line of the photospheric magnetic field, with their merging and subsequent evolution finally resulting in a set of high coronal loops. This first, ordinary, stage of development of the nonthermal process is followed by a more intensive one (stages *b* and *c* of the development of a γ-ray burst in Fig. 1 from [14]; see also [19, 20]). There are grounds to suppose that extremely efficient particle acceleration takes place in a structure between points on the opposite sides of the spot. Several such field lines (a loop?) at time 10:28:06 UT are shown in Fig. 5 in [22]; they connect two bright points located inside the main hard source in the 171 Å picture (HXR/M2, 10:27:00 UT). We emphasize that, according to this same work [22], the height of the top of this loop exceeds 20″—a factor of two to three higher than in typical impulsive nonthermal sources. Another, weaker part of the hard source was localized in the ribbon on the opposite side of the neutral line.

One characteristic observable manifestation of the development of efficient nonthermal processes for the July 14, 2000, flare was the propagation of emission (a jet) in the TRACE picture obtained at 10:24:25 UT (Fig. 3b). This jet was subsequently transformed into a thin line, almost parallel to the neutral line but shifted significantly to the east (Fig. 3c). Later, this line formed the feet of loops in the regions of north polarity of the longitudinal field, and another ribbon appeared on the opposite side of the neutral line.

The character of the white-light flare of July 14, 2000, was appreciably different from most similar phenomena (Fig. 4b). The optical continuum emission usually makes up two or more bright points located in the penumbra or just near the spots. However, in this case, another component dominated after 10:19 UT: the ribbon emission, which is usually very weak even in the most powerful impulsive events. We should make the following qualification here: white-light emission was detected in this event only by the TRACE satellite. There are grounds to suppose that the TRACE spectral band is more sensitive to the blue part of the spectrum (the flares detected at the Earth in green and red light are not identical to those detected by TRACE), so that it includes the contribution of numerous lines formed in the lowest layers of the chromosphere. Nevertheless, the absence of well-defined bright points in the corresponding TRACE pictures probably points toward the specific nature of the particle propagation in the loop (the source of acceleration); namely, most of the particles (especially, the relativistic particles) did not disappear into the feet, but instead filled the entire large-scale coronal structure. The influence of the accelerated electrons and considerable heat fluxes on the dense chromospheric layers near the feet of the coronal loops resulted in the white-light emission of the ribbons.

Therefore, the extensive observational data for the flare of July 14, 2000, enable us to localize the source of acceleration of the relativistic particles. There is clear evidence that its height exceeded the heights of loops in typical sources of accelerated electrons in ordinary impulsive events. We will briefly discuss a possible scheme for the development of such a nonthermal process below.

4. COMPARATIVE ANALYSIS OF PARTICLE ACCELERATION IN POWERFUL FLARES

It has recently become possible to carry out a statistical study of proton events. The IZMIRAN database contains 1143 proton enhancements detected over approximately 28 years of observations of particles in near-Earth space. The fact that 1152 flares stronger than M4 took place during this period suggests that the proton acceleration is associated with events of such power. Over half of these proton enhancements (618) were firmly identified with solar flares. The other events correspond to phenomena that occurred behind the limb (at least 20%), as well as to cases when individual flares are superposed in periods of very high activity and to gaps in the X-ray observations in 1975–1985.

We will discuss here only a few results, which are considered in detail in [23]. It is clear that the more powerful the flare, the higher the probability that it is a proton flare. We compared the characteristics of all 409 reliably localized proton flares at longitudes from 20° E to 90° W with a group of control flares that occurred in the same longitude range and had the same mean power (M6 and brighter) but did not display enhanced proton fluxes. We found that the duration dt of the proton events was 80 ± 4 min, while that for the control group was 57 ± 4 min.

Therefore, the relation between the total number of protons accelerated to $E \geq 10$ MeV and the overall energetics or X-ray magnitude of the flare is the most clearly expressed. The duration of the entire event probably also has some influence, but this is expressed weakly and cannot be used as a serious argument in support of the posteruptive acceleration of particles in a vertical current sheet or by a shock.

By considering 110 proton flares more powerful than X1, we were able to reveal the following tendency [24]: large fluxes of protons with energies of about 10 MeV "feel" the development of giant arcs after CMEs, while the relative number of relativistic particles (or the 10–100 MeV spectrum in the acceleration region) is related more closely to the general development of nonthermal processes during or near the impulsive phase. The posteruptive phase is identified both via direct observations of giant loop systems and from an analysis of the temperature behavior derived from the ratio of the signals from the two channels of the GOES satellites [25]. On the other hand, impulsive events are characterized, first and foremost, by the fluence of nonthermal emission with $E \geq 30$ keV.

These general conclusions for X flares are valid as well for the data compiled in the table, which represent events with the most efficient acceleration of relativistic protons. Of course, all the tabulated flares in the western and central parts of the disk were included among the above proton X flares and were simply the events with the greatest fluxes of relativistic protons. In a few cases, the events in the table were preceded by flares with approximately the same efficiency of relativistic-particle generation, but the fluxes in the second event were enhanced due to the interaction of a CME in interplanetary space (between the two shocks in the process of "cannibalizing" the CME), as took place on August 4, 1972.

Returning to the effect of the flare duration on the acceleration, we note that the table contains only a few short-lived, compact flares, on March 22, 1991, May 24, 1990, and November 6, 1997, whereas the total durations of the soft X-ray emission in other events exceed one hour. However, in most cases, the impulsive phase was clearly expressed, although in the powerful events under consideration it may not occur at the onset of the phenomenon. Thus, we believe that the role of the duration of the events in the generation of relativistic particles was overestimated in previous studies, and the most efficient acceleration is associated with the powerful, impulsive episodes of large flares.

It has been shown using a database of over 55 000 X-ray flares that even powerful, impulsive events on the Sun are usually not followed by an appreciable flux of accelerated protons. This is due to the fact that the source of the acceleration is localized in loops of a closed magnetic configuration near the chromosphere. In this case, the hard X-ray and γ-ray radiation is maximum, but it is very difficult for accelerated particles to escape. In the rare cases when a powerful short flare results in the formation of even a small region where the magnetic field lines extend outwards to the sphere of solar-wind source, the accelerated particles escape into interplanetary space. One such example is the event of May 24, 1990, for which the flux of solar neutrons detected at the Earth was maximum.

In our previous paper [26], we analyzed data on the detection of neutrons by the FEBUS apparatus on the GRANAT satellite [9]. We used a simple kinematic model to determine the duration of neutron generation, $\delta t = 230 \pm 15$ s; the exponent of the power-law spectrum, $\gamma = -1.94 \pm 0.08$; and the total number of particles escaping into interplanetary space, about 2×10^{28} neutrons with energies $E >$ 100 MeV. The theoretical results of [27] indicate that $N_p(E > 30\,\mathrm{MeV}) = 2 \times 10^{33}$ protons are necessary to generate this number of neutrons; this is typical for other solar CRs. The γ-ray burst profile [28] enables one to refine the adopted approximation for the observed flux of solar neutrons during the registration of a large flux of hard photons. The resulting estimate for the duration of the interval of relativistic-neutron

generation is less than 100 s. Therefore, the required fluxes of relativistic particles are generated on the Sun within a few minutes, so that the duration of the flare as a whole is not a determining factor.

Let us now consider the long-lived powerful flares presented in the table. One typical case is the event of June 15, 1991. This magnitude-X12.5 flare began at 8:10 UT, had coordinates 36° N and 70° W, and was characterized by a powerful impulse and the subsequent development of a set of giant loops, which existed until the end of that day. Powerful optical continuum emission was observed at 8:15–8:23 UT in compact arcs about $10''$ in size [29]. The development of bright points near spots is characteristic of white-light flares. One characteristic feature of this particular white-light flare is the presence of a neutral-carbon absorption line, demonstrating that the electron beam penetrated into fairly deep layers of the photosphere.

The enhancement of the flux of energetic protons was almost isotropic, probably due to the extremely perturbed state of interplanetary space at this time. Under such conditions, we can use the diffusive model for propagation of the particles [30], modified for the case of prolonged injection. This model describes the observed temporal profiles of the event fairly well, including the very beginning of the enhancement, which is usually not possible when analyzing solar cosmic rays. The observed profiles and the diffusive approximation for them can be used to derive the integrated fluxes and fluences of particles with various energies and to estimate the total number of accelerated protons in near-Earth orbit. The integrated fluence for energies over 30 MeV was $(4 \pm 2) \times 10^8$ cm^{-2}, and the total number of particles was $N_{tot}(\geq 30\ \text{MeV}) = (1.7 \pm 0.9) \times 10^{33}$.

The injection time is determined fairly accurately for energies of 80–350 MeV. The results for the exit of the proton flux are presented in Fig. 5. Models with instantaneous and continuous injection are compared in [21]. It was found that protons with energies of 175–350 MeV began to leave the Sun at $8{:}23 \pm 0{:}01$ UT, with their emission continuing for 6 ± 1 min. This time coincides with the end of the impulsive microwave burst—the end of the impulsive phase, when the Hα ribbons deviate sharply from each other. In previous studies, this was called the flash phase. The acceleration of protons with higher energies of 1.3–4 GeV takes place at almost the same time, $8{:}25 \pm 0{:}03$ UT, and continues for an even shorter time.

Therefore, the acceleration of relativistic protons with a wide range of energies begins just after the impulsive phase of the flare of June 15, 1991, and continues for several minutes.

Thus, for all the events listed in the table, there are grounds to suspect an additional acceleration of relativistic particles, whose localization and duration are close to those of the impulsive acceleration. Detailed gamma-ray measurements were carried out for some of the flares in the table. They indicate that the durations of the main relativistic-particle acceleration were no longer than a few minutes. This was confirmed for some events from an analysis of data on protons and neutrons detected at the Earth and in interplanetary space.

Recall that we consider here the acceleration of the largest fraction of the relativistic particles. The observation of photons with energies from 30 MeV to a few GeV by the GAMMA-1 telescope beginning on June 15, 1991, after 8:37 UT demonstrated the existence of energetic solar particles at the beginning of the posteruptive energy release. The high level of γ-ray emission within about 10 min after the impulse was a general characteristic of all the powerful flares of June 1991. In their analysis of the GAMMA-1 data for the flare of June 15, 1991, Akimov *et al.* [31] suggested that the second maximum of the microwave burst (at ≈8:30 UT) corresponded to an additional particle-acceleration event rather than particles being confined in a magnetic trap.

It stands to reason that not all of the cases presented in the table had impulses as powerful as that of June 15, 1991. Moreover, the impulsive energy release and the hard burst occur in powerful flares within tens of minutes after their onset, as in the flare of July 14, 2000, and the similar event of November 4, 2001. The character of the white-light flare also varies slightly: instead of ordinary bright points in the penumbra of a spot (or spots), 1600 Å continuum emission by ribbons was observed in the events of July 14, 2000, and November 4, 2001. Such emission in the lower chromosphere could be due to the filling of high coronal loops with relativistic particles and their disappearance in the feet of the arc system.

Concluding this section about powerful flares, we note that, in most cases, it is not necessary to assume that the most efficient acceleration of relativistic particles occurs in more than one process. On August 4, 1991, there were two successive energetic events, as follows from observations of two maxima of the intensity of the 2.22 MeV line [32] and of the continuum emission formed by the decay of neutral π^0 mesons; however, such cases seem to be the exception rather than the rule.

5. CONCLUSIONS

We have carried out a morphological study of the proposed source of relativistic-particle acceleration. This has become feasible due to the availability of

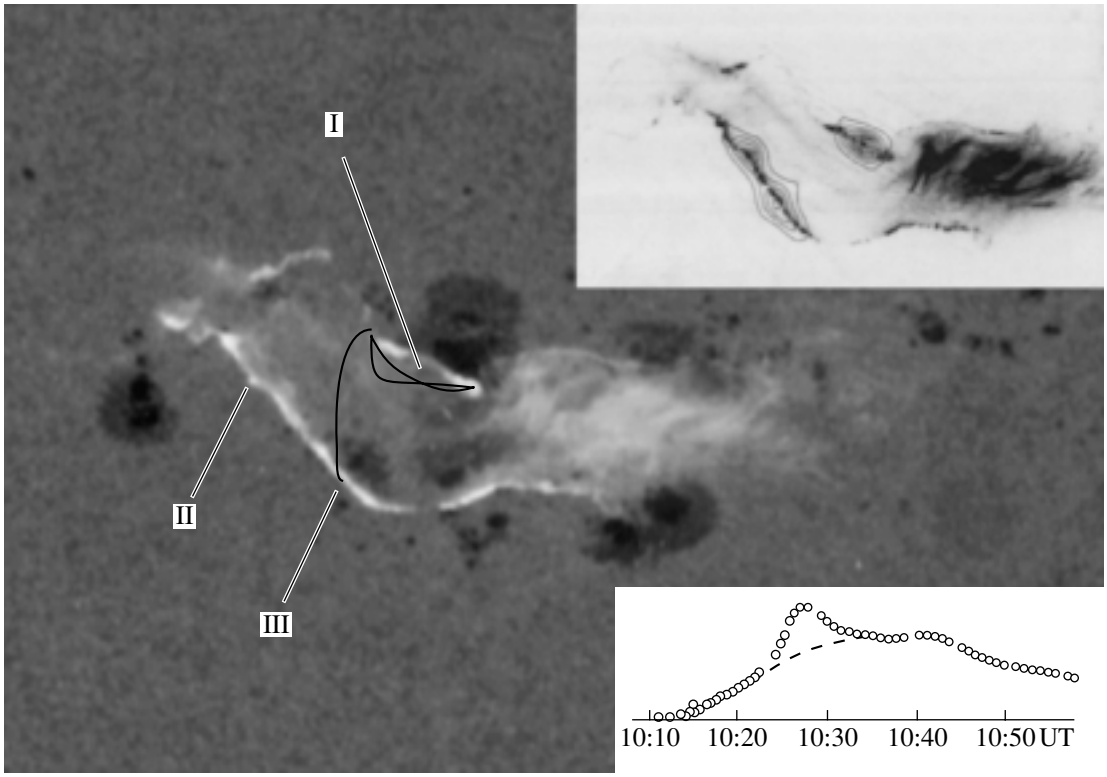

Fig. 6. Schematic comparison of effects near the maximum of the γ-ray burst of July 14, 2000. The ribbons of the flare are superimposed onto the picture of the spots (at 195 Å, and with the brightness of the western part reduced). The positions of sources of hard X-ray emission are marked by the Roman numbers I, II, and III; contours of equal brightness are presented in the top right corner (in accordance with Fig. 8 in [19]). The position of the loops satisfying the calculations [22] of the behavior of the magnetic field lines in the region of hard emission is shown schematically. The brightness variations at 171 Å above the entire flare region are presented in the bottom right corner, in accordance with [20]; the dashed curve denotes the brightness of the same region without the area covered by the hard source I.

observations of hard electromagnetic radiation from nonstationary processes on the Sun with very high temporal and spatial resolution, as well as the direct detection of accelerated particles in interplanetary space and at the Earth. We have presented arguments in support of the idea that, in approximately 30 powerful events occurring over the last three solar cycles, there was an additional source accelerating particles to relativistic energies much more efficiently than the acceleration that occurs in purely impulsive flares and in the posteruptive phase of powerful events.

Modern observations make it possible to localize the source of this efficient acceleration using the fact that the relativistic electrons and protons are accelerated simultaneously. As an example, we discussed the flare of July 14, 2000, in detail in Section 3. The result of this analysis is very simple: the location of the acceleration coincides with the source of the emission at $h\nu = 33$–53 keV at the time of the maximum of the γ-ray burst. To within an accuracy of $3''$, the same location is associated with the emission of an excess number of photons in short-wavelength lines of ions with ionization temperatures of 1.2–1.5 MK. This is illustrated in Fig. 6, which shows that the position of the brightest source I in the M2 band of the HXT hard X-ray telescope touches the eastern edge of the largest spot of the group. This schematic also shows the positions of loops calculated in [20] for both the loop located inside the source I and the higher loop, which connects the source with another flare ribbon (II and III). The relative brightness of source I at 171 Å shows that extra emission appears during the acceleration [20] (see also [19] for 195 Å). This probably implies that a source of intense hard emission is localized at a higher height than the sources of small, impulsive flares, i.e., above the chromospheric Hα bulge over this zone of spots. As follows from the calculated magnetic fields and observations of large active regions at the limb at 171 and 195 Å, this height can be estimated to be 15 ± 5 thousand kilometers. This corresponds to a mean scale between the characteristic sizes of the chromospheric and coronal loops. Therefore, the acceleration really can be more

efficient in a more extended source that remains in a region of strong magnetic fields and field gradients.

In addition, the escape of the accelerated particles and filling of the coronal loops are facilitated by the higher location of the source. This probably initiated the development of the powerful posteruptive phase in the flare of July 14, 2000. The entire sequence of events after 10:24 UT shows that the propagation and drift of the accelerated particle in the magnetic field must be included in flare models. In this particular case, the first brightenings at 195 Å appeared at opposite ends of the northern ribbon, separated by a distance of over 200 000 km, within less than one second. Consequently, the speeds involved must be comparable to the speed of light. Heating near the top of the arcade began within a few seconds, but the arcade decayed into the postflare X-ray loops only approximately 10 min after the large-scale ejection and particle acceleration. Note also that the large-scale disturbance that occurs in some flares is perpendicular to both the field (i.e., the field lines) and the height gradient, i.e., along the drift velocity in the nonuniform magnetic field.

Our analysis demonstrates that there was an additional source of acceleration of energetic particles in all the events listed in the table (except for the event of November 8, 2000), as well as in the powerful flares of February 23, 1956, and August 4, 9, 11, and 15, 1972. At least one characteristic feature of these events seems to be clearly demonstrated: the process took place in groups of spots covering a small area and, for recent flares, just near the penumbra. One dominant acceleration process seems to be sufficient for most prominent proton enhancements in which protons with energies of 1 GeV and large fluxes of protons with energies over 100 MeV were observed. The peculiarities of the observed temporal profiles of the particles can be reasonably explained by the conditions for the subsequent evolution of the beams in the corona and in interplanetary space.

The entire set of accumulated data on proton events makes it possible to outline the basic features of the acceleration process producing large fluxes of relativistic protons. The effective acceleration should take place in a fairly strong magnetic field, which can trap the charged particles during the acceleration. On the other hand, this field must not be extremely strong, so that the accelerated particles can leave after they are generated. The height of the acceleration region cannot be small, since a large density of matter at a small height would be unfavorable for particle acceleration. On the other hand, a height that is too large would also be unfavorable, since the solar magnetic fields become increasingly simple and weak with increasing height.

The solar relativistic "accelerator" operates rarely and only for short periods, during the impulsive phases of the most powerful flares. The duration of the acceleration (or its most efficient phase) does not exceed a few minutes. In other words, the Sun works as a relativistic accelerator only 10^{-6} of the time. This provides evidence that efficient acceleration is associated with some type of rapid changes in the solar magnetic fields, which can be considered catastrophic. The time when the solar accelerator operates coincides with the onset of a powerful ejection of plasma (eruption). The eruptive reconstruction of the solar magnetic field simultaneously accelerates the particles and facilitates their escape from the generation region.

Of course, not every eruption results in the efficient generation of solar cosmic rays. When this takes place, acceleration to relativistic energies occurs only within a small fraction of the extensive eruption zone, namely, in the region where strong magnetic fields with large gradients initially existed. Such conditions are satisfied near spots, i.e., precisely where the discussed flare episode was observed.

Therefore, the generation of energetic particles on the Sun is associated with both CMEs and flares. There is no reason to think CMEs produce the relativistic particles, but we believe that CMEs and energetic solar CRs are formed by the same complex of interconnected sporadic solar phenomena. The relation to flares is even closer than that to CMEs. The acceleration of charged particles is a natural part of a flare, and the accelerated relativistic particles facilitate a faster spatial expansion of the flare process. We believe that our general conclusions are valid not only for the largest but also for ordinary proton events, in which the flux of protons with energies over 100 MeV is small or absent. Of course, the relative contributions of impulsive acceleration, the additional (eruptive) acceleration discussed here, and posteruptive acceleration may be different.

We especially emphasize that the coincidence between the region of acceleration of relativistic particles and the source of hard photons found for the flare of July 14, 2000, seems to occur frequently but not always. An alternative is the situation when another foot of the coronal loop is brighter and more extended in the hard X-ray and γ-ray radiation (in Fig. 6, this corresponds to a larger flux from sources II and III compared to source I). This situation was detected for the first time in a stellar flare observed by the ASTRON satellite, where the response to the action by the accelerated particles on the chromosphere was most powerful after 40 s rather than at the onset of the flare, probably in another foot of the loop. In other words, the conditions for the propagation of the accelerated particles downward, to dense layers just

under the source of the acceleration, may be less favorable, so that hard bremsstrahlung photons will be formed here over less area than in another foot of the loop. If intense emission appears in the foot opposite to the acceleration region, then the propagation of accelerated particles with various energies can lead to some difference in the positions of the photon sources.

A discussion of the theoretical problems related to flares and the acceleration of particles is beyond the scope of this paper. Nevertheless, we emphasize that the crucial point for the theory is whether the additional acceleration takes place along moderate-scale loops or in a strong current flowing along the separatrix. Unfortunately, the calculations of [22] indicate only the possible presence of loops in the hard source I; however, this was not confirmed by observations and cannot be considered to be an established fact. Thus, this problem requires more careful study.

ACKNOWLEDGMENTS

We are grateful to the researchers conducting observations with the TRACE, SOHO, and Yohkoh satellites (the solar images) and the GOES and IMP-8 satellites (data on the solar proton enhancements) who have provided free access to the corresponding results. We thank S.B. Rinchinov for permission to use a figure from his dissertation and I.M. Chertok, V.V. Grechnev, and V.G. Kurt for discussions of this study. This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17086, 03-07-90389, and 04-02-16763).

REFERENCES

1. A. I. Sladkova, G. A. Bazilevskaya, V. A. Ishkov, *et al.*, *Catalogue of Solar Proton Events 1987–1996* (Univ. Press, Moscow, 1998), p. 246.
2. A. V. Belov and E. A. Eroshenko, Radiat. Meas. **26** (3), 461 (1996).
3. A. B. Struminskiĭ, Astron. Zh. **80**, 992 (2003) [Astron. Rep. **47**, 916 (2003)].
4. G. Rank, J. Ryan, H. Debrunner, *et al.*, Astron. Astrophys. **378**, 1046 (2001).
5. E. Rieger, *Solar and Stellar Flares*, Ed. by R. Haisch and M. Rodono (Kluwer, Dordrecht, 1989), p. 323.
6. E. V. Vashenyuk, Doctoral Dissertation in Mathematical Physics (Apatity, 2000).
7. M.-B. Kallenrode, in *Proceedings of the Solar Encounter: The First Solar Orbiter Workshop*, ESA **SP-493**, 23 (2001).
8. B. B. Akimov, N. G. Leikov, A. B. Belov, *et al.*, in *High Energy Solar Phenomena: New Era of Spacecraft Measurements*, AIP Conf. Proc. **294**, 130 (1994).
9. O. V. Terekhov, R. A. Syunyaev, A. V. Kuznetsov, *et al.*, Pis'ma Astron. Zh. **19**, 163 (1993) [Astron. Lett. **19**, 65 (1993)].
10. K. Watanabe, Y. Muraki, Y. Matsubara, *et al.*, astro-ph/0304067 (2003).
11. A. M. Gal'per, B. I. Luchkov, Yu. V. Ozerov, *et al.*, Astron. Zh. **77**, 447 (2000) [Astron. Rep. **44**, 389 (2000)].
12. S. B. Rinchinov, Candidate's Dissertation in Mathematical Physics (Moscow, 2003).
13. K.-L. Klein, G. Trottet, P. Lantos, and Delaboudiniere, Astron. Astrophys. **373**, 1073 (2001).
14. G. H. Share, R. J. Murfhy, A. J. Tylka, *et al.*, Solar Phys. **204**, 43 (2001).
15. A. B. Belov, J. W. Bieber, E. A. Eroshenko, *et al.*, in *Proceedeings of the 27th ICRC* (Copernicus Gesellschaft, Hamburg, 2001), Vol. 8, p. 3446.
16. J. W. Bieber, W. Dröge, P. A. Evenson, and R. Pyle, Astrophys. J. **567**, 622 (2002).
17. I. M. Chertok, V. V. Fomichev, A. A. Gnezdilov, *et al.*, Solar Phys. **204**, 141 (2001).
18. V. I. Ishkov, in *Contemporary Problems of Astrophysics*; Tr. Gos. Astron. Inst. Shternberga **67**, part 2, 264 (2001).
19. L. Fletcher and H. Hudson, Solar Phys. **204**, 71 (2001).
20. M. J. Ashwanden and D. Alexsander, Solar Phys. **204**, 93 (2001).
21. A. V. Belov, E. A. Eroshenko, and M. A. Livshits, in *Proceedings of the Eighth International Symposium on Solar Terrestrial Physics* (Japan, Sendai, 1994), p. 26.
22. Y. Yan, M. J. Ashwanden, S. Wang, and Y. Deng, Solar Phys. **204**, 29 (2001).
23. V. V. Akimov, H. Garcia, K. V. Getman, *et al.*, Astron. Lett. **25**, 42 (1999).
24. A. R. Osokin, A. V. Belov, and M. A. Livshits, Astron. Vestn. (2005, in press).
25. M. A. Livshits, O. G. Badalyan, and A. V. Belov, Astron. Zh. **79**, 659 (2002) [Astron. Rep. **46**, 597 (2002)].
26. A. V. Belov and M. A. Livshits, Pis'ma Astron. Zh. **21**, 42 (1995) [Astron. Lett. **21**, 37 (1995)].
27. R. Ramaty, in *Physics of the Sun*, Ed. by P. A. Sturrock *et al.* (Reidel, 1986), Vol. 2, p. 291.
28. N. Vilmer, A. N. Mac Kinnon, G. Trottet, and C. Barat, Astron. Astrophys. (2004, in press).
29. A. N. Babin and A. N. Koval', Pis'ma Astron. Zh. **18**, 727 (1992) [Sov. Astron. Lett. **18**, 294 (1992)].
30. S. M. Krimigis, J. Geophys. Res. **70**, 2943 (1965).
31. B. B. Akimov, P. Ambroz, A. B. Belov, *et al.*, Solar Phys. **166**, 107 (1996).
32. A. Struminsky, M. Matsuoka, and K. Takahashi, Astrophys. J. **429**, 400 (1994).

Translated by Yu. Dumin

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 678–687. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 746–756.
Original Russian Text Copyright © 2004 by Badalyan, Obridko.

Solar Magnetic Fields and the Intensity of the Green Coronal Line

O. G. Badalyan and V. N. Obridko

Institute of Terrestrial Magnetism, Ionosphere, and Radiowave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia

Received December 25, 2003; in final form, January 9, 2004

Abstract—Synoptic maps of the intensity of the λ530.5 nm FeXIV green coronal line and maps of computed coronal magnetic fields for the period 1977–2001 are compared. For quantitative comparisons, the correlation coefficients r for the correlation between these two parameters at corresponding points of the synoptic maps are calculated. This coefficient exhibits cyclic variations in the spot-formation zone, $\pm 30°$ and the zone above 30° and is in antiphase in these two zones. In the low-latitude zone, the correlation coefficient is always positive, reaches its maximum at activity minimum, and strongly decreases by activity maximum. Above 30°, r reaches maximum positive values at activity maximum and then gradually decreases, passing through zero near the beginning of the phase of activity minimum and becoming negative during this phase. A Fourier analysis of r as a function of time reveals a wavelike variation with a period close to 1.3 yr (known also from helioseismological data for the tachoclinic region of magnetic-field generation), as well as a pronounced wave with a period of about 5 yr. The latitude dependence of r seems to be related to variations in the contributions from local, large-scale, and global fields. Our analysis suggests an approach to studying the complex problem of mechanisms for coronal heating. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

It is quite clear that the activity of the Sun and stars of late spectral types similar to the Sun is undoubtedly related to magnetic fields. In the absence of magnetic fields, stars would be "lifeless" spheres without appreciable manifestations of activity. However, the magnetic fields have different relations to different active processes, and the mechanisms for these relationships have not been fully studied. To elucidate the influence of the magnetic field on physical processes in the corona, the relationships between various indices and the magnetic field must be quantitatively evaluated. Such studies that are also based on reasonably extensive statistical data are small in number.

One very important index of solar activity is the intensity of the so-called green coronal line. This is the FeXIV λ 530.5 nm forbidden line, which is formed in the lower corona at a temperature of ~ 2 MK. Its emission characterizes the level of activity in the solar corona, which is, in turn, related to the activity of underlying layers of the solar atmosphere. The regions of brightest emission in the green line are dense loops and loop clusters, since the intensity of this line is proportional to the square of the density. The existence of such regions is associated with coronal magnetic fields and related to the general problem of coronal heating. Regions of reduced green-line emission are genetically related to coronal holes [1, 2]. Thus, studies of the spatial and temporal distributions of the green-line brightness and comparison of these distributions with parameters of the magnetic field provide a very promising tool for analyzing variations in solar activity and mechanisms for coronal heating.

One big advantage of the index characterizing the green-line emission is that it can be determined almost simultaneously for all heliographic latitudes. This makes it possible to analyze solar activity based on uniform data for the entire solar surface. Thus, this index can provide a uniform and continuous digital field for all points on the disk over long time intervals. In this respect, it stands out, for example, from the Wolf number describing low-latitude activity or the number of polar (high-latitude) facula.

The currently available series of systematic observations of the green coronal line covers almost six solar-activity cycles. These observations have been carried out by a small number of coronal stations, with each of them operating in its own photometric scale. This makes it difficult to combine all available observations into a single system. We will use here the database prepared by Sýkora [3–5].

Since the green-line emission is very closely related to magnetic fields on various scales, studies of the time evolution of the spatial-intensity distribution in the green corona can be used to trace the evolution of coronal magnetic fields. It is especially important in this context that the intensity of the green coronal

1063-7729/04/4808-0678$26.00 © 2004 MAIK "Nauka/Interperiodica"

line can be directly measured, in contrast to the parameters of the coronal magnetic field, which can only be calculated based on photospheric observations applying certain assumptions. We also emphasize that detailed and systematic observations of photospheric magnetic fields cover a substantially shorter time interval than do observations of the green line. Therefore, conclusions derived from joint investigations of magnetic fields and the emission in the green coronal line can be extrapolated to past times when magnetic-field measurements were obtained only episodically and on various scales.

Unfortunately, few direct comparisons between the parameters of coronal-line emission and the magnetic field have been carried out. We note in this context the paper of Guhathakurta *et al.* [6], who made such comparisons based on data for 1984–1992. They calculated the coronal temperature from the ratio of the intensities of the red (λ637.4 nm FeX) and green (λ530.3 nm FeXIV) lines and then compared this temperature with the white-corona emission and the line-of-sight component of the photospheric magnetic field measured at the Kitt Peak Observatory. The temperature distribution obtained in [6] is very different from the magnetic-field distribution. The spatial and temporal distributions of the green- and red-corona brightnesses are more interesting. The distribution of the green-corona emission fairly closely resembles the distribution of the magnetic field. In particular, midlatitude bands in which the magnetic-field strength and coronal-line intensities increase can be noted in the growth phase of the cycle in both hemispheres. As the maximum of the cycle is approached, these bands expand toward both high and low latitudes, with a single equatorial band forming toward the cycle minimum.

A different technique for comparing the emission in the green coronal line with magnetic-field parameters was employed by Wang *et al.* [7], who analyzed data from the SOHO LASCO C1 coronagraph obtained over five days. The structure of the magnetic field lines was calculated in a potential approximation using data of the Stanford Wilcox Solar Observatory. The structures of the magnetic field and green-line emission agreed well. It was shown that the density at the base of magnetic flux tubes is related to the strength of the magnetic field as $n_{foot} \propto \langle B_{foot} \rangle^{0.9}$.

Fairly extensive soft X-ray and EUV observations of the solar disk have now been accumulated. Yohkoh, SOHO, Trace, and CORONAS images of the corona cover a period of about one activity cycle, from 1991. Such observations make it possible to compare coronal images obtained in various X-ray lines with daily observational maps of the magnetic field. However, these spacecraft have different equipment, resulting in nonuniformity of the observational material and, accordingly, in some problems with such comparisons.

The radiation at λ195 Å FeXII recorded on the spacecraft is emitted by virtually the same regions where the green coronal line is emitted. Images taken in this line show that the coronal emission is enhanced above active regions and reduced above coronal holes. Comparisons of green-line synoptic maps with short-wavelength coronal images [8] show that they are in good agreement. Discrepancies seem to be due to transient phenomena that are not reflected in the synoptic maps.

Our previous analysis of the spatial and temporal distributions of the green coronal-line intensity for 1943–2001 [8] confirmed that the green-corona brightness is largely determined by the strength of the coronal magnetic field for this long time interval. There was good agreement in the spatial structures of several synoptic maps of the green-corona brightness and maps of the coronal magnetic field at the heights of the green-line emission during the decline phase of activity cycle 21. This led us to conclude that the green-line intensity and the strength of the coronal magnetic field are, in general, positively correlated at low latitudes. In addition, we tested in [8] a technique for quantitatively estimating correlations between synoptic maps of the green-line intensity and the coronal magnetic field.

Here, we present a detailed comparison between synoptic maps of the green-line intensity and of the magnetic field for 1977–2001 (activity cycles 21, 22, and 23—the current cycle). For quantitative comparisons, we calculated the correlation coefficients between the green-line intensity and magnetic-field strength at corresponding points of the green-corona and magnetic-field synoptic maps. The calculations were done for both the entire latitude range $\pm 70°$ for which the magnetic field is computed and for individual latitude zones. This analysis makes it possible to draw certain conclusions about the extent to which magnetic fields on various scales affect the green-corona emission at various phases of the solar activity cycle. This approach enables us to closely approach a solution to the very complex problem of coronal-heating mechanisms.

2. THE MATERIAL USED

2.1. Database of Intensities of the Green Coronal Line

Monitoring observations of the FeXIV λ 530.5 nm green coronal line were initiated in 1939 at the Pic du Midi and Arosa stations. During the first several years, test observations were carried out, and observational data are available for only a few days

of those years. Later, systematic observations were started on a small network of coronal stations, and the yearly number of observational days increased. However, different stations employed different observational techniques based on inevitably different photometric scales, and carried out their measurements at different heights over the limb. This raised the nontrivial task of reducing the observations of different stations to a unified system. Some issues related to this procedure are considered in detail by Sýkora [3–5].

Several photometrically uniform databases of green-corona intensities are available, which differ mainly in their choice of the coronal station adopted as a photometric standard [1, 9, 10]. We use here the database compiled by J. Sýkora (Slovak Republic); a detailed description of this database can be found in [11]. The observations are represented in the form of an array in which the intensities of the green line along the solar limb are given for each day in position-angle steps of 5°. Thus, synoptic maps with a resolution of $\sim 13°$ in longitude and 5° in latitude can be directly constructed from the observational data. The green-line intensity is expressed in absolute coronal units (acu, one millionth of the intensity at the solar-disk center within 1 Å in the continuum next to the line) and refers to a height of 60″ above the limb.

The database can be used to consider cyclic variations in the green-line intensity both in the entire latitude range from $+90°$ to $-90°$ and for individual latitude zones, with various time resolutions. This enables one to investigate north–west asymmetries with respect to heliographic latitude, study the brightest (active) regions in the corona separately, and examine regions of reduced green-line emission that appear to be related to coronal holes [12]. Analyses of active longitudes and coronal-rotational velocities at various latitudes are also possible [13].

2.2. Calculations of the Coronal Magnetic Field

The coronal magnetic-field strengths were calculated in a potential approximation based on measurements at the photospheric level carried out at the Wilcox Solar Observatory (the data were obtained via the Internet). We used observations of the line-of-sight component of the photospheric magnetic field summarized in synoptic maps for each Carrington rotation as the source data.

We calculated the coronal magnetic field using the widely known method described in [14, 15] and employed a code that allows the calculation of all components of the magnetic field everywhere from the photospheric surface to the source surface [16, 17]. Synoptic maps of these components can be calculated for any selected time that is taken as the position of the central meridian. For the subsequent analysis, we calculated the total magnetic-field strength B (the square root of the sum of the squares of the radial, B_r, and tangential, B_t, components) at a height of $1.1R_\odot$, which roughly corresponds to the data for the green coronal line. The calculations summed over ten harmonics and introduced a polar correction to take into account the lower reliability of magnetic-field measurements near the poles [18]. Both the original and calculated data on the magnetic field are restricted to the latitude range $\pm 70°$.

3. COMPARISONS BETWEEN THE SYNOPTIC MAPS OF THE GREEN-LINE INTENSITY AND CORONAL MAGNETIC FIELDS

Cyclic variations in the brightness distribution of the green coronal line were considered in [8] based on synoptic maps for 1943–2001 subjected to a running average over six Carrington rotations with a step of one rotation. Such synoptic maps make it possible to trace gradual changes in the spatial distribution of the brightness of the green coronal line during the solar activity cycle. Smoothing over six Carrington rotations enables more reliable identification of long-lived features on fairly large scales.

Similar synoptic maps were also constructed for the magnetic-field strength. The maps calculated for each Carrington rotation were averaged over six rotations with a step of one rotation. All the calculations were done for a distance of 1.1 $R_\odot$.

Figure 1 presents three pairs of synoptic maps for the magnetic-field strength in μT (upper row) and the green-line intensity in acu (lower row). In a given map, the chosen range for the quantity displayed is divided into eight gradations, with the largest values shown as black, and the smallest, as white. Not all eight gradations are always shown in a map, in order to make clearer cyclic variations in the parameters studied (for example, the decreases in the green-line brightness at the activity minimum). In particular, white is missing in the first and second maps of the lower row (green line), and the two highest gradations of black are missing from the third map of the same row. The contours delimiting the darkest areas correspond to 240, 480, and 280 μT in the upper row of maps (from left to right) and to 150, 170, and 40 acu in the lower row of maps. Thus, the contour increments are 30, 60, and 35 μT for the maps in the upper row and 18.75, 13.75, and 5 acu for the maps in the lower row. The horizontal and vertical axes plot heliographic longitude and latitude, and time runs from left to right. The heading over any pair of maps indicates the corresponding averaging interval.

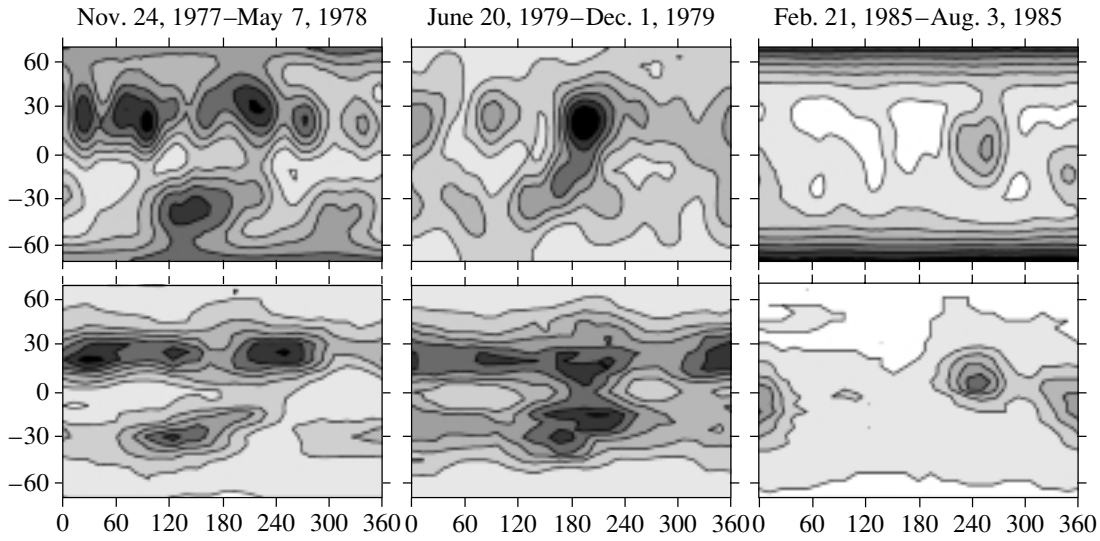

Fig. 1. Comparison of synoptic maps of the magnetic-field strength (top) and green-line intensity (bottom). The magnetic-field strength is measured in μT, and the green-line intensity, in acu. In each map, the total range of the corresponding quantities is divided into eight gradations, so that the largest values are shown as black, and the smallest, as white. The contour increments (from left to right) are 30, 60, and 35 μT for maps in the top row and 18.75, 13.75, and 5 acu for maps in the bottom row. Longitude is plotted along the horizontal axis, and latitude, along the vertical axis, and time runs from right to left.

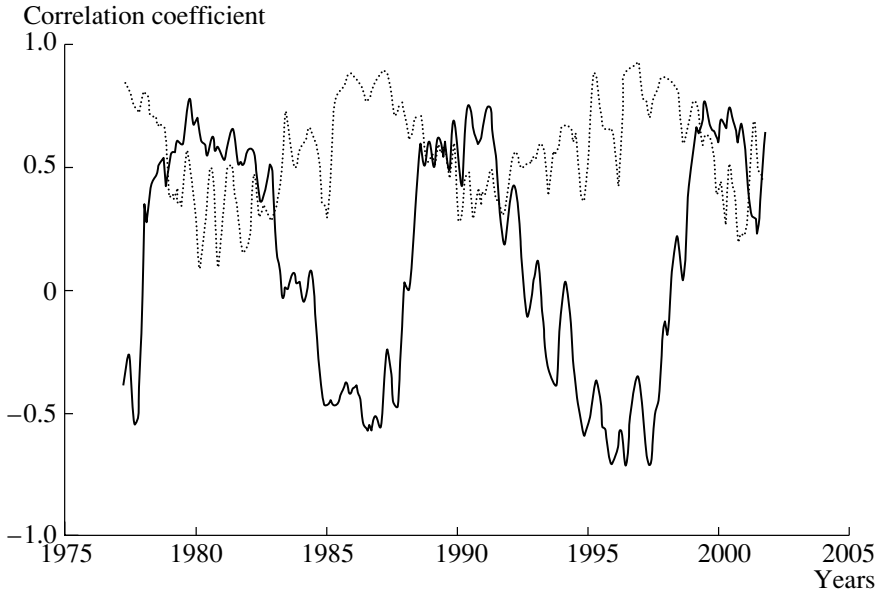

Fig. 2. Time variations in the correlation coefficient r for the correlation between the magnetic-field strength and the intensity of the green line measured at spatially coincident points of the maps in the spot-formation zone, $\pm 30°$ (dotted curve), and at high latitudes, above 30° (solid curve).

The maps in Fig. 1 were constructed for various phases of cycle 21. The left pair of maps refers to the growth phase of the activity cycle. We can see that the green-line intensity and the magnetic-field strength are generally in good agreement. Regions of both large and small values of the corresponding parameters coincide well, especially in the low-latitude spot-formation zone. The next pair of maps corresponds to a time interval near the cycle maximum. Here, the agreement between the maps is worse for the spot-formation zone and better for the higher latitudes. The third pair of maps was constructed for a period near the minimum of the activity cycle. The intensity of the green coronal line is substantially reduced, and the magnetic field is much more uniform (less structured). However, the rela-

tively rare regions of enhanced magnetic fields in the low-latitude zone (sometimes, there is only one such region) normally coincide with regions of enhanced green-line emission. At the same time, a negative correlation between the green-line intensity and the magnetic-field strength can be noted at high latitudes. The field strength increases toward the poles, substantially exceeding the B values in the equatorial zone; in contrast, the green-line intensity decreases toward the poles. Note that, combined with our Fig. 1, the three pairs of maps for the declining phase of cycle 21 presented in [8] suggest gradual changes in the correspondence between the parameters studied at low and high latitudes during the activity cycle.

To quantify similarities or differences between the synoptic maps, it was suggested in [8] to compute cross-correlation coefficients for corresponding points in maps of the green corona and magnetic fields. For this purpose, the magnetic-field strengths were calculated for those points in the synoptic map for which data on the green line were available; there are 27 such points in longitude (one rotation) and 29 points in latitude (in steps of 5° in the range ±70°, where the magnetic field was calculated). Such cross correlations can be determined both for the entire range of latitude ±70° and for selected, narrower latitude zones.

Even the first estimates based on this technique [8] indicated that the intensity of the green-corona emission is closely related to the magnetic-field strength. However, this relationship depends strongly on the latitude zone and phase of the cycle. The correlation coefficient and even its sign are different in the polar zone and in the low-latitude spot-formation zone. It was concluded that the correlation coefficient for the high-latitude corona (above 30°) increases near the cycle maximum, when the magnetic field at high latitudes is relatively low. Further, as the minimum is approached, the correlation coefficient gradually decreases and changes its sign. This can be seen in Fig. 1.

Here, we present a much more detailed comparison for all available synoptic maps of the green-line intensity and magnetic field. This enables us to trace cyclic changes in the relationship between the green-line intensity and magnetic-field strength in various latitude zones.

Figure 2 displays the time variations of the correlation coefficient, r, for two latitude zones—the spot-formation zone ±30° and a high-latitude zone. The high-latitude zone refers to regions with latitudes exceeding 30° in both the northern and southern hemispheres. First and foremost, we can see that r exhibits cyclic variations in both zones. Further, the r values in the two zones seem to vary in antiphase. In the low-latitude zone, r is always positive, reaching a maximum at the activity minimum and strongly decreasing by the activity maximum. The situation is different at latitudes above 30°. The correlation coefficient r reaches its maximum positive values at the maximum of the activity cycle, after which it gradually decreases, passes through zero, and, by the activity minimum, reaches negative values whose absolute magnitudes are nearly equal to the positive values achieved at activity maximum. At the same time, the absolute values of r at the high latitudes never become as large as they do in the spot-formation zone near minima of the solar-activity cycle. The cross-correlation coefficients for the maps presented in Fig. 1 are (from left to right) 0.812, 0.574, and 0.800 for the zone ±30° and 0.274, 0.755, and −0.458 for the zone above 30°.

The sign change of r at high latitudes coincides with the reference points of the cycle t_{mA} (the beginning of the growth phase) and t_{Dm} (the end of the decline phase and beginning of the minimum phase) to within a few months, so that negative values of r fairly precisely border the phase of minimum activity. (We use here the calendar dates for the reference points determined in [19].)

The spatial and temporal variations in r over the entire latitude range ±70° are shown in Fig. 3. All features of the variations of r with time and latitude can be traced clearly in this diagram. The regions of relatively high positive correlation coefficients, $r \geq 0.5$, span a fairly broad latitude range during periods of solar-activity decline and alternate with small intervals where this r is reduced, also over a broad range of latitudes. The increases in size of the polar region with negative r values occur in a similar wavelike fashion during periods that are close to the solar-activity minimum. It is interesting that there is a latitude zone, 30°−40°, where r is close to 0.5 and exhibits virtually no cyclic variations. This means that zones displaying different behavior of the correlation coefficient are roughly separated by a latitude belt of quiescent filaments.

The nature of the latitude dependence of the correlation coefficient r is probably related to variations in the contributions of the local, large-scale, and global fields, each having their own dependence on the phase of the cycle.

4. DEPENDENCE OF THE CORRELATION COEFFICIENT ON THE PHASE OF THE CYCLE

The time interval for which the cross correlation between the synoptic maps of the green-line intensity and the coronal magnetic-field strength can be studied (in other words, the interval for which the magnetic field can be calculated from the Stanford data)

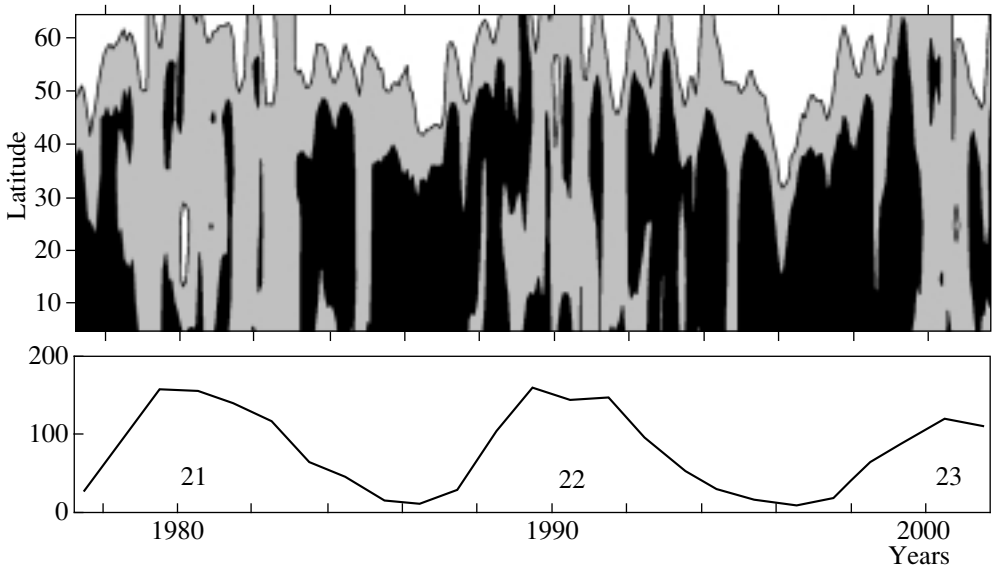

Fig. 3. General latitude–time diagram for the correlation coefficient r. Regions with $r > 0.5$ are shown black; those with $-0.5 < r < 0.5$, gray; and those with $r < -0.5$, white. The bottom panel plots the mean annual Wolf numbers along the vertical axis; the numbers of the corresponding activity cycles are indicated.

encompasses activity cycles 21, 22, and (partially) 23. It is interesting to trace changes in r from cycle to cycle. This can be done by superposing the time variations in r for various cycles on one graph.

Figure 4 shows such a superposition for latitudes $\pm 30°$. The time shift was chosen so that the coincidence (pairwise correlation) of all three curves was best. The year is plotted along the horizontal axis on a scale in which, for the first cycle considered (cycle 21, heavy solid curve), zero corresponds to 1976, and 12, to 1988. The next two cycles (22 and 23) are shifted by one and two chosen time shifts, respectively.

Both Fig. 2 and the general map in Fig. 3 clearly display a certain quasiperiodic component superposed on the overall time variations in r. The Fourier analysis described in detail below shows that this component has a period close to 1.5 yr. The effect of this wave is fairly significant for our choice of the best time shift in Fig. 4. We ultimately took this shift to be 133 Carrington rotations, slightly less than 10 yr. We can see in Fig. 4 both the generally good coincidence in the cyclic variations of r and the agreement of many features associated with the high-frequency component. The correlations between the curves in Fig. 4 for cycles 21 and 22, 21 and 23, and 22 and 23 calculated in the latitude zone $\pm 30°$ for the chosen time shift are 0.755, 0.768, and 0.728, respectively.

We carried out a Fourier analysis to reveal quasiperiodic variations in r more reliably. Calculations were done separately for ten-degree latitude zones and for the entire latitude range. The green-corona brightness and the magnetic-field strength themselves, averaged over six rotations for latitudes $\pm 30°$ and above 30°, were also subjected to a Fourier analysis. The sum of all the squares of the Fourier amplitudes is equal to the standard deviation of the correlation coefficient r from its mean over the time interval studied. The Fourier analysis confirmed the presence of a high-frequency wave (with a period of 1–1.5 yr) in the r values for all latitudes, as well as a fairly pronounced 5-yr wave. Oscillations with periods of 1–1.5 yr and 5 yr are distinctly visible in both r and the indices themselves.

Figure 5 shows a Fourier periodogram for r for all latitudes and the period range 0.5–3 yr. The range of the amplitude variations for oscillations of r was chosen to be 0.015–0.06 and is divided into six gradations; thus, the contour increment in Fig. 5 is 0.0075.

Oscillations with periods of 1–1.5 yr are clearly visible in Fig. 5; oscillations with periods of about 1 yr dominate at high latitudes, while 1.3-yr oscillations dominate at medium and low latitudes. Note that this same period, 1.3 yr, was detected helioseismologically in the tachocline region of the magnetic-field generation [20]. The 1.3-yr oscillations are also distinct at the tachocline level at medium and low latitudes, while these oscillations seem to be weaker or to have a shorter period at high latitudes. The well-known quasi-biennial oscillations are very weak in Fig. 5. Since we are analyzing the correlation coefficient for correlations between the intensity of the green coronal line and the magnetic-field strength, it is natural that not all significant oscillations in these activity indices will be manifest in the correlation between them. In our case, the quasi-biennial oscillations can be easily distinguished in the magnetic-field strength

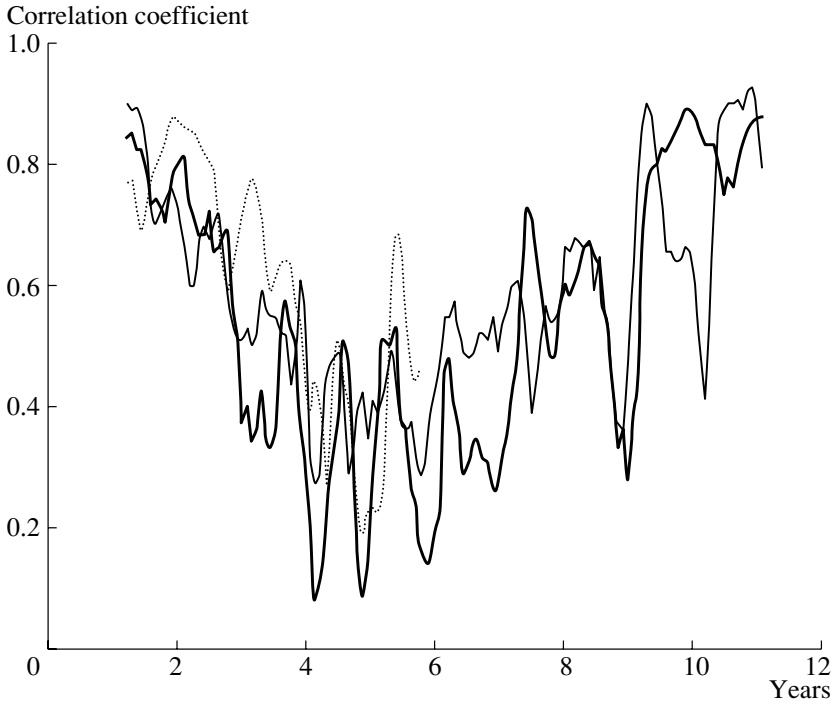

Fig. 4. Time variations in the correlation coefficient r during cycles 21 (heavy solid curve), 22 (light solid curve), and 23 (dotted curve) for latitudes $\pm 30°$. The year is plotted along the horizontal axis on a scale in which zero corresponds to 1976 and 12 corresponds to 1988 for cycle 21, and the scales for cycles 22 and 23 are shifted by 133 and 266 rotations, respectively. The shift by 133 Carrington rotations is chosen to achieve the best pairwise correlation of all three curves. A quasi-annual wave superposed onto the cyclic variations in r is clearly visible in all three curves.

and green-line intensity. These oscillations are even more pronounced in the asymmetry of various solar-activity indices [21, 22].

In contrast to periods of 1–1.5 yr, oscillations in r with periods of about 5 yr gradually increase toward high latitudes (the low-frequency region is not shown in Fig. 5). At middle latitudes, oscillations with this period form a distinct spectral peak.

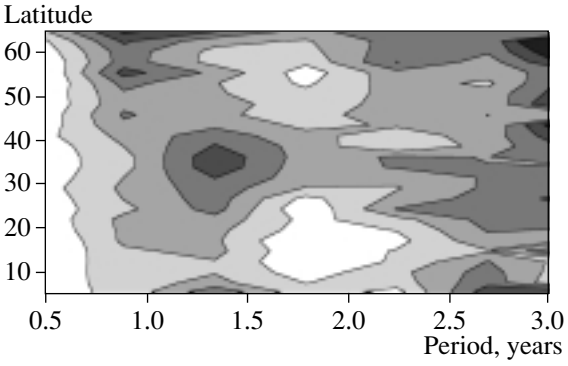

Fig. 5. Fourier periodogram of the correlation coefficient r for periods from 0.5 to 3 yr. The contour increment for variations in the amplitude of r is 0.0075. Oscillations in the range 1–1.5 yr are clearly visible, while the quasi-biennial oscillations are very weak.

At higher latitudes, this peak gradually changes on its low-frequency side into a general increase in the oscillation amplitude right up to the 11-yr period. At low latitudes, the 5-yr oscillations are relatively weak, and the oscillation amplitude gradually grows at low frequencies.

The low-frequency part of the Fourier spectrum of the correlation coefficient is substantially weakened at latitudes of 30°–40°. This agrees with the fact that, as was noted above in the discussion of Fig. 3, the 11-yr cycle is weakly expressed in the behavior of the correlation coefficient at these latitudes.

Figure 6 represents r for the correlation between the coronal magnetic-field strength and the intensity of the green coronal line as a function of the cycle phase. The cycle phase here is calculated according to [23], as $\Phi = (T - m)/(|M - m|)$, where T is the current time and M and m are the times of the nearest maximum and minimum of the 11-yr cycle, respectively. According to this definition, the phase is zero at the cycle minimum, positive on the ascending branch of the cycle, and negative on the descending branch.

Figure 6 most clearly displays overall variations in r with the cycle phase for the spot-formation zone (upper curve, open circles) and the zone above 30° (lower curve, solid circles). All the features of the time

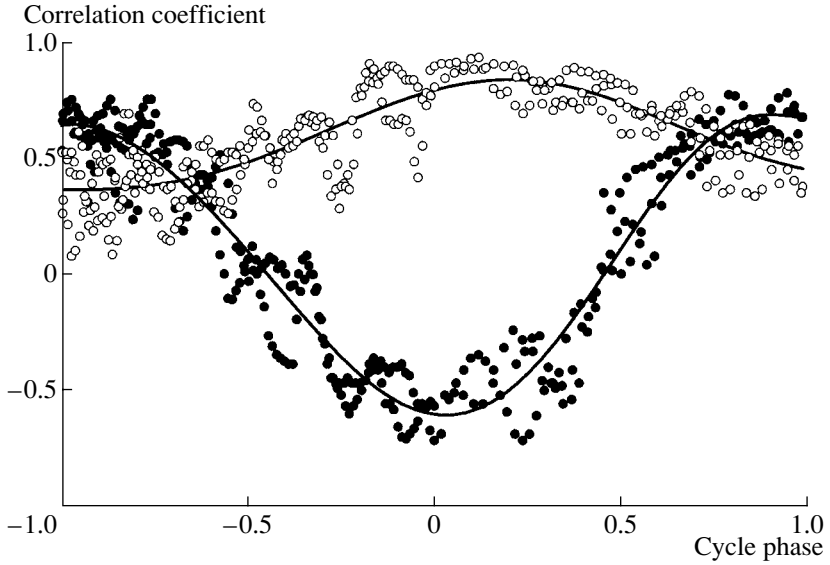

Fig. 6. Variation in the correlation coefficient r with phase of the cycle for the spot-formation zone (upper curve, open circles) and the zone above 30° (lower curve, solid circles).

variations of r in these two latitude zones discussed above are clearly visible in this figure. The mean correlation coefficient in the spot-formation zone varies from ≈ 0.35 at the activity maximum to ≈ 0.85 at the minimum, and it varies in the high-latitude zone from ≈ 0.6 to ≈ -0.6. The effect of the high-frequency wave is also appreciable, leading to corresponding departures of the actual variations in r from the solid curves approximating the cyclic variations in r.

Figure 7 presents the relationship between r and the magnetic-field strength itself at a height of 1.1 $R_\odot$ for latitudes ±30° (upper panel) and latitudes above 30° (lower panel). The following features can be noted. Most importantly, there is a decrease in r with increasing magnetic-field strength in the spot-formation zone and, vice versa, an increase in r with B at latitudes above 30°. The magnetic field itself in the spot-formation zone varies dramatically during a cycle, by a factor of 10–15, and approaches values $B \sim 20$ μT at the cycle minimum. Recall that this is the field averaged over six Carrington rotations for the given latitude zone. At higher latitudes (lower graph), the mean magnetic field varies much less during the cycle, by only a factor of 1.5–2, reaching 70–100 μT at the activity maximum (negative values of r) and no more than 200 μT at the activity minimum.

A joint consideration of Figs. 6 and 7 clearly shows that the green-line intensity in the spot-formation zone is most closely related to the magnetic-field strength near the activity minimum. The field in the spot-formation zone decreases and becomes more ordered, and the enhancement of the field in some regions results in a corresponding enhancement of the green-line emission. In contrast, the field at high latitudes increases as the minimum is approached, while the intensity of the green line decreases (Fig. 1).

At activity maximum, when the low-latitude magnetic field is strong, there is no such relation, as is reflected by a substantial decrease in r. We should emphasize that the general cyclic variations in both the green-line intensity and the magnetic-field strength in the spot-formation zone correspond well with the well-known Wolf-number curve. In contrast, the field strength for latitudes above 30° is in antiphase with the Wolf numbers, while the intensity of the green line continues to follow the cyclic Wolf-number curve.

Thus, Fig. 6 and, especially, Fig. 7 indicate that the relationship between the green-line intensity and the coronal magnetic field is not determined by a single universal law. The magnetic fields on various scales influence the brightness of the green corona in different ways.

5. CONCLUSIONS

We have made a detailed comparison between synoptic maps of the intensity of the λ530.5 nm green coronal line and the strength of the coronal magnetic field for 1977–2001 (activity cycles 21, 22, and 23, the current cycle). Maps of the green coronal-line intensity were constructed using a photometrically uniform database [3, 11], applying a running average over six Carrington rotations in steps of one rotation, making it possible to reveal large-scale, long-lived features. The magnetic-field strength was calculated from measurements of the line-of-sight field component at the photospheric level carried out at the Stanford John Wilcox Observatory. The calculations

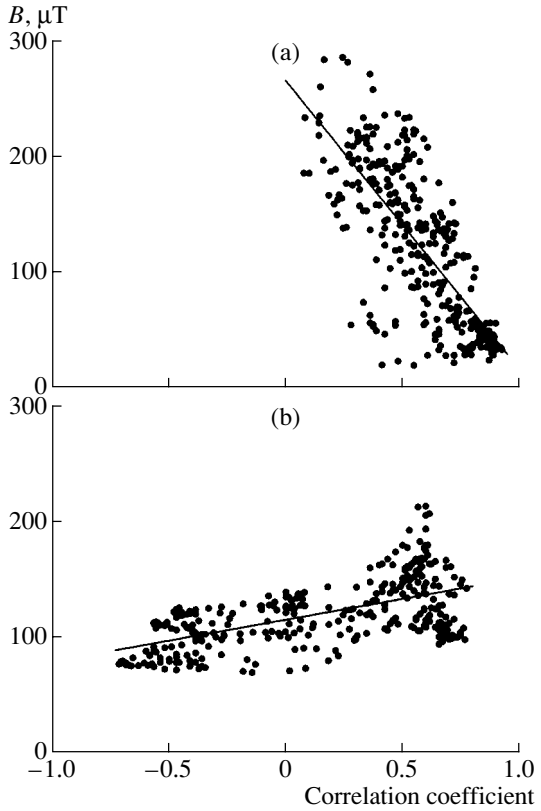

Fig. 7. Relationship between r and the magnetic-field strength in the spot-formation zone (upper panel) and the zone above 30° (lower panel).

were done for a distance of 1.1 $R_\odot$ in a potential approximation for latitudes within ±70°, and the resulting synoptic maps were likewise averaged over six rotations with a step of one rotation.

To quantify the comparison of the green-line intensity and magnetic-field strength in each pair of maps, we calculated the correlation coefficient r for the correlation between these parameters at points of the synoptic maps for which measurements of the green-line intensity were available. This correlation coefficient was calculated for the entire latitude range ±70°, for narrow ten-degree zones, and for the spot-formation zone ±30° and high latitudes above 30°.

We found that r varies cyclically in zones of both low and high latitudes. There are virtually no such cyclic variations between these zones, at latitudes of 30°–40°, close to the zone of quiet filaments. The correlation coefficients r in the spot-formation zone ±30° and at latitudes above 30° vary in antiphase. In the low-latitude zone, r is always positive, reaches maximum values at the activity minimum, and decreases strongly by the activity maximum. In contrast, at latitudes above 30°, r reaches its maximum positive values at the maximum of the activity cycle, after which it gradually decreases and becomes negative by the phase of activity minimum.

The correlation coefficient r increases with the magnetic-field strength in the spot-formation zone and decreases at latitudes above ±30°. The magnetic-field strength averaged over six Carrington rotations varies by a factor of 10–15 in the spot-formation zone and reaches values of $B \sim 20\ \mu$T at the activity-cycle minimum. At latitudes above ±30°, the magnetic field varies by, on average, only a factor of 1.5–2 during the activity cycle, always remaining in the range 50–60 μT at the activity maximum and not exceeding 200 μT at the phase of activity minimum.

A Fourier analysis of the time variations in r reveals a wave with a period close to 1.3 yr; the period of this wave is reduced somewhat at higher latitudes. Helioseismological data likewise indicate 1.3-yr variations in the tachocline region of magnetic-field generation at middle and low latitudes. A 5-yr period also stands out in the Fourier spectrum of the oscillations of r. The amplitude of the oscillations is considerably reduced at latitudes of 30°–40°, while the 11-yr cycle is clearly expressed at higher and lower latitudes.

The dependence of r on the cycle phase provides direct evidence for differences in the influence of fields on various scales on the coronal intensity. As can easily be shown, a universal linear relationship between the magnetic field and the coronal intensity would imply that there should be no dependence of r on the mean magnetic-field strength and, therefore, on the phase of the cycle. If the relationship is nonlinear, r can vary, but only within narrow limits. In our case, however, not only cyclic variations but also changes in the sign of r are observed at high latitudes. This testifies to the action of different mechanisms for the formation of the corona for magnetic fields on small, medium, and large scales. Our results can be used to quantitatively test various models for coronal heating.

The simplest pattern is observed at the cycle minimum, when the large-scale, quasi-dipolar magnetic field at the poles of the Sun reaches a maximum. Polar coronal holes are present in the corona. The coronal intensity and magnetic-field strength at the poles vary in antiphase. In the equatorial zone, the large-scale field virtually disappears, and only isolated complexes of local activity remain. The relationship between the coronal brightness and magnetic fields in the equatorial zone is especially clearly expressed during this period, and the correlation coefficient r reaches a maximum.

The situation is much more complex at activity maximum. Local magnetic fields spread to relatively high latitudes, and tongues of large-scale field (accompanied by coronal holes) penetrate into the equatorial region. Fields on intermediate characteristic scales (the background and floccular fields, loops over

large-scale neutral lines) are clearly expressed. Furthermore, ephemeral active regions living one to three days, which affect the physical processes occurring in the corona but are not reflected in synoptic maps of the magnetic field, as well as nanoflares, are manifest most effectively at the phase of activity maximum. This substantially reduces r in the equatorial zone during the period of cycle maximum.

Thus, the relationship between the magnetic field and the brightness of the green corona and, therefore, the mechanisms for coronal heating seem to be different for fields on different scales. In relatively small-scale fields (local fields) at low latitudes, the heating of the corona is controlled by low flux tubes and nonstationary processes in and near these tubes. Because the magnetic field at the cycle maximum is a complex mixture of arches and loops with various heights and a very complex configuration, the correlation coefficients r cannot be large, even though the brightness of the green corona increases with the field strength. Such a mixture of variously scaled fields is also present at higher latitudes. As the minimum is approached, the field structure becomes more ordered at both high and low latitudes, individual activity complexes emerge in the spot-formation zone, and extended regions of large-scale field appear at high latitudes. The correlation coefficient r increases everywhere, but the relationship between the magnetic-field strength and the green-line intensity is opposite in sign for these latitude zones.

This simple scheme describes the general cyclic behavior of the relationship between the green coronal-line emission and the magnetic-field strength but cannot naturally explain some details. The interpretation of a number of finer effects revealed by our analysis must be elaborated more carefully and in greater detail.

ACKNOWLEDGMENTS

We are grateful to J. Sýkora for discussions of certain points touched upon in this paper. This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16199) and INTAS (grant 2000-840).

REFERENCES

1. J. Sýkora, Solar Phys. **140**, 379 (1992).
2. V. Letfus, L. Kulčár, and J. Sýkora, in *Solar and Interplanetary Dynamics*, Ed. by M. Dryer and E. Tandberg-Hanssen (Reidel, Dordrecht, 1980), p. 49.
3. J. Sýkora, Bull. Astron. Inst. Czec. **22**, 12 (1971).
4. J. Sýkora, Contrib. Astron. Obs. Skalnaté Pleso **22**, 55 (1992).
5. V. Lelfus and J. Sýkora, *Atlas of the Green Corona Synoptic Charts for the Period 1947–1976* (Veda, Bratislava, 1982).
6. M. Guhathakurta, R. R. Fisher, and R. C. Altrock, Astrophys. J. **414**, L145 (1993).
7. Y.-M. Wang, N. R. Sheeley, Jr., S. H. Hawley, *et al.*, Astrophys. J. **485**, 419 (1997).
8. O. G. Badalyan, V. N. Obridko, and J. Sýkora, Solar Phys. (submitted).
9. R. N. Ikhsanov and V. G. Ivanov, in *Proceedings of the Conference on The Sun During the Magnetic Field Mark Change*, Ed. by V. I. Makarov and V. N. Obridko (Gos. Astron. Obs., 2001), p. 175.
10. A. G. Tlatov, S. A. Guseva, and Kim Gunder, in *Proceedings of the Conference on The Sun During the Magnetic Field Mark Change*, Ed. by V. I. Makarov and V. N. Obridko (Gos. Astron. Obs., 2001), p. 385.
11. O. G. Badalyan, V. N. Obridko, and J. Sýkora, Solar Phys. **199**, 421 (2001).
12. L. Kulčár and J. Sýkora, Contrib. Astron. Obs. Skalnaté Pleso **24**, 79 (1994).
13. J. Sýkora, Adv. Space. Res. **14** (4), 73 (1994).
14. J. T. Hoeksema and P. H. Scherrer, *The Solar Magnetic Field—1976 through 1985*, WDCA Report UAG-94, NGDC, Boulder (1986).
15. J. T. Hoeksema, *Solar Magnetic Fields—1985 through 1990*, Report CSSA-ASTRO-91-01 (1991).
16. K. G. Ivanov and A. P. Kharshiladze, Geomagn. Aeron. **34**, 22 (1994).
17. J. Sýkora, O. G. Badalyan, and V. N. Obridko, Solar Phys. **212**, 301 (2003).
18. V. N. Obridko and B. D. Shelting, Solar Phys. **184**, 187 (1999).
19. V. N. Obridko and B. D. Shel'ting, Astron. Zh. **80**, 1034 (2003) [Astron. Rep. **47**, 953 (2003)].
20. R. Howe, J. Christensen-Dalsgaard, F. Hill, *et al.*, Science **287**, 2456 (2000).
21. O. G. Badalyan, V. N. Obridko, J. Rybák, and J. Sýkora, in *Proceedings of the ISCS 2003 Symposium on Solar Variability as an Input to the Earth's Environment*, ESA SP-535, Ed. by A. Wilson (Noordwijk, Netherlands, 2003), p. 63.
22. O. G. Badalyan and V. N. Obridko, in *Proceedings of the Conference on Climatic and Ecological Dimensions of Solar Activity*, Ed. by V. I. Makarov and V. N. Obridko (Gos. Astron. Obs. Ross. Akad. Nauk, St. Petersburg, 2003), p. 33.
23. S. A. Mitchell, Handb. Astrophys. **4**, 231 (1929).

Translated by A. Getling

Astronomy Reports, Vol. 48, No. 8, 2004, pp. 688–698. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 8, 2004, pp. 757–768.
Original Russian Text Copyright © 2004 by M. Eselevich, V. Eselevich.

Additional Plasma Streams with Steep Fronts in Structures of Enhanced Brightness in the Coronal Streamer Belt

M. V. Eselevich and V. G. Eselevich

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

Received January 28, 2003; in final form, March 15, 2004

Abstract—The presence of additional plasma streams with enhanced density in bright rays of the solar coronal-streamer belt is demonstrated. The streams have steep fronts, whose widths $\delta \approx 0.10\ R_\odot$ (where $R_\odot$ is the solar radius) are comparable to the spatial resolution of the LASCO C3 instrument of the SOHO satellite. The additional streams are similar to streams of the slow, quasi-steady solar wind in the streamer belt in terms of their plasma density, directional velocity, and lifetime and are apparently one of the main sources of the slow solar wind. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Our analysis of uncalibrated solar-corona observations made with the SOHO/LASCO C2 and C3 instruments (processing level L0.5) [1] shows that, in the absence of coronal mass ejections, the streamer belt is made of a set of radial rays of enhanced brightness. The minimum angular diameters of individual rays are $d \approx 2.0°–3.0°$. Both the quasi-steady solar wind, whose parameters vary slowly, on characteristic time scales of up to 10 days or more, and sporadic (unsteady) plasma motions with characteristic lifetimes of several hours are present in the rays [2]. Such sporadic plasma motions include (a) matter inhomogeneities that randomly appear in some rays moving away from the Sun, known as "blobs" [3, 4], and (b) recently discovered sporadic plasma streams directed either away from or toward the Sun, which result from the disintegration of a streamer—presumably due to magnetic-field-line reconnection [5, 6]. In some cases, the formation of the quasi-steady solar wind includes a stage when an additional plasma stream with enhanced density directed away from the Sun fills a separate ray [2]; this stage is also associated with the unsteady solar wind. The lifetimes of the additional streams vary from several hours to several days. Therefore, some can be considered sporadic and others quasi-steady streams of the solar wind. Preliminary studies have shown that the leading fronts of such streams can be extremely narrow. Our aim is to verify the existence of additional streams with steep fronts in rays of the streamer belt based on an analysis of calibrated data from the SOHO/LASCO C2 and C3 instruments (processing level L1).

2. SOURCE DATA. IDENTIFICATION OF RAYS IN THE STREAMER BELT. ANALYSIS TECHNIQUE

The source data for our analysis were white-light images of the solar corona obtained with the LASCO C2 and C3 telescopes on the SOHO spacecraft. We analyzed calibrated 1024×1024 images using processing level L1 (see http://lasco-www.nrl.navy.mil/). These data are free of scattering, vignetting, etc., and the intensities are given in units of the mean solar-disk intensity (P_{mbs}). The C2 coronagraph yields images of the white corona in the region $R = 2–6\ R_\odot$ and the C3 coronograph in the region $R = 3.7–30\ R_\odot$.

Using these data, a narrow ray ($d \approx 2°–3°$) in the streamer belt can easily be identified visually in coronal images within the field of view of the coronagraph if the ray is located in a section of the streamer belt that is stretched along a meridian [7]. It is possible to identify and trace individual rays in other sections of the streamer belt over fairly long times due to the following important property of the images of such rays in the plane of the sky. The apparent latitude of the ray in the plane of the sky (the angle between the visible image of the ray in the plane of the sky and the equatorial plane), Λ, varies with the position of the ray relative to the plane of the sky (the longitude of the ray) as the Sun rotates [8]. The law for this variation depends on the real latitude of the ray on the Sun (i.e., the angle between the ray and the solar equatorial plane), λ, and the heliographic latitude of the center of the Sun (or the heliographic latitude of the Earth), B_0. Accordingly, the ray will describe some curve in a synoptic map (usually an arc with its

bulge facing the equator). The equation of this curve (in terms of latitude and Carrington longitude) as a function of λ was obtained by Hundhausen [8] for the particular case of zero heliographic latitude of the Earth, $B_0 = 0°$, and by us [1] for $B_0 \neq 0°$.

If $\lambda \neq 0°$, then, due to this effect, rays aligned with a parallel and, therefore, merging into one ray become distinguishable as they move away from the plane of the sky. The rays end up at different angular deviations from the plane of the sky and, accordingly, have different apparent latitudes Λ. A visual analysis of synoptic maps of the coronal-brightness distribution based on LASCO data and available via the Internet showed that the most pronounced rays revealed in this way are located at the tip of the bend in the streamer belt—the point most distant from the solar equator in the northern or southern direction [2]. Rays in belt sections elongated in longitude (i.e., parallel to the solar limb) are less intense but also clearly separated when they pass near the western or eastern limb. Our technique for extracting quantitative information from white-light coronal images included the following operations. For each image of the white-light corona obtained using the C2 or C3 coronograph, we constructed distributions of the coronal intensity $P(\Lambda)$ or $P(R)$ for a fixed heliocentric distance R or apparent latitude Λ, respectively, separately for the eastern or western limb. We represent P in units of P_{msb}.

We studied the dynamics of the $P(\Lambda)$ and $P(R)$ distributions with time t. We characterized the brightness of an individual ray using P_R—the "amplitude of the ray brightness," which was defined in [1]. The introduction of the concept of ray brightness P_R is made possible by the fact that a ray stands out against the background of the intensity profile $P(\Lambda)$ due to the slope of the two lines forming the ray and can be considered to be straight from the tip P_M—the intensity maximum in the ray—to the inflection points A and B. This enables us to define the ray brightness as $\Pi(t) = P(t) - P_S(t)$; accordingly, the amplitude of the ray brightness at a given time t is $P_R(t) = P_M(t) - P_S(t)$ and the angular size is $d \approx 2°–3°$, as is shown, e.g., at the bottom of Fig. 1 in [1]. Here, $P_S(t)$ is the "background" intensity, i.e., the average of $P(t)$ over angles smaller than 7° at a given time t. To verify our results and acquire new information on the physics of the processes under study, we also used another parameter—the time-differenced ray brightness, $P(t - t_0)$. We employed the following procedure to determine the time-differenced brightness functions $P(\Lambda, t - t_0)$ and $P(R, t - t_0)$. The intensity profiles $P(\Lambda, t_0)$ and $P(R, t_0)$ for some initial time $t_0 < t$ were subtracted from the profiles $P(\Lambda, t)$ and $P(R, t)$ to obtain the time-differenced brightness functions $P(\Lambda, t - t_0) = P(\Lambda, t) - P(\Lambda, t_0)$ and $P(R, t - t_0) = P(R, t) - P(R, t_0)$, respectively. These were used to determine the time-differenced amplitude functions of the ray brightness for individual rays in accordance with the relationships $P_R(\Lambda, t - t_0) = P_M(\Lambda, t) - P(\Lambda, t_0)$ and $P_R(R, t - t_0) = P_M(R, t) - P(R, t_0)$, where P_M is the intensity maximum in the ray. If the ray brightness $\Pi(t, \Lambda, R)$ and the time-differenced brightness function $P(t - t_0, \Lambda, R)$ are defined in this way, the contribution of the F corona is eliminated. This makes it possible to determine the plasma density N_R averaged over the angular diameter of the ray d at a given distance R without using data on the polarization intensity, which are mainly intended for eliminating the effects of the F corona.

We will now obtain a formula to estimate N_R treating the Sun like a point source. According to [8], the intensity of the scattered radiation in the direction toward the Earth normalized to the intensity of the photospheric emission is

$$\frac{P}{P_{mbs}} \approx \frac{\pi\sigma}{2}\left(1 - \frac{u}{3}\right)\frac{R_\odot^2}{R}\int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} N(1 + \sin^2\theta)d\theta, \quad (1)$$

where $\sigma = 7.95 \times 10^{-26}$ cm^2 sr^{-1} is the differential cross section for Thomson scattering, u is the degree of limb darkening, R is the projected distance in the plane of the sky from the center of the Sun to the point at which the scattering of the photospheric radiation is considered (or the distance between the line of sight and the Sun's center), and θ is the deviation angle of the ray from the plane of the sky, taken to be positive in the direction of the solar rotation.

We neglect limb darkening ($u = 0$) and assume that only plasma confined in a narrow ray with angular size $d \ll 1$ rad lying in the plane of the sky contributes to scattering. We can then obtain an estimate for N_R in the ray at a given distance R using (1):

$$N_R \approx \frac{P_R}{P_{mbs}}\frac{2}{\pi\sigma}\frac{R}{R_\odot^2 d} \quad (2)$$

$$= \frac{1.15 \times 10^{14}}{d}\frac{P_R}{P_{mbs}}\frac{R}{R_\odot} \quad \text{cm}^{-3},$$

where d is expressed in radians.

Formula (2) is valid if the ray axis lies in the plane of the sky ($\theta = 0°$.) As the ray leaves the plane of the sky due to the Sun's rotation, its intensity $P_R(\theta)$ decreases with increasing $|\theta|$. According to [9], the law for this decrease depends on the law for variations of the density N as a function of the true heliocentric distance r. (As the deviation from the plane of the sky increases, segments of the ray with progressively increasing heliocentric distances $r > R$ end up in the

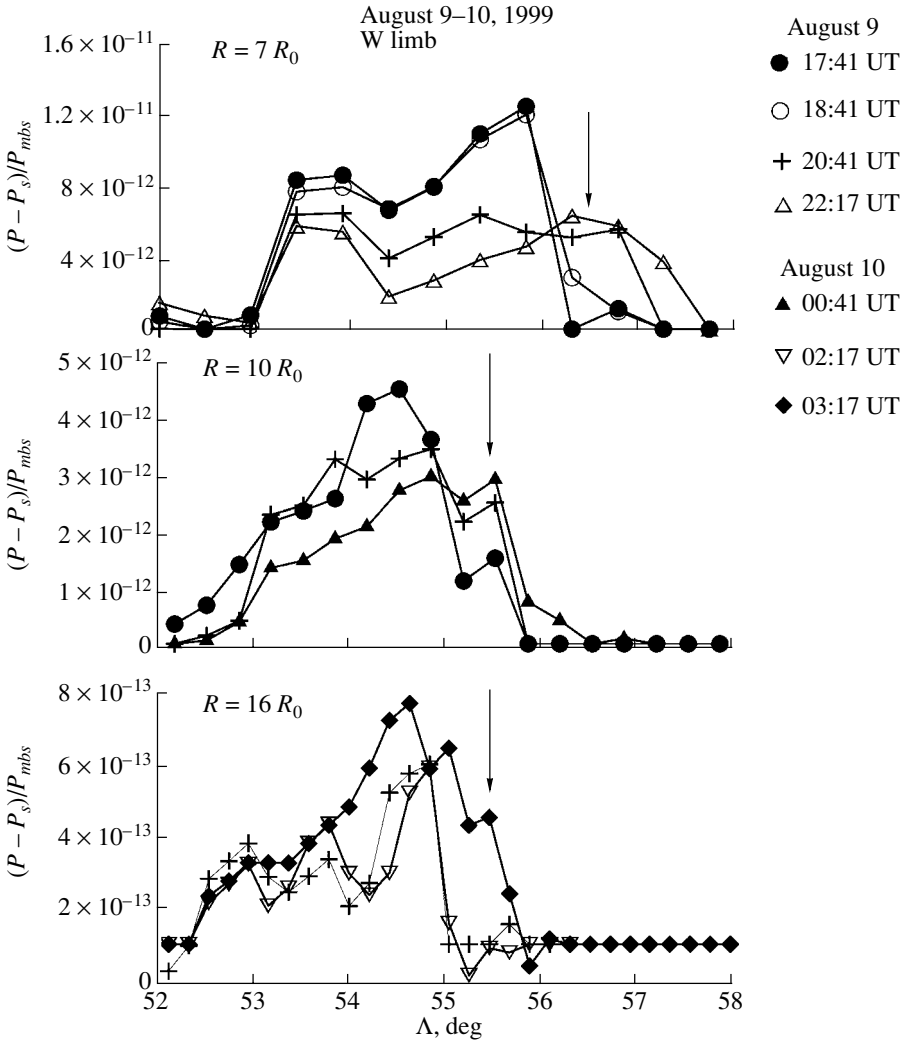

Fig. 1. Intensity of the ray $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$ as a function of the apparent latitude Λ at $R = 7\,R_\odot$ (top), $R = 10\,R_\odot$ (center), and $R = 16\,R_\odot$ (bottom) for successive times on August 9–10, 1999 (LASCO C3 data).

line of sight). According to (1), we obtain for a narrow ray with the density-variation law $N \sim 1/r^\alpha$

$$P_R(\theta) \approx P_R(0)(1 + \sin^2\theta)\cos^\alpha\theta, \qquad (3)$$

where $P_R(0) = P_R$ is the brightness of the ray in the plane of the sky ($\theta = 0°$). Since the rotational axis of the Sun is inclined to the ecliptic plane, the angle θ does not coincide with the longitude Ψ_L, which is also measured from the limb plane in the direction of the solar rotation but along the equator. The two angles are related as follows [1]:

$$\sin\theta = \cos\lambda\cos B_0\sin\Psi_L + \sin\lambda\sin B_0; \qquad (4)$$

here, B_0 is the heliographic latitude of the center of the solar disk (or the heliographic latitude of the Earth).

3. PROPERTIES OF THE ADDITIONAL STREAMS OF SOLAR-WIND PLASMA IN RAYS OF THE STREAMER BELT

We will now investigate the additional plasma streams in the rays of the streamer belt in two typical and simple situations, when the section of the streamer belt under study is (1) in the plane of the sky or (2) normal to the plane of the sky. The time interval used for the analysis should not exceed ≈ 0.5 day. These conditions must be satisfied so that we can eliminate the effect of solar rotation, which can shift the ray in apparent latitude Λ during the observations. (It can readily be seen that such shifts were virtually absent in the cases considered below.)

3.1. A Streamer Belt in the Plane of the Sky

We now consider the coronal images for August 9–10, 1999. During this period, the section of the

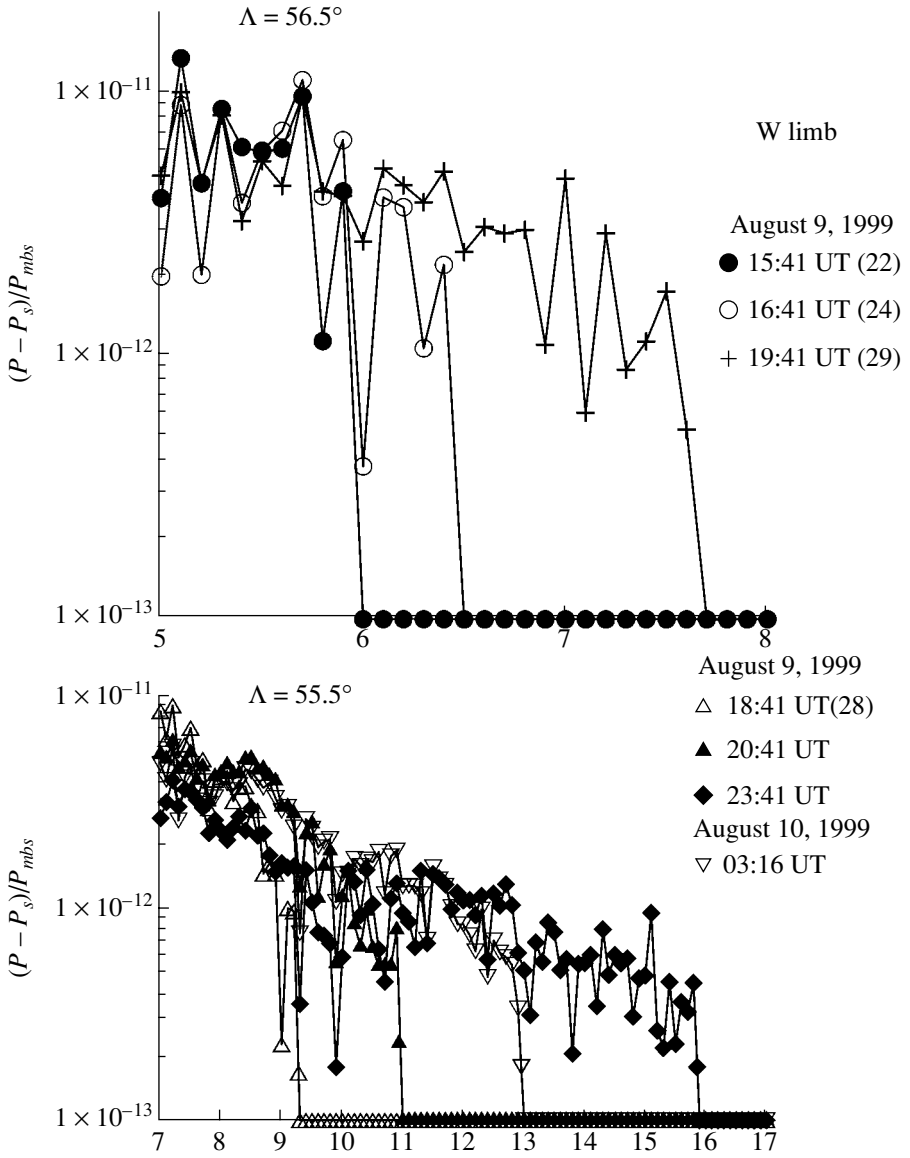

Fig. 2. Amplitude of the ray brightness $P_R(R,t) = P_M(R,t) - P_S(R,t)$ as a function of R at the latitudes $\Lambda \approx 56.5°$ (top) and $\Lambda \approx 55.5°$ (bottom) (see Fig. 1) at successive times during August 9–10, 1999.

streamer belt located at the western limb was extended in longitude (i.e., along the solar limb) and situated nearly in the plane of the sky. Thus, a sequence of bright rays forming the pattern of the streamer belt should be observed at the western limb in this case.

We will analyze the dynamics of the intensity in space and time for one of these two rays, whose intensity maximum is located at an apparent latitude of $\approx 55°$, at the western limb. Figure 1 shows the intensity profiles in this ray, $\Pi(\Lambda,t) = P(\Lambda,t) - P_S(\Lambda,t)$, as a function of Λ for various distances R at successive times. Here and below, negative values in curves for $\Pi(R,\Lambda,t)$ and $P(R,\Lambda,t-t_0)$ are everywhere set equal to $10^{-13}P_{mbs}$ for C3 data and to $10^{-11}P_{mbs}$ for C2 data. This means that we are studying only positive ray intensities and their variations. This considerably simplifies the appearance of the curves and facilitates the analysis.

We can see from Fig. 1 that the temporal variations in the $\Pi(\Lambda,t)$ profile are very similar at all three distances R: the profile broadens toward large Λ, while the intensity in the ray itself decreases (except for the radius $R = 16\ R_\odot$, where the brightness in the ray itself even grows slightly). On the whole, the position of the ray (and its left half in particular) remains virtually unchanged. The expansion of the ray profile at large R starts when t is large: after $t =$ 18:41 UT at $R = 7\ R_\odot$ (upper graph in Fig. 1) and after $t =$ 20:41 UT at $R = 16\ R_\odot$ (lower graph in Fig. 1). The ray is not strictly radial: at distances from $R = 7\ R_\odot$ to $R = 16\ R_\odot$,

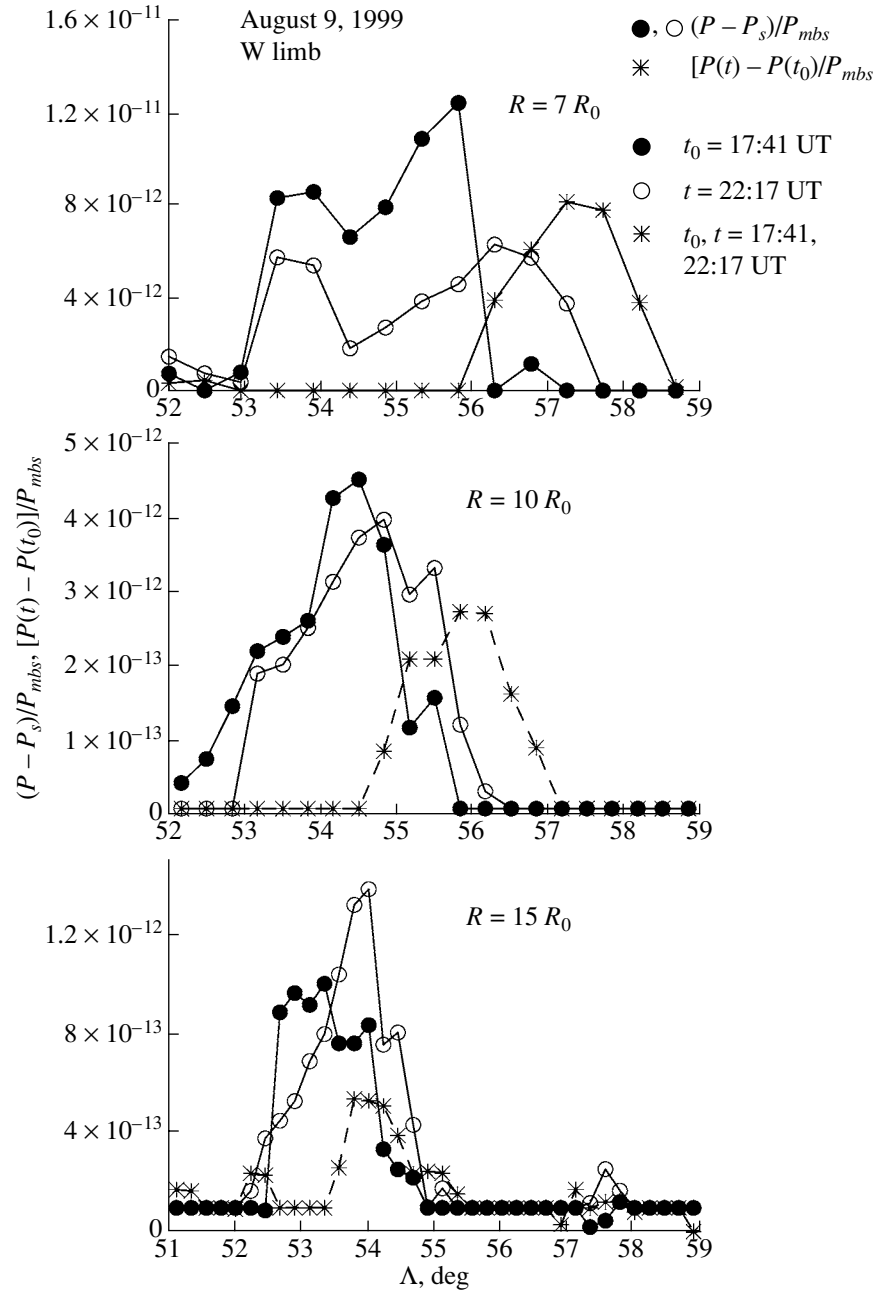

Fig. 3. Ray brightness $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$ (solid and open circles) and changes in brightness with time $P(\Lambda, t - t_0)$ (asterisks) as functions of the apparent latitude Λ at $R = 7\ R_\odot$ (top), $R = 10\ R_0$ (center) and $R = 16\ R_0$ (bottom) for successive times during August 9–10, 1999. The time $t_0 = 17{:}41$ UT (August 9, 1999).

its maximum moves from $\Lambda = 55.5°$ to $\Lambda = 54.5°$, i.e., by approximately 1°. A similar Λ displacement is exhibited by the maximum of the broadened section of the ray (marked with an arrow) outside the original, unperturbed ray profile.

With allowance for this weak nonradiality of the ray, Fig. 2 shows the R dependences of the ray-intensity amplitude $P_R(R)$ (marked with an arrow in Fig. 1) for successive times at distances $R = 5–8\ R_\odot$ for a latitude of $\Lambda \approx 56.5°$ (upper graph) and at distances $R = 5–8\ R_\odot$ for a latitude of $\Lambda \approx 55.5°$ (lower graph). When plotting the curves in Fig. 2, we removed noise and other nonuniformities propagating ahead of the front. Figure 2 shows that the broadened part of the profile propagating from the Sun forms a steep front whose speed is $V \approx 100–150$ km/s.

According to (2), the ray brightness P_R is uniquely related to the plasma density N_R (since there is no contribution from the F corona); therefore, an additional stream of the solar-wind plasma propagates

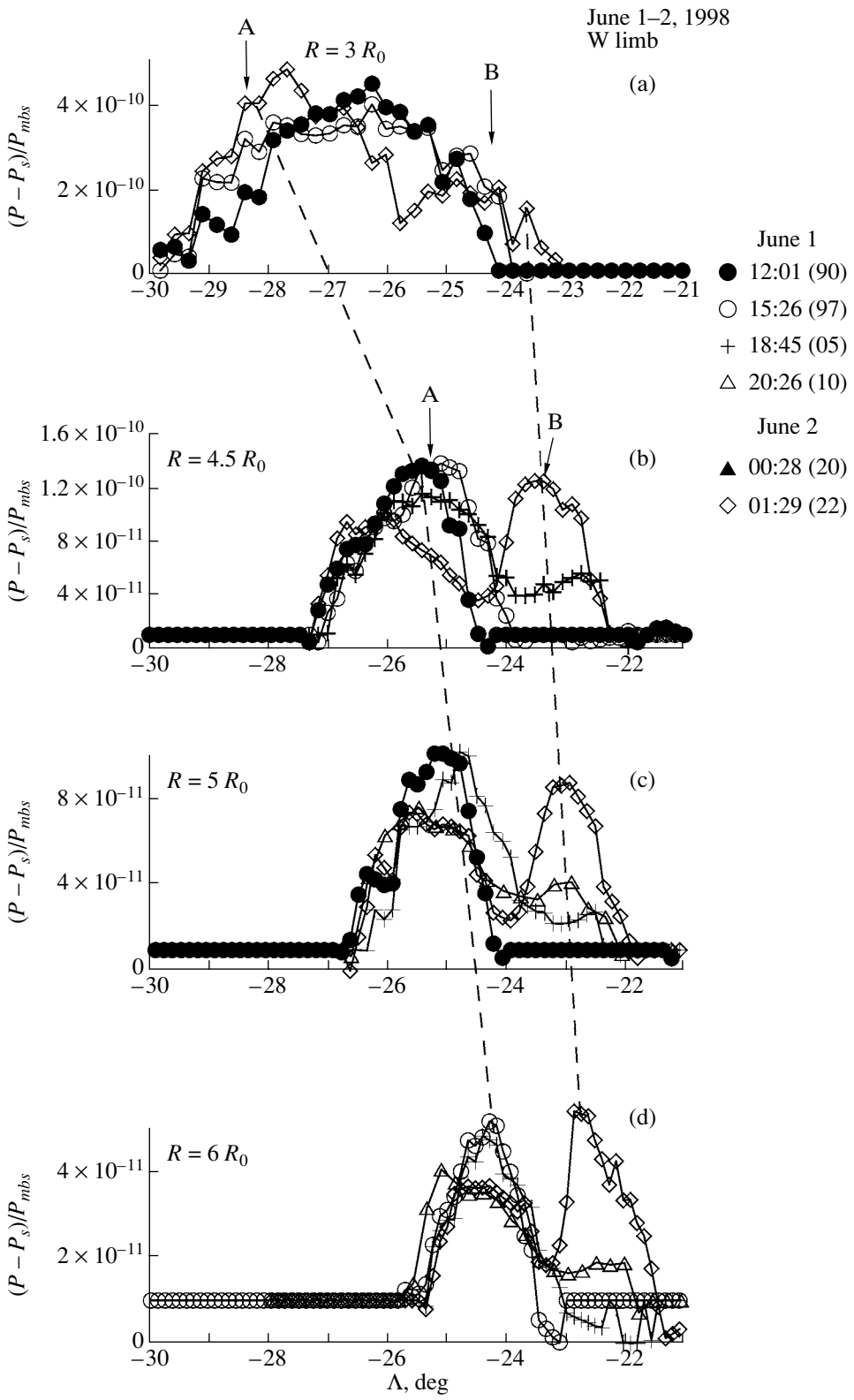

Fig. 4. Ray brightness $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$ as a function of the apparent latitude Λ at $R =$ (a) 3 $R_\odot$, (b) 4.5 $R_\odot$, (c) 5 $R_\odot$, and (d) 6 $R_\odot$ for a time sequence during June 1–2, 1998 (LASCO/C2 data).

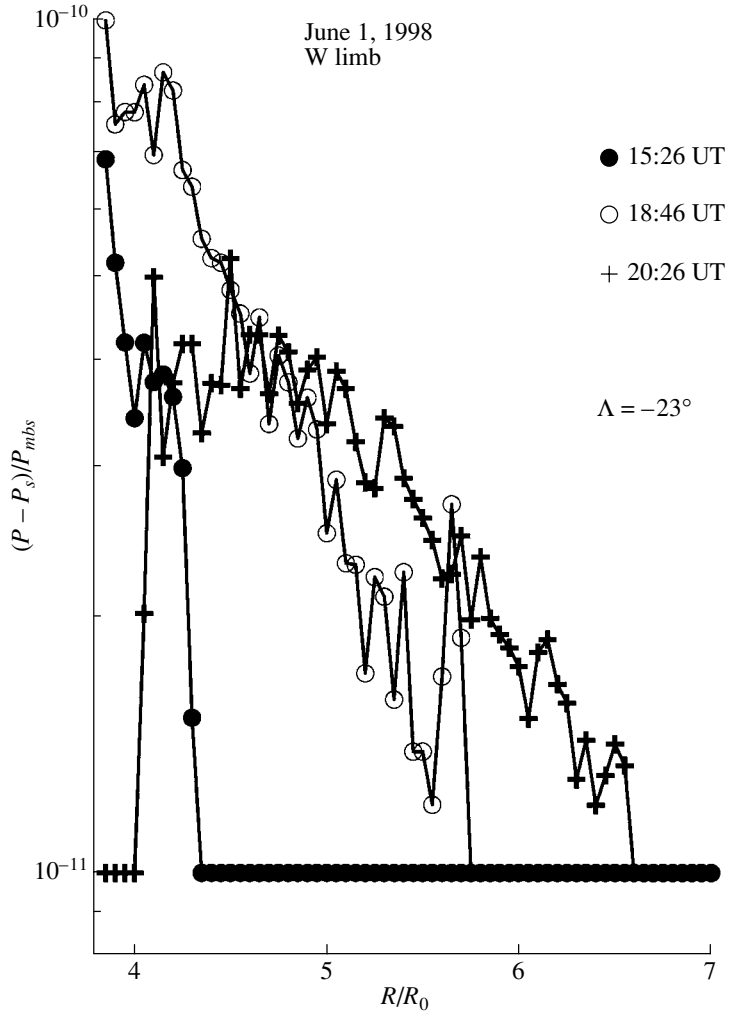

Fig. 5. Amplitude of the ray brightness $P_R(R,t) = P_M(R,t) - P_S(R,t)$ as a function of distance R at a latitude of $\Lambda \approx -23°$ for successive times on June 1, 1998 (LASCO/C2 data).

from the Sun in this case. The width of the front, $d \approx 0.1R_0$, is comparable to the spatial resolution of the C3 instrument. It is quite possible that the plasma also flows inside the ray (e.g., near the maximum of P_M), but this flow is difficult to detect, since its brightness declines with time. Let us compare for the broadened section of the ray ($\Lambda > 55°$) in Fig. 1 the time-differenced ray brightness $P(\Lambda, t - t_0)$ with the difference between the ray brightness at time t, $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$, and the ray brightness at time t_0, $\Pi(\Lambda, t_0) = P(\Lambda, t_0) - P_S(\Lambda, t_0)$.

Figure 3 illustrates this comparison for times $t_0 =$ 17:41 UT and $t =$ 22:17 UT. Ideally, if $P_S(\Lambda, t_0)$ were equal to $P_S(\Lambda, t)$, we would find that $\Delta\Pi(\Lambda, t - t_0) = P(\Lambda, t) - P(\Lambda, t_0) = P(\Lambda, t - t_0)$. We can see from Fig. 3 that the profile $P(\Lambda, t - t_0)$ (asterisks) and the profile $\Delta\Pi(\Lambda, t - t_0)$ (difference between the open and solid circles) in the upper and central graphs are qualitatively similar for $\Lambda > 56°$ and differ in amplitude and angular size by less than $\approx \pm 50\%$. This difference seems to result from the difference between $P_S(\Lambda, t_0 =$ 17:41 UT) and $P_S(\Lambda, t =$ 22:17 UT).

3.2. A Streamer Belt Normal to the Plane of the Sky

Let us consider the coronal image for June 1–2, 1998. During that time, a small section of the streamer belt located at an apparent latitude of $\Lambda \approx -20°$ to $-30°$ near the western limb was almost normal to the plane of the sky. A streamer belt with this orientation can be observed as either one or two rays, which can have different or equal brightnesses [10, 11]. Figure 4 shows the profiles $\Pi(\Lambda)$ within the given section for various R and at various times t. We can see how one ray (A) expands and another (B) forms.

Figure 4a presents the profile $\Pi(\Lambda)$ at $R = 3\ R_\odot$, i.e., below the apex of the helmet, which is located at

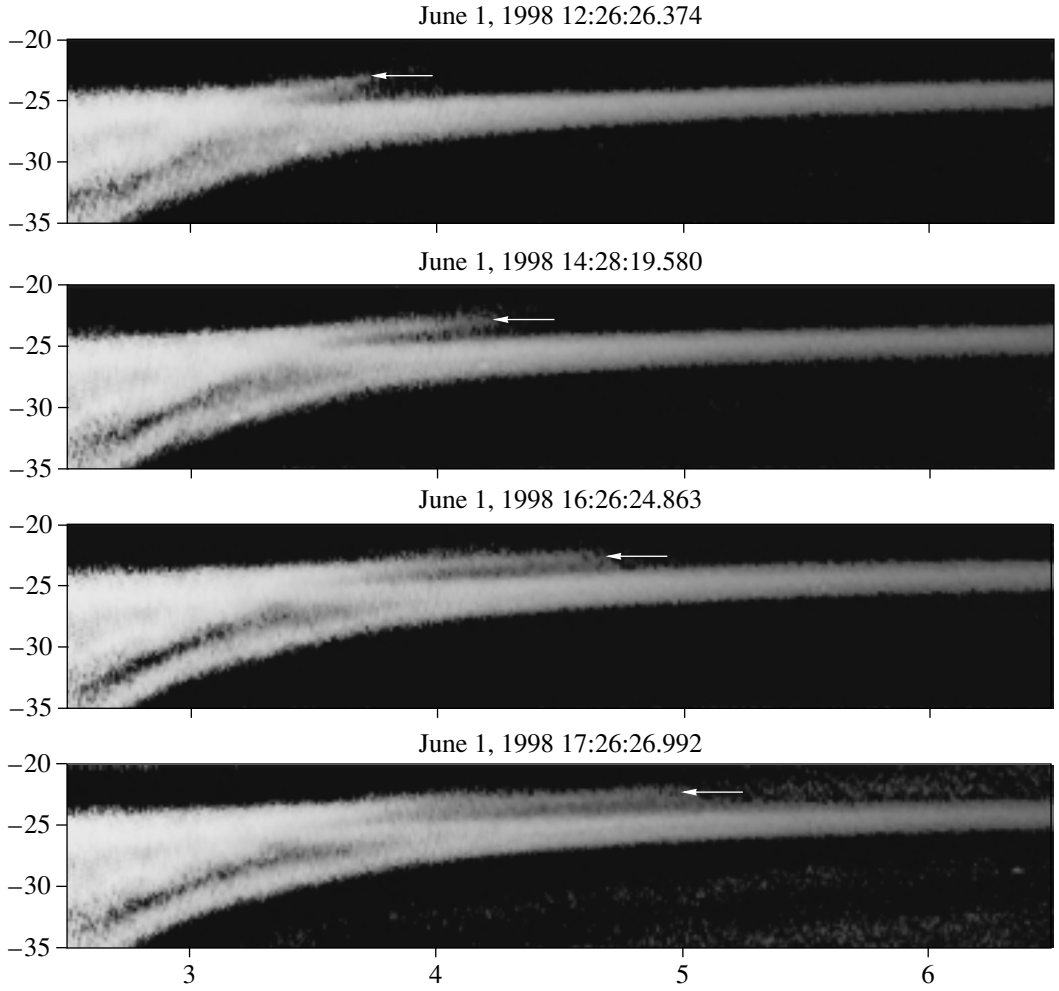

Fig. 6. Two-dimensional map of the ray brightness $\Pi(\Lambda, R) = P(\Lambda, R_\odot) - P_S(\Lambda, R_\odot)$ in the vicinity of rays A and B for successive times on June 1, 1998. The arrows mark the positions of the leading front of the additional stream filling ray B at several times. The lower ray A is clearly visible during the entire period (W limb; LASCO/C2 data).

$R \approx 3.5\ R_\odot$. At the initial time, $t = 12{:}01$ UT (solid circles), the profile $\Pi(\Lambda)$ is broad and smooth and has a maximum near a latitude of $\Lambda \approx -26°$, where the apex of the helmet is located. At subsequent times (open circles and diamonds), brightening to the left and right of the distribution maximum and an intensity decrease in the central part of the distribution are observed. According to [10, 11], two rays, which, in general, have different intensities, extend on both sides of the helmet. The brightness increases shown in Fig. 4a occur in these rays, which are labeled A and B. In Figs. 4b–4d, ray A is clearly visible during the first three hours (solid and open circles) at all distances to $R = 6\ R_\odot$, whereas ray B remains barely visible. Beginning at $t = 15{:}26$ UT, ray A begins to broaden slightly toward the right at $R = 4.5\ R_\odot$ (open circles in Fig. 4b). This process intensifies with time.

The broadening of ray A is similar at all distances but with a time lag. In particular, an appreciable broadening at $R = 6\ R_\odot$ is recorded only after $t = 18{:}46$ UT (crosses and open triangles in Fig. 4d). After $t = 18{:}46$ UT, the brightness of ray B starts growing. By $t = 01{:}29$ UT (June 2), a two-ray structure with nearly equal intensities P_R for rays A and B is already well developed at all distances to $R = 6\ R_\odot$. At the same time, the position of the ray itself, including its left half, remains unchanged during the entire period.

As can be seen from the graphs in Fig. 5, the time lags between the expansion of ray A and the formation of ray B at various distances R (see Fig. 4) are due to the motion of an additional, steep-fronted plasma stream from the Sun at a latitude of $\Lambda \approx -23°$. A two-dimensional map of the ray brightness, $\Pi(\Lambda, R) = P(\Lambda, R_\odot) - P_S(\Lambda, R_\odot)$, in the vicinity of rays A and B is shown for several successive times in Fig. 6. The arrows mark the positions of the leading front of the additional stream filling ray B at succes-

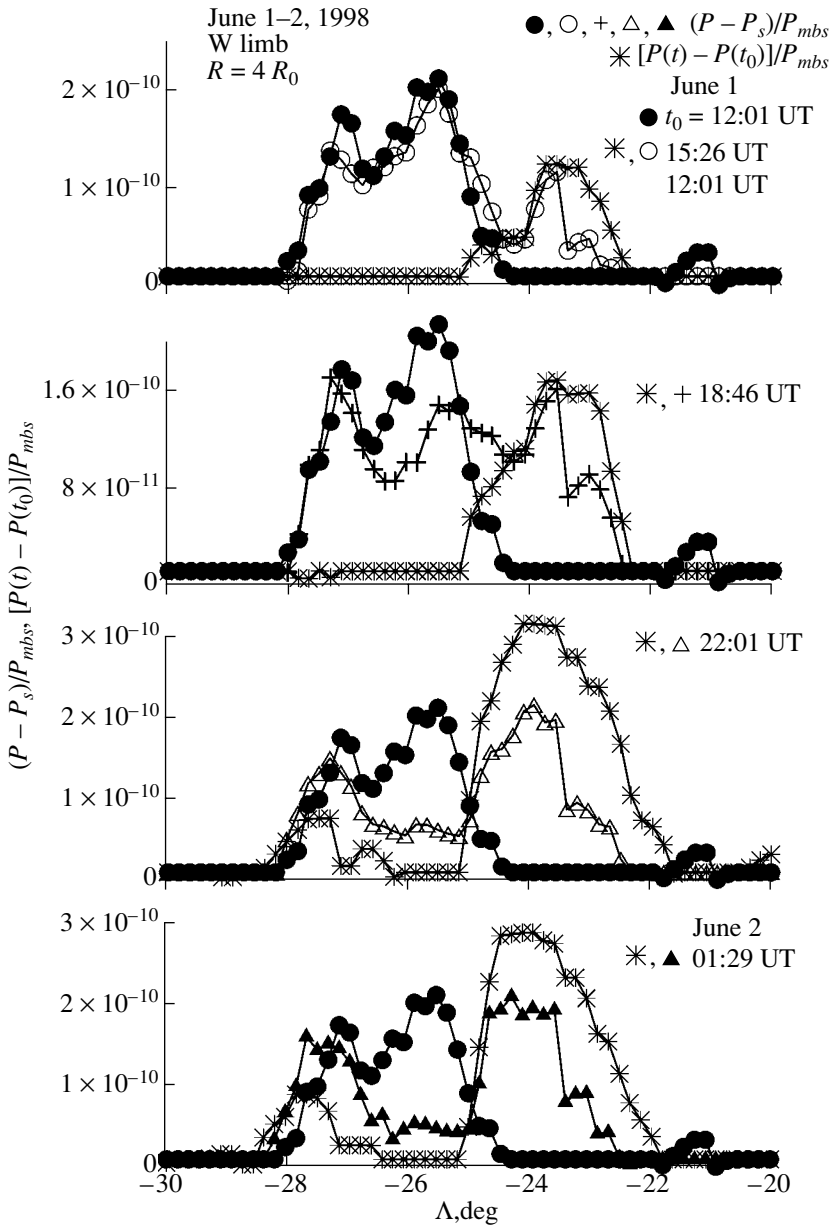

Fig. 7. Ray brightness $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$ (open circles, crosses, open and solid triangles) and the time change in the brightness $P(\Lambda, t - t_0)$ (asterisks) as functions of the apparent latitude Λ at $R = 4\ R_\odot$ for successive times during June 1–2, 1998. The time $t_\odot$ = 12:01 UT (June 1, 1998).

sive times. The lower ray A is clearly visible during the entire period of the observations.

Figure 6 shows that the additional stream moves radially within a narrow layer (magnetic flux tube) and has sharply defined edges. The narrowing leading front of the stream also has a fairly sharp edge. Let us compare the time-differenced brightness $P(\Lambda, t - t_0)$ with the difference $\Delta\Pi(\Lambda, t - t_0)$ between the ray brightness at time t, $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$, and at time t_0, $\Pi(\Lambda, t_0) = P(\Lambda, t_0) - P_S(\Lambda, t_0)$, for the broadened part of ray A (at $-23° > \Lambda > -25°$) and the forming ray B ($-22° > \Lambda > -24°$) in Fig. 4.

This is done in Figs. 7 and 8 for time t_0 = 12:01 UT and eight times t. Recall that, if $P_S(\Lambda, t_0) = P_S(\Lambda, t)$, then $\Delta\Pi(\Lambda, t - t_0) = \Pi(\Lambda, t) - \Pi(\Lambda, t_0)$. We can see from Figs. 7 and 8 that the profile $P(\Lambda, t - t_0)$ (asterisks) and the profile $\Delta\Pi(\Lambda, t - t_0)$ (difference between the open and solid circles) in the upper and middle graphs are qualitatively similar for $\Lambda > -25°$ and differ in amplitude and angular size by no more than several tens of percent. Figure 8

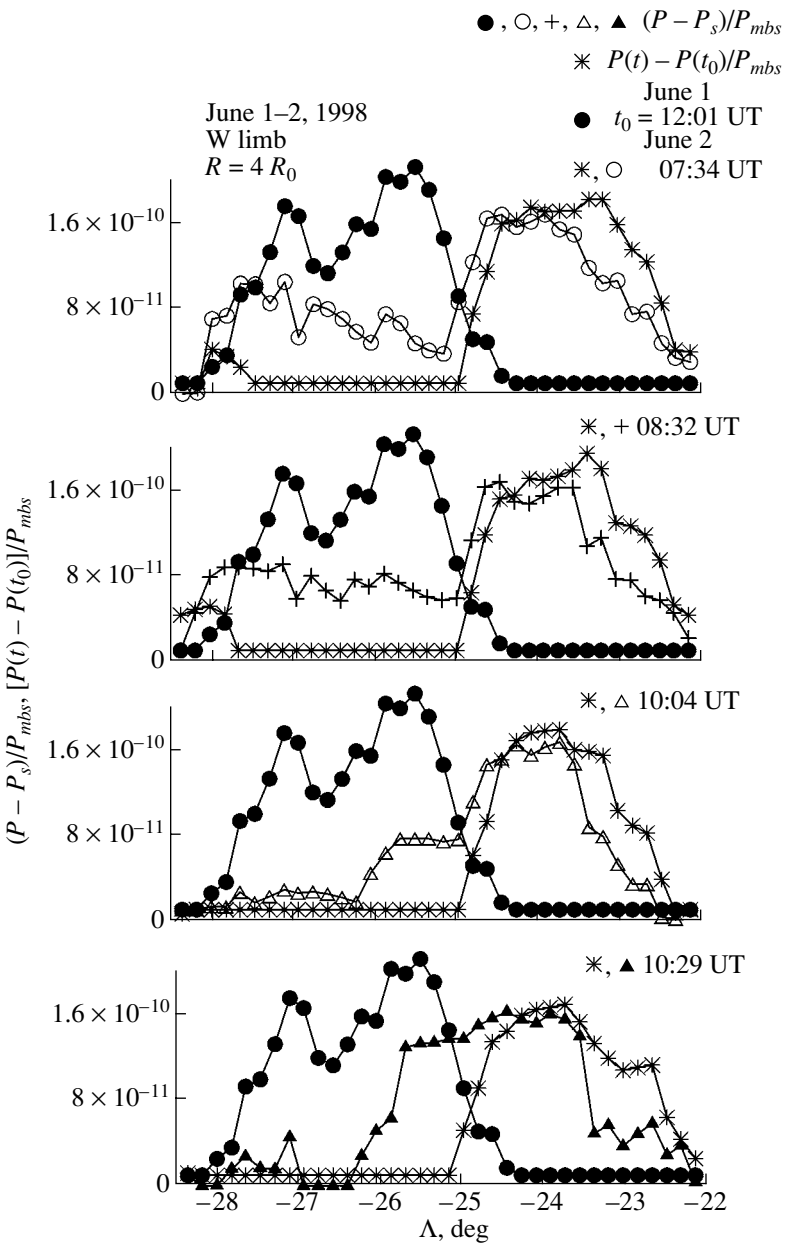

Fig. 8. Ray brightness $\Pi(\Lambda, t) = P(\Lambda, t) - P_S(\Lambda, t)$ (open circles, crosses, open and solid triangles) and the time change in the brightness $P(\Lambda, t - t_0)$ (asterisks) as functions of the apparent latitude Λ at $R = 4\ R_\odot$ for successive times on June 2, 1998. The time $t_0 = 12{:}01$ UT (June 1, 1998).

shows especially good agreement, although the time difference can reach almost a day (e.g., for the lower graph in Fig. 8). This testifies to the very high quality and stability of LASCO coronal images obtained over at least half a day.

Let us estimate the plasma density at the peak of the front of the additional stream at a distance of $R = 4\ R_\odot$. In the case illustrated in Fig. 4, the front is located at $R \approx 4\ R_\odot$ at time 15:26 UT (solid circles); at the peak, $P_R \approx\ 4 \times 10^{-11} P_{mbs}$. Substituting this value into (2) and using the fact that $d \approx 2.5°$, we obtain $N \approx 4 \times 10^5$ cm^{-3}. By $t = 01{:}29$ UT (June 2), the well-established intensity of ray B ($\Lambda \approx -23°$) at $R = 4\ R_\odot$ is $P_R \approx 2 \times 10^{-10} P_{mbs}$, and the density of the well-established plasma flow in this ray is accordingly $N \approx 2 \times 10^6$ cm^{-3}. Let us compare these figures with the mean density in the streamer belt determined from measurements of the polarization brightness of the corona during solar eclipses. According to [12], such measurements carried out in various years yielded $N \approx 4 \times 10^5 - 4 \times 10^6$ cm^{-3} at $R = 4\ R_\odot$, which is comparable to the plasma

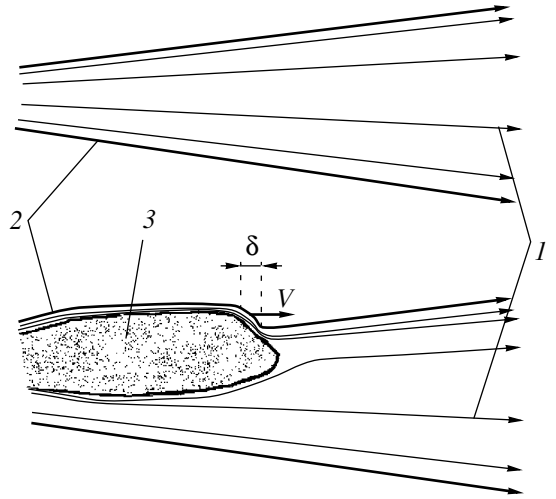

Fig. 9. Schematic of the expansion of the bright ray and the formation of the sharp front of the additional plasma stream moving within the ray.

density in the additional streams. Therefore, the additional streams seem to be among the main sources replenishing the slow quasi-steady solar wind in the streamer belt. Note also that additional plasma streams in rays of the streamer belt are a very widespread phenomenon, whose role should be investigated and become more fully understood in the near future.

4. DISCUSSION

Fist, note that at least at distances $R < (3{-}4)\,R_\odot$, the front velocity of the additional stream is lower than the speed of sound (for temperature $T \approx 10^6$ K, it is $V_s \approx 150$ km/s). Therefore, there is no shock wave at the front. Accordingly, the expansion of the bright ray and the related formation of the steep front can have an explanation as shown Fig. 9. The ray of enhanced brightness is a flux tube of radially stretched magnetic field lines with a quasi-steady stream of the solar-wind plasma inside it (top of Fig. 9). The additional plasma stream flowing into the tube has an increased total pressure, $P = (2kNT + B^2)/8\pi$. As a result, the magnetic flux tube expands, and a component of the magnetic field normal to the velocity of the stream develops at its front. As a consequence, the steep front forms and moves together with the stream. To obtain a crude estimate of its width d, we can assume that the ion Larmor radius for the highly conducting coronal plasma is $d \sim V/\omega_h$. We also assume that, at $R = 5\,R_\odot$, the magnetic field is $B \sim 5(R_\odot/R)^2$ G ~ 0.2 G and the velocity is $V \sim 100$ km/s; then, $d \sim V/\omega_h \sim 50$ m. Therefore, the intrinsic width of the front is very small, and the observed widths of the additional streams are determined by the spatial resolution of the C2 and C3 instruments.

5. CONCLUSIONS

(1) We have demonstrated the presence of additional plasma streams with enhanced density in bright rays of the streamer belt. The streams have steep fronts, whose widths, $d \approx 0.10\,R_\odot$, are comparable to the spatial resolution of the LASCO C3 instrument.

(2) The additional plasma streams are similar to streams of the slow, quasi-steady solar wind in the streamer belt in terms of their plasma density, directional velocity, and lifetime. They appear to be one of the main sources replenishing the slow, quasi-steady solar wind in the streamer belt.

ACKNOWLEDGMENTS

The SOHO/LASCO data used here were obtained by a consortium including the Naval Research Laboratory (United States), Max-Planck-Institut für Aeronomie (Germany), Laboratoire d'Astronomie (France), and the University of Birmingham (Great Britain). This work was supported by the Program "Leading Scientific Schools of Russia" and the Federal Science and Technology Program in Astronomy.

REFERENCES

1. V. G. Eselevich and M. V. Eselevich, Solar Phys. **188**, 299 (1999).
2. V. G. Eselevich and M. V. Eselevich, Solar Phys. **195**, 319 (2000).
3. N. R. Sheeley, Jr., Y.-M. Wang, S. H. Hawley, *et al.*, Astrophys. J. **485**, 472 (1997).
4. Y.-M. Wang, Jr., N. R. Sheeley, J. H. Walters, *et al.*, Astrophys. J. **498**, L165 (1998).
5. Y.-M. Wang, N. R. Sheeley, Jr., R. A. Howard, *et al.*, Geophys. Res. Lett. **26**, 1203 (1999).
6. Y.-M. Wang, N. R. Sheeley, Jr., R. A. Howard, *et al.*, Geophys. Res. Lett. **26**, 1349 (1999).
7. V. G. Eselevich, J. Geophys. Res. **103**, 2028 (1998).
8. A. J. Hundhausen, J. Geophys. Res. **98**, 13177 (1993).
9. V. G. Eselevich, V. G. Fainshtein, and M. V. Eselevich, Solar Phys. **200**, 259 (2001).
10. V. G. Eselevich, V. G. Fainshtein, and M. V. Eselevich, Solar Phys. **197**, 101 (2000).
11. V. G. Eselevich and M. V. Eselevich, Solar Phys. **203**, 165 (2001).
12. K. P. Dere, G. E. Brueckner, R. A. Howard, *et al.*, Solar Phys. **175**, 601 (1997).

Translated by A. Getling