Astronomy Reports, Vol. 47, No. 11, 2003, pp. 889–902. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 963–977.

Spectral Observations of the Symbiotic Mira Variable V407 Cyg in 1993–2002

A. A. Tatarnikova[1], P. M. Marrese[2], U. Munari[2], T. Tomov[3], and B. F. Yudin[1]

[1] *Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119899 Russia*

[2] *Astronomical Observatory, Padua, Asiago, Italy*

[3] *Astronomical Center, Copernicus University, Torun, Poland*

Received March 30, 2003; in final form, May 8, 2003

Abstract—We present high- and low-resolution spectral observations of the symbiotic Mira variable V407 Cyg obtained in 1993–2002. Both emission and absorption spectra of the star are analyzed in detail. We present the basic spectral parameters for three states of the hot component of V407 Cyg—its passive and quiescent states and an outburst in 1998—and study the evolution of the emission-line profiles during transitions from one state to another. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

V407 Cyg (Nova Cygni 1936) belongs to a small group of symbiotic stars whose cool component is a Mira variable. This variable was discovered by Hoffmeister [1] during a nova-like outburst. In August 1936, at the outburst maximum, its photographic magnitude was $m_{pg} \approx 14$. Its brightness then decreased slowly, with its photographic magnitude reaching $m_{pg} \approx 16$ after two years. Further, Meinunger [2] detected pulsations of the star with the period $P \approx 745^{\rm d}$ and amplitude $\Delta m_{pg} \geq 3.5$.

In 1998, the hot component of V407 Cyg underwent another intense outburst, the second in its photometric history, which has not yet ended. At that time, the V brightness of the hot component flickered with an amplitude of $\sim 0.2^m$ on time scales of several minutes [3]. Photometric and low-resolution spectral observations of V407 Cyg have indicated that its hot component can be in three qualitatively different states. If we consider the star as a rapidly rotating magnetic white dwarf, these three states may be associated with passive (with a very low accretion rate), ejection, and accretor states [3].

Here, we prefer to denote the last two states as "quiescent" and "outburst," respectively. In the quiescent state, the brightness of the hot component is lower than in outburst. In addition, these two states are qualitatively different: the HeII λ4686 Å line is seen in the spectrum of V407 Cyg in the quiescent state but not in outburst. In other words, the temperature of the hot component in the quiescent state exceeds that in outburst. In many ways, provided it starts from its quiescent state, an outburst of the hot component of V407 Cyg looks the same as an outburst of a classical symbiotic star, such as BF Cyg. When the hot component is in its passive state, it is nearly invisible against the background of the Mira star, and does not excite bright emission lines in the nebula.

We present here high-resolution spectra of V407 Cyg obtained in 1993–2002. During this interval, the hot component spent time in all three states, whose photometric characteristics were analyzed in [3].

Let us first briefly review the spectral evolution of V407 Cyg before 1993.

The spectrum of V407 Cyg was first recorded photographically using an objective prism by Merrill and Burwell in 1950 [4]. Their survey was targeted at stars with appreciable Hα emission. In their catalog, V407 Cyg was labeled MHα 289–90. Referring to the work of P. Minkovsky, Merrill and Burwell [4] noted that the spectrum had a combined (symbiotic) character.

The next spectrum of V407 Cyg was obtained by Herbig [5] on October 10, 1952. The only emission line visible was Hδ. Even Hγ was not detected, though the spectrograph was centered precisely on its wavelength.

In 1973, Welin [6] announced the detection of a very strong Hα line in the spectrum of V407 Cyg. However, Gieseking and Schumann [7] did not see any trace of this line in 1976, at a brightness phase close to the pulsation minimum of the Mira star. Hα emission was likewise not detected in spectra obtained in 1984–1988 [8]; however, the presence of forbidden [NII] $\lambda\lambda$6548, 6584 Å lines was noted.

1063-7729/03/4711-0889$24.00

Two attempts were made to detect UV radiation from V407 Cyg using the IUE satellite, on August 29, 1982 and May 26, 1991. Both attempts failed, though the 1991 IUE spectrum had an exposure of 24 300 s! No emission lines were seen in a low-resolution optical spectrum obtained in October 1991 [9]. The survey of line profiles of symbiotic stars [10] presents echelle spectra of V407 Cyg obtained on September 23, 1988 and August 2, 1991. In 1988, neither Hα nor [OIII] λ5007 Å lines were detected, while weak multicomponent [NII] emission and Hα absorption were seen. Weak, asymmetrical Hα emission ($W(\mathrm{H}\alpha) \leq 4$ Å) appeared in 1991 [9]. It is likely that the Mira star itself could contribute substantially to the Hα emission during this time. We will denote the state of the hot component of V407 Cyg when the emission spectrum of V407 Cyg is extremely weak as the "passive" (inactive) state.

Only in August 1994 was an optical spectrum of V407 Cyg obtained for the first time. It resembled a typical symbiotic-star spectrum, displaying HI, HeI, HeII, [NII], [OIII] emission lines against the background of the radiation of the cool M6–M7 component [9]. A similar low-resolution spectrum was obtained in October 1995 [11]. In both cases, the Mira was in an intermediate phase between its maximum and minimum brightness. In March 1995, near its brightness maximum, it dominated at optical wavelengths, and only Hα emission was seen in its spectrum [11]. We will denote this state of the hot component the "quiescent" state. It differs from the outburst state not only in the lower optical brightness of the hot component, but also in the higher degree of excitation of the spectrum, and, in particular, in the presence of HeII λ4686 Å emission.

2. OBSERVATIONS

High-resolution spectra ($R = \lambda/\Delta\lambda \approx 18\,500$) of the symbiotic Mira V407 Cyg were obtained with an echelle spectrograph at the Cassegrain focus of the 1.82-m telescope of the Asiago Astronomical Observatory (Italy). The spectra were obtained in two wavelength intervals, which in total cover the range from 4300 to 9000 Å; the interval from 7000 to 9000 Å is presented only in the spectra obtained starting from 1999. Table 1 contains the dates and exposure times of the observations, as well as the observation times given in fractions of the B-band pulsation period of the Mira:

$$\varphi = (\mathrm{JD} - \mathrm{JD(max)})/P, \qquad (1)$$

where JD(max) = 2445975 is the Julian date corresponding to the maximum of the B-band pulsations and $P = 762.9$ days is the pulsation period [3].

Several low-resolution spectra of V407 Cyg at 4000–7500 Å with a dispersion of 8 Å/element, which we used to estimate $W(\mathrm{H}\alpha)$, were obtained with a spectrograph mounted on the 1.25-m telescope of the Crimean Station of the Sternberg Astronomical Institute [12].

The high-resolution spectra were processed using the IRAF package, together with the SPE code developed at the Crimean Astrophysical Obseratory. The errors in the radial velocities of single lines do not exceed 5 km/s. The spectra can be found at http://infra.sai.msu.ru/. The photometric data presented below to provide a more detailed description of the states of V407 Cyg were taken from [3, 9].

During the 1998–2002 outburst of the hot component, numerous emission lines were observed in the spectrum of V407 Cyg. Bright HI, HeI, and FeII emission lines, forbidden [OIII], [OI], and [NII] lines characteristic of planetary nebulae and symbiotic novae, and forbidden lines of singly ionized metals ([FeII], [CaII]) were seen against the background of intense TiO and VO absorption bands. At the same time, HeII λ4686 Å emission was absent until the Autumn of 2002.

3. EMISSION SPECTRUM OF V407 Cyg

3.1. Hydrogen Lines

Figure 1 presents the B light curve of V407 Cyg and the variation of the Hα equivalent width $W(\mathrm{H}\alpha)$ as a function of time, given in fractions of the pulsation period [see (1)]. The brightness values are taken from [3, 9].

In the interval from $\varphi = 5$ to $\varphi = 6.5$, the variations of the Hα equivalent width correspond to the pulsed oscillations of the star's brightness: the decrease of $W(\mathrm{H}\alpha)$ by a factor of ten in the transition from $\varphi = 5.5$ to $\varphi = 6.0$ is associated with an increase in the brightness of the star in this spectral interval by $\sim 2.5^m$; i.e., with an increase of the continuum level. However, at phases $\varphi > 6.5$, this dependence is disrupted—a dramatic decrease in the Hα equivalent width was observed at the minimum brightness of the star (Fig. 1).

These events were recorded on August 20, 1996, August 30, 1996 ($\varphi \approx 6.70$), and July 30, 1997 ($\varphi \approx 7.14$). Both the brightness of the star ($B \approx 17$) and the Hα equivalent width ($W(\mathrm{H}\alpha) \leq 4$ Å) reached their minima at that time. In other words, the hot component was in the passive state, and the Mira star dominated completely in the optical, despite the fact that its brightness was minimum. A similar state of V407 Cyg was observed in October 1991 (Fig. 1) [9].

The onset of an outburst of the hot component was detected during the first photometric and spectral

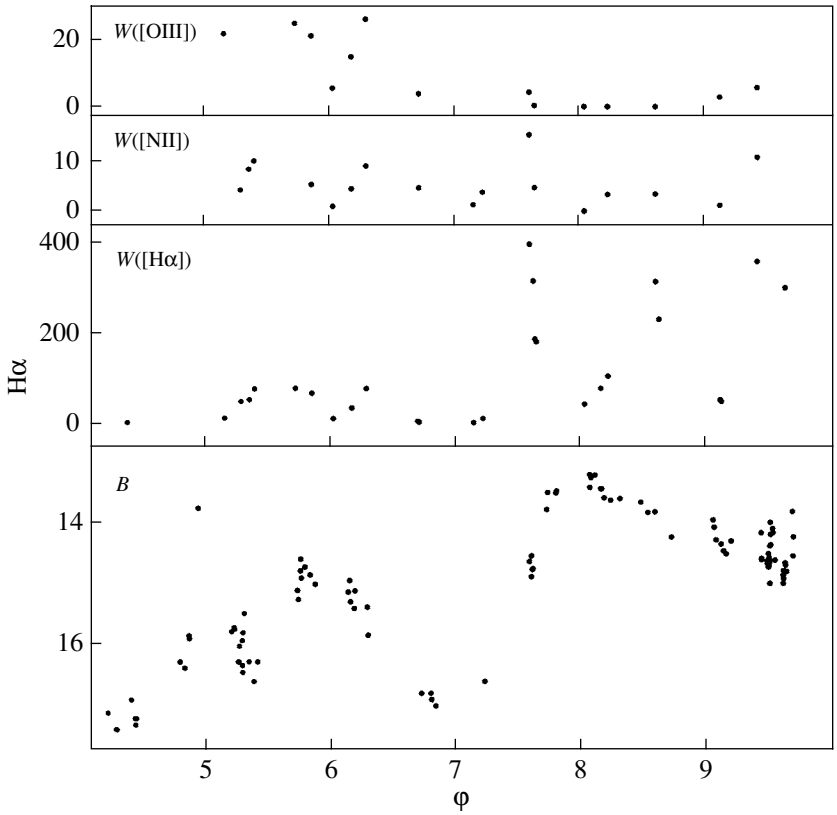

Fig. 1. B brightness of V407 Cyg and equivalent widths of the Hα, [NII] λ6583 Å, and [OIII] λ5007 Å lines as a function of time, given in fractions of the pulsation period φ [see (1)].

observations of V407 Cyg in July 1998. The star was unusually bright ($B \approx 14.7$) for the corresponding brightness phase of the Mira ($\varphi \approx 7.59$), and the Hα line was unusually intense ($W(\mathrm{H}\alpha) \approx 395$ Å; see Fig. 1). Comparing the spectrum obtained in July 1998 with those obtained in 1993–1995, when the hot component was in its quiescent state, we can note a considerable broadening of the Hα line wings and an absence of the HeII λ4686 Å line.

In the approximation, the Hα and Hβ line profiles can be considered to be the sum of three components (excluding the spectra obtained on September 25, 1997 and April 23, 2002): two emission components (narrow and broad) and an absorption component. Their equivalent widths (and, accordingly, relative intensities), radial velocities (including the relative velocities), and full widths at half maxima (FWHM) varied with time. In some cases, we were able to reproduce the emission using a single component (a Lorentz profile); however, an additional broad Gaussian had to be introduced to reproduce the line wings. On all the observation dates (with the exception of September 25, 1997), the narrow component was more strong than broad; i.e., its equivalent width was larger.

To illustrate the variations of the Hα line profile, Fig. 2 presents normalized line profiles $I_{\mathrm{norm}} = (I_\lambda - I_{\mathrm{cont}})/(I_{\mathrm{max}} - I_{\mathrm{cont}})$. The individual panels indicate the dates of the observations, the phase of the Mira in the B band, and W(Hα). However, this way of representing the Hα line and the scale used hinder visibility of the line wings, whose presence and width depend on the state of the hot component. Figure 3 displays Hα line profiles normalized to the continuum (with their lower parts given on an enlarged scale) in July 1995, August 1996, and August 1998. On these dates, the hot component of V407 Cyg was in its quiescent state, in its passive state, and at the beginning of an outburst, respectively.

Figure 4 presents variations of the heliocentric radial velocity (which we will call simply the radial velocity) of the narrow emission and absorption components of the Hα line in 1993–2001. No regular trend is seen in the variations of this parameter, in either the emission or absorption component. For all our observing dates except for those in 2002, the average radial velocities of the Hα narrow emission and absorption components were $\langle V \rangle \approx -38.5$ and -50.1 km/s, respectively, while we find for the Hβ line $\langle V \rangle \approx -38.9$ and -45.0 km/s. The radial velocity of

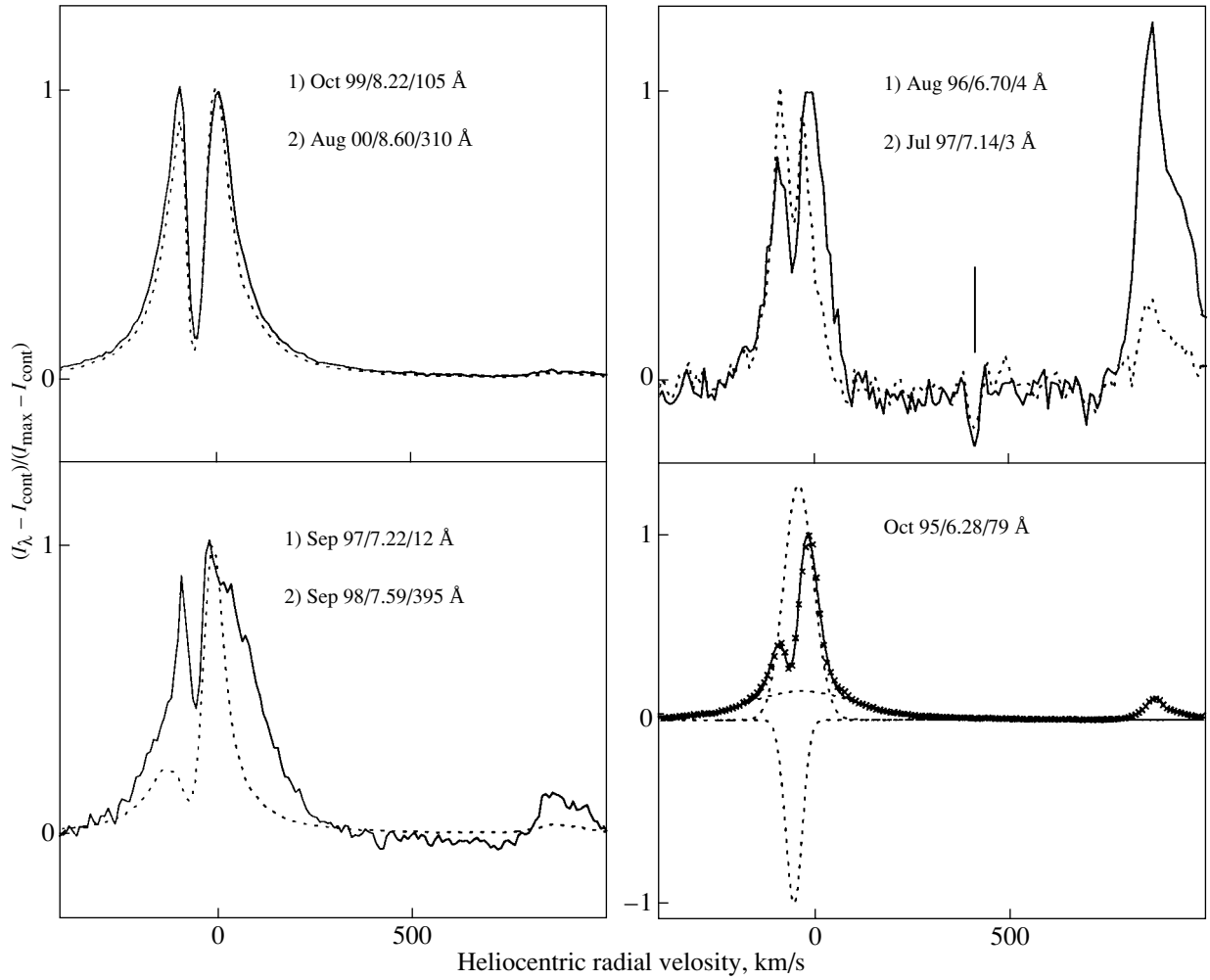

Fig. 2. Normalized Hα line profiles. The individual panels indicate the dates of the observations, φ, and W(Hα). The lower right panel presents an approximation to the Hα line profile using three Gaussians (the crosses show the observed profile, the dashed curves the individual components, and the solid curve the sum of the three Gaussians). In the remaining panels, parameter set "1" corresponds to the solid curve and set "2" to the dashed curve. The vertical line in the upper right panel marks the position of the CaI λ6573 Å absorption line.

the cool component of V407 Cyg is $\langle V \rangle \approx -41$ km/s (see Section 4 for more detail). Thus, the average radial velocity of the narrow Hα emission component is close to that of the Mira. This may imply that the narrow emission component is formed primarily in outer (relative to both components) parts of the circumstellar envelope of V407 Cyg.

The Hα line profile varied appreciably during our observations. In 1993–1995, when the hot component of V407 Cyg was in its quiescent state, the line had a P Cygni profile (Fig. 2). The average FWHM of the absorption component and the narrow and broad emission components were ~ 45, ~ 90, and ~ 300 km/s, respectively (right lower panel in Fig. 2). The line wings extended to 500 km/s (Fig. 3). The equivalent width of the narrow component exceeded that of the broad component by a factor of two to three.

In August 1996 and July 1997, the hot component was in its passive state. At this time, the broad component disappeared from both the Hα line and wings, and acquired a saddle-like shape (Figs. 2 and 3). At the same time, the FWHM width of both the absorption and narrow emission components did not change substantially. The equivalent width of the Hα line was ≤ 4 Å, and the CaI λ6573 Å absorption line, which originates in the atmosphere of the Mira (see below), was clearly visible nearby (Fig. 2).

Subsequent observations of V407 Cyg were carried out in September 1997 (Table 1), only two months after our July observing run. Over this time, the Hα profile had undergone dramatic changes (Fig. 2), although the brightness of the star was still

Table 1. Observations with the echelle spectrograph

Date	Julian date	Exposure time, s	φ
September 5, 1993	2449236.4	2000	5.26
October 27, 1993	2449288.4	600	5.34
November 27, 1993	2449319.3	720	5.38
November 12, 1994	2449669.4	1200	5.84
March 21, 1995	2449797.7	600	6.01
July 14, 1995	2449912.6	1000	6.16
October 11, 1995	2450002.4	1800	6.28
August 20, 1996	2450315.6	1518	6.69
August 30, 1996	2450326.5	2400	6.70
July 30, 1997	2450660.4	900	7.14
September 25, 1997	2450716.5	1200	7.22
July 8, 1998	2451002.6	1800	7.59
August 8, 1998	2451034.3	500	7.63
June 9, 1999	2451338.6	554	8.03
October 29, 1999	2451481.3	3600	8.22
August 14, 2000	2451771.4	6600	8.60
August 16, 2000	2451772.5	7200	8.60
August 17, 2000	2451773.6	2400	8.60
September 9, 2001	2452162.4	3600	9.11
April 23, 2002	2452387.6	3600	9.41

low ($B \approx 16.6$ [9]) and the equivalent width of the Hα line was only ~ 12 Å.

We fit this Hα profile using four Gaussians: one absorption component and three emission components. Its shape was specified by two broad (FWHM $\approx$ 230 and 460 km/s) components whose total equivalent width exceeded that of both of the narrow ($\sim$60 km/s) emission and absorption ($\sim$ 35 km/s) components by a factor of ~ 7.6.

Such variations of the Hα line profile over only two months indicate that the broad component of the line is formed appreciably closer to the hot companion than is the narrow component. This relative positioning of regions with high and relatively low velocity dispersions is also consistent with our general understanding of the structure of symbiotic stars whose hot components display accretion activity. The disappearance of the broad base in the Hα profile when the hot star goes into its passive state, along with its low brightness, indicates a dramatic decrease in accretion activity at that time.

Table 2. Heliocentric radial velocities (in km/s) of emission (FeII and [FeII]) and absorption (LiI λ6708 Å, CaI λ6573 Å) lines*

Date	φ	V(FeII)	V([FeII])	V(LiI)	V(CaI)
1	2	3	4	5	6
September 5, 1993	5.27	N	N	−34	N
October 27, 1993	5.34	N	N	UE	N
November 27, 1993	5.38	N	N	UE	N
November 12, 1994	5.84	−45	−48	–	N
March 21, 1995	6.01	N	N	–	−45
July 14, 1995	6.16	−43	−39	–	N
October 11, 1995	6.28	−40	−45	–	N
August 30, 1996	6.70	N	N	–	−42
July 30, 1997	7.14	N	N	−40	−42
September 25, 1997	7.21	N	N	–	−39
July 8, 1998	7.54	−53	−50	–	N
August 8, 1998	7.63	−55	−51	–	N
June 9, 1999	8.03	−52	−50	−47	N
October 29, 1999	8.22	−51	−47	N	N
August 14, 2000	8.60	−50	−49	N	N
September 9, 2001	9.11	−41	−39	−42	−36
April 23, 2002	9.41	−32	−32	N	N
Average		−46.2	−45.0	−40.8	−41.2

* Columns 3 and 4 contain the average radial velocities identified for the FeII (narrow component) and [FeII] lines for each date. In columns 3–6, "–" denotes spectral interval not represented; "UE" an insufficient exposure; and "N" no real line seen.

The radical changes in the Hα profile in September 1997 can be interpreted as the onset of activity of the hot component, which is first manifest only hydrodynamically (a region with a large velocity dispersion is formed), without any substantial increase in the energy release or, accordingly, the star's brightness. Note also that no appreciable wings were visible in the Hα profile at that time. However, they appeared immediately as soon as the enhanced brightness of the hot component began to be manifest in the star's activity in July–August 1998 (Fig. 3).

During the outburst of the hot component (1998–2002), the Hα line displayed a P Cygni profile in 1998 and 2001, while it was saddle-like in 1999–2000 (Fig. 2). In the latest spectrum obtained, on April 23, 2002, the Hα line displays a complex P Cygni profile with a series of absorption components located

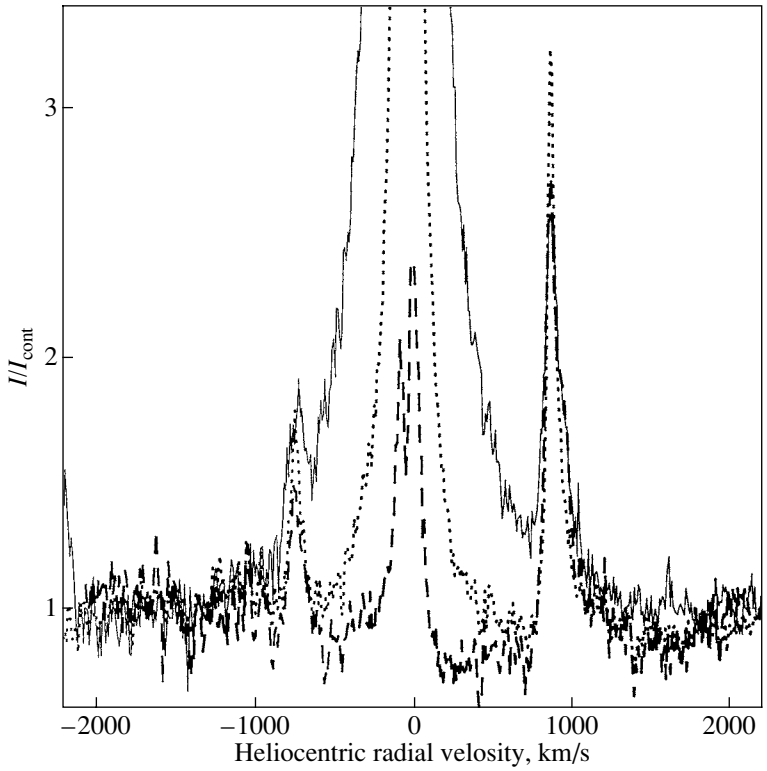

Fig. 3. Hα line wings during a deep minimum of the hot component on August 30, 1996 (dashed curve), its quiescent state on July 14, 1995 (dotted curve), and an outburst on August 8, 1998 (solid curve).

at radial velocities of −65, −160, and −225 km/s. The Hα profile was broadest in 1999–2000. At that time, the FWHM of the narrow and broad components were ∼ 200 and ∼ 600 km/s, respectively.

Thus, during the outburst of the hot component of V407 Cyg, the Hα line width increased by about a factor of two over its width in 1993–1995, when the star was in its quiescent state. At the same time, the width of the wings nearly tripled, reaching ∼1500 km/s (Fig. 3). The broad wings may be due to a rapid wind from the hot star in outburst.

In August 2000, the Mira was close to its pulsation minimum ($\varphi \approx 8.60$), and the hot companion in outburst dominated in the *BV* bands [3]. At that time, the spectrum of V407 Cyg displayed clearly visible Paschen emission lines. Two of these were blended with CaII lines, however, the P(14) λ8598 Å line was free of blending. The equivalent width of this line was 0.65 Å, and its profile can be fit fairly well by a single Gaussian. Recall that the Hα profile at that time was saddle-like, and the heliocentric velocity of its absorption and emission components was about −42 km/s. At the same time, the heliocentric radial velocity of the P(14) line was about −53.9 km/s.

3.2. HeII λ4686 Å Line

In high-resolution spectra, the HeII λ4686 Å emission line is strongly underexposed, and this spectral interval is completely missing from many of the spectra. This line was detected confidently only in the spectrum obtained in October 1995 ($\varphi \approx 6.28$), and was also detected in a low-resolution spectrum obtained at that same time [11]. This line was also detected in August 1994 [9].

During the outburst maximum, the HeII λ4686 Å line was not observed in either high- or low-resolution spectra. It may be present only in the last low-resolution spectrum obtained on October 9, 2002 ($\varphi \approx 9.68$). Its equivalent width was fairly small (∼ 0.2 Å), and further observations are required to confirm the presence of this line in the spectrum of V407 Cyg.

3.3. FeII and [FeII] Lines

The profiles of the FeII and [FeII] emission lines are appreciably different. The FeII profile displays a narrow component and a broad base, and can be reproduced as a sum of two Gaussians shifted in wavelength. The forbidden lines have only a narrow component (Fig. 5). In June 1999, the FWHM of the narrow and broad components of the FeII lines (with the

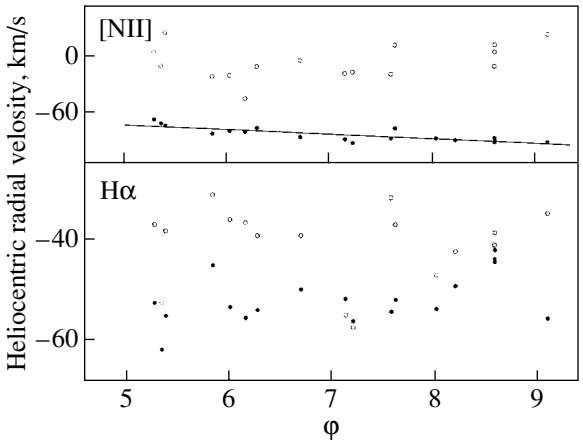

Fig. 4. Heliocentric radial velocities of the narrow emission (circles) and absorption (dots) components of the Hα line, and of the central (dots) and red (circles) components of the [NII] λ6583 Å line, as functions of φ. The solid curve indicates the linear trend in the radial velocity of the central [NII] component.

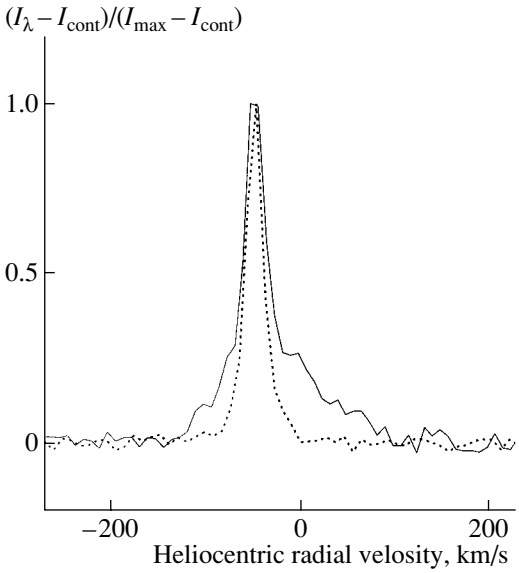

Fig. 5. Normalized FeII λ6456 Å (solid curve) and [FeII] λ7155 Å (dotted curve) line profiles in August 2000.

radial velocities of the individual components given in parentheses) were ∼ 30 km/s (−52 km/s) and ∼ 170 km/s (−25 km/s), respectively, while these widths were ∼ 20 km/s (−50 km/s) and ∼ 115 km/s (−32 km/s) in August 2000.

The radial velocities of the narrow emission component of the FeII lines roughly coincide with those of the [FeII] lines (Table 2). The radial velocity of the FeII narrow component averaged over the total observation time is $\langle V \rangle \approx -46.2$ km/s, while the average radial velocity of the [FeII] lines is $\langle V \rangle \approx -45.0$ km/s. We can therefore assume that they are formed in a single region in which there are no substantial gas motions, whereas the broad FeII bases originate in regions with higher velocity dispersions, which are presumably located closer to the hot companion.

3.4. Nebular [NII] and [OIII] Lines

Figure 1 presents the equivalent width of the [NII] λ6584 Å line as a function of the optical-brightness phase of the Mira. At phases φ = 5–6.5, the variations of W([NII]) are similar to those of W(Hα), and occur with the same amplitude (∼ 10). However, after φ ≈ 6.5, when the Hα equivalent width decreased by a factor of ∼ 30 (in August 1996 and July 1997), the [NII] lines remained fairly bright. In the spectrum obtained on August 30, 1996, the [NII] λ6584 Å line was even brighter than the Hα line (Fig. 2). A similar situation was observed in 1985 and 1987 [8].

This occurs when the hot component goes into the passive state, and the source of ionizing radiation dramatically decays. This results in the partial recombination of the nebula. The recombination time depends on the electron density in the nebula and the recombination factor of the corresponding ion: $\tau_{rec} = 1/(\alpha_{rec} n_e)$. It is likely that the structure of the nebula in V407 Cyg is complicated and inhomogeneous, as in other symbiotic stars. The nebular [NII] lines are formed in regions with $n_e \leq 10^6$ cm^{-3}, whereas the main contribution to the Hα line flux comes from regions with substantially higher density, if they exist in the envelope. Therefore, the recombination time in the region of formation of the nebular [NII] lines can appreciably exceed that in the dense layers of the envelope.

Accordingly, the [NII] lines will decay much more slowly than Hα and Hβ. This provides evidence for a substantial density decrease in the envelope of V407 Cyg. In this case, the disappearance of the broad component in the Hα profile in August 1996 and July 1997 may be in part related to the fact that the hydrogen ions recombine in relatively dense layers, while they remain ionized in less dense layers. This situation is fairly likely, since the envelope is probably far from spherically symmetric in the vicinity of the accreting hot component.

The nebular [NII] lines display complex profiles that consist of two or three emission components. The central component, which is accompanied by a red component in all our spectrograms (Fig. 6), is the most intense. A faint blue component shifted from the central component by ∼80 km/s was detected only in 1993 and 2000. Figure 6 presents the [NII] λ6584 Å line profile in September 2001.

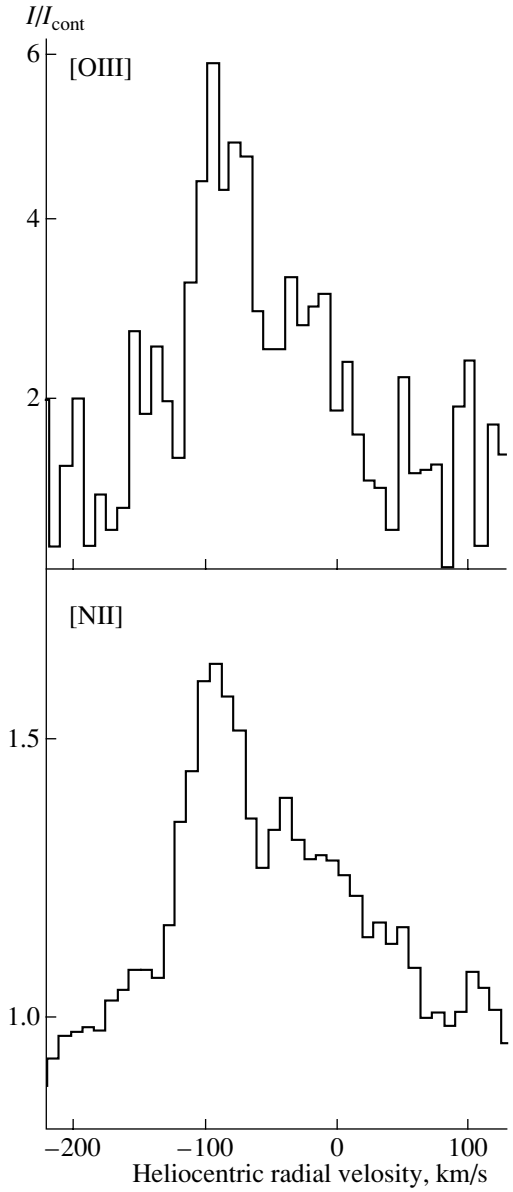

Fig. 6. Normalized [NII] λ6583 Å and [OIII] λ5007 Å line profiles in September 2001.

Figure 4 displays variations of the radial velocity of the central and red components of the [NII] λ6584 Å line. A linear trend can be noted in the variations of the radial velocity of the central component: $V \approx -75$ km/s in 1993, while $V \approx -95$ km/s in 2001. The difference between the radial velocities of the central component of the [NII] line and the narrow component of the Hα line exceeds 30 km/s, testifying that these lines, indeed, originate in different parts of the circumstellar envelope of V407 Cyg. The average radial velocities of the central and red components of the [NII] line at the time of our observations were $\langle V \rangle \approx -85$ km/s and -7 km/s, respectively. The average of these two values is ~ -40 km/s, close to the average radial velocity of the narrow component of the Hα line and the red giant. This suggests that the structure of the region of formation of the [NII] lines is bipolar, as, for example, in CH Cyg [13].

Before the outburst, the brightness of the [OIII] λ5007 Å line was comparable to that of the Hβ line, placing it among the brightest lines in the spectrum of V407 Cyg. At phases $\varphi = 5$–6.5, the amplitude of the equivalent-width oscillations of this line is half the amplitude in the Hα and [NII] lines. The equivalent width of the [OIII] line decreased dramatically in August 1996 and July–September 1997, though the line was still visible. It disappeared at the outburst maximum (1999), appearing again only in 2001.

The presence of the [OIII] line in the passive state of the hot component, when the source of ionizing radiation is switched off, is explained in the same way as the apparent enhancement of the [NII] line against the background of the dramatically weakened Hα line. Its absence during the outburst of the hot component (1999–2000) may be due to the fact that O^{++} ions originating during the quiescent state of the hot component had already recombined by that time, whereas the temperature of the hot component in outburst was not sufficient to doubly ionize oxygen. The presence of SiII $\lambda\lambda$6347, 6371 Å lines in the spectrum of V407 Cyg (see below) at that time supports this suggestion. These lines are characteristic of classical symbiotic stars during outbursts of their hot components, when they observationally resemble A–F supergiants.

Before the outburst, the [OIII] λ5007 Å line could be fit fairly well by a single Gaussian. Its radial velocity was $\langle V \rangle \approx -70$ km/s, which is ~ 10 km/s higher than the average radial velocity of the central component of the [NII] λ6584 Å line for this time interval. When the [OIII] line reappeared in 2001, its profile displayed a three-component structure (Fig. 6), with the radial velocity of its central component being roughly the same as that of the central component of the [NII] line.

4. ABSORPTION SPECTRUM OF V407 Cyg

4.1. Absorption Spectrum of the Mira Star

The detection of the ^{7}LiI λ6708 Å resonance doublet in the spectrum of the Mira is of great interest. The presence of lithium in the atmosphere of this star, which, judging from its period, has progressed appreciably in its evolution along the asymptotic giant branch, is explained by the penetration of the convective envelope into the hydrogen shell source (see [14] and references therein for details). In a spectrum

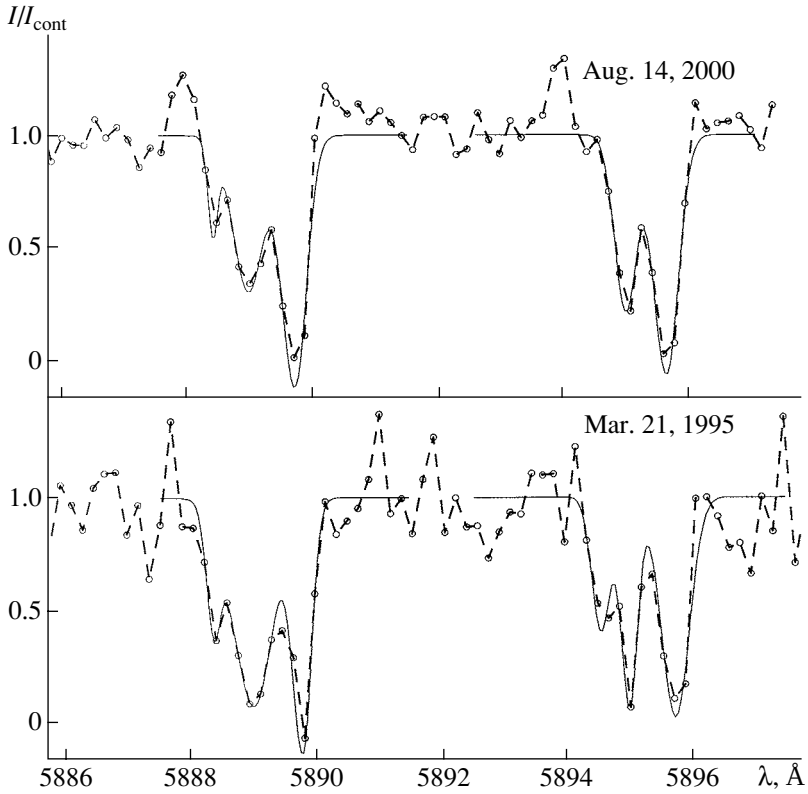

Fig. 7. Profiles and model fits (thin solid curves) of the NaI D1 and D2 lines when the dominant contribution to the radiation of V407 Cyg in this wavelength interval was made by the Mira (1995) and by the hot component (2000).

obtained in July 1997, when the hot component was in its passive state, the equivalent width of the lithium line was ~ 0.34 Å. On other dates when this line was detected, the hot component was in its active state, manifest as a decrease in the line equivalent width. However, intrinsic variability of the line itself in the atmosphere of the Mira—depending, for example, on the brightness phase—cannot be totally ruled out.

Table 2 presents estimates of the radial velocity of this line. Our resolution is not sufficient to separate the doublet lines (the difference of their wavelengths is ~ 0.15 Å). Therefore, we used the average wavelength of these lines as the laboratory wavelength when estimating the radial velocities. The average radial velocity over four observation dates when the LiI line was detected is $\langle V \rangle \approx -40.8$ km/s.

Table 2 also presents the radial velocities of the CaI λ6573 Å absorption line. This line is situated in the red wing of the Hα line, and is clearly visible when the Mira dominates in this spectral interval (see the upper right panel in Fig. 2). As was the case for the lithium line, the maximum equivalent width of the calcium line was detected during the quiescent state of the hot component, in August 1996 and July 1997: $W(\mathrm{CaI}) \approx 0.17$ Å. The average radial velocity over five observation dates when the calcium line was detected is $\langle V \rangle \approx -41.2$ km/s, coincident with the analogous value for the lithium line. Thus, we can adopt -41 km/s as the average radial velocity of the Mira. Note that, on average, the [FeII] lines indicate a slightly more negative radial velocity, although the difference is only ~ 4 km/s, while the narrow components of the Hα and Hβ lines display a slightly more positive velocity, by ~ 2.5 km/s.

Among late-type red giants, R Cmi, SU Mon, R Cyg, and HR 8714 have roughly the same lithium-line equivalent widths [15]. All these stars belong to type S. At the same time, the ratio of the equivalent widths of the lithium and CaI λ6573 lines in V407 Cyg is somewhat higher than in these stars. This probably indicates a higher lithium abundance in the atmosphere of V407 Cyg. A similar lithium-line equivalent width is displayed in the LMC by the S star HV 5584, whose effective temperature is ~ 3200 K; this corresponds approximately to a spectral type of M6, while the lithium abundance in its atmosphere is $\log \varepsilon(^7\mathrm{Li}) \approx 2.8$ [16].

The spectrum of V407 Cyg displays numerous, strong TiO bands, while VO bands are also visible in the red (λ7334–7534 Å and others). The long pulsation period of the Mira and the presence of the

Table 3. Heliocentric radial velocities and equivalent widths of the NaI λ5890 Å line

Date	Blue component		Red component	
	W, Å	V, km/s	W, Å	V, km/s
November 12, 1994	0.60	−46.6	0.45	−5.6
March 21, 1995	0.74	−46.3	0.42	−5.7
June 9, 1999	0.49	−46.3	0.50	−10.1
October 29, 1999	0.34	−31.6	0.42	−6.2
August 14, 2000	0.43	−48.2	0.55	−11.2
September 9, 2001	0.36	−44.2	0.52	−12.0
Average	–	−43.8	–	−8.4

Table 4. Heliocentric radial velocities and equivalent widths of the KI λ7699 Å line

Date	Blue component		Red component	
	W, Å	V, km/s	W, Å	V, km/s
June 9, 1999	0.27	−51.4	0.14	−15.0
October 29, 1999	0.21	−45.5	0.16	−11.8
August 14, 2000	0.29	−42.4	0.18	−18.5
September 9, 2001	0.26	−44.1	0.18	−14.3
April 23, 2002	0.5	−38.0	0.20	−9.4
Average		−44.3		−13.8

Table 5. Equivalent width of the λ5780 Å diffuse absorption band and corresponding color excess

Date	W, Å	$E(B-V)$
August 8, 1998	0.51	1.0
June 9, 1999	0.28	0.6
October 29, 1999	0.45	0.9
August 16, 2000	0.53	1.0
September 9, 2001	0.36	0.7
April 23, 2002	0.42	0.8
Average	0.42	0.8

lithium line in its spectrum suggest that the red giant in V407 Cyg is an S or MS star; i.e., its atmosphere is highly enriched with s-process elements as a result of the third dredge-up.

However, no ZrO molecular bands were detected at 6474 and 5751 Å. The presence of these bands in low-resolution spectra provides grounds for classifying a red giant as an MS or S star, with stars of these types being divided into subtypes depending on their intensity [17]. The absence of these bands implies that the V407 Cyg Mira does not belong to one of these types. Note that, due to the absence of ZrO λ6474 Å absorption, the nearby TiO λ6479 Å band appears rather intense, and is clearly visible in the low-resolution spectra.

Jurdana–Sepic and Kotnik–Karuza [18] found very faint ZrO $\lambda\lambda$5718, 6412 Å bands and an absence of ZrO $\lambda\lambda$5551, 6474 Å bands in a high-resolution spectrum of the symbiotic star CH Cyg, which contains an M7 red giant. Weak absorption features are also visible in the spectrum of V407 Cyg at 5718 and 6412 Å, which may likewise be due to ZrO molecular bands. However, they are not observed in the low-resolution spectra of V407 Cyg.

Accordingly, the V407 Cyg Mira should be assigned to spectral type M [17]. The spectral type estimated from the short-wavelength drop-off of the λ7053 Å TiO band when it dominated in this spectral interval was M6–M7. During the passive state of the hot component in July 1997, the same estimate was obtained from the λ5448 Å TiO band. Thus, these supplementary data confirm that variations of the spectral type of the V407 Cyg red giant are insignificant in the course of the brightness variations of the Mira [9].

We have also identified a number of lines of rare-earth elements in the spectrum of the Mira. For example, BaII λ4334 Å and LaII λ4827 Å lines are seen at green wavelengths, while ZrI λ7440 Å and NdII λ7514 Å lines are visible in the red. Overall, the high-resolution spectrum of the Mira is very similar to that of the cool component of the symbiotic star CH Cyg, which also has spectral type M6–M7. The only substantial difference is that the spectrum of CH Cyg does not display any lithium lines.

4.2. Absorption Lines of the Circumstellar Envelope

The NaI $\lambda\lambda$5890, 5895 Å (Fig. 7) and KI λ7699 Å absorption lines display a pronounced two-component structure. Tables 3 and 4 present the radial velocities and equivalent widths of individual components of the NaI λ5890 Å and KI λ7699 Å lines on dates when it was possible to measure them. The average radial velocities of the blue components of the NaI λ5890 Å and KI λ7699 Å lines are $\langle V\rangle \approx -43.8$ and -44.3 km/s, respectively. Since $\langle V([\mathrm{FeII}]\approx -45$ km/s and $\langle V(\mathrm{CaI})\rangle \approx -41.2$ km/s, this suggests that the blue components of the NaI and

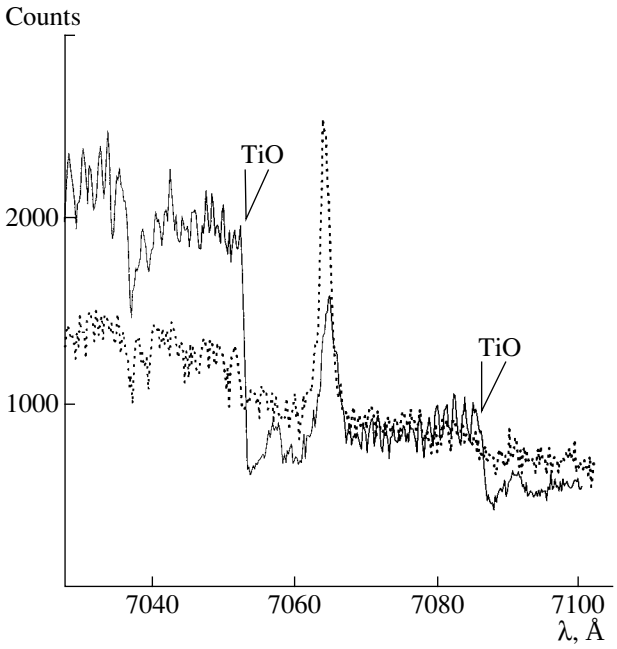

Fig. 8. TiO λ7053 Å molecular band in spectra obtained on August 14, 2000 (dotted curve) and September 9, 2001 (solid curve).

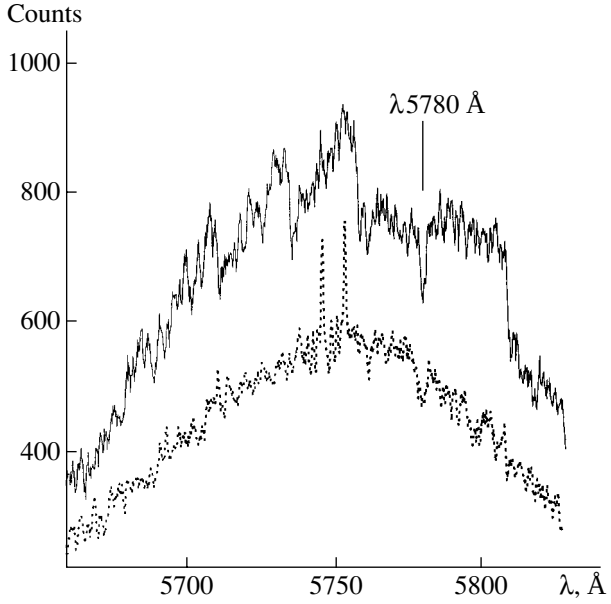

Fig. 9. The λ5780 Å diffuse absorption band in the spectra obtained on June 9, 1999 (solid curve) and August 14, 2000 (dotted curve).

KI lines are formed in the circumstellar envelope of V407 Cyg, while the red component ($\langle V\rangle \approx -10$ km/s) presumably forms in the interstellar medium.

The NaI D1, D2 lines sometimes indicate the presence of a faint third absorption component at radial velocities $V \approx -80$ km/s (Fig. 7). In the case of the NaI λ5890 Å line, this component may be the TiO λ5889.4 Å molecular band; however, the NaI λ5896 Å is free of molecular bands. It is also possible that faint emission components appeared in the NaI D1, D2 lines in April 2002.

The equivalent widths of the NaI doublet are so high that they fall on the horizontal section of the W(NaI)–$E(B-V)$ dependence [19]. Likewise, the equivalent width of the blue component of the KI line is so high that it does not correspond to any real color excess. In addition, it is obvious for a number of reasons that the W(NaI)–$E(B-V)$ (W(KI)–$E(B-V)$) dependence obtained for the interstellar medium is not suitable for circumstellar gaseous dust envelopes. In particular, this may be due to the fact that the dust of the interstellar medium is essentially a mix of silicate and carbon particles covered with ice mantles, while the dust grains in the envelopes of M red giants consist of silicates without ice mantles. The average color excess for the interstellar reddening estimated from the equivalent width of the red component of the KI line is $E(B-V) \approx 0.7$.

Since there is no significant difference in the average radial velocities of the [FeII], NaI, and CaI lines, our spectral resolution is insufficient to detect the fine structure of the blue component of the Na line and determine to what degree it is formed in various parts of the circumstellar envelope of V407 Cyg. However, indirect data suggest that its formation is not purely due to absorption of the radiation of the Mira in its immediate vicinity.

This is evident from a comparison of the data obtained in August 2000 and September 2001 (Table 3). In August 2000, the Mira was close to its minimum brightness ($\varphi \approx 8.60$), and the active hot component dominated in the optical radiation of V407 Cyg. This is confirmed by the fact that the TiO λ5448 Å band was not visible in the spectrum, while the TiO λ7053 Å band corresponded to a spectral type of M1. In September 2001, the Mira was close to its maximum brightness ($\varphi \approx 9.11$). Although the hot component was still in its active state [3], it did not yet dominate in the optical. The TiO λ5448 Å band was rather intense (Fig. 8), and corresponded to a spectral type of M5, while the TiO λ7053 Å band corresponded to spectral type M6.

Consequently, if the blue component of the Na line formed exclusively due to absorption of the Mira's radiation in its circumstellar envelope, the equivalent width in August 2000 should have been substantially lower than in September 2001, whereas in reality it was slightly higher (see Table 3). In other words,

this component does not disappear against the background of the radiation of the hot component and ionized envelope. Thus, the NaI D1, D2 doublet forms due to the absorption of radiation both from the Mira and from its hot companion.

In August 2000, absorption lines of the OI $\lambda\lambda$7771.96, 7774.18, 7775.40 triplet were detected in the spectrum of V407 Cyg. The OI absorption lines are probably blended with fainter emission lines; however, it was not possible to separate the components. The lines are only visible when the hot continuum is very strong even at red wavelengths. The OI triplet lines are most likely formed in regions that are closer to the hot components than to the cool giant. The average radial velocity of these lines is $V \approx -25$ km/s, close to that of the broad FeII component.

In August 1998 and August 2000, the SiII $\lambda\lambda$6347, 6371 Å lines were detected in the spectrum of V407 Cyg. These displayed an inverse P Cygni profile that consisted of two emission components and a sharp absorption component. Such lines are typical for the A–F supergiant phase observed in many classical symbiotic stars during bright outbursts. The radial velocity of the absorption component of the SiII lines was $V \approx -15$ and -22 km/s in 1998 and 2000, respectively, close to that of the OI triplet.

In the spectrum obtained in September 2001, two lines of the CaII $\lambda\lambda$8542, 8662 Å triplet display inverse P Cygni profiles with sharp emission and absorption components. The emission components are at radial velocities $V \approx -69.5$ and -61.4 km/s, while the absorption components are at $V \approx -39.2$ km/s and -40.1 km/s. In July 1999, the CaII triplet consisted of sharp absorption lines without any trace of emission components. The heliocentric radial velocities of these lines were -50.2 km/s (for λ8542 Å) and -48.4 km/s (for λ8662 Å). These lines may also have had P Cygni profiles in August 2000, however, the CaII lines were blended by much stronger Paschen emission lines at that time.

5. DIFFUSE ABSORPTION BANDS

Finally, the last constituent of the V407 Cyg absorption spectrum are its diffuse absorption bands. These are formed when the radiation of the star passes through the interstellar medium, however, details of their origin remain unclear.

The λ6283.9 Å band is most clearly visible in the spectrum of V407 Cyg. It has a complex multicomponent structure. Unfortunately, this band is not very suitable for determining the interstellar reddening. Reasonably good calibration curves are available only for bluer diffuse absorption bands [20]. A rough color excess estimated from the equivalent width $W(\lambda 6284) \geq 1$ Å is $E(B-V) \geq 0.6^m$.

The λ5780 Ådiffuse absorption band in spectra of V407 Cyg obtained in 1993–1995 is strongly distorted by noise, but, it was clearly seen in 1998–2002 (Fig. 9). Table 5 presents the equivalent widths and corresponding color excesses of this band. The average color excess is $E(B-V) \approx 0.8$, in good consistency with the $E(B-V)$ estimate obtained above from the equivalent width of the KI λ7699 Å line.

6. CONCLUSION

High-resolution spectral observations of the symbiotic Mira variable V407 Cyg in 1993–2002 found the hot component in all three of its possible states. These observations appreciably supplemented the description of these states given in [3] based on $UBVJHKLM$ photometry.

During the passive state of the hot component, all its optical manifestations disappear, except for the faint Hα line and forbidden [OIII] and [NII] lines. The Hα line profile consists of an absorption and a narrow (~ 90 km/s) emission component, and can be reproduced by a sum of two Gaussians. In the transition to the passive state, when the source of ionizing radiation becomes dramatically fainter, the ratio of the [NII] λ6583 Å and Hα line fluxes substantially increases. This indicates that the circumstellar envelope of V407 Cyg contains regions with high (10^8–10^9 cm^{-3}) densities, where the bulk of the Hα and Hβ fluxes are formed. The forbidden [NII] lines are formed in more rarefied outer parts of the circumstellar envelope, which may have a bipolar structure on these scales.

The HeII λ4686 Å line is observed in the quiescent state of the hot component. The flux in this line corresponds to the temperature of the star, which is equal to $\sim 6 \times 10^4$ K [9]. The broad (~ 300 km/s) component and wings in the Hα line profile extend to velocities of 500 km/s appear.

During the outburst of the hot companion, the FWHM of the narrow and broad components of the Hα line doubles, and the wings sometimes extend to 1500 km/s. At this time, the HeII λ4686 Å line disappears from the spectrum of V407 Cyg, while the SiII $\lambda\lambda$6347, 6371 Å lines become visible. These display inverse P Cygni profiles with two emission components and a sharp absorption. These lines are characteristic of the A–F supergiant phase observed in many classical symbiotic stars during bright outbursts. The color temperature of the hot companion

in the UBV bands is ~ 7000 K [3]. Thus, the temperature of the hot star decreases when it is in outburst.

If the outburst of the hot component of V407 Cyg had started from its quiescent rather than its passive state, it would photometrically and spectrally have resembled the outburst of a classical symbiotic star, such as BF Cyg. In particular, during this transition in classical symbiotic stars, the bolometric flux of the hot components increases by a factor of only a few (see [3] and references therein). Note that Merrill [21] classified the hot component of the symbiotic Mira R Aqr in what we call the quiescent and outburst states as an Ofp and Bep star, respectively.

The emission-line profiles (with the exception of that for [FeII]) display a variable multi-component structure. During our observations, the radial velocity of the central component of the [NII] λ6583 Å line gradually decreased from -75 km/s in 1993 to -95 km/s in 2002. At the same time, no regular trend in the variations of the radial velocities of either the absorption or emission components of the Hα line was found. The average radial velocity of the narrow emission components of the Hα and Hβ lines is close to that of the red giant, and they probably form in outer (relative to both components of V407 Cyg) parts of the stellar envelope.

The broad component of the Hα line is formed close to the hot companion during periods of accretion activity. The star's departure from the passive state is probably initially manifest in the appearance of the broad component, which then dominates the Hα profile (hydrodynamically rather than energetically). At this time, the brightness of the hot star remains low. When this brightness increases, the narrow component again begins to dominate the Hα profile, and the broad wings appear.

The LiI resonance doublet has been detected in the spectrum of V407 Cyg. At the same time, none of the intense ZrO bands that would be visible in low-resolution spectra were seen. In particular, the high-resolution spectra have enabled us to identify an absorption band near 6476 Å visible in low-resolution spectra with the TiO λ6479 Å band rather than the ZrO λ6474 Å band. Thus, the cool component of V407 Cyg is an oxygen Mira. The M6–M7 spectral type of the Mira does not display any substantial variations during its brightness pulsations.

The color excesses estimated from the equivalent widths of the component of the KI λ7699 Å line that originates in the interstellar medium and of the λ5780 Å diffuse absorption band are $E(B-V) \approx 0.7$ and 0.8, respectively.

Three subtypes are now recognized among the entire collection of symbiotic stars, according to the type of outbursts of their hot components: classical symbiotic stars (Z And), symbiotic novae (V1016 Cyg), and recurrent novae (T CrB). However, there also exist symbiotic stars whose outbursts differ from these. These include such well-known objects as CH Cyg, R Aqr, and MWC 560, as well as the perhaps less known V407 Cyg. The photometric and spectral manifestations of their activity are fairly similar. It is therefore possible that they may represent another subtype of symbiotic star, which could be called R Aqr-type stars.

The main distinguishing features of these stars are the following.

(1) The hot component undergoes a passive state, when it is essentially invisible against the background of the Mira-type star radiation and does not excite bright emission lines in the nebula.

(2) During an outburst of the hot component, its bolometric luminosity is only several percent of the bolometric luminosity of the cool component. The Hα line displays broad (>1000 km/s) wings. No HeII λ4686 Å line is seen, and the star can formally be classified as a hot giant.

(3) There exists an intermediate state between the passive and outburst states, which we have called here the quiescent state. It differs from the outburst state not only in the lower optical brightness of the hot component, but also in the substantially higher excitation of the emission spectrum, in which HeII $\lambda\lambda$1640, 4686 Å lines appear. The Hα line wings are present, but are appreciably less extended than during outburst.

The radiation of the hot component displays rapid variability (flickering).

Note that these characteristics must be considered in combination, since, individually, they may also be typical of other subtypes of symbiotic stars. The unique properties of R Aqr-type symbiotic stars are the low luminosity of the hot component relative to that of the red giant, even during outbursts of the hot component, and also the presence of these three qualitatively different states.

R Aqr and V407 Cyg clearly demonstrate a third property, whose presence can be detected only when the cool component has sufficiently low optical brightness that it does not mask the emission spectrum of the hot component. The cool components of R Aqr and V407 Cyg are Mira stars with large amplitudes for their optical brightness variations, so that this condition is satisfied in the vicinity of the brightness minimum of the Mira. At the same time, the cool components of CH Cyg and MWC 560 are not Miras, and the existence of three possible states of their hot components has not been detected for these symbiotic stars. This problem may be solved

after more detailed studies of these stars using already available observational data.

7. ACKNOWLEDGMENTS

The authors thank the referee for useful comments. This study was supported by the Russian Foundation for Basic Research (project nos. 02-02-16235, 03-02-16622, 03-02-06765) and The Polish Research Committee (grant KBN 5P03D00320).

REFERENCES

1. C. Hoffmeister, Veroff Sterne. Sonneberg. **1**, 295 (1949).
2. L. Meinunger, Mitt. Veranderl. Sterne. **87**, 111 (1966).
3. E. A. Kolotilov, V. I. Shenavrin, S. Yu. Shugarov, and B. Yudin, Astron. Zh. **80**, 700 (2003) [Astron. Rep. **47**, 777 (2003)].
4. K. M. Merrill and C. G. Burwell, Astrophys. J. **112**, 72 (1950).
5. G. H. Herbig, Astrophys. J. **131**, 632 (1960).
6. G. Welin, Astron. Astrophys., Suppl. Ser. **9**, 183 (1973).
7. F. Gieseking and J. D. Schumann, Astron. Astrophys., Suppl. Ser. **26**, 367 (1976).
8. V. F. Esipov and B. L. Yudin, Astrofizika **29**, 285 (1988).
9. E. A. Kolotilov, U. Munari, A. A. Popova, *et al.,* Pis'ma Astron. Zh. **24**, 526 (1998) [Astron. Lett. **24**, 451 (1998)].
10. R. G. Ivison, M. F. Bode, and J. Meaburn, Astron. Astrophys., Suppl. Ser. **103**, 201 (1994).
11. U. Munari and T. Zwitter, Astron. Astrophys. **383**, 188 (2002).
12. V. F. Esipov, private communication (2002).
13. R. M. L. Corradi, U. Munari, M. Livio, *et al.*, Astrophys. J. **560**, 912 (2001).
14. A. A. Tatarnikova, P. M. Maresse, U. Munari, *et al.*, Pis'ma Astron. Zh. (2003, in press).
15. A. M. Boesgaard, Astrophys. J. **161**, 1003 (1970).
16. V. V. Smith, B. Plez, and D. L. Lambert, Astrophys. J. **441**, 735 (1995).
17. P. C. Keenan, Astrophys. J. **120**, 484 (1954).
18. R. Jurdana-Sepic and D. Kotnik-Karuza, Astron. Astrophys. Trans. (2003, in press).
19. U. Munari and T. Zwitter, Astron. Astrophys. **318**, 269 (1997).
20. K. Josafatsson and T. P. Show, Astrophys. J. **319**, 436 (1987).
21. P. W. Merrill, Astrophys. J. **81**, 312 (1935).

Translated by K. Maslennikov

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 903–915. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 978–991.

Spectra, Optical Identifications, and Statistics of a Complete Sample of Radio Sources at Declinations 10°–12°30′

A. G. Gorshkov[1], V. K. Konnikova[1], and M. G. Mingaliev[2]

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

[2] *Special Astrophysical Observatory, Nizhniĭ Arkhyz, Russia*

Received April 7, 2003; in final form, May 8, 2003

Abstract—The results of 0.97, 2.3, 3.9, 7.7, 11.1, and 21.7 GHz observations of a complete sample of radio sources obtained on the RATAN-600 radio telescope are presented. The sample is comprised of sources from the 4.85-GHz MGB survey, and contains all sources at declinations 10°–12°30′ (J2000) with Galactic latitudes $|b| > 15°$ and flux densities $S_{4.85} > 200$ mJy. Optical identifications have been obtained for about 86% of the radio sources with flat spectra and 59% of those with steep spectra. The spectra of the flat-spectrum sources have been decomposed into extended and compact components.

1. ВВЕДЕНИЕ

We are currently studying two complete samples of radio sources. The first sample, which was derived from the Zelenchuk Survey at 3.9 GHz [1], contains all sources in this survey with flux densities $S_{3.9} >$ 200 mJy at declinations 3°30′–6° (B1950); the right ascensions are in the range 0–24^h, and the Galactic latitudes are $|b| > 10°$. This sample contains 160 objects, and we have been studying it since 1984.

The second sample, derived from the GB6 catalog at 4.85 GHz [2], contains all sources from the catalog with flux densities $S_{4.85} > 200$ mJy at declinations 10°–12°30′ (J2000); the right ascensions are in the range 0–24^h, and the Galactic latitudes are $|b| > 15°$. In this sample of 153 objects, 83 have flat spectra with $\alpha_{(3.9-7.7)} > 0.5$ ($S \propto \nu^\alpha$) and 70 have steep spectra with $\alpha_{(3.9-7.7)} < -0.5$.

The main goals of our investigations of these samples are the following:

(1) studying the variability of the sample sources on time scales from several days to several years (observations over the wide frequency range 0.97–21.7 GHz can be used to derive the main characteristics of the variability—its time scale and amplitude, as well as the spectrum of the variable component and the time dependence of its amplitude–frequency characteristics);

(2) deriving the statistical properties of the radio-source spectra;

(3) searching for interesting objects with unusual characteristics in both the radio and optical ranges;

(4) investigating the cosmological evolution of quasars (which requires that redshifts be found for the majority of the optical objects identified with the radio sources).

Daily multi-frequency observations on the RATAN-600 telescope in 1998–1999 confirmed the presence of flux-density variability on time scales of about four days detected earlier [3–5]—so-called intraday variability (IDV). Our studies have shown that such variability is inherent to virtually all flat-spectrum sources, and that IDV is characterized by a flat frequency spectrum with a mean modulation index of about 2% at 2.3–21.7 GHz. Roughly 20% of the objects display significant IDV, as is clearly visible in their structure functions. Several sources show cyclic variability on characteristic time scales of 4–25 days. The statistical properties of the long-term variability of sources in the complete sample and the properties of individual flares for the most active sources have been derived [1, 6].

Various statistical characteristics of the spectra of the sample sources have also been obtained [7, 8]. The unique radio source 0527 + 0331, found to have the most prominent long-term variability, was discovered and studied (all source names are comprised of the first four digits of their right ascension and declination for epoch J2000) [4, 9, 10]. Optical observations aimed at obtaining spectra and redshifts of objects identified with the sample sources to 21^m are ongoing. Studies of the second sample were begun to enable confirmation of the results obtained for the first sample on the basis of a different statistically independent ensemble of radio sources. Up through 2002, three

1063-7729/03/4711-0903$24.00

series of daily multi-frequency observations were carried out for most of the sample sources over 80 days in 2000, 104 days in 2001, and 98 days in 2002.

We present here the first results obtained for the second sample of radio sources.

2. OBSERVATIONS

Meridian observations were carried out on the Northern sector of the RATAN-600 radio telescope at 0.97, 2.3, 3.9, 7.7, 11.1, and 21.7 GHz in June–November 2001. The parameters of the receivers used are presented in [11]. The observations were carried out in a fixed-focus regime, as described in [12]. Readjustment of the main mirror was possible within $\pm 1.25^\circ$ of the center of the declination zone. An equal number of panels of the main mirror were mounted at all altitudes, in order to reduce the influence of variations in the radiation of the outer panels as the curvature of the circular reflector changed. The effective area remained constant over the entire frequency range.

The calibrator for the observations was 1347 + 1217, whose angular size is much smaller than the horizontal cross section of the antenna beam right up to 21.7 GHz. We took the flux density of 1347 + 1217 to be 6.15, 4.12, 3.23, 2.36, 1.99, and 1.46 Jy at 0.97, 2.3, 3.9, 7.7, 11.1, and 21.7 GHz, respectively.

The observations were reduced using a program package that yielded both the flux densities for individual observations and the mean flux density over an entire observing series. The reduction was based on optimal filtration of the input data using the method described in detail in [13]. Before this optimal filtration, non-linear filters were used to clean the input data of impulsive noise, jumps, and trends with time scales longer than the scale of the antenna beam in right ascension. When deriving the mean flux density over an entire observing interval, we used only those recordings for which the noise dispersion at the location of the source belonged to a single general population; the method used to reduce such recordings is described in [14].

The mean flux densities were determined via optimal filtration of the mean recording, whose ith point was the median value of all the ith points of the cleaned input recordings. As a check, we also derived the mean flux density from the relation

$$\bar{S} = \left(\sum_i^n S_i\right)/n, \qquad (1)$$

where S_i is the flux density of the ith observation and n is the number of observations. Introducing a weighting function is superfluous in this case, since only recordings belonging to a single general population were summed.

It is clear that the flux densities obtained in these two ways should be similar, and appreciable differences between them suggest the presence of bad recordings that were not removed during the initial filtration. Our experience demonstrated that such differences arose fairly rarely, testifying to the correctness of the filtration algorithm applied. Bad recordings that were not excluded were primarily those corresponding to observations made with an incorrect antenna setup. If a significant difference was observed, all the recordings were inspected visually and any considered suspicious were excluded, after which the entire reduction procedure was repeated.

The errors in the measurements were determined in two ways:

$$\sigma_\Sigma = \left(\sigma^2 / \sum_i A_i^2\right)^{\frac{1}{2}},$$

where σ^2 is the dispersion of the residual noise in a mean recording after excluding the detected source and A_i is the tabulated antenna beam, and

$$\sigma_s = \left(\left(\sum_i^n (S_i - \bar{S})^2\right)/n(n-1)\right)^{\frac{1}{2}},$$

where $\bar{S}$ is the mean flux density given by (1).

These two estimates should be similar. If they belonged to different general populations (according to the Fisher criterion), we likewise searched for bad recordings. In any case, the larger of the two estimates was adopted as the error in the measured flux density. In this approach to estimating the errors, the errors also include the rms variation of the flux density associated with intrinsic variability of the source during the series of observations.

For several steep-spectrum sources for which significant linear polarization was found in [15], we constructed the flux density using the formula [16]

$$S = 0.5 S_0 [1 + p \cos 2(q - \chi - \psi)],$$

where S_0 is the total flux density of the source, p and q are the degree of linear polarization and the parallactic angle, respectively, χ is the angle between the plane of linear polarization of the receiver and the vertical, and ψ is the position angle of polarization.

3. OPTICAL IDENTIFICATIONS

We used accurate radio coordinates taken primarily from the JVAS (Jodrell Bank–VLA Astrometric Survey) catalog at 8.4 GHz [17] (an rms coordinate error of $0.014''$) and the NVSS (NRAO VLA

Sky Survey) catalog at 1.4 GHz [18] (mean rms coordinate errors of about 0.11″ and 0.56″ for right ascension and declination, respectively). The optical coordinates and B magnitudes were taken from the USNO astrometric survey [19] or APM [20]. A radio source was considered to be identified with an optical object if the difference between the radio and optical coordinates was less than 3σ of the error in the radio range coordinates.

A significant number of objects identified with sources in the sample had been classified earlier. All redshifts and object classifications presented here were obtained from the catalog of quasars and active galactic nuclei [21] published in 2001, the NED extragalactic database [22], and our earlier work [23–25]. Thirteen objects identified with sources in the sample have not yet been classified.

4. RESULTS

4.1. Radio Sources with Steep Spectra and the Flux-density Scale

The sample includes 70 sources with steep spectra. We did not include double-lobed sources whose total flux density at 4.85 GHz was less than 400 mJy in the sample. We did not observe ten sources because either they were larger than the size of the antenna beam at most of the frequencies observed or they were double sources that were unresolved in our observations. Table 1 presents for these sources (1)–(2) their coordinates, (3) classification (as quasars, Q, or galaxies, G), (4) redshifts, (5) references for the redshifts, and (6) B magnitudes.

The spectrum of the calibrator $1347+1217$ at 0.97–21.7 GHz was approximated by a power law, $\log S = A + B\log\nu$. The spectra of 68% of the sources could be approximated well with such linear relations (S spectra), while the spectra of the remaining sources were approximated using parabolas of the form $\log S = A + B\log\nu + C\log^2\nu$.

The spectra of three sources flatten at high frequencies (type C_+), probably due to the presence of a compact component, and the spectra of 15 sources flatten at low frequencies due to synchrotron self-absorption (type C_-); this flattening is most evident at frequencies below 2.3 GHz, while the spectra remain close to linear power laws at higher frequencies.

The flux densities for 36 of the steep-spectrum sources were below our detection limit at 21.7 GHz. The measured flux densities for some of the sources are underestimated, mainly at frequencies higher than 7.7 GHz, because the size of the source was comparable to the size of the antenna beam in right ascension. This underestimation is consistent with the angular sizes given in the Texas 0.365 GHz survey [26].

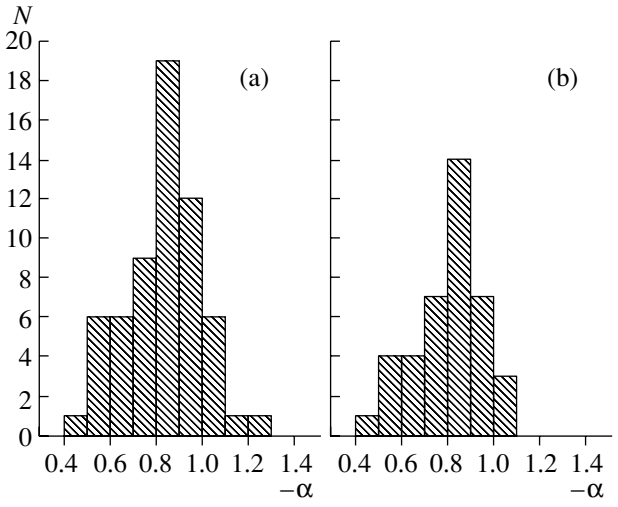

Fig. 1. Distribution of spectral indices for (a) all sources with steep spectra at 3.9 GHz and (b) sources with S spectra.

For these sources, the spectra were approximated using only the data at frequencies where the source angular size was smaller than the antenna beam.

Table 2 presents information about the 60 sources with steep spectra. The columns give the (1)–(2) J2000 coordinates of the objects, (3) optical identifications (quasar, Q, or galaxy, G; a "+" indicates that the source has been identified but a spectrum is not yet available), (4) redshifts, (5) references for the redshifts, (6) B magnitudes, and (7)–(9) coefficients A, B, and C for the approximations to the measured spectra. A star in the last column indicates sources whose angular sizes are comparable to the size of

Table 1. Coordinates and optical identifications of extended radio sources in the sample

Radio coordinates (J2000.0)		Id	z	Ref.	V
R.A.	DEC.				
$02^h17^m07.62^s$	11°04′10.59″	Q	0.408	[21]	15.9
03 58 57.71	10 27 17.76	G	0.031	[21]	15.4
04 13 42.14	11 11 44.82	G	0.306	[21]	19.5
09 14 19.36	10 06 38.28	G	0.311	[25]	19.9
12 29 51.84	11 40 24.09	G	0.083	[22]	16.0
12 30 48.27	12 23 33.07	G	0.004	[21]	12.9
14 16 53.51	10 48 40.06	G	0.024	[22]	13.0
15 14 49.50	10 17 00.74	EF			
22 49 54.59	11 36 30.84	G	0.026	[22]	9.4
23 15 34.40	10 27 18.40	G	0.255	[25]	17.6

Table 2. Optical identifications and approximations of the sample constants

Radio coordinates (J2000.0)		Id	z	Ref.	B	Parameters Approximation			Comments (see text)
R.A.	DEC.					A	B	C	
$00^h29^m08.953^s$	+11°36′28.20″	EF				2.984	−0.792	−0.086	
00 34 56.157	+10 27 52.03	G	0.057	[22]	15.0	3.021	−0.636		*
00 40 50.354	+10 03 23.14	G	0.188	[21]	18.7	3.872	−0.984		
00 44 34.653	+12 11 19.32	G	0.228	[21]	18.9	2.934	−0.832		*
02 20 47.458	+11 21 42.40	EF				3.278	−0.846	−0.108	*
02 38 30.865	+10 10 07.66	G			19.6	3.171	−0.938		
02 45 14.608	+10 47 01.57	EF				2.740	−0.796		
03 15 21.039	+10 12 43.12	G	0.222	[25]	19.6	3.331	−0.925	0.240	
03 27 23.109	+12 08 35.71	EF				3.191	−0.815	0.081	
04 40 12.413	+11 34 03.83	EF				2.998	−0.861	−0.087	
07 28 32.882	+12 10 10.39	EF				3.152	−0.891		
07 45 28.211	+12 09 28.85	EF				3.007	−1.067		
08 04 47.974	+10 15 22.73	Q	1.956	[21]	17.6	3.415	−1.030		
09 06 04.192	+11 03 27.61	G		[25]	19.2	3.011	−0.874		*
10 14 16.028	+10 51 06.16	+			18.7	3.048	−0.673	−0.224	
10 34 05.090	+11 12 31.95	EF				3.175	−0.908		
11 00 47.733	+10 46 12.98	Q	0.420	[21]	17.9	2.954	−0.936		*
11 04 34.802	+11 03 23.90	EF				3.006	−0.792	−0.138	
11 09 46.038	+10 43 43.21	EF				3.244	−0.562	−0.377	
11 26 27.190	+12 20 33.10	G			19.2	3.123	−0.910		
11 30 19.246	+10 15 26.30	EF				2.694	−0.505		
11 40 27.693	+12 03 07.44	G			15.0	3.301	−0.913		
11 53 03.107	+11 07 20.29	EF				3.079	−1.043	−0.116	
11 59 29.114	+10 46 01.00	EF				2.953	−0.920		
12 04 26.711	+11 29 09.68	EF				2.893	−0.766		
12 23 08.875	+10 29 01.05	EF				2.843	−0.643	−0.132	
12 28 36.804	+10 18 41.69	Q	2.305	[21]	19.0	2.941	−0.885		
12 31 19.866	+11 22 45.03	EF				2.979	−0.709	−0.104	
13 06 19.248	+11 13 39.79	G	0.084	[24]	15.4	2.684	−0.605		
13 09 05.161	+10 29 39.91	EF				2.826	−0.786		
13 21 18.844	+11 06 49.25	Q	2.175	[21]	18.4	3.475	−0.778	−0.052	
13 41 04.302	+10 32 05.96	EF				2.902	−0.988		
13 47 33.425	+12 17 23.94	G	0.121	[21]	14.9	3.783	−0.463		
13 52 56.363	+11 07 07.57	EF				3.295	−0.739	−0.353	
14 23 30.103	+11 59 51.24	Q	1.611	[21]	17.5	3.077	−0.635		
15 11 29.436	+10 01 43.73	EF				2.935	−0.874		
15 22 12.151	+10 41 30.35	G	0.204	[25]	21.0	2.783	−0.823		*
15 23 27.563	+11 30 23.76	+			19.1	2.697	−0.555		
15 23 56.936	+10 55 44.02	+			21.1	3.137	−0.786	−0.115	
15 59 06.913	+12 10 26.95	EF				2.839	−0.814		
15 59 16.840	+11 15 46.11	EF				2.926	−0.746		
16 04 05.732	+11 27 59.88	EF				2.871	−0.892		
16 21 10.388	+10 46 13.88	Q	1.305	[25]	21.8	3.107	−0.743	+0.118	

Table 2. (Contd.)

Radio coordinates (J2000.0)		Id	z	Ref.	B	Parameters Approximation			Comments
R.A.	DEC.					A	B	C	(see text)
$16^h31^m45.247^s$	+11°56′02.99″	Q	1.792	[21]	18.2	3.306	−0.582		
16 38 22.118	+10 35 07.74	EF				3.144	−0.614		
16 40 47.989	+12 20 02.08	EF				3.382	−0.414	−0.064	
17 09 35.035	+11 40 27.72	+			18.8	2.865	−0.780		*
17 13 43.382	+10 37 24.86	+			19.3	2.869	−0.749	−0.084	
17 27 53.762	+10 42 56.50	Q	0.833	[21]	20.2	2.784	−0.891		
20 22 08.547	+10 01 10.81	Q	0.469	[21]	19.4	3.651	−0.941		*
21 38 26.222	+11 58 04.15	EF				2.979	−0.940		*
21 51 04.053	+12 19 50.45	EF				3.114	−0.958	−0.070	*
22 01 16.687	+10 23 47.59	Q	1.729	[21]	18.1	2.832	−0.848		
22 03 45.543	+12 17 15.63	+			18.9	2.895	−0.844		
22 29 57.456	+11 27 37.73	G	0.239	[23]	21.2	2.794	−0.783		
22 41 34.421	+11 45 44.84	EF				3.062	−0.841		*
22 54 10.450	+11 36 37.90	Q	0.325	[21]	15.6	3.283	−0.739		
23 11 17.868	+10 08 15.35	Q	0.432	[21]	16.2	2.950	−0.897		*
23 12 10.467	+12 24 03.46	Q	1.285	[25]	19.1	2.842	−0.745		
23 29 41.090	+11 17 28.60	G	0.119	[22]	18.4	2.758	−0.507		

the antenna beam in right ascension at frequencies higher than 7.7 GHz. The flux density of 1140 + 1203 at 3.9 GHz is appreciably lower than the values approximated using the remaining data.

Figure 1 shows the distribution of spectral indices for (a) all the sources and (b) the sources with type S spectra. The mean spectral index for the sources with S spectra is $\bar{\alpha} = -0.808$ with an rms deviation of $\sigma = 0.15$, while the mean spectral index at 3.9 GHz for all the sources is $\bar{\alpha} = -0.815$ with an rms deviation of $\sigma = 0.17$. In [7], we obtained for the steep-spectrum sources in a sample at declinations 3°30′ − 6° (B1950) the mean spectral index $\bar{\alpha} = -0.857$ with an rms deviation of $\sigma = 0.14$. After taking into account sources that were obviously resolved by the antenna beam, this value decreased to $\bar{\alpha} = -0.81$, which is virtually identical to the mean spectral index obtained for our second sample in a different range of declinations.

The mean ratio of our 1.4-GHz flux densities and the flux densities from the NVSS survey [18] for all the steep-spectrum sources with sizes much less than the size of the antenna beam below 11.1 GHz is $S_{app1.4}/S_{\rm NVSS} = 1.019 \pm 0.004$. The mean ratio of our 4.85-GHz flux densities and the flux densities of the GB6 survey is $S_{app4.85}/S_{\rm GB6} = 1.09 \pm 0.01$. On average, our flux-density estimates for 0.365 GHz are in agreement with those in the Texas survey, although the accuracy of our approximations outside the range of our observed frequencies is lower than the accuracy within this range.

Of the 70 steep-spectrum radio sources, 21 are identified with galaxies (17 of which have measured redshifts), 14 are identified with quasars, and 29 are in empty fields to 21^m. The spectra of six objects have not yet been obtained. Figure 2 shows the redshift distributions of the galaxies and quasars. The mean redshift of the galaxies is $\bar{z}(\mathrm{G}) = 0.15$ ($\sigma = 0.1$), while that of the quasars is $\bar{z}(Q) = 1.22$ ($\sigma = 0.7$).

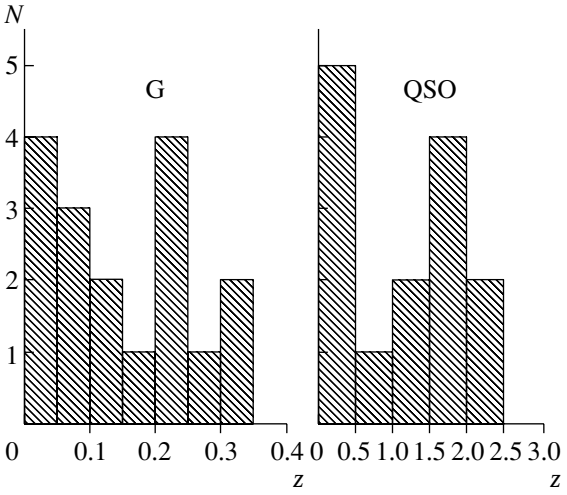

Fig. 2. Redshift distribution of galaxies and quasars identified with steep-spectrum sources.

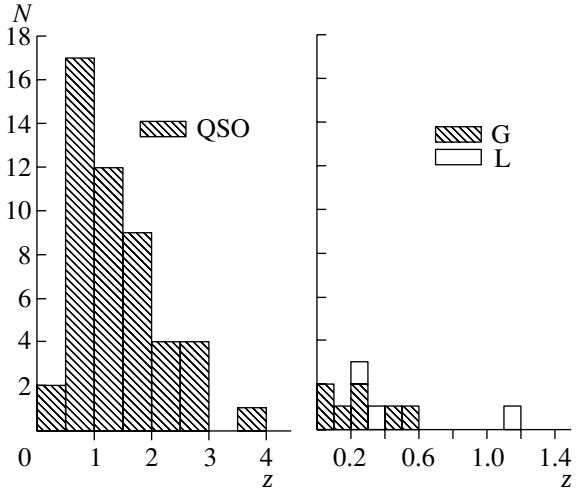

Fig. 3. Redshift distribution for the quasars, galaxies, and BL Lac objects identified with flat-spectrum sources in the sample.

4.2. Flat-spectrum Radio Sources ($\alpha > -0.5$)

Table 3 presents for the flat-spectrum sources their (1)–(2) J2000 radio coordinates, (3) optical identifications, (4) redshifts, (5) references for the redshifts, (6) B magnitudes, and (7)–(18) flux densities and rms measurement errors. Flux densities were obtained at six frequencies for all these sources except $2203+1007$, whose flux density at 0.97 GHz does not exceed 70 mJy, and $0448+1127$, whose 7.7 GHz flux density we were not able to estimate because the neighboring strong source $0449+1121$ fell in the second feed horn. The flux densities of $0449+1121$ and $1728+1215$ are significantly and rapidly variable, and the table presents the spectra obtained on June 5, 2001 and August 30, 2001, respectively.

Of the 83 flat-spectrum objects in the complete sample, 72 (86%) have optical identifications, and 11 are in empty fields to 21^m. The spectrum of the optical object identified with $1603+1105$ is stellar, probably because the real optical object is blended with a star. Forty-nine of the sources are classified as quasars and eight as galaxies (seven of which have measured redshifts). Seven sources are BL Lac objects (three of which have measured redshifts), among them $0409+1217$, $0757+0956$, and $1309+1154$, which have high degrees of polarization in the radio. Six objects have not yet been classified.

Figure 3 presents the redshift distribution of the quasars, galaxies, and BL Lac objects. The mean redshift of the quasars is $\bar{z}(\mathrm{Q}) = 1.40$ (rms deviation $\sigma(\mathrm{Q}) = 0.74$), of the BL Lac objects is $\bar{z}(\mathrm{L}) = 0.55$ ($\sigma(\mathrm{L}) = 0.40$), and of the galaxies is $\bar{z}(\mathrm{G}) = 0.26$ ($\sigma(\mathrm{G}) = 0.19$).

Figure 4 shows the redshift dependence of the absolute spectral luminosity at 11.1 GHz for the flat-spectrum sources in the sample calculated for a uniform, isotropic cosmological model with a zero cosmological constant, deceleration parameter $q = 0.5$, and $H = 50$ km s^{-1}Mpc^{-1}. The solid line represents the minimum luminosity that can be detected for the given flux-limited sample.

Most of the flat-spectrum sources are composed of at least two components—an extended one and a compact one.

To study the characteristics of the spectra of compact components, we must isolate these spectra by excluding the contribution of any extended components. Here, when we refer to the "compact component," we mean the total radiation of the jet, which is made up of clouds of relativistic electrons that have formed as a result of propagating shock waves and are located in various stages of their evolution. The formation and evolution of these clouds under the action of shocks in the relativistic jet plasma is thought to give rise to the observed variability of extragalactic radio sources [27–29]. Valtaoya *et al.* [30] present a semi-quantitative analysis of a "generalized model" that can be applied to trace the evolution of the spectral and temporal characteristics of flares occurring at various stages of the interaction of a shock front and cloud of relativistic electrons.

Spectra covering more than six decades in frequency are constructed in [31] based on nearly coordinated, simultaneous observations carried out from 20 cm to 1400 Å. The spectra of active sources are smooth, and are fit well by a parabolic dependence on a logarithmic scale. At the same time, the spectra of sources with low activity are less smooth and poorly

Table 3. Identifications and flux densities of flat-spectrum sources

Radio coordinates (J2000.0)		Id	z	Ref.	B	Flux density and its error, mJy											
R.A.	DEC.					0.97 GHz		2.3 GHz		3.9 GHz		7.7 GHz		11.1 GHz		21.7 GHz	
$00^h07^m55.710^s$	$+10°27'43.89''$	EF				310	25	300	11	285	4	262	6	255	8	225	10
00 10 31.007	+10 58 29.51	G	0.089	[21]	14.1	90	15	148	10	182	2	257	3	354	4	650	20
00 36 23.767	+10 07 57.43	Q	1.909	[21]	17.2	822	25	553	10	501	7	483	10	490	10	500	26
00 37 26.042	+11 09 50.91	EF				235	20	235	14	230	9	215	4	223	4	226	10
00 38 18.017	+12 27 31.25	Q	1.395	[21]	16.5	1196	26	781	10	683	6	564	6	539	6	440	22
00 42 44.371	+10 09 49.19	EF				257	20	252	18	243	6	215	6	230	7	210	20
01 21 29.001	+11 27 00.53	Q	2.487	[23]	19.7	88	15	150	15	186	6	210	6	181	11	150	20
01 21 41.595	+11 49 50.42	Q	0.570	[21]	19.2	1532	20	1704	17	1779	18	1948	13	2030	14	1996	45
01 43 31.090	+12 15 42.95	Q	1.178	[25]	19.7	348	15	220	10	173	3	136	4	128	2	110	10
02 03 46.657	+11 34 45.39	Q	3.610	[21]	19.4	682	24	882	11	902	5	984	7	1039	8	1060	32
02 11 13.177	+10 51 34.79	L		[23]	15.1	160	15	297	9	317	5	333	5	364	5	420	20
02 25 41.910	+11 34 25.47	Q	0.924	[21]	17.9	380	22	361	8	333	4	280	5	282	4	303	15
02 42 29.171	+11 01 00.72	Q	2.694	[25]	19.3	1690	25	1341	15	1166	7	1156	9	1266	9	1357	43
03 02 30.548	+12 18 56.77	EF				514	29	536	18	501	5	415	7	361	10	232	20
03 09 03.625	+10 29 16.34	Q	0.863	[21]	18.4	459	24	544	12	651	8	835	9	1018	10	1500	30
03 21 53.104	+12 21 13.95	Q	2.670	[21]	19.4	2000	36	1844	19	1748	4	1442	10	1471	10	1487	40
03 45 01.317	+12 18 48.77	Q	0.901	[23]s	19.5	430	15	296	10	205	3	158	3	149	3	147	14
03 55 45.553	+12 31 46.14	Q	1.616	[21]	18.2	960	21	595	10	527	5	443	5	398	5	333	17
04 09 22.009	+12 17 39.85	L	1.020	[21]	19.2	1003	26	1076	15	938	9	814	7	775	7	723	20
04 44 12.467	+10 42 47.29	Q	2.400	[25]	19.0	820	32	596	18	500	13	377	5	340	8	250	30
04 48 50.413	+11 27 54.39	Q	1.375	[25]	19.4	92	15	178	15	230	11	–	–	280	11	260	20
04 49 07.672	+11 21 28.63	Q	1.207*	[21]	19.9	689	24	1279	15	1732	21	2573	29	3235	30	4377	115
05 09 27.457	+10 11 44.59	L		[25]	19.2	663	20	552	12	464	5	467	10	490	8	589	21
05 16 46.646	+10 57 54.77	Q	1.580	[25]	19.1	1330	20	940	12	818	5	655	6	583	6	470	20
07 45 33.060	+10 11 12.69	EF				2503	35	3930	30	3726	23	2956	22	2502	20	1831	40
07 49 27.385	+10 57 33.12	G	0.214	[25]	19.1	219	43	180	15	182	4	169	9	165	11	126	15
07 50 52.047	+12 31 04.83	Q	0.889	[21]	17.7	1401	34	1385	39	1516	31	1875	32	2370	21	3252	70
07 57 06.640	+09 56 34.80	L	0.280	[21]	14.7	1089	41	1121	12	1130	9	1320	11	1433	17	1536	56
07 58 07.658	+11 36 46.05	G	0.573	[25]	16.0	557	30	393	20	346	5	298	7	294	7	250	20
08 27 06.513	+10 52 24.15	Q	2.295	[21]	17.8	185	20	144	12	141	5	138	5	127	7	150	25

Table 3. (Contd.)

Radio coordinates (J2000.0)		Id	z	Ref.	B	Flux density and its error, mJy											
R.A.	DEC.					0.97 GHz		2.3 GHz		3.9 GHz		7.7 GHz		11.1 GHz		21.7 GHz	
$08^h32^m38.478^s$	$+10°40'19.68''$	EF				193	24	254	15	210	8	191	5	183	10	230	20
08 33 14.368	+11 23 36.25	Q	2.979	[21]	18.3	724	30	351	7	351	3	290	3	274	3	253	10
09 45 49.860	+12 05 31.32	+			19.3	553	28	358	17	308	4	246	6	250	8	262	24
09 46 35.069	+10 17 06.13	Q	1.007	[24]	18.7	358	20	352	10	338	4	280	3	250	3	181	16
09 47 45.857	+11 13 53.99	Q	1.760	[21]	18.1	278	25	230	15	215	5	187	6	168	10	134	15
10 01 57.735	+10 15 49.70	Q	1.530	[21]	17.7	352	20	321	11	299	3	267	3	266	4	226	11
10 02 52.846	+12 16 14.59	+			19.6	207	20	250	11	302	6	360	8	401	13	405	20
10 15 44.024	+12 27 07.07	L		[24]	19.2	185	31	232	11	277	3	367	4	446	5	532	27
10 42 44.530	+12 03 31.73	Q	1.028	[21]	17.5	4191	33	2435	20	1750	12	1230	10	1049	9	766	11
11 03 03.530	+11 58 16.61	Q	0.917	[24]	18.9	305	20	293	7	320	4	345	6	388	5	474	20
11 18 57.302	+12 34 41.72	Q	2.118	[21]	17.9	2410	25	1836	17	1690	15	1534	15	1591	14	1543	38
11 32 59.491	+10 23 42.63	Q	0.540	[21]	16.9	1154	25	617	10	443	4	364	4	375	4	364	19
12 07 12.625	+12 11 45.88	Q	0.896	[24]	19.0	112	20	224	10	260	9	265	9	281	5	318	18
12 18 26.094	+11 05 05.27	Q	1.403	[21]	19.0	247	33	222	8	217	2	283	3	352	3	335	18
12 54 38.256	+11 41 05.89	Q	0.870	[21]	16.1	730	24	706	12	682	10	746	6	888	7	924	30
13 09 33.933	+11 54 24.56	L		[21]	18.5	850	23	990	12	1062	6	1170	8	1238	10	1248	30
13 15 01.853	+12 20 52.63	G	0.261	[24]	18.7	298	20	245	18	219	8	205	10	210	7	215	20
13 27 54.465	+12 23 11.16	Q	0.950	[25]	19.5	560	30	420	15	411	9	405	10	409	8	415	20
14 30 09.739	+10 43 26.86	Q	1.710	[21]	17.8	220	30	624	09	889	5	916	7	911	8	778	29
14 44 50.736	+11 31 56.40	Q	0.851	[21]	17.9	367	20	211	10	160	4	120	7	110	7	98	10
14 53 44.241	+10 25 57.57	Q	1.770	[25]	20.5	383	15	361	10	295	4	195	4	161	2	102	14
14 55 55.418	+11 51 45.86	EF				450	30	327	10	222	6	172	8	146	6	115	18
15 04 24.980	+10 29 39.20	Q	1.833	[21]	18.8	1687	29	1637	14	1745	10	2136	16	2388	18	2540	76
15 07 21.882	+10 18 44.99	G			13.5	340	30	295	15	252	4	176	6	144	7	98	15
15 25 02.936	+11 07 44.09	Q	0.331	[21]	17.4	450	15	361	07	223	2	284	3	283	3	282	18
15 50 43.595	+11 20 47.45	Q	0.436	[21]	16.3	1042	20	544	13	390	4	279	5	252	8	220	20
15 55 43.044	+11 11 24.38	L	0.360	[21]	14.5	314	15	300	08	274	3	328	3	355	4	419	17
16 03 41.930	+11 05 48.68				18.3	185	25	194	12	215	7	237	4	254	6	257	30
16 08 46.203	+10 29 07.78	Q	1.226	[21]	17.8	1671	34	2217	20	2711	17	3286	27	3714	37	4258	133
16 27 37.032	+12 16 07.11	Q	1.216	[25]	18.2	348	30	294	13	275	7	259	11	252	7	220	20
16 40 58.892	+11 44 04.23	G	0.078	[22]	15.6	481	36	315	16	276	9	231	6	219	14	230	36

Table 3. (Contd.)

Radio coordinates (J2000.0)		Id	z	Ref.	B	Flux density and its error, mJy											
R.A.	DEC.					0.97 GHz		2.3 GHz		3.9 GHz		7.7 GHz		11.1 GHz		21.7 GHz	
$16^h45^m54.675^s$	+11°13′52.64″	EF				583	20	365	10	301	4	246	6	239	8	250	30
17 06 20.498	+12 08 59.81	+			19.0	115	17	159	07	187	2	199	3	197	3	173	11
17 22 44.582	+10 13 35.77	Q	0.732	[25]	21.6	380	30	378	18	374	7	370	6	369	12	351	34
17 28 07.051	+12 15 39.48	Q	0.588	[25]	21.7	256	20	405	08	524	6	611	8	635	10	680	29
17 46 56.965	+11 27 17.35	EF				555	20	345	10	290	4	260	3	249	4	195	10
20 31 54.995	+12 19 41.34	Q	1.215	[21]	18.2	875	17	876	13	902	6	850	8	914	8	1034	20
20 35 22.334	+10 56 06.78	Q	0.601	[21]	16.6	1156	25	780	20	785	11	842	10	942	10	1040	47
20 49 45.865	+10 03 14.40	EF				185	20	434	08	500	3	480	4	492	4	511	22
21 23 13.359	+10 07 54.96	Q	0.932	[21]	18.1	930	30	646	10	513	3	403	4	368	3	336	18
21 45 18.776	+11 15 27.30	Q	0.550	[21]	18.0	364	18	368	07	413	4	460	4	499	5	571	16
21 57 12.862	+10 14 24.80	Q	0.761	[21]	18.4	215	20	272	12	267	3	269	4	318	2	428	16
22 00 07.933	+10 30 07.90	+			21.7	300	20	269	12	296	8	319	5	344	10	358	26
22 03 30.953	+10 07 42.58	EF						220	13	325	3	276	3	215	3	102	10
22 22 52.991	+12 13 49.82	+			20.9	290	15	260	15	252	2	230	3	228	3	200	10
22 32 36.409	+11 43 50.89	Q	1.037	[21]	17.1	7822	47	6297	26	5316	28	4266	36	4264	40	4686	70
22 33 58.450	+10 08 52.10	Q	1.854	[23]	17.6	395	25	320	24	360	13	350	10	315	18	275	20
23 00 18.317	+10 37 54.08	Q	2.816	[23]	19.3	79	10	130	10	168	4	159	9	161	7	148	15
23 10 28.517	+10 55 30.68	G	0.494	[25]	18.9	303	18	274	8	447	5	357	4	395	4	427	22
23 30 09.952	+12 28 28.60	G	0.144	[23]	17.5	510	26	343	19	269	6	206	5	190	8	153	20
23 30 40.853	+11 00 18.71	Q	1.489	[21]	17.8	1180	20	1255	20	1183	10	1053	10	1026	10	923	14
23 47 36.406	+11 35 17.89	+			18.3	468	40	267	20	220	5	183	5	186	6	153	20
23 50 02.031	+11 06 36.71	+			19.7	275	15	256	6	257	2	269	3	279	3	280	17

The optical spectrum of the object identified with the source 0449+1121 was obtained on the 6-m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences. The spectrum shows only a continuum, with no lines, and we have classified the object as a BL Lac object [25]. The references to the redshift presented in the 2001 catalog of quasars and active galaxies ($z = 1.207$) do not include observational data for this object.

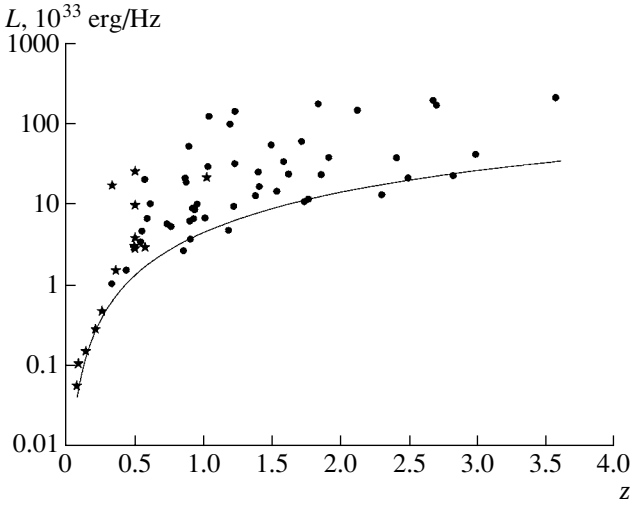

Fig. 4. Redshift dependence of the absolute spectral luminosity at 11.1 GHz for flat-spectrum sources from the sample. The stars denote BL Lac objects and galaxies, while the circles denote quasars. See text for an explanation of the solid curve.

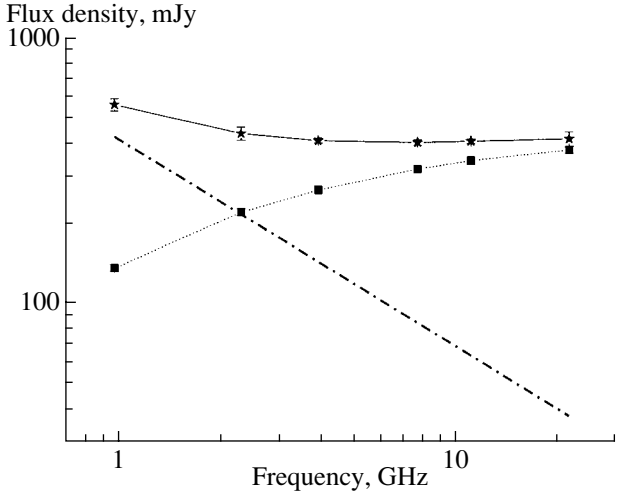

Fig. 5. Decomposition of the spectrum of 1327 + 1223 in June 2001. The solid curve shows the observed spectrum, the dotted curve the spectrum of the compact component, and the dot–dash curve the spectrum of the extended component.

described by parabolic fits. The spectra of variable sources at various stages of development of the variability are studied in [32] based on data obtained in 1965–1973 at 6.6 and 10.7 GHz [33], in 1965–1984 at 4.8, 8.0, and 14.5 GHz [34], and from 1978 to the present at millimeter wavelengths [35]. These spectra are also approximated well by logarithmic parabolas.

In our subsequent analysis, we assume that the spectra of the radio sources are comprised of two components: a power-law component of the form

$$\log S = S_0 + \alpha \log \nu$$

and a compact component that can be represented by a quadratic function of the form

$$\log S = C + 0.5(\log \nu - B)^2/A,$$

where C is the logarithm of the peak flux density, B is the logarithm of the peak frequency, and A is the logarithmic interval from the peak frequency to the frequency at which $d\log S/d\log \nu = 1$ (the curvature parameter of the spectrum) [31].

We decomposed the spectra into components by finding the solution with the minimum residual

$$\sum [S_{\nu i} - (S^s_{\nu i} + S^c_{\nu i})]^2,$$

where $S_{\nu i}$ is the measured flux density at the given frequency, $S^s_{\nu i}$ the flux density of the power-law component, and $S^c_{\nu i}$ the flux density of the compact component. We went through this procedure for nearly all the sources using the data at 0.365 GHz [26].

We believe that the spectral characteristics of the extended components of flat-spectrum and steep-spectrum sources are similar. Accordingly, we considered the decomposition to be successful if the spectral index for the extended component lay in the range $\alpha = -0.5 \ldots -1.1$. The spectra of the extended components were assumed to remain constant over all observing epochs.

We present the characteristics of the spectra at epochs June–November 2001 below. We considered variations of these characteristics within only a narrow time interval from this epoch (one year).

The spectra of the sample sources can be divided into four groups.

(1) The spectra of 28 sources can be described with two components: an extended component with a power-law spectrum and a compact component that is approximated well by a logarithmic parabola with its maximum at a frequency no higher than 25 GHz. The spectrum of the extended component can be determined more accurately. The mean spectral index for the extended components is $\bar{\alpha} = -0.79$, which is close to the mean spectral index obtained for the steep-spectrum sources in the sample, confirming the correctness of the procedure used for the spectral decomposition.

Figure 5 shows the spectral decomposition for 1327 + 1223 as an example, while Fig. 6 show the decompositions for 2035 + 1056 at epochs (a) June 2002 and (b) June 2001.

For 23 objects, the extended component comprises from 18% (1706 + 1208) to 100% (1042 + 1203) of the total flux density at 0.97 GHz. The extended

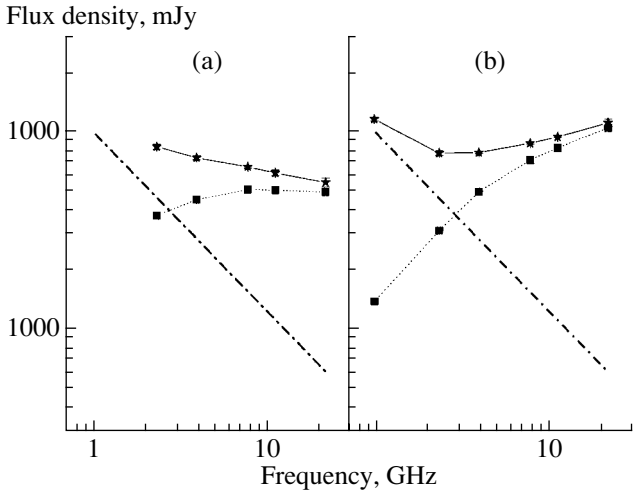

Fig. 6. Same as Fig. 5 for 2035 + 1056 at epochs (a) June 2002 and (b) June 2001.

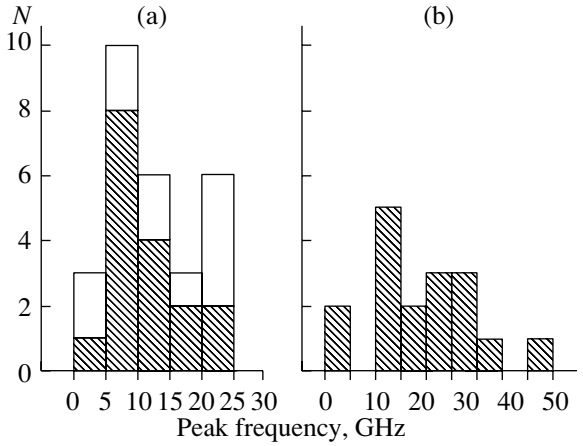

Fig. 7. Distribution of the peak frequencies of the compact components (a) in the rest frame of the observer and (b) in the rest frame of the source. The shaded region in the left diagram corresponds to sources with measured redshifts.

components of five of the objects, such as 0121 + 1127 and 2203 + 1007, are small.

The mean redshift of the quasars of this group is $\bar{z} = 1.58$, with an rms deviation of $\sigma = 0.61$. Figure 7 shows the distribution of the peak frequencies of the compact components in the rest frames of the observer and source. The observed quasar peak frequencies show a dependence on redshift: $\bar{\nu}_{\max} = 8.5$ GHz for quasars with redshifts $z > 1.3$ and $\bar{\nu}_{\max}$=14 GHz for those with $z < 1.3$. In the rest frame of the source, this dependence disappears, and the mean peak frequencies for these redshifts become 25 and 26 GHz, respectively. We did not find any statistically significant correlations between the approximation coefficients for the compact components in the source rest frame and the dependences of these coefficients on the absolute spectral radio luminosity of the source. This same result was obtained for the compact components of sources with declinations 3°30′–6° [8].

The flux densities of most of the sources in this group vary slowly, and their variability indices $V = (S_{\max} - S_{\min})/(S_{\max} + S_{\min})$ do not exceed 0.05. An exception is 1722 + 1013, whose flux density displays appreciable variability, with the variability index at 21.7 GHz being $V = 0.32$ over six months. In November 2001, the observed spectrum of this source rose with frequency, and the spectrum maximum was located beyond our observing interval. The time variations in the spectra of most sources are characterized by a shift of the maximum toward lower frequencies during the evolution of an isolated flare.

(2) The spectra of 22 sources can be decomposed into a power-law component and a parabolic component whose peak frequency is appreciably higher than our highest frequency. The mean spectral index and rms deviation for the extended components of this group do not differ from those for the previous group. The mean redshift for the quasars and BL Lac objects of this group is $\bar{z} = 0.99$, with an rms deviation of $\sigma = 0.57$. The accuracy of the approximation is not sufficiently high to permit quantitative estimates. This group includes more active sources, a large fraction of which have their spectral maxima at frequencies above 30 GHz. The time variations of the spectra of such sources are characterized by variations in the spectral index corresponding to the section of the parabola that grows with frequency. For example, the spectral index of the compact component of 1015 + 1227 between 7.7 and 11.1 GHz varied from +0.38 to +0.55 over three years. The peak frequencies of the compact components of the three sources of this group with the highest redshifts [1118 + 1234 ($z = 2.118$), 1504 + 1029 ($z = 1.883$), and 1608 + 1029 ($z = 1.226$)] are at frequencies lower than 25 GHz at other observing epochs. The spectrum of the compact component of the galaxy 0010 + 1058 in September 2000 can also be approximated by a parabola with its maximum within our frequency range (10 GHz). Seven sources in this group have prominent extended components in our frequency range, and we can detect the compact component only at our highest frequencies.

There is no fundamental difference between the first and second group of sources, and the peak frequencies are distributed fairly uniformly when observations over a broader range of frequencies are considered. The higher activity of sources in the second group is associated with their higher peak frequencies [30].

(3) The spectra of 17 sources cannot be decomposed using two-component models. We believe that their spectra include contributions from several compact components. The measured points in the spectra are insufficient to enable an unambiguous decomposition into components. Only for two sources whose extended components were weak and for which there were observations at 0.365 GHz were we able to separate their spectra into two components, each of which has a parabolic spectrum. For example, the spectrum of $1453+1025$ is the superposition of two parabolas with their maxima at 1.3 and 20 GHz. Most sources in this group remained complex at all observing epochs. The mean redshift of the quasars is $\bar{z}=1.76$, with rms deviation $\sigma=0.87$. Many sources display appreciable flux-density variability at high frequencies, $V_{21.7}>0.13$ over a year. Several sources also display appreciable variability at low frequencies, comparable in amplitude to the variability at high frequencies. For example, the variability index of $0409+1217$ at 2.3 GHz is $V=0.33$, while the variability index at 21.7 GHz is $V=0.29$. More than half of the BL Lac objects belong to this group.

(4) The 0.365–21.7 GHz spectra of 11 sources either can be approximated by a power law with an index from $\alpha=-0.04$ $(0037+1109)$ to $\alpha=-0.5$ $(1455+1151,\ 1507+1018)$ or flatten at frequencies above 3.9 GHz. This flattening is probably due to a compact component that is present at higher frequencies, although it is not possible to separate out the spectral components of these sources using the model considered. Sixty per cent of the identified sources in this group are galaxies, and sources that were observed more than once do not display significant variability.

We were not able to interpret the spectra of $0833+1123$ and $2310+1055$. There are no flux-density measurements for the former source at frequencies below 0.97 GHz, while our observations show a flux density at 0.97 GHz of 724 mJy. In the case of the latter source, a nearby source falls into the antenna beam at several frequencies.

5. CONCLUSION

We have derived flux densities in the range 0.97–11.1 GHz for all the sample sources with steep spectra and up to 21.7 GHz for one third of the spectra. The spectra of 68% of the sources can be approximated with power laws over this entire frequency range, and the spectra of 28% of the sources display self-absorption at low frequencies. The spectra of three sources flatten toward high frequencies, probably due to the presence of compact components whose peak frequencies are above our studied frequency range.

We have also derived 0.97–21.7 GHz flux densities for all the sample sources with flat spectra, and divided their spectra into extended and compact components. The extended components can be approximated by logarithmic power laws with spectral indices $\alpha=-0.5\ldots-1.1$, while the compact components were described using logarithmic parabolas. We were able to obtain a satisfactory spectral decomposition for 50 sources, consistent with available flux-density measurements at low frequencies. The contribution of the extended component at 0.97 GHz can range from 0 to 100% for various sources. The spectra of 17 sources could not be decomposed using such two-component models, probably because they include the contributions of several compact components.

We found no statistically significant correlations between the spectral parameters and the absolute spectral radio luminosities of compact components whose peaks are below 25 GHz in the source rest frame.

All the sources have been optically identified. A significant number of the optical objects associated with the sample sources were classified earlier; work on obtaining optical spectra of the remaining optical objects is ongoing.

Fifty-nine per cent of the steep-spectrum sources are identified to 21^m; among these, 52% are galaxies with a mean redshift of $\bar{z}=0.15$, 34% are quasars with $\bar{z}=1.22$, and the spectra of six objects have not yet been obtained.

Eighty-six per cent of the flat-spectrum sources have optical identifications; among these, 68% are quasars with a mean redshift of $\bar{z}=1.40$, 10% are BL Lac objects with $\bar{z}=0.55$, 11% are galaxies with $\bar{z}=0.26$, and the remaining objects have not yet been classified.

6. ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-16331), a grant of the "Universities of Russia" program (project no. УР.02.03.005), and a grant from the State Science and Technology Program "Astronomy."

REFERENCES

1. A. G. Gorshkov and V. K. Konnikova, Astron. Zh. **72**, 291 (1995) [Astron. Rep. **39**, 257 (1995)].
2. P. C. Gregory, W. K. Scott, K. Douglas, and J. J. Condon, Astrophys. J., Suppl. Ser. **103**, 427 (1996).
3. D. S. Heeschen, Astron. J. **89**, 1111 (1984).
4. A. G. Gorshkov, V. K. Konnikova, and M. G. Mingaliev, Astrophys. Space Sci. **278**, 93 (2001).
5. A. G. Gorshkov, V. K. Konnikova, M. G. Mingaliev, *et al.*, Astron. Zh. (in preparation).

6. A. G. Gorshkov, V. K. Konnikova, and M. G. Mingaliev, Astron. Zh. (in preparation).
7. A. M. Botashev, A. G. Gorshkov, V. K. Konnikova, and M. G. Mingaliev, Astron. Zh. **76**, 723 (1999) [Astron. Rep. **43**, 631 (1999)].
8. A. G. Gorshkov, V. K. Konnikova, and M. G. Mingaliev, Astron. Zh. **77**, 407 (2000) [Astron. Rep. **44**, 353 (2000)].
9. A. G. Gorshkov and V. K. Konnikova, Astron. Zh. **74**, 374 (1997) [Astron. Rep. **41**, 325 (1997)].
10. A. G. Gorshkov, V. K. Konnikova, and M. G. Mingaliev, Astron. Zh. **77**, 188 (2000) [Astron. Rep. **44**, 161 (2000)].
11. A. B. Berlin, A. A. Maksyasheva, N. A. Nizhel'skiĭi, *et al.*, in *Abstracts of the XXVII Radio Astronomy Conference, St. Petersburg* (1997), Vol. 3, p. 115.
12. N. S. Soboleva, A. V. Temirova, and T. V. Pyatununa, Preprint Spets. Astrofiz. Obs. (1986).
13. A. G. Gorshkov and O. I. Khromov, Astrofiz. Issled. (Izv. SAO) **14**, 15 (1981).
14. A. G. Gorshkov and V. K. Konnikova, Astron. Zh. **73**, 351 (1996) [Astron. Rep. **40**, 314 (1996)].
15. H. Tabara and M. Inoue, Astron. Astrophys., Suppl. Ser. **39**, 379 (1980).
16. A. D. Kuz'min and A. E. Solomonovich, *Radio Astronomical Methods for the Measurement of Antenna Parameters* [in Russian] (Sov. Radio, Moscow, 1964).
17. I. W. A. Browne, Mon. Not. R. Astron. Soc. **293**, 257 (1998).
18. J. J. Condon, W. D. Cotton, E. W. Greisen, *et al.*, Astron. J. **115**, 1693 (1998).
19. D. Monet, A. Bird, B. Canzian, *et al.*, *USNO-SA1.0* (U.S. Naval Obs., Washington, 1996).
20. R. L. Pennington, R. M. Humphreys, S. C. Odewahn, *et al.*, Publ. Astron. Soc. Pac. **105**, 103 (1993).
21. M. P. Veron-Cetty and P. Veron, Astron. Astrophys. **374**, 92 (2001).
22. NASA/IPAC Extragalactic Database. http://nedwww.ipac.caltech.edu.
23. V. Chavushyan, R. Mujica, J. R. Valdes, *et al.*, Astron. Zh. **79**, 771 (2002) [Astron. Rep. **46**, 697 (2002)].
24. V. L. Afanas'ev, S. N. Dodonov, A. V. Moiseev, *et al.*, Astron. Zh. **47**, 6 (2003) [Astron. Rep. **47**, 458 (2003)].
25. V. L. Afanas'ev, S. N. Dodonov, A. V. Moiseev, *et al.*, Astron. Zh. (in preparation).
26. J. N. Douglas, Bull. Am. Astron. Soc. **19**, 1048 (1987).
27. R. D. Blandford and A. Königl, Astrophys. J. **232**, 34 (1979).
28. A. H. Marscher and W. K. Gear, Astrophys. J. **298**, 114 (1985).
29. P. A. Hughes, H. D. Aller, and M. F. Aller, Astrophys. J. **374**, 57 (1991).
30. E. Valtaoya, H. Terasranta, S. Urpo, *et al.*, Astron. Astrophys. **254**, 71 (1992).
31. R. Landau, B. Golisch, T. J. Jones, *et al.*, Astrophys. J. **308**, 78 (1986).
32. A. G. Gorshkov, in *Abstracts of the XXVII Radio Astronomy Conference, St. Petersburg* (1997), Vol. 1, p. 176.
33. B. H. Andrew, J. M. Macleod, G. A. Harvey, and W. J. Medd, Astron. J. **83**, 863 (1978).
34. H. D. Aller, M. F. Aller, G. Latimer, and P. E. Hodge, Astrophys. J., Suppl. Ser. **59**, 513 (1985).
35. H. Terasranta, M. Tornikoski, E. Valtaoja, *et al.*, Astron. Astrophys., Suppl. Ser. **94**, 121 (1992).

Translated by D. Gabuzda

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 916–924. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 992–1001.

Evidence for the Prolonged Existence of >100 MeV Solar Protons in Coronal Structures

A. B. Struminskiĭ

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

Received January 23, 2003; in final form, March 27, 2003

Abstract—Solar proton events (SPEs) of 1989–1992 accompanied by ground-level enhancements of the cosmic-ray intensity are analyzed. A diffusive-propagation model can well approximate the solar proton intensity at energies of 84–200 MeV measured on board the GOES-7 satellite during the first 15–20 hours of the development of an SPE, assuming several episodes of proton injection into interplanetary space over some hours. The detection of π^0-decay gamma-ray emission during the flares of June 11 and 15, 1991 on board the CGRO and GAMMA-1 space observatories is the sole experimental evidence for the existence of protons with energies > 100 MeV near the Sun over such extended times. The number of protons in the source is estimated for various times based on the propagation model. Moreover, this value is in qualitative and quantitative agreement with the number of protons required to generate the observed gamma-ray emission for the June 15, 1991 event (and in qualitative agreement for the June 11, 1991 event). This provides evidence that processes of prolonged trapping and/or acceleration of protons in coronal structures may be possible in other such events. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The characteristics of solar proton fluxes derived from solar gamma-ray observations and those measured at low (~10 MeV), middle (~100 MeV) and high (~1000 MeV) energies in interplanetary space reflect different features of the same phenomenon—solar proton events (SPEs). This phenomenon includes solar cosmic-ray acceleration, storage, and propagation. After 60 years of SPE studies, it may appear that all possible scenarios for their development have been investigated or, at least, suggested [1–4]. However, we still do not have strong evidence in favor of one particular scenario that can operate over the entire energy range.

Based on the duration of the soft X-ray emission of solar flares, two groups of events can be identified—impulsive and gradual. In the current paradigm, different acceleration mechanisms dominate during impulsive (less than one hour) and gradual (more than one hour) X-ray events [2]. Stochastic acceleration operates directly in the flare region during impulsive events, but a shock wave driven by a coronal mass ejection (CME) accelerates particles high into the corona and interplanetary space. Differences between acceleration mechanisms lead to differences in the characteristic acceleration times, maximum energies, and chemical and charge-state abundances of ions in interplanetary space. It is thought that during gradual events, the particles generating the solar gamma-ray emission and those propagating in interplanetary space are different and that processes occurring in the lower and middle corona do not contribute significantly to the fluxes of interplanetary particles.

This scheme has encountered serious difficulties (see [5–6]):

(1) it is doubtful whether shocks can accelerate solar protons to energies of several MeV [3, 6];

(2) some ground-level enhancements (GLEs) of solar cosmic rays are correlated with impulsive flares [7];

(3) cases of prolonged solar-flare gamma-ray emission with pronounced impulsive and gradual phases are known [8];

(4) the characteristics of the gamma-ray emission of impulsive and gradual events are the same [9];

(5) impulsive events are accompanied by CMEs [10];

(6) the chemical and charge-state abundance of ions measured during gradual events varies from the characteristic composition for impulsive events to that for gradual events [11].

These experimental facts support another view of this problem. Processes in the lower and middle corona could be a source of both the interacting and interplanetary particles. In this case, the CME shock is not the main accelerator, but the CME lifts loops and opens closed magnetic-field lines in various regions of the corona, creating favorable conditions for

1063-7729/03/4711-0916$24.00

the release and propagation (or acceleration) of particles during the post-eruptive phase [12–15]. SOHO observations of several events in 1997 suggest that the ability of a CME to supply particles to the interplanetary medium depends on the evolution of the flare below $\sim 2R_\odot$ [16]. Bazilevskaya *et al.* [17] have underlined the special role played by the magnetic-field configuration in active regions in the formation of > 100 MeV solar proton fluxes.

It is possible that one key to the problem lies at middle energies. This part of the spectrum should reflect physical limitations on the acceleration, escape, and propagation of particles [3]. Because the effective energy range of high-latitude neutron monitors (NMs) has nearly no overlap with the high-energy range of satellite detectors, our knowledge of the differential spectrum of solar protons at middle energies is limited. In spite of their more than 44-year history, stratospheric measurements provide data with large gaps [17], and do not enable investigation of the dynamics of SPEs with the desired time resolution.

The purpose of our current work is to show that SPEs of the 22*nd* solar cycle [18] accompanied by GLEs detected by the neutron-monitor network [19] at middle energies (> 100 MeV) are consistent with simple diffusive-propagation models if a prolonged injection of particles from a source on the Sun into interplanetary space is assumed.

The detection of prolonged gamma-ray emission on June 11, 1991 by EGRET [20] and on June 15, 1991 by GAMMA-1 [12] and COMTEL [21] is the sole experimental fact demonstrating that the injection of > 100 MeV protons over hours is possible in principle. As is noted in the review [11], beginning with the flare of June 3, 1982 and now, especially, with the series of the flares observed in June 1991, discussions of particle acceleration and propagation increasingly use the terms "proton trapping" and "second–phase acceleration," used in an earlier stage of SPE studies.

Following the adopted classification of SPEs, we will first consider impulsive events, then mixed events with well defined impulsive and gradual components of the solar gamma-ray and radio emission; finally, we will briefly discuss gradual events.

2. DATA AND METHODS USED

We use a simple diffusion model to approximate the time profiles of the solar proton intensities measured at 84–200 MeV by the GOES-7 satellite. These data are available on the Internet (http//spidr.ngdc.noaa.gov). Table 1 presents the list of events considered. The events of August 12, 1989 and October 30, 1992 had intensities at middle energies comparable to those of many GLEs, and were therefore also included in our list.

The solution to the diffusion equation in the case of three-dimensional isotropic diffusion for instantaneous particle injection from a source is well known [1]. The particle density at some time and at some distance from the source (normalized to the total number of particles in the source) is given by the expression

$$f(r,t) = \frac{1}{4\Gamma(3/2)}(K_0 t)^{-3/2} \exp\left(-\frac{r^2}{4K_0 t}\right). \quad (1)$$

where $K_0 = \lambda V/3$ is the diffusion coefficient, λ is the mean free path, and V is the velocity. In the case of prolonged injection with the time profile $Q(\tau)$, the convolution

$$U(r,t) = \int_0^t Q(t-\tau) f(r,\tau)\, d\tau \quad (2)$$

determines the particle density at time t a distance r from the source.

In the considered proton energy range, there is agreement between the theoretical and experimental values of the mean free path in interplanetary space [22]. The mean free paths for protons with energies of about 100 MeV should be 0.08–0.3 AU; in our calculations, we adopted the value 0.11 AU. As is noted in [1], the physical meaning of λ is far from clear in very simple diffusion models, but the condition $\lambda = \text{const}$ for various events indicates that some properties of the interplanetary magnetic field do not vary from event to event.

An estimate of the isotropic solar cosmic-ray flux in interplanetary space is

$$j(r,t) \approx U(r,t)V/3. \quad (3)$$

The required injection-time profiles $Q(\tau)$ for the considered events were obtained by demanding that the approximating curve be within the errors of the one-minute GOES data for at least the first 15–20 hours after the onset of the X-ray flare. The time step in the numerical calculations was five minutes. A constant injection rate was assumed at each stage, and this rate was changed step-wise between stages so as to give a good fit to the time profiles observed in interplanetary space.

Table 2 presents the results of this procedure: the onset times, durations, and relative intensities of the injection episodes. The derived injection profiles yield only a lower limit for the total proton number in the source, $N_p(E_p > 100\,\text{MeV})$, since not all the accelerated particles may escape into interplanetary space. The value $N_p(E_p > 30\,\text{MeV})$) is usually estimated from data on solar-flare gamma-ray emission, and can be found via extrapolation based on an adopted

Table 1. Some characteristics of parent solar flares for powerful SPEs of the 22*nd* solar cycle

Date	X-ray event			Coordinates	Maximum NM (%) [19]
	Magnitude	Onset	Max.		
July 25, 1989	X2.6, imp	08:38	08:43	N26W85	GLE40, 5.5%
Aug. 12, 1989	X2.6, grad	13:58	14:27	S16W38	
Aug. 16, 1989	>X12.5, grad	00:54	01:18	S15W85	GLE41, 16%
Sept. 29, 1989	X9.8, grad	10:47	11:33	?	GLE42, 372%
Oct. 19, 1989	>X12.5, grad	12:32	12:58	S25E09	GLE43, 53%
Oct. 22, 1989	X2.9, grad	17:38	18:05	S27W32	GLE44, 179%
Oct. 24, 1989	X5.7, grad	17:36	18:31	S29W57	GLE45, 123%
Nov. 15, 1989	X3.2, imp	06:52	06:59	N11W28	GLE46, 12%
May 21, 1990	X5.5, imp	22:10	22:19	N34W37	GLE47, 22%
May 24, 1990	X9.3, imp	20:44	20:51	N36W76	GLE48, 43%
May 26, 1990	X1.4, imp	20:45	20:58	?	GLE49, 8.8%
May 28, 1990	C1.4, imp	04:28	04:33	?	GLE50, 6%
	C9.7, imp	05:15	05:21		
June 11, 1991	X12.5,imp+grad	01:56	02:09	N32W15	GLE51, 8%
June 15, 1991	X12.5,imp+grad	08:10	08:21	N36W70	GLE52, 28%
June 25, 1992	X3.9, imp+grad	19:51	20:14	N09W69	GLE53, 7%
Oct. 30, 1992	X1.7, grad	16:59	18:16	S26W63	
Nov. 2, 1992	X9.0, grad	02:31	03:08	S23W90	GLE54, 3.6%

proton-energy spectrum. For example, for an integrated power-law index of 3 or 3.5, the values in the right column of Table 2 should be multiplied by factors of 37 or 68, respectively.

3. RESULTS AND DISCUSSION

3.1. Impulsive Events

The events considered in this section are typical impulsive events in terms of the duration of the soft X-ray emission [7] and the relative abundance of Fe nuclei observed in interplanetary space [23].

The time profiles of the proton intensity at 84–200 MeV (Fig. 1) for the events of July 25, 1989 and November 15, 1989 are well fit by a model with the impulsive injection of about 50% of the particles, with the subsequent escape of the remaining particles over several hours. If the mean free path is decreased to 0.08 AU, it is no longer necessary to consider prolonged injection with an intensity of < 1 relative unit during these events.

To describe the slow rise of the proton intensity at the Earth's orbit on May 21–22, 1990 (Fig. 2b), we must suppose that there were two episodes of prolonged and relatively weak proton injection into interplanetary space. The event maximum is associated with a third injection of particles, which was shorter but more intense than the first two. On the contrary, during the event of May 24, 1990 (Fig. 2b), the first, nearly impulsive injection was most intense, with about 25% of the accelerated particles escaping into interplanetary space. The remaining particles were injected over an extended period, possibly right up until the next powerful flare on May 26.

Since the fluxes of solar neutrons, protons, and gamma-rays were measured on May 24, 1990, this event has been well studied [3, 24–25]. According to [24], approximately $N_p(E > 30\ \mathrm{MeV}) = 7 \times 10^{32}$ protons were injected into interplanetary space (our estimate is 1×10^{33}), which is less than 20% of the number of protons required to generate the observed gamma-ray emission, 5×10^{33}–1×10^{34}. Debrunner *et al.* [24] concluded that the populations of interacting and interplanetary protons were different, and the protons observed in interplanetary space were accelerated by a shock.

Table 2. Time profiles for proton injection from a solar source into interplanetary space and total number of > 100 MeV protons used to fit the GOES-7 84–200 MeV proton intensity

Date	Time since onset of X-ray flare	τ_i, min	Q_i, rel. units	Fraction, %	N_p (>100 MeV)
July 25, 1989	0	5	60	63	2.6×10^{30}
	5	180	1	37	
Aug. 12, 1989	0	270	0.45	1	4.6×10^{31}
	270	260	4.5	13	
	530	70	20	16	
	600	70	5	4	
	670	170	35	66	
Aug. 16, 1989	0	30	1	0.4	4.4×10^{31}
	30	35	85	33	
	65	170	25	48	
	235	165	1	2	
	400	100	5	5.6	
	500	1000	1	11	
Sept. 29, 1989	0	55	1	0.1	3.6×10^{32}
	55	205	150	43	
	260	170	50	12	
	430	35	420	21	
	465	220	10	3	
	685	5	1000	7	
	690	240	40	14	
Oct. 19, 1989	0	5	250	3	2.5×10^{32}
	5	75	1	0.2	
	80	385	75	60	
	465	270	50	28	
	735	415	5	4.3	
	1150	115	20	4.5	
Oct. 22, 1989	0	5	1200	17	1.8×10^{32}
	5	65	1	0.2	
	70	115	80	26.6	
	185	400	40	46.2	
	585	665	5	10	
Oct. 24, 1989	0	30	2	0.2	1.3×10^{32}
	30	340	55	73.4	
	370	670	10	26.4	

Table 2. (Cont.)

Date	Time since onset of X-ray flare	τ_i, min	Q_i, rel. units	Fraction, %	N_p (>100 MeV)
Nov. 15, 1989	0	5	35	49	1.9×10^{30}
	5	180	1	51	
May 21, 1990	0	240	8.5	64	1.7×10^{31}
	240	85	1	3	
	325	35	30	33	
May 24, 1990	0	5	150	25	1.5×10^{31}
	5	55	1	2	
	60	200	4	27	
	260	540	2	36	
	800	630	0.5	10	
May 26, 1990	0	5	50	13	9.6×10^{30}
	5	75	1	4	
	80	180	8.5	83	
May 28, 1990	0	35	3	6	9.8×10^{30}
	35	1050	1.5	83	
	1085	420	0.5	11	
June 11, 1991	0	10	5	0.4	7×10^{31}
	10	50	30	11	
	60	210	5	8	
	270	430	25	80.6	
June 15, 1991	0	5	50	3	5×10^{31}
	5	10	700	83	
	15	270	5	14	
June 25, 1992	0	5	105	70	3.9×10^{30}
	5	460	0.5	30	
Oct. 30, 1992	0	135	1	3	2.4×10^{31}
	135	165	5	19	
	300	50	15	17	
	350	230	5	26	
	580	170	9	35	
Nov. 2, 1992	0	60	1	0.3	1×10^{32}
	60	40	250	50	
	100	135	1	0.7	
	235	230	42	49	

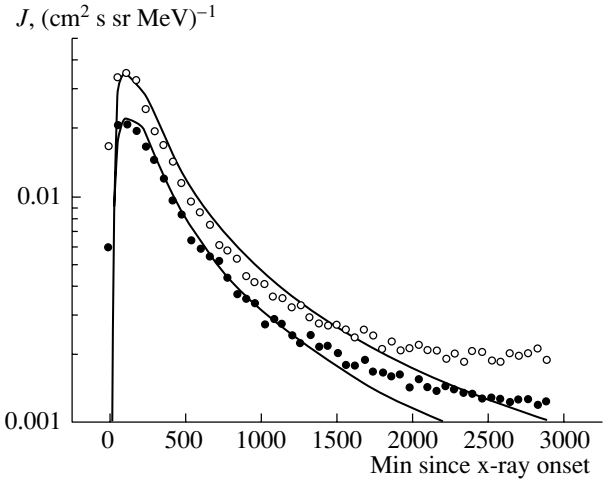

Fig. 1. Hourly intensities of 84–200 MeV protons obtained from the one-minute GOES-7 data starting from the onset of the X-ray flare for the SPEs of July 25 (open circles) and November 15, 1989 (filled circles) and their model approximations (curves).

It seems that, with some minor exceptions, the scenario [25] is applicable to all the GLEs of May 1990. Several loops were filled by accelerated particles after the initial energy release, and the disruption of these loops resulted in several injections of particles into interplanetary space. Kocharov *et al.* [25] attribute the delay of the particle injection by several hours after the impulsive phase to shock acceleration. This suggestion appears reasonable for the events of May 24, 26, and 28, but not for the event of May 21, 1990. It is very unlikely that the efficiency of the shock acceleration could increase by a factor of a few at a considerable distance from the Sun.

3.2. Gradual Events with an Impulsive Phase

The events of June 11 and 15, 1991 are homologous X-ray flares, which have both an impulsive and gradual phase according to the radio and gamma-ray observations. Nearly equal numbers of protons with similar energy spectra were accelerated [21]. The lower panel of Fig. 2 shows time profiles of the proton intensity measured at the Earth's orbit and their model approximations.

Four phases of proton injection on the Sun can be distinguished on June 11, 1991, in consistency with the γ-ray observations [20–21]. The main feature of the June 11, 1991 flare—γ-ray emission from π^0 decays more than eight hours after the impulsive phase—is well pronounced in the time-profile of the > 100 MeV proton intensity. According to the model estimates, only 20% of the accelerated protons left the Sun in the first 4.5 hrs after the onset of the flare, but the injection of protons into interplanetary space

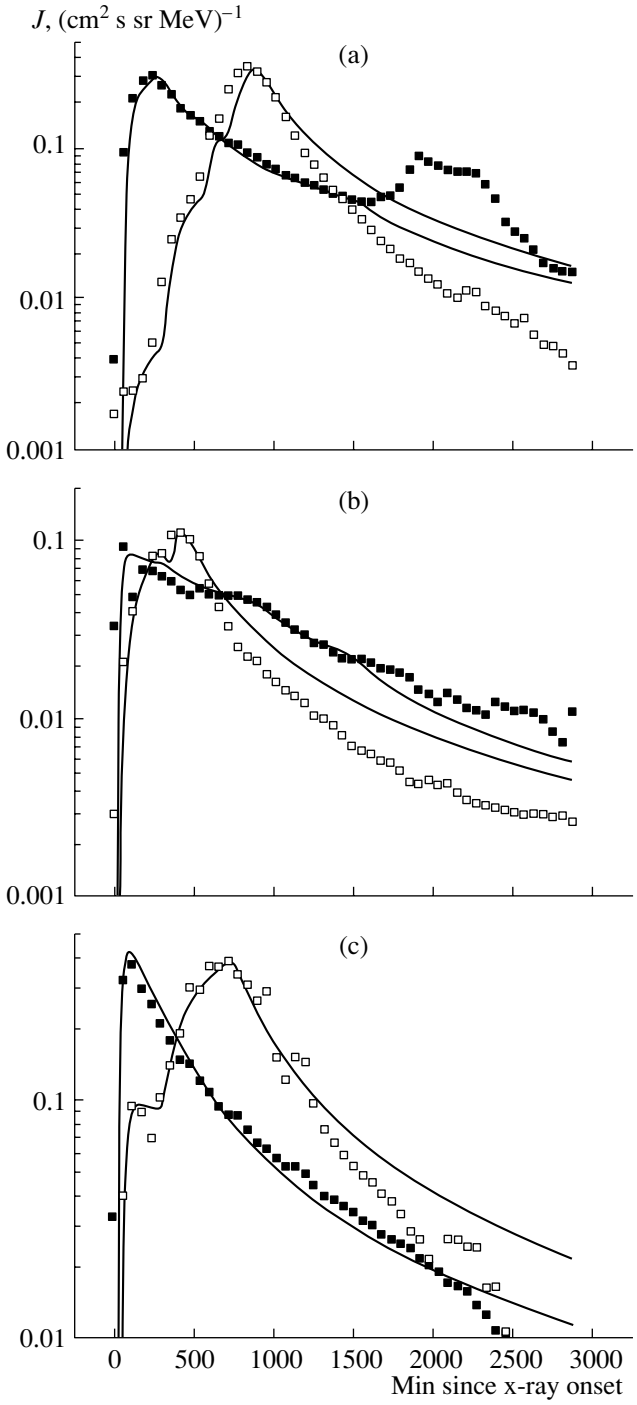

Fig. 2. Same as Fig. 1 for the SPEs of (a) Aug. 12, 1989 (open squares) and Aug. 16, 1989 (filled squares), (b) May 21, 1990 (open squares) and May 24, 1990 (filled squares), and (c) June 11, 1991 (filled squares) and June 15, 1991 (open squares).

stopped simultaneously with the end of pion generation near the Sun. This provides direct evidence that the same population of protons interacted in the Sun

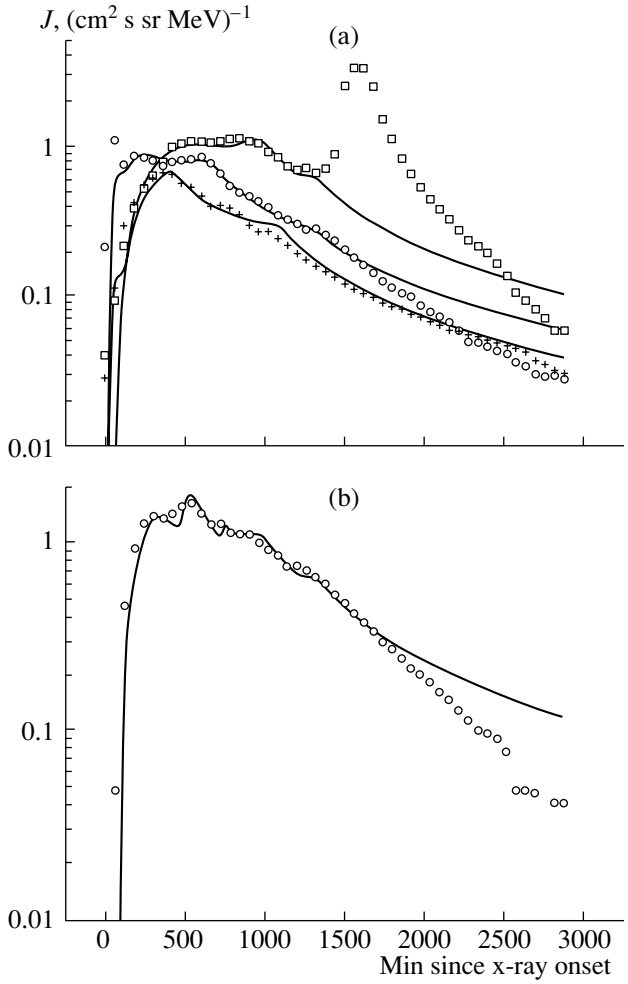

Fig. 3. Same as Fig. 1 for the SPEs of (a) September 29, 1989, (b) Oct. 19, 1989 (open squares), Oct. 22, 1989 (open circles), and Oct. 24 , 1989 (crosses).

and propagated in interplanetary space on June 11, 1991.

In contrast, the June 15, 1991 event is characterized by a rapid release of the accelerated protons into interplanetary space. About 86% of the accelerated protons were injected during the first 15 minutes after the onset of the flare. Unfortunately, no γ-ray observations were performed at that time [12, 21]. According to the estimates of [26], approximately $N_p\,(E > 30\,\mathrm{MeV}) = (2.9 \pm 0.4) \times 10^{32}$ protons must remain close to the Sun in order to produce the γ-rays observed in the period 08:26–09:25 UT. Let us compare this value with results of our calculations: in all, about $N_p\ (E > 100\ \mathrm{MeV}) = 5.0 \times 10^{31}$ protons were accelerated, but only 14% of them remained near the Sun at 08:26 UT; i.e., for an integrated proton spectrum with a power-law index of 3–3.5, we have $N_p\,(E > 30\,\mathrm{MeV}) = 2.6(4.8) \times 10^{32}$. Therefore, the number of protons interacting in the Sun and the number of protons in interplanetary space agree. In addition, the spectrum of the interacting protons obtained in [26] for this time interval shows that most of the high-energy protons have escaped the region of γ-ray emission.

Note that the ratio of ^{3}He/^{4}He fluxes at energies of 50–110 MeV/nucleon measured by the CRRES satellite for the June 1991 events was one to three orders of magnitude higher than in the solar corona, in contradiction to the shock-acceleration model [27]. This ratio does not depend on the scale of the event, and is determined by the structure of the corresponding active region.

3.3. Gradual Events

The total number of particles in the source is one to two orders of magnitude higher for the gradual events presented in Table 2. The time profiles of the solar proton intensity observed in interplanetary space clearly show effects associated with prolonged and multiple injection. We extended the two-source model for the GLE of September 29, 1989 using neutron-monitor network data [4]. Arguments in favor of solar cosmic-ray acceleration by shocks at distances of up to $5–10R_\odot$ are presented in [28]. On the other hand, analysis of the optical and radio observations [13] shows that post-flare activity in the lower corona continued at that time, possibly associated with additional acceleration/trapping of particles near the Sun. The ion abundance observed in interplanetary space was enriched in Fe in all the gradual events considered [23].

The upper panel in Fig. 2 presents time profiles of the proton intensity for the events of August 12 and 16, 1989 (the second maximum observed on August 17, 1989 is associated with a new solar flare [18]). The injection characteristics of this pair are similar in many respects to those of the pairs of May 21 and 24, 1990, June 11 and 15, 1991, and October 30 and November 2, 1992. Thus, Figure 2 shows that both relatively rapid and slow injections of > 100 MeV protons are possible, independent of the type of X-ray event and the position of the flare on the solar disk.

The main feature of the gamma-ray emission of the September 29, 1989 flare, which occurred behind the solar limb, is that the 2.223 MeV line indicates the presence of an extended source on the visible side of the Sun [4, 12]. Giant coronal loops filled by accelerated protons can provide the conditions needed for such a source. To describe the proton intensity observed at the Earth's orbit on September 29–30, 1989 (lower panel, Fig. 3), we must suppose that a series of such loops were disrupted over more than ten hours.

The upper panel in Fig. 3 shows time profiles of the proton intensity observed in a series of large SPEs in October 1989. These SPEs were previously modeled assuming shock acceleration of protons [29]. However, the $^3He/^4He$ ion ratio in the source of gamma-ray emission on October 24, 1989 was strongly enhanced, as is characteristic of impulsive flare acceleration [12]. As is noted above, the diffusive-propagation model provides only a lower limit for the number of particles in the source, since some fraction of the particles may not be released into interplanetary space and another fraction may propagate via some non-diffusive mechanism. The event of October 19, 1989 provides an answer to the question of where and how these particles might propagate. It is argued in [30–31] that the second maximum observed on October 20, 1989 was due to solar protons entering a trap in the solar wind.

4. CONCLUSION

Assuming a prolonged injection of particles from the Sun into interplanetary space, a simple diffusion model for the propagation of solar cosmic rays fits well the time profiles of the > 100 MeV proton intensity observed at the Earth's orbit in the first 15–20 hours after the onset of the X-ray flare. Independent of the type of parent X-ray event and its heliolongitude from ~E10 to ~W110, both rapid (impulsive) and prolonged (gradual) injection can be realized. The discrepancy by a factor of 1.5–2 between the modeled and observed intensities at the beginning of some events is possibly associated with strong anisotropy of the solar cosmic-ray flux, which is not properly accounted for when calculating the proton intensity from the GOES detector count rate. The global transformation of the inner heliosphere resulting from the propagation of CMEs and associated solar-wind structures can explain the discrepancies observed during the late phases of some events (about one day after their onset).

Extension of the model to lower proton energies (~10 MeV) is complicated by the large uncertainty of the mean free path (a factor of 5–10 discrepancy between the theoretical and experimental values [22]) and the need to include the effects of convection and adiabatic deceleration [32]. A proper treatment of these effects should make it possible to obtain injection time profiles for low-energy protons.

Extension of the model to solar proton energies of > 500 MeV, where only neutron-monitor network data are available, requires better knowledge of the neutron-monitor sensitivities at these energies, the geomagnetic cutoffs, and the acceptance cones of the observation points. In addition, correct interpretation of high- and middle-latitude neutron-monitor data requires that solar-proton spectra measured onboard satellites and those estimated from neutron-monitor data be properly normalized relative to each other [33].

The injection time profiles considered here were individually selected, and so can be considered only qualitative evidence for prolonged proton injection from the source into interplanetary space. Correctness of the choice of injection function can be demonstrated *a priori* by agreement between the calculated and observed anisotropies of the solar particles. Unfortunately, no experimental data on the anisotropy of 100 MeV protons are available, and it is not appropriate to carry out comparisons with experimental data for lower and higher energies, since the duration of the injection might be appreciably different for these energies.

Currently, it is usual to associate the prolonged injection [28] and independence of the SPE characteristics on heliolongitude [17] with shock acceleration in the upper corona and interplanetary medium. From this point of view, the obtained injection functions can be considered imaginary sources on the Sun corresponding to acceleration at a shock front. However, the realization of several acceleration episodes with intensities differing by a factor of a few by the action of a single shock seems very unlikely. In addition, we have found that the mean free path does not depend on the intensity of the proton flux propagating in the interplanetary medium. Apparently, plasma waves generated by solar protons are not important, at least for energies of > 100 MeV, which is inconsistent with theoretical models of shock acceleration in the interplanetary medium.

Therefore the prolonged trapping and/or acceleration of protons in the vicinity to the Sun during the decay phases of powerful solar flares is typical, and we must consider the possible existence of radiation belts of activity complexes. The term "local solar radiation belts" was first introduced in [34]; this refers to quasi-stationary coronal traps of solar cosmic rays formed by the loop structures of an active-region magnetic field.

The detection of prolonged gamma-ray emission on June 11, 1991 by EGRET [20] and on June 15, 1991 by GAMMA-1 [12] and COMTEL [21] is the sole experimental fact demonstrating that the injection of > 100 MeV protons over some hours is possible in principle. Only the high sensitivity of these instruments, designed for purely astrophysical tasks, made it possible to perform such measurements. As is shown above, the numbers of protons interacting in the Sun and propagating in interplanetary space are consistent with each other. According to our estimates, similar numbers of protons must be trapped close to the Sun over several hours in other events (for example, those of August 16, 1989 and November 2,

1992), making the generation of similar gamma-ray emission likely. However, this cannot be tested, since no gamma-ray observations were performed at that time.

The question of whether prolonged and varied proton injection is associated with prolonged and multiple acceleration or with simple trapping remains open [8–9, 12, 15, 21, 24, 26], and we have not considered it here. One possible active-region magnetic-field topology for impulsive and gradual events was proposed in [35]. The minor role of flares in generating interplanetary particle fluxes was underlined by Reames [35], who emphasized that most of the flare particles would be trapped in the Sun. This last statement appears to be incorrect. In impulsive events, the release of particles with less than some particular rigidity is depressed, and most of them die in the Sun (the event of May 24, 1990 is an example). In gradual events due to the effects of CMEs, virtually all the particles are released from the Sun in some way (the event of October 19, 1989 is an example).

Studies of events of the current (23*rd*) solar cycle and searches for possible counterparts of prolonged injections in the X-ray and radio ranges are goals for future investigations.

ACKNOWLEDGMENTS

This work has been partially supported by the Russian Foundation for Basic Research (project nos. 01-02-17580, 02-02-16992, 02-02-17086) and the State Program in Science and Technology "Astronomy." The author thanks A.V. Belov, R.T. Guschina, M.A. Lifshits and L.I. Miroshnichenko for their help and interest in this work.

REFERENCES

1. L. I. Dorman, and L. I. Miroshnichenko, *Solar Cosmic Rays* [in Russian] (Nauka, Moscow, 1968).
2. D. V. Reames, Space Sci. Rev. **90**, 413 (1999).
3. J. M. Ryan, J. A. Lockwood, and H. Debrunner, Space Sci. Rev. **93**, 35 (2000).
4. L. I. Miroshnichenko, *Solar Cosmic Rays* (Kluwer Acad., Dordrecht, 2001).
5. E. W. Cliver and H. V. Cane, EOS Trans. Am. Geophys. Union **83** (7), 61 (2002).
6. M.-B. Kallenrode, J. Phys. G (2003, in press).
7. S. W. Kahler, M. A. Shea, D. F. Smart, and E. W. Cliver, in *Proceedings of 22nd ICRC* (1991), Vol. 3, p. 21.
8. J. M. Ryan, Space Sci. Rev. **93**, 581 (2000).
9. R. Ramaty and N. Mandzhavidze, in *Proceedings of IAU Symposium No. 195: Highly Energetic Physical Processes and Mechanisms for Emission from Astrophysical Plasmas* (Astronomical Society of the Pacific, San Francisco, 2000), p. 123.
10. S. W. Kahler, D. V. Reames, and N. R. Sheeley, Jr., Astrophys. J. **562**, 588 (2001).
11. A. J. Tylka, D. V. Reames, and C. K. Ng, Geophys. Res. Lett. **26**, 2141 (1999).
12. V. V. Akimov, P. Ambroz, A. V. Belov, *et al.*, Solar Phys. **166**, 107 (1996).
13. K.-L. Klein, E. L. Chupp, G. Trottet, *et al.*, Astron. Astrophys. **348**, 271 (1999).
14. T. Laitinen, K.-L. Klein, L. Kocharov, *et al.*, Astron. Astrophys. **360**, 729 (2000).
15. K.-L. Klein, G. Trottet, P. Lantos, and J.-P. Delaboudiniere, Astron. Astrophys. **373**, 1073 (2001).
16. J. Torsti, A. Anttila, L. Kocharov, *et al.*, Geophys. Res. Lett. **25**, 2525 (1998).
17. G. A. Bazilevskaya, M. B. Kraĭinev, V. S. Makhmutov, *et al.*, Geomagn. Aeron. (2002, in press).
18. A. I. Sladkova, G. A. Bazilevskaya, V. N. Ishkov, *et al.*, *Catalogue of Solar Proton Events 1987–1996* [in Russian], Ed. by Yu. I. Logachev (Moscow Univ. Press, Moscow, 1998).
19. M. A. Shea, D. F. Smart, L. C. Gentale, *et al.*, in *Proceedings of 24th ICRC* (1995), Vol. 4, p. 244.
20. G. Kanbach, D. L. Bertsch, C. E. Fichtel, *et al.*, Astron. Astrophys., Suppl. Ser. **97**, 349 (1993).
21. G. Rank, J. Ryan, H. Debrunner, *et al.*, Astron. Astrophys. **378**, 1046 (2001).
22. J. W. Bieber, W. H. Matthaeus, C. W. Smith, *et al.*, Astrophys. J. **420**, 294 (1994).
23. W. Dietrich and C. Lopate, in *Proceedings of 26th ICRC* (1999), Vol. 6, p. 71.
24. H. Debrunner, J. A. Lockwood, C. Barat, *et al.*, Astrophys. J. **479**, 997 (1997).
25. L. G. Kocharov, G. A. Kovaltsov, J. Torsti, *et al.*, in *Proceedings 24th ICRC* (1995), Vol. 4, p. 159.
26. L. G. Kocharov, H. Debrunner, G. Kovaltsov, *et al.*, Astron. Astrophys. **340**, 257 (1998).
27. T. G. Guzik, J. Chen, and J. P. Wefel, in *Proceedings of 24th ICRC* (1995), Vol. 4, p. 123.
28. S. W. Kahler, Astrophys. J. **428**, 837 (1994).
29. A. Anttila, L. G. Kocharov, J. Torsti, *et al.*, Ann. Geophys. **16**, 921 (1998).
30. A. Struminsky, Ann. Geophys. **20**, 1247 (2002).
31. D. Lario and R. B. Decker, Geophys. Res. Lett. **29**, 31 (2002).
32. M.-B. Kallenrode, J. Geophys. Res. **106**, 24989 (2001).
33. A. Struminsky, Preprint No. 10(1153) (Inst. Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Russ. Acad. Sci., 2002).
34. G. P. Lyubimov, Kosm. Issled. (2002, in press).
35. D. V. Reames, Astrophys. J. Lett. **571**, L63 (2002).

Translated by A. Struminskiĭ

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 925–933. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 1002–1012.

Global Repeating Events in the History of the Earth and the Motion of the Sun in the Galaxy

G. N. Goncharov[1] and V. V. Orlov[2]

[1]*Geochemistry Department, St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 199034 Russia*

[2]*Sobolev Astronomical Institute, St. Petersburg State University, St. Petersburg, Russia*

Received March 25, 2003; in final form, May 8, 2003

Abstract—Chronological analyses of correlations between certain global repeating events (mass extinctions of marine organisms, meteorite impacts, and flashes in the frequency of geomagnetic reversals) during the Phanerozoic Eon and the motion of the solar system in the Galaxy are presented for five rotationally symmetrical models for the regular Galactic gravitational field. Thirteen of sixteen mass-extinction events can be described by a repetition interval of 183 ± 3 million years. This is in agreement with the anomalistic period (interval between two subsequent passages of the Sun through the apocenter of its Galactic orbit) in the model of Allen and Martos. The positions of the minima and maxima in Gaussian functions approximating the frequency distribution for geomagnetic reversals also agree with the times of passage of the Sun through the apocenter and pericenter, respectively, of its Galactic orbit in this model. The maximum in the distribution of the deviations of the dates of mass extinctions from the nearest dates of impacts of large, crater-forming bodies is close to zero, providing evidence that many such events are correlated. As a rule, extinctions follow impact events. The impacts of large bodies have occurred most often when the solar system passes through the Galactic plane, while mass extinctions occur more often at some distance from the Galactic plane (about 40 pc). As a rule, intervals of increases in the frequency of geomagnetic reversals coincide with dates of impacts of large bodies. At the same time, these intervals do not show a clear correlation with the dates of mass extinctions. The intensity of mass extinctions, like the energy released by impacts, is consistently higher in periods when the Sun is moving from the apocenter toward the pericenter of its orbit, than when it is moving from the pericenter toward the apocenter. Thus, there is evidence for a variety of relationships between repeating global events in the Phanerozoic and the motion of the Sun in the Galaxy. Long-period variations in the frequency of geomagnetic reversals are correlated with the orbital motion of the Sun, and increases in the frequency of geomagnetic reversals are correlated with impacts. Mass extinctions are correlated with the impacts of large bodies, whose motions may have been perturbed by clouds of interstellar material concentrated toward the Galactic plane and by the shock front associated with the Perseus spiral arm, through which the solar system passes. The velocity of the Sun relative to the spiral pattern is estimated. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The existence of epochs of mass extinctions in the geological history of the Earth has been verified on the basis of extensive statistical data for marine organisms by Raup and Sepkoski [1–3] and Sepkoski [4], and also by Benton [5] for both marine and land organisms. They are associated with global environmental disturbances that clearly must be initiated by powerful sources. Published results do not provide sufficient information to identify these sources with certainty. Nevertheless, according to a number of studies [6, 7, 9, 10], correlations between the dates of mass extinctions and of numerous prominent impact events [6, 7], and also the dates of some peaks in the frequency of geomagnetic polarity reversals [6, 8], may provide evidence that these phenomena are associated with collisions of the Earth with massive bodies whose orbits were disturbed by Galactic tidal forces [6].

Rampino and Stothers [9, 11] were probably the first to draw attention to the possibility that mass extinctions are related to periodic passages of the solar system through the Galactic plane. Orbital disturbances of comets in the Oort cloud by material concentrated in the Galactic plane could initiate "cometary showers" onto the Earth's surface (see, for example, [10–12]). Agreement between the periodicities of the above phenomena in the geological historical record would represent an important argument in favor of a Galactic origin for mass extinctions. In fact, the application of statistical tests to the Phanerozoic Eon or, with more precise dating, specifically to the Mesozoic and Cenozoic Eras, frequently reveals a

1063-7729/03/4711-0925$24.00

periodicity of 30 million years in the dates of the most prominent crater-creating impact events [6, 13], which is similar to the periodicity of geomagnetic reversals over the last 165 million years [6, 8].

A periodicity of 26–27 million years is often detected in the history of mass extinctions of marine organisms [2, 7]. The difference in the periodicities of these two types of events is small, and could be due to the use of different dating techniques [14]. However, some other origins that could result in time lags between these events are also possible. The existence of non-periodic events that are not consistent with the period of 26–27 million years and form a significant fraction of the original data provide evidence supporting this possibility. As a consequence, some studies have argued against the presence of strict periodicities in the history of mass extinctions. In particular, Benton [5] has drawn attention to seven peaks of mass extinctions during the last 250 million years separated by time intervals from 20 to 60 million years.

Recently, passages of the solar system through the Galactic spiral arms have also been considered as a possible origin of critical periods in the history of the Earth (see, for example, [15]). Using the angular rotational velocity of the spiral pattern, 13.5 km/s/kpc, Khristoforova [15] obtained an interval of 370 million years for successive passages through the spiral arms, which cannot explain the short-period component of these events.

Searching for evidence of Galactic effects acting on geological processes requires taking into account the component associated with the anomalistic period (the time interval between successive passages through the orbit apocenter). Analyzing the possible responses of the Earth to this component of motion may help in investigating details of the structure and evolution of the Galactic disk during the last 4.5 billion years.

In the current paper, we search for correlations between terrestrial indicators that are most sensitive to external effects (mass extinctions, impact events, and peaks in the frequency of geomagnetic reversals) and passages of the solar system through regions of various densities in the Galactic disk.

2. CHRONOLOGIES OF MASS EXTINCTIONS, IMPACT EVENTS, AND GEOMAGNETIC INVERSIONS

In their studies of changes in the bio-diversity of marine animals during the Phanerozoic Eon based on an analysis of fossilized remains, Raup and Sepkoski [1, 3] and Sepkoski [4] found that the most statistically well founded epochs of mass extinctions occurred in the Llanvirnian (O_{2l}, 468.6), Ashgillian (O_{3as}, 439), and Ludlovian (S_{2ld}, 411), and also at the boundaries separating the Frasnian and Famennian (D_{3f}, 367), Serpukhovian and Bashkirian (C_{1s}, 322.8), and Gzhelian and Asselian (C_{3g}, 290) Ages. They are also clearly detected later in the Tartarian (P_{2t}, 250) and Upper Norian (T_{3n}, 209.5), then in the Pliensbachian (J_{1p}, 187), Callovian (J_{2c}, 157.1), Tithonian (J_{3tt}, 145.6), Aptian (K_{1a}, 112), and Cenomanian (K_{2c}, 90.4), and also between the Maastrichtian and Danian (K_{2m}, 65) Stages and in the Late Eocene (P_2^3, 35.4), Middle Miocene (N_1^2, 10.4) and Pliocene (N_2^2, 1.6). Here and below, the indices of the stratigraphic stages with which the maxima of the marine extinctions are associated and the ages corresponding to their upper boundaries in millions of years are indicated in parentheses, in accordance with the data of Harland *et al.* [16, 17]. The indicated age of the P_{2t} boundary (250 million years) was taken from Renne *et al.* [18]. Table 1 presents the percentage of species of marine organisms dying out based on the data of Sepkoski [4], after removing the background extinctions using the baseline technique. The mean errors in these quantities are about 2%.

In addition to Sepkoski's [4] data, Table 1 also presents the dates of two other extinction events. Abrupt changes in the marine biota have recently been revealed between the Sakmarian and Artian Ages (P_{1s}, 268.8), although they were not identified by Sepkoski [4], possibly due to large uncertainties in the stratigraphic correlations for various regions in the Permian Period, resulting in calculations being realized for only five time intervals in the Permian [19]. An additional extinction event between the Paleocene and Eocene (P_1^2, 56.5)—which was rather brief (no longer than 400 thousand years, according to Ortiz [21]), but intense and, in contrast to others, accompanied by climactic warming—is also indicated in some works [20, 21].

We can see in Table 1 that three (O_{3as}, P_{2t}, and K_{2m}) of the four most significant events (these plus T_{3n}) characterized according to Sepkoski [4] by the extinction of 37% or more of marine organisms, occurred at intervals of 185–189 million years, anticipating the anomalistic period of the solar motion in the Galaxy.

Analyses of the most precise radiological and biostratigraphic ages of impact craters presented both in recent reviews and in the original studies [13, 22–25] provide information about the epochs of impacts of 72 bodies over the last 450 million years. Table 1 also presents the impact epochs for the 32 most massive bodies, which gave rise to astroblemes with diameters exceeding 9 km; the only exception we have made is for the two oldest astroblemes, whose diameters are

Table 1. Dates and intensities (percentage of species dying out) of mass extinctions in marine organisms, solar passages through the Galactic plane, impacts of the most massive meteorites, and peaks in the frequency of geomagnetic polarity reversals (in millions of years)

Extinction events			AM model	Impact events	Geomagnetic reversals (±2.5 million years)
Index	Date	Intensity			
				0.90 ± 0.10	
N_2^2	2	5	3	1.03 ± 0.02	
				3.5 ± 0.5	
N_1^2	10	6		14.8 ± 0.7	7.5*
				21.5 ± 1.2	
P_2^3	35	10	36	35 ± 5	37.5*
				37 ± 2	
				38 ± 4	
P_1^2	56			50.5 ± 0.8	52.5*
				55 ± 5	
				58 ± 2	
K_{2m}	65	42	76	64.98 ± 0.05	62.5
				65 ± 2	
				67.27 ± 1.97	
				73.8 ± 0.3	72.5
				77 ± 0.4	
K_{2c}	90	11		88 ± 3	
				89 ± 2.7	
				95 ± 7	
K_{1a}	112	8	117	117 ± 8	
				121 ± 2.3	
				128 ± 5	132.5*
J_{3tt}	146	11	151	142.5 ± 0.5	152.5*
J_{2c}	157	7			
				169 ± 6	167.5*
J_{1p}	187	15	182	186 ± 8	192.5*
T_{3n}	210	41	215	214 ± 1	222.5
				217 ± 3	
P_{2t}	250	59	255	247 ± 5.5	≥ 247.5
P_{1s}	269				
C_{3g}	290	8	296		
C_{1s}	323	15	331	323 ± 5	
D_{3f}	367	30	362	368 ± 1.1	
S_{2ld}	411	10	395	397 ± 3	
O_{3as}	439	37	435	440 ± 2	
O_{2l}	469	7	472		

* Peaks significant at the 2σ level.

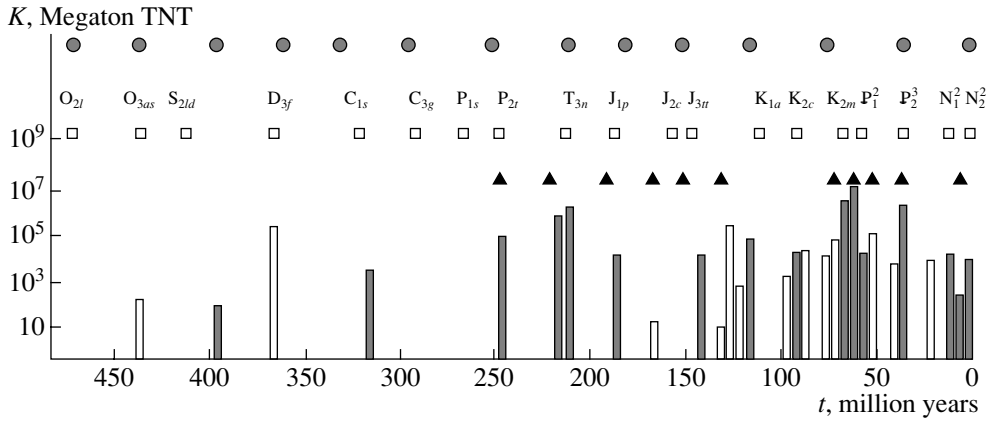

Fig. 1. Dates of mass extinctions (squares), impact events (histogram), peaks in the frequency of geomagnetic reversals (triangles), and solar passages through the Galactic plane in the AM model (circles). Shaded histogram columns indicate impact events that coincide with mass extinctions.

5–6 km, since their initial diameters must have been considerably larger.

We have weighted each of the 72 impact events using the energy released during the impact. This energy K in megatons of TNT ($1\ \mathrm{MT} = 4.2 \times 10^{15}$ J) is calculated via the expression proposed by Basilevskiĭ *et al.* [26]

$$K = 2.4(g/g_0)^{0.4} D_\tau{}^{3.4}, \qquad (1)$$

where D_τ is the crater diameter in km and g and g_0 are the free-fall accelerations on the planet and the Earth, respectively. According to Basilevskiĭ *et al.* [26], this formula is valid for impact craters with diameters up to $(1–2) \times 100$ km, and yields a lower limit for larger craters. In addition, we must take into account the erosive disappearance of small craters and the reduction in the diameters of large craters due to their geological ages. Consequently, the estimated energy releases for older events are lower than their real values. Consideration of this effect is necessary if we wish to estimate the possible effects on the environment and biota resulting from a large number of small impacts along with the more prominent major impact events.

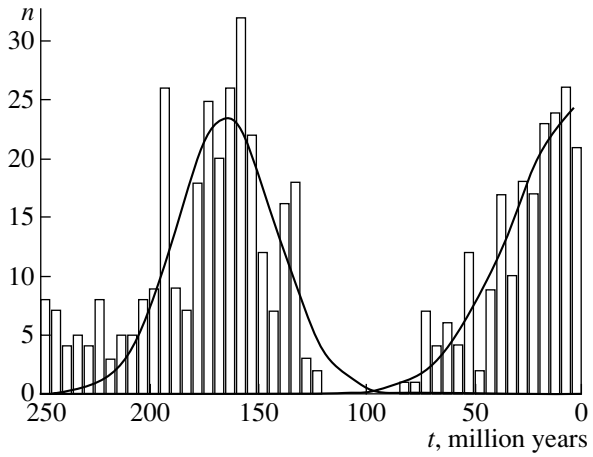

Fig. 2. Distribution of 506 geomagnetic reversals in geological time. The smooth curves show an approximation using Gaussians.

Figure 1 presents a histogram of the energy releases of the 72 impact events binned in intervals of 5 million years. We can see that many of the epochs of mass extinction coincide with epochs of substantial energy release, predominantly due to massive bodies impacting the Earth's surface. The only exceptions are the extinctions O_{2l}, C_{3g}, P_{1s}, and J_{2c}, for which no significant synchronous impact events are revealed. However, massive impacts without global biota crises are also encountered, especially over the last 170 million years (Fig. 1). This may provide evidence that global mass extinctions do not always accompany collisions between the Earth and massive bodies. In addition, it is well known that the environmental changes resulting from an asteroid impact have a duration of several years [27, 28], whereas some epochs of mass extinction have continued over hundreds of thousands or millions of years.

To establish the epochs of maxima in the frequency of geomagnetic reversals, which may reflect the Earth's reaction to its passage through regions of various densities in the Galaxy, we used the master list of reversals over the last 250 million years compiled by Shreider [29, 30], which corresponds to the geological scale time proposed by Harland *et al.* [17]. Figure 2 shows a histogram of the geomagnetic-reversal frequency binned in intervals of 5 million years constructed using these data. There are two broad maxima approximately Gaussian in shape separated by 160 ± 30 million years, which is close to the

Table 2. Solar orbital parameters in five Galactic models

Model	$R_{\min}$, kpc	$R_{\max}$, kpc	e	$\lvert z\rvert_{\max}$, kpc	P_a, 10^6 yrs	P_z, 10^6 yrs
AM	7.96	9.15	0.070	0.085	183	36
CI	8.18	8.62	0.026	0.081	152	35
KO	8.19	9.23	0.060	0.038	168	14
MSO	7.45	8.53	0.068	0.048	136	18
N	8.47	10.2	0.095	0.119	192	53

anomalistic period. In addition, some shorter peaks in the frequency of geomagnetic reversals are detected at the 2σ significance level with respect to their neighboring minima, as well as less significant peaks (Table 1). These maxima coincide with epochs of intense energy release associated with impacts (Fig. 1), with the most massive bodies playing a dominant role in this correlation (Table 1). The only exception is the maximum of the geomagnetic-reversal frequency at 152.5 million years, for which no corresponding impact events have been detected, possibly due to incompleteness in the list of such events.

3. COMPARISON OF GLOBAL EVENTS WITH THE PARAMETERS OF THE SUN'S MOTION IN THE GALAXY

In a comparative analysis of global repeating events in the history of the Earth and the solar motion in the Galaxy, we must take into account a considerable number of components included in modern models of the regular Galactic field. To construct the Sun's orbit in the Galaxy, we must assume a model for the mass distribution or a potential for the regular field in the Galaxy. A number of well known Galactic models can be used for this purpose (for example, the AM [31], CI [32], KO [33], MSO [34], and N [35] models).

Usually, we assume that the Galaxy consists of several subsystems, with the main subsystems being the nucleus, bulge, disk, and halo. For each subsystem or superposition of subsystems, we choose a specific form of potential that depends on some set of parameters. These parameters are estimated using data on the Galactic rotation and the structural–kinematic characteristics of the solar neighborhood.

It is quite natural that the solar orbits in different models of the Galactic field are somewhat different. To estimate these differences, we constructed solar orbits for the five Galactic models noted above. We numerically integrated the equations of the solar motion 6×10^8 yrs into the past. The orbit in the comoving plane is box-like for all five models. Some of the orbital parameters are presented in Table 2. The columns indicate the model, minimum ($R_{\min}$) and maximum ($R_{\max}$) distances from the Galactic rotational axis, orbital eccentricity

$$e = \frac{R_{\max} - R_{\min}}{R_{\max} + R_{\min}}, \qquad (2)$$

maximum distance from the Galactic plane $|z|_{\max}$, mean anomalistic period P_a (interval between two successive passages through the orbit apocenter), and mean period P_z for oscillations about the Galactic plane. The quantities $R_{\min}$, $R_{\max}$, and $|z|_{\max}$ are in kiloparsecs (kpc) and the periods P_a and P_z are in millions of years.

Table 2 shows that the different models yield different parameters of the solar orbit. The relative differences in the oscillations in the z direction are especially appreciable. The parameters of the solar motion in the Galactic plane, in particular the anomalistic period P_a, are more stable from model to model.

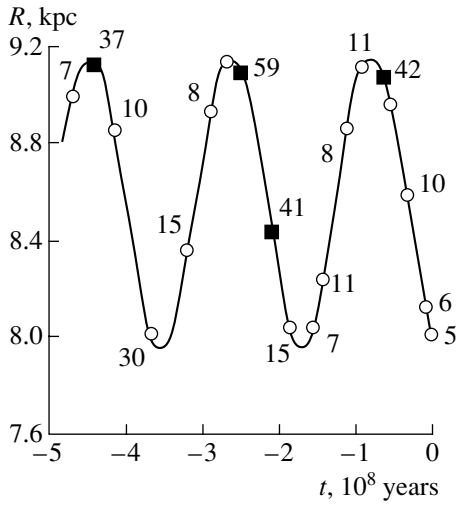

Fig. 3. Time evolution of the distance between the Sun and the Galactic rotational axis in the AM model and dates for mass extinctions (circles). The four most significant extinctions are marked by dark squares. The numbers next to the symbols indicate the percentage of species of marine organisms dying out.

Table 3. Significance levels for the Kolmogorov–Smirnov criterion

Sample	α_{KS}, %
ME–Imp	0.0085
Rev–Imp	0.0365
Rev–ME	>20
ME–AM	2.0
Imp–AM	11
AM–Rev	>20

Note: ME denotes mass extinctions, Imp—prominent impact events, Rev peaks in the frequency of geomagnetic reversals, and AM—passages through the Galaxy plane in the AM model.

Note that three of the four most intense epochs of mass extinction, occurring 65, 250 and 439 million years ago, are separated by intervals of $\Delta t = (185–190) \times 10^6$ yrs, which is close to the anomalistic period $P_a \approx 183 \times 10^6$ yrs obtained for the AM model (Fig. 3). In this model, the solar passages through the orbit apocenter (81, 263, and 446 million years ago) are likewise close to three of the four indicated epochs of intense mass extinctions.

If we take into account only the mass extinctions, translating their dates with a step close to the anomalistic period, we can synchronize thirteen of the sixteen events presented in Table 1. The events that repeat with this step are separated by the following time intervals (from young to old):

$$N_2^2 - J_{1p}(185); \quad P_2^3 - T_{3n}(175); \quad K_{2m} - P_{2t}(185);$$
$$K_{2c} - P_{1s}(179); \quad K_{1a} - C_{3g}(178);$$
$$J_{3tt} - C_{1s}(177); \quad J_{1p} - D_{3f}(180);$$
$$T_{3n} - S_{2ld}(201); \quad P_{2t} - O_{3as}(189);$$
$$C_{3g} - O_{2l}(179).$$

The intervals in millions of years are presented in parentheses. The average interval for these ten differences is 183 ± 3 million years, in good agreement with the anomalistic period in the AM model. The extinction events N_1^2, P_1^2, and J_{2c} fall out of this sequence.

To characterize the coincidence of various events, we studied the distributions of deviations Δ from the nearest dates of sets of compared events. We constructed histograms of these distributions (Fig. 4) and determined the significance level α_{KS} for their deviations from a uniform, random distribution according to the Kolmogorov–Smirnov criterion (Table 3). The number k of intervals for the histograms is determined in accordance with [36] via the formula

$$k = 3.3 \log N + 1,$$

where N is the number of measurements. The maximum of the distribution of deviations of mass-extinction dates from dates of impacts of massive crater-creating bodies is close to zero (Fig. 4a). This deviation is significant ($\alpha_{KS} = 0.0085\%$), indicating a high degree of correlation between the dates of mass extinctions and of massive-body impacts. The massive-body impacts preceded the nearest extinctions at the $\alpha_{BS} = 5\%$ significance level for binomial test statistics.

The dates of peaks in the frequency of geomagnetic reversals and of massive impacts are also close, although the deviations between them are somewhat higher (Fig. 4b). No regularity in their sequence has been detected. It is possible that this is due to large errors in the estimates of the dates of these peaks. At the same time, the distribution of deviations is not random, with a high level of significance ($\alpha_{KS} = 0.037\%$). The peaks in the geomagnetic reversals are distributed randomly with respect to the dates of mass extinctions (Fig. 4c, Table 3).

The dates of mass extinctions are coincident with passages of the solar system through the Galactic plane in the AM model ($\alpha_{KS} = 2.0\%$). The main mode of deviation distribution is shifted somewhat from zero (by approximately 6 million years, which corresponds to a distance from the plane of about 40 pc; Fig. 4d). If the model is correct, this may indicate that the majority of mass extinctions occurred at a certain distance from the Galactic symmetry plane. The statistics of the signs of Δ does not reveal any regularity in these events.

There is a mode near zero in the distribution of the deviations of prominent impact events from the AM epochs for the solar system's passages through the Galactic plane (Fig. 4e); however, in general, the distribution does not significantly differ from a random distribution ($\alpha_{KS} = 11\%$). Nevertheless, we can note some increase in the frequency of impacts in the plane.

The distribution of deviations of the dates of peaks in the frequency of geomagnetic reversals from epochs of the Sun's passages through the Galactic plane is random (Fig. 4f, Table 3).

In addition to the indicated regularities, there is a correlation between the intensities of some global events and the Sun's motion along individual sections of its Galactocentric orbit in the AM model. Namely, the mean values $\bar{X}$ and rms deviations S of the intensity of mass extinctions (Table 1, Fig. 3) during the Sun's motion from the orbital apocenter to the pericenter exceed those for the motion from the pericenter to the apocenter (Table 4). All four "great" Phanerozoic mass-extinction events (O_{3as}, P_{2t}, T_{3n}, and K_{2m}) occurred during the apocenter–pericenter sections.

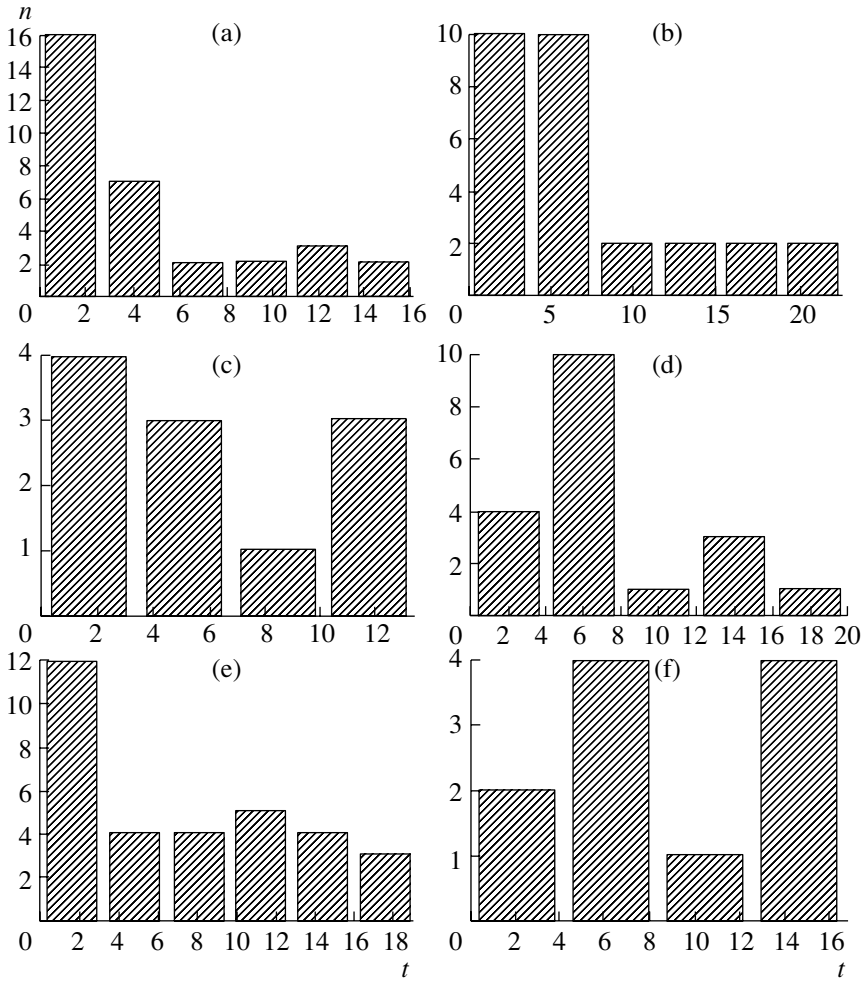

Fig. 4. Histograms of deviations of dates for various events: (a) mass extinctions and impact events; (b) reversals and impact events; (c) mass extinctions and reversals; (d) mass extinctions and Galactic-plane passages; (e) Galactic-plane passages and impact events; (f) Galactic-plane passages and reversals.

The significance of the difference between the mean intensities of mass extinctions determined separately for the sections of the solar motion from the apocenter to the pericenter ($\bar{X}_1 = 25.5\%$, $n_1 = 10$) and from the pericenter to the apocenter ($\bar{X}_2 = 9.3\%$, $n_2 = 7$) can be estimated by applying the t criterion to the entire available data set starting from O_{2l} (Fig. 3, Table 1) and taking into account the corresponding difference in the variances ($S_1 = 18.8\%$, $S_2 = 3.3\%$) [37]. The resulting significance level, $\alpha_t = 0.014$, indicates the high trustworthiness of this difference.

The energy releases calculated for the entire sample of 72 dated Phanerozoic impact events are also greater for those occurring when the Sun is moving from the apocenter to pericenter than in the opposite case. The ratios of the energy releases obtained from the data of Fig. 1 and Table 4 for successive identical sections of the orbit of duration 180 million years are close to each other, and equal to 7.1 ± 0.3 on average. This coefficient apparently represents a "clearing of the geological memory" of the intensity of impact events by roughly a factor of seven over one anomalistic period of the solar motion in the Galaxy.

The ratio of the energy releases of impacts occurring during the Sun's motion from the apocenter to the pericenter and vice versa is approximately 54. Taking into account the "clearing factor" for half of the orbit, we find that the intensity of impact energy releases in the orbital section from the apocenter to the pericenter exceeds the intensity for the pericenter–apocenter section by a factor of 15. It is interesting that, allowing for a small time shift, the long-period component of the frequency of geomagnetic reversals increases in the section of the orbit from the apocenter to pericenter (Fig. 2). In the opposite section, this tendency changes its sign, and the frequency of reversals decreases.

Table 4. Statistical parameters for mass-extinction and impact events in various sections of the solar orbit in the AM model

Orbit section, 10^6 yrs	$\bar{X}$, %	S, %	M, 10^3 Megaton TNT
0–81	16	18	158000
81–172	9	2	3000
172–263	38	22	23000
263–354	12	5	420
354–446	26	14	3100

Note: The first, third, and fifth rows correspond to motion from the apocenter toward the pericenter, while the second and fourth rows correspond to motion from the pericenter toward the apocenter. $\bar{X}$ and S are the arithmetic mean and rms deviation of the intensity of mass extinctions. M is the total energy release calculated for the entire sample of 72 impact events.

4. DISCUSSION AND CONCLUSIONS

Our study of correlations between certain global repeating events (mass extinctions of marine organisms, peaks in the frequency of geomagnetic reversals, and impacts of celestial bodies) during the Phanerozoic Eon and the Sun's motion in the Galaxy for five rotationally symmetrical models of the regular Galactic gravitational field enables us to identify more precisely the mechanisms behind the cause–effect relationships between various global processes in the cosmo-geological history of the Earth.

A repetition interval of 183 ± 3 million years is revealed for thirteen of sixteen mass-extinction events. This interval coincides with the anomalistic period of the Sun's motion in the Galaxy in the model of Allen and Martos [31]. The minimum and maximum of the frequency distribution of geomagnetic reversals approximated by two Gaussians (Fig. 2) also coincides with the Sun's passages through the orbital apocenter and pericenter in this model. Therefore, we can assert with some confidence that events of mass extinction and the long-period component of the frequency of geomagnetic reversals during the Phanerozoic are associated with the solar motion in the Galaxy described by the AM model.

The behavior of various parameters of the terrestrial magnetic field (the reversal frequency, polarity, strength, variations in its direction and strength) studied by Galyagin *et al.* [38] using wavelet analysis for a long time interval of 1700 million years and in more detail for the Phanerozoic Eon does not reveal any well defined periodicities. This provides evidence for a Neogene instability in the geomagnetic field, which cannot be "tuned" to periodic processes related to the solar motion in the Galaxy. One exception may be the oscillations of the polarity indicated in [38] for four periods of about 180 million years in the interval from 1500 to 800 million years ago, which are consistent with the anomalistic period. On the other hand, there are numerous short flashes with durations close to one entire oscillation. A more detailed study of the distribution of these reversal peaks over the last 250 million years with a resolution of 5 million years shows (Fig. 2, Table 3) that they are in agreement with the impacts of celestial bodies.

The main maximum of the distribution of the deviations of the dates of mass extinctions from the nearest impacts of massive crater-creating bodies is close to zero, indicating a correlation between many of these events. In addition, mass extinctions usually follow the impacts of massive bodies.

The uncertainty noted above in the reaction of the biosphere to specific impact events may be due to the impacting bodies having a variety of origins: asteroids and cometary nuclei, for example. It is plausible that cometary nuclei saturated with toxic components exert more disastrous effects on the biosphere compared to asteroids, although the energies released during the impacts of both a cometary nucleus and an asteroid of the same size could be comparable if the differences in their velocities and densities roughly compensate each other. Some support for this idea may be provided by the behavior of the peaks in the frequency of geomagnetic reversals, which, as a rule, coincide with the dates of impacts of massive bodies. At the same time, these peaks do not show any clear correlation with dates of mass extinctions. Increased concentrations of osmium and iridium and other signs of impacts of massive asteroids are likewise not typical of many sediments stored during epochs of mass extinction (see, for example, [7]). This may also provide indirect support for a "cometary" origin of mass extinctions.

The dates of mass extinctions of organisms are close to epochs of the solar system's passages through the Galactic plane in the AM model. This correlation may be associated with the increased concentration of molecular clouds at the Galaxy plane. At the same time, there is no significant correlation between impact events and solar passages through the Galactic plane. Taking into account the correlation of mass extinctions with numerous impact events, together with the fact that the impact of cometary nuclei may be a more probable origin for mass extinctions than asteroid impacts, we suggest that the impact of cometary nuclei is associated with an increase in the frequency of impact events during cometary showers.

The facts testifying to the existence of variations in the intensity of external effects during the solar

motion along different sections of its Galactocentric orbit, while the translational time sequence of global events is preserved, require special attention and discussion. On average, stronger effects are revealed during the orbital sections corresponding to motion from the orbit apocenter to the pericenter. This asymmetry may be associated with passages through the Perseus arm, near the orbit apocenter (see, for example, [39]).

Dates of entries into the Perseus arm may be correlated with the most intense extinction events, such as O_{3as} (439 million years ago), P_{2t} (250 million years ago), and K_{2m} (65 million years ago). Exits from the arm may also be correlated with the intense extinction events D_{3s} (367 million years ago), T_{3n} (310 million years ago), and P_1^2 (56 million years ago). If we associate these extinctions with epochs of entries into and exits from the Perseus spiral arm, this suggests interaction with the arm shock front as the mechanism initiating perturbations. The shortening of the interval between entries and exits during each successive orbit (72, 40, and 9 million years, respectively) may indicate that the arm is "running away" from the solar system with each new approach near the solar orbital apocenter. If the spiral angle is about 20° (see, for example, [40]), we can estimate the velocity of the solar motion with respect to the spiral pattern to be 3 km/s; for an angle of 5° [39], this velocity is about 13 km/s.

REFERENCES

1. D. M. Raup and J. J. Sepkoski, Science **215**, 1501 (1982).
2. D. M. Raup and J. J. Sepkoski, Proc. Nat. Acad. Sci. **81**, 801 (1984).
3. D. M. Raup and J. J. Sepkoski, Science **231**, 833 (1986).
4. J. J. Sepkoski, Geotimes **39**, 15 (1994).
5. M. J. Benton, Science **268**, 52 (1995).
6. S. Yabushita, Earth, Moon, Planets **72**, 343 (1996).
7. M. R. Rampino and B. M. Haggerty, Earth, Moon, Planets **72**, 441 (1996).
8. D. M. Raup, Nature **314**, 341 (1985).
9. M. R. Rampino and R. B. Stothers, Science **226**, 1427 (1984).
10. M. R. Rampino and B. M. Haggerty, *Hazards due of to Asteroids and Comets,* Ed. by T. Gehrels (Univ. Arizona Press, Tucson, 1994), p. 827.
11. M. R. Rampino and R. B. Stothers, Nature **308**, 709 (1984).
12. P. Nurmi, M. J. Valtonen, and J. Q. Zheng, Mon. Not. R. Astron. Soc. **327**, 1367 (2001).
13. R. A. F. Grieve and L. J. Pesonen, Earth, Moon, Planets **72**, 357 (1996).
14. R. B. Stothers, Geophys. Res. Lett. **16**, 119 (1989).
15. N. N. Khristoforova, *Proc. All-Russian Astron. Conf., St. Petersburg* (2001), p. 188.
16. W. B. Harland, R. L. Armstrong, A. V. Cox, *et al., A Geologic Time Scale 1989* (Cambridge Univ. Press, Cambridge, 1990).
17. W. B. Harland, R. L. Armstrong, L. E. Craig, *et al., A Geologic Time Scale 1989. Chart with Modifications and Additions* (Cambridge Univ. Press, Cambridge, 1990).
18. P. R. Renne, Zichao Zhang, M. A. Richards, *et al.*, Science **269**, 1413 (1995).
19. A. S. Alekseev, Doctoral Dissertation (Mosk. Gos. Univ., Moscow, 1998).
20. K. Kaiho, Paleogeogr. Paleoclimatol. Paleoecol. **111**, 45 (1994).
21. N. Ortiz, Mar. Micropaleontol. **26**, 341 (1995).
22. V. I. Fel'dman, *Petrology of Meteorites* [in Russian] (Izd. Mosk. Gos. Univ., Moscow, 1990).
23. L. L. Koshkarova, M. A. Nazarov, and N. N. Kononkov, Geokhimiya, No. 7, 928 (1994).
24. M. A. Nazarov, Doctoral Dissertation (Inst. Geochem., Ross. Akad. Nauk, Moscow, 1995).
25. L. J. Pesonen, Earth, Moon, Planets **72**, 377 (1996).
26. A. T. Bazilevskiĭ, B. A. Ivanov, K. P. Florenskiĭ, *et al., Impact Craters on the Surface of the Moon and Planets* [in Russian] (Nauka, Moscow, 1983).
27. L. W. Alvarez, W. Alvarez, F. Asaro, and H. V. Michel, Science **208**, 1095 (1980).
28. W. S. Wolbach, R. S. Lewis, and E. Anders, Science **230**, 167 (1985).
29. A. A. Shreĭder, Fiz. Zemli, No. 4, 3 (1993).
30. A. A. Shreĭder, Fiz. Zemli, No. 11, 29 (1997).
31. C. Allen and M. A. Martos, Rev. Mex. Astron. Astrofis. **13**, 137 (1986).
32. R. G. Carlberg and K. A. Innanen, Astron. J. **94**, 666 (1987).
33. S. A. Kutuzov and L. P. Osipkov, Astron. Zh. **66**, 965 (1989) [Sov. Astron. **33**, 498 (1989)].
34. M. Miyamoto, C. Satoh, and M. Ohashi, Astron. Astrophys. **90**, 215 (1980).
35. S. Ninković, Astron. Nachr. **313**, 83 (1992).
36. B. L. van der Waerden, *Mathematische Statistik* (Springer, Berlin, 1957; Inostr. Lit., Moscow, 1960).
37. G. K. Kanji, *100 Statistical Tests. New Edition* (Sage, London, 1999).
38. D. K. Galyagin, M. Yu. Reshetnyak, L. M. Pecherskiĭ, *et al.*, Fiz. Zemli **4**, 82 (2000).
39. A. M. Mel'nik, Pis'ma Astron. Zh. (2003, in press).
40. I. I. Nikiforov and T. V. Shekhovtsova, *Stellar Dynamics: from Classic to Modern*, Ed. by L. P. Ossipkov and I. I. Nikiforov (St. Petersburg Univ. Press, St. Petersburg, 2001), p. 88.

Translated by V. Badin

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 934–945. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 1013–1025.

Large-scale Dimmings Produced by Solar Coronal Mass Ejections According to SOHO/EIT Data in Four EUV Lines

I. M. Chertok and V. V. Grechnev

Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
Institute for Solar–Terrestrial Physics, P.O. Box 4026, Irkutsk, 664033 Russia

Received March 26, 2003; in final form, May 8, 2003

Abstract—SOHO/EIT data are used to analyze dimmings, or transient coronal holes (regions of reduced soft-X-ray and EUV emission), which are observed on the solar disk after halo-type coronal mass ejections (CMEs). Simultaneous observations in the 171 Å FeIX/X, 195 Å FeXII, and 284 Å FeIX coronal lines, which are sensitive to temperatures of $T_e \approx 1.2$, 1.5, and 2.0 MK, respectively, are considered, together with the 304 Å HeII transition-region line ($T_e \approx (0.02–0.08)$ MK). Difference images taken at intervals of six and twelve hours and compensated for solar rotation indicate that dimmings are normally strongly pronounced and have similar large-scale structures in the moderate-excitation-temperature 171 Å and 195 Å coronal lines, while the higher-temperature 284 Å line mainly display the deepest portions of the dimmings. In addition, clear dimmings with relatively small areas are visible in the 304 Å transition-region line during many CMEs, in particular, in regions adjacent to the source of the eruption. Moreover, dimmings in the transition region without coronal counterparts are observed during some events. These results suggest that the opening of magnetic-field lines and the resulting density reduction that occur during a CME can also involve cold plasma of the transition region. In addition, the effects of temperature variations cannot be ruled out for some dimming structures. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Global restructuring and strong perturbations of the magnetic field during coronal mass ejections (CMEs) on the Sun are accompanied by large-scale phenomena, such as dimmings, coronal waves, post-eruptive arcades, etc. (see, e.g., the reviews [1–4] and references therein). Dimmings, or transient coronal holes [5–11], are regions of reduced soft X-ray and EUV emission with lifetimes of several hours to one and a half days. They are formed after the CME near the eruption center, for example, at the periphery of the pre-eruptive sigmoid structure, and can cover a large portion of the solar disk. Analyses of Yohkoh/SXT [12] and SOHO/EIT [13] heliograms suggest that dimmings should mainly be interpreted as a result of the complete or partial opening of the coronal magnetic fields in these structures, which results in the evacuation of material and a corresponding reduction in the intensity of their emission. Direct evidence for outflows from dimmings located near an eruption center was obtained in [10] based on the Doppler shifts of several lines recorded with the SOHO/CDS spectrometer [14].

A coronal wave is a emitting front that can fairly often be observed ahead of a developing dimming and can propagate from the eruption center at a speed of several hundred km/s [8, 15–22]. This wave is interpreted either as a MHD disturbance (shock wave)—a coronal analog of the chromospheric Moreton wave observable in Hα [23]—or as a result of the compression of plasma at the advancing boundary of the dimming—a region of opening magnetic-field lines [24, 25].

If the structure of the global solar magnetosphere is relatively simple (as is typical near the minimum of the activity cycle or during its growth phase, when only one active region is present on the solar disk), dimmings and coronal waves are more or less isotropic, and develop in a wide-angle sector symmetrically about the eruption source [8]. However, as is shown in [26], when the global solar magnetosphere is complex (as is typical near the cycle maximum, when several active regions, filaments, and other structures are observed on the disk), strongly anisotropic, channel-shaped ("channeled") dimmings can coexist with relatively compact dimmings adjacent to the eruption center. The channeled and compact dimmings have comparable contrasts, but the channeled dimmings stretch along several narrow, extended features (channels) and can span nearly the entire visible disk. In such cases, coronal waves are either not observed at all or are also

1063-7729/03/4711-0934$24.00

anisotropic and propagate within a restricted angular sector.

In some cases, so-called twin dimmings are observed, with the strongest intensity decreases occurring in two regions adjacent to the eruption center located symmetrically about the polarity-reversal line and the posteruptive arcade [8, 9]. Such twin dimmings are probably the bases of a large-scale twisted structure that exhibits eruptive behavior during the CME (see also [2, 27]). The post-eruptive arcade is formed when the strongly CME-disturbed magnetic field in a vast coronal region relaxes to an equilibrium state, reconnecting in extended current sheets high in the corona, and accompanied by prolonged energy release, particle acceleration (sometimes, to very high energies), and the formation of a system of rising loops [28].

To understand the nature of dimmings and coronal waves, it is important to analyze observations of these disturbances in various spectral lines and wavelength ranges. Almost all studies of currently known dimmings (and coronal waves) have been based on Yohkoh/SXT observations in soft X-rays corresponding to temperatures $T_e > 2.5$ MK or/and SOHO/EIT observations in the EUV FeXII line (195 Å), which are sensitive to coronal plasma at $T_e \approx 1.5$ MK [6–9, 11, 29]. These data are used precisely because they represent full-disk images obtained over fairly short time intervals—for example, usually 12 min for the 195 Å line. Meanwhile, full-disk SOHO/EIT heliograms are also observed in the 171 Å FeIX/X and 284 Å FeXV coronal lines excited at $T_e \approx 1.2$ and 2.0 MK, respectively, as well as with a 304 Å filter [13, 30], which includes the HeII transition-region line ($T_e \approx 0.02$–0.08 MK) and the far less intense SiXI coronal component ($T_e \approx 1.0$ MK). However, the use of 171, 284, and 304 Å heliograms is substantially limited by the large time interval between solar images, usually six hours. It is obvious that coronal waves cannot be studied using such heliograms; however, the situation is different for dimmings, since they have long lifetimes. Observations in the same four EUV lines with much higher time resolution are being carried out by the TRACE space observatory [31], but, like the SOHO/CDS spectrometer data [14], these data have been used little in studies of large-scale dimmings and coronal waves due to the restricted field of view of these instruments [10, 32, 33]. Thus far, the analysis of large-scale dimmings simultaneously either by Yohkoh/SXT soft X-ray data and 195 Å SOHO/EIT UV heliograms or by SOHO/EIT UV data in three coronal lines 171, 195, and 284 Å has been carried out for a few events only. Just these studies have revealed similarities between dimmings observed in coronal lines corresponding to different temperatures and in soft X-rays, and have demonstrated that the opening of magnetic structures and decreases in the plasma density in these structures play a more dominant role in dimmings than do temperature variations. To our knowledge, no observations of dimmings on the solar disk using a 304 Å filter have been analyzed to date. In [36], 304-Å heliograms were used to analyze only one eruptive event at the limb.

We analyze dimmings here using data for all four SOHO/EIT wavelengths, including the band containing the HeII transition-region line. This requires special methodological approaches, which are described in Section 2. Four events are used in Section 3 as examples to illustrate the multi-temperature properties of dimmings observed in the FeIX/X, FeXII, and FeXV coronal lines and the HeII transition-region line. We discuss our results in Section 4.

2. METHODOLOGICAL NOTES

In the standard SOHO/EIT observing regime, heliograms in a $45' \times 45'$ field of view with a 1024×1024 frame and $2.6''$ pixels are recorded once every 12 minutes in the high-priority 195 Å line and four times each day (at about 1, 7, 13, and 19 UT) in the 171, 284, and 304 Å lines [30]. Only during rare and relatively short special programs is the twelve-minute regime implemented for lines other than the 195 Å line. In such cases, however, data for the analysis of some particular event in all four lines remain limited and nonuniform. Moreover, the probability of recording major eruptive events during such special, short-term programs is not high. At the same time, detailed analyses of the development of dimmings and coronal waves in lines other than the high-priority line are obviously of considerable interest, and should be the subject of careful studies. We have used here data of standard SOHO/EIT observations. The corresponding FITS files are available in the EIT catalog at http://umbra.nascom.nasa.gov/eit/eit-catalog.html.

Events in which a CME and the main stages of the development of dimmings occur during a six-hour interval between observations in the 171, 284, and 304 Å lines are the most suitable for the desired analysis, provided that no other major eruptive events take place during this interval. The analysis in each of the four lines includes comparisons of heliograms taken before and after the event. The dimmings are most pronounced in difference images obtained by subtracting the pre-event heliogram from the corresponding heliogram taken during or after the event.

In such images, dimmings appear as dark features with reduced intensity. However, a simple subtraction of two images, especially over a six-hour interval, will inevitably give rise to artifactual light and dark features (spurious dimmings), since the solar rotation leads to the displacement of features on the disk. To avoid such effects, we employed an IDL procedure before the subtraction to compensate for the solar rotation, as is done in [26]—the images at each wavelength are all rotated to a single reference position, usually corresponding to a time before the event.

In particular, this procedure for rotating (aligning) the images made it possible to analyze events that culminated near one of the observational times in the 171, 284, and 304 Å lines (i.e., 1, 7, 13, or 19 UT), as long as the events were isolated. In such cases, the difference images were constructed using frames taken twelve hours apart, and both the reference and subtracted frames were rotated to the time of a central frame taken during the event; i.e., they were shifted by six hours. For example, if the event occurred near 13 UT, the difference images were obtained from the 7 UT and 19 UT heliograms, which were first rotated to 13 UT.

For adequate comparison of dimmings in all four lines, rotated (aligned) images of six-hour or twelve-hour intervals were formed not only at the 171, 284, and 304 Å wavelengths, but also in the 195 Å line. The much more detailed twelve-minute data at this wavelength and the running- and fixed-difference images obtained from them after compensation for the solar rotation were used to gain information on the development of dimmings and accompanying phenomena. The running difference images are produced by subtracting the preceding heliogram from each current heliogram, while the fixed-difference images are created by subtracting the same reference image of the disk (usually obtained before the event) from all subsequent heliograms. The running difference images emphasize changes in the brightness, location, and structure of sources that have appeared during the interval between successive frames. Changes that have occurred during the event, including dimmings, are clearly visible in the fixed-difference images (naturally, relative to the chosen pre-event reference frame). Obviously, in forming difference images over a six-hour interval, there is no reason to associate all brightness variations on the disk with the eruptive event considered, even if it is an isolated event, since weak independent activity is almost constantly present in other regions of the disk, and also manifests itself in the difference heliograms.

As in [26], the difference images were formed in a restricted intensity range to make the dimmings better defined against the background of other structures. As a result, the brightest sources (in particular, those arising during a flare and in the posteruptive phase) are usually saturated. Naturally, the outlines and some structural details of the dimmings vary to some extent, depending on the choice of intensity range. One more remark concerns Grid-like patterns—systems of periodic vertical and horizontal strips—that are fairly common in the original SOHO/EIT images and have an instrumental origin [13, 30]. In the rotated difference images, the horizontal strips disappear but the vertical strips are enhanced inside the disk, especially if the heliograms are rotated by a large angle, corresponding to a six-hour interval, as in this case.

This technique for the formation of six-hour or twelve-hour SOHO/EIT difference images compensated for solar rotation enabled us to analyze simultaneous manifestations of dimmings in the four EUV lines.

3. ANALYSIS OF PARTICULAR EVENTS

We employed the above technique to analyze several dozen eruptive events occurring in 1997–2002, mainly related to halo CMEs recorded by the SOHO/LASCO coronagraph [37] in which the source of eruption was located in the central zone of the visible disk. Some of these events such as the rotated six-hour or twelve-hour difference images in the 171, 195, 284, and 304 Å lines, detailed data on the development of accompanying disturbances observed at 195 Å presented in the form of JAVA computer movies containing running- and fixed-difference images, and other illustrations, including data on CMEs and the time profiles and relative depths of dimmings, can be found at http://helios.izmiran.troitsk.ru/lars/Chertok/Dimming/general.html.

To demonstrate the multi-wavelength characteristics of the dimmings, we consider below the same three events described by us in [26] in the context of studies of channeled dimmings at 195 Å, and supplemented them by another event with well-defined dimmings in the transition-region line. Thus, our study can be considered as a continuation of [26], where information on the first three events is given in more detail; in particular, there we describe the development of dimmings in the 195 Å line based on twelve-minute heliograms. For each of the events considered, in addition to the rotated difference images showing the dimmings in all four wavebands arranged in order of increasing temperature, the corresponding figures of this section include the original 284 and 304 Å heliograms illustrating the background against which the eruption took place.

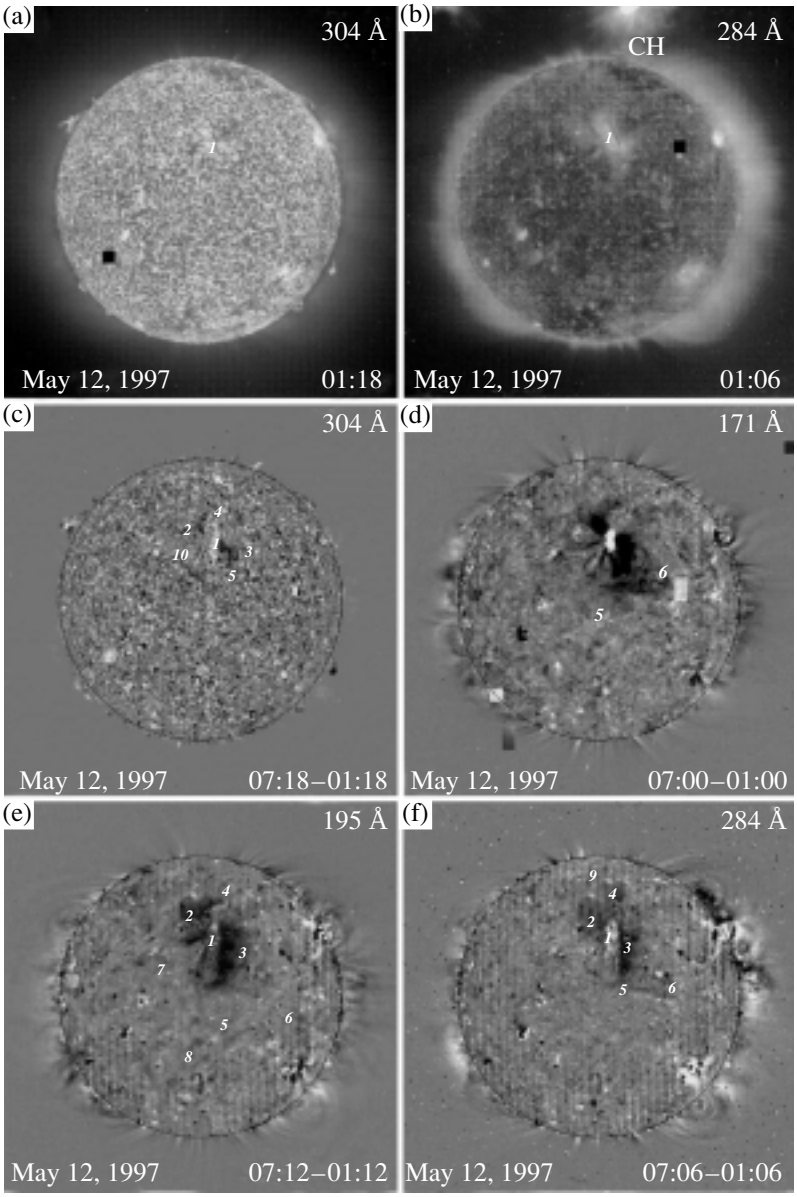

Fig. 1. Event of May 12, 1997: (a), (b) pre-event SOHO/EIT heliograms in the 304 and 284 Å lines and (c)–(f) rotated six-hour difference images in the 304, 171, 195, and 284 Å lines, illustrating the location and structure of dimmings in the transition region and corona.

3.1. Event of May 12, 1997

The characteristics of the halo CME and related quasi-isotropic large-scale disturbances in this well-known eruptive event have been analyzed many times [3, 8, 26, 27, 38]. The overall similarity in the locations and shapes of the deepest dimmings near the eruption center in the 171, 195, and 284 Å coronal lines was demonstrated in [8] using the original (not difference) SOHO/EIT heliograms. It is interesting to analyze the manifestation of the dimmings of this event in the six-hour difference heliograms, using both the coronal lines and the 304 Å transition-region line.

The eruption source was located in AR 8038 (region *1* in Fig. 1), which was observed in the central zone of the northern half of the disk (N21, W09) and was essentially the only active region on the visible hemisphere (Figs. 1a, 1b); the event as a whole was associated with a long-lived flare of class 1F/C1.3 with its maximum near 04:50 UT. The CME recorded by the SOHO LASCO coronagraph was classified as a full halo emitting everywhere around the occulting disk, but its brightness was maximum over the eastern, western, and northern limb sectors [38].

According to the 195 Å observations, the mod-

erately deep dimmings and coronal wave (in this case, under the conditions of a relatively simple global solar magnetosphere) were quasi-isotropic and covered the visible disk almost entirely [3, 8, 26]. Under the chosen restriction of the intensity range the deepest dimmings can be seen in the difference heliogram between 07 and 01 UT (Fig. 1e), first of all, as two darkened regions with similar sizes and shapes (twin dimmings) situated symmetrically about the eruption center, northeast and southwest of the bright, compact posteruptive arcade 1. In addition to the main dimmings 2 and 3 directly joining to the posteruptive arcade, the intensity is also considerably reduced in the fairly extended branches 2-4 and 3-5, which emerge from these dimmings and stretch toward the northern polar coronal hole (CH) and the heliographic equator, respectively. Branch 3-5 of the dimming broadens near its southern end and has a faint westward extension along the equator, to point 6. Two other narrow dimming branches, 1-7 and 1-8, originate from the eastern edge and southern end of the posteruptive arcade, with the southern branch 1-8 even intersecting the heliographic equator.

Similar dimmings can be seen in the corresponding 171 Å heliogram (Fig. 1d); i.e., at a somewhat lower temperature of the coronal plasma. Some differences can be noted in the southern branch of the dimming, which splits in region *5*, while its westward equatorial continuation 5-6 appears as two parallel strips with a more pronounced darkening (deeper dimming) than in the 195 Å line. Another difference is the lack of the transequatorial dimming branch 1-8 observed in the 195 Å line.

The twin structure of the main dimmings 2-4 and 3-5 is also reproduced in the high-temperature 284 Å coronal line (Fig. 1f). Of the dimming branches, the southern component of the westward, near-equatorial dark strip 5-6 is also visible here. In addition to showing the dimmings, the emitting chain 1-9 is most pronounced in this line; it has developed by this time (07:06 UT) along the western edge of the northeastern dimming 2-4, from the posteruptive arcade to the southern boundary of the northern polar coronal hole. As is shown in [8] and follows from analyses of the available heliograms (including difference heliograms), within the next two to twelve hours, this chain had stretched over the entire southern boundary of the coronal hole, and was observed in the 171, 284, and 304 Å lines as well as in the 195 Å line. As in other similar cases [39], the formation of this chain seems to result in connection with energy released due to the interaction between the open field lines of the coronal hole and large-scale magnetic structures involved in the CME process.

It is important that manifestations of dimmings are also pronounced in the six-hour difference heliogram taken in the 304 Å transition-region line (Fig. 1c). Here, some analogs of the main coronal twin dimmings, 1-2-4 and 1-3-5, are observed on either side of the posteruptive arcade. The considerably smaller width and extent of the dimmings in the 304-Å line, as well as their locations, suggest that the transition region is reached by the deepest, central regions of the coronal dimmings. Moreover, some dimming fragments that are not visible in the coronal lines are revealed in the 304 Å line, such as the branch 2-10 that originates from point 2 toward the southwest. It will become clear below that, with some variations, such situations are also typical of other eruptive events.

Note that, as seen from the rotated difference heliograms taken at even longer time intervals, the twin dimmings described above and the brightening along the southern boundary of the northern polar coronal hole recorded in the three coronal lines and the transition-region line were observed over at least 15 hours, until 19 UT, and largely preserved their spatial structure during this time.

3.2. Event of January 18, 2000

In contrast to the preceding event, this eruptive event occurred in a fairly complex global solar magnetosphere, when several active centers, filaments, and coronal holes were observed on the disk (Figs. 2a, 2b). In particular, it included a prolonged class 1N/M3.9 flare with its maximum at 17:27 UT in the southeastern active region AR 8831 (S19 E11; region 1 in Fig. 2) and a well-defined halo CME, which initially developed over the eastern limb and subsequently extended over all position angles. Note that the eruptive southeastern prominence seen in the 304 Å pre-event heliogram (Fig. 2a) and its counterpart in the 304 and 171 Å difference heliograms (Figs. 2c, 2d) are associated with the preceding near-limb activity.

As is shown in [26] and can be seen from the six-hour difference image in the 195 Å line (Fig. 2e), the corresponding dimmings displayed an anisotropic structure in the form of extended, channeled transequatorial features. An annular dimming is visible in the immediate vicinity of the eruption center and posteruptive arcade 1, whose deepest parts are the northern and southeastern elements 2 and 3. The latter feature is connected with active center 4, located in the northern half of the disk near the central meridian, by a pronounced, narrow transequatorial dimming strip, and has a westward extension 4-5. The other eastern transequatorial channeled dimming 6-7, which is fragmentary in places and less intense,

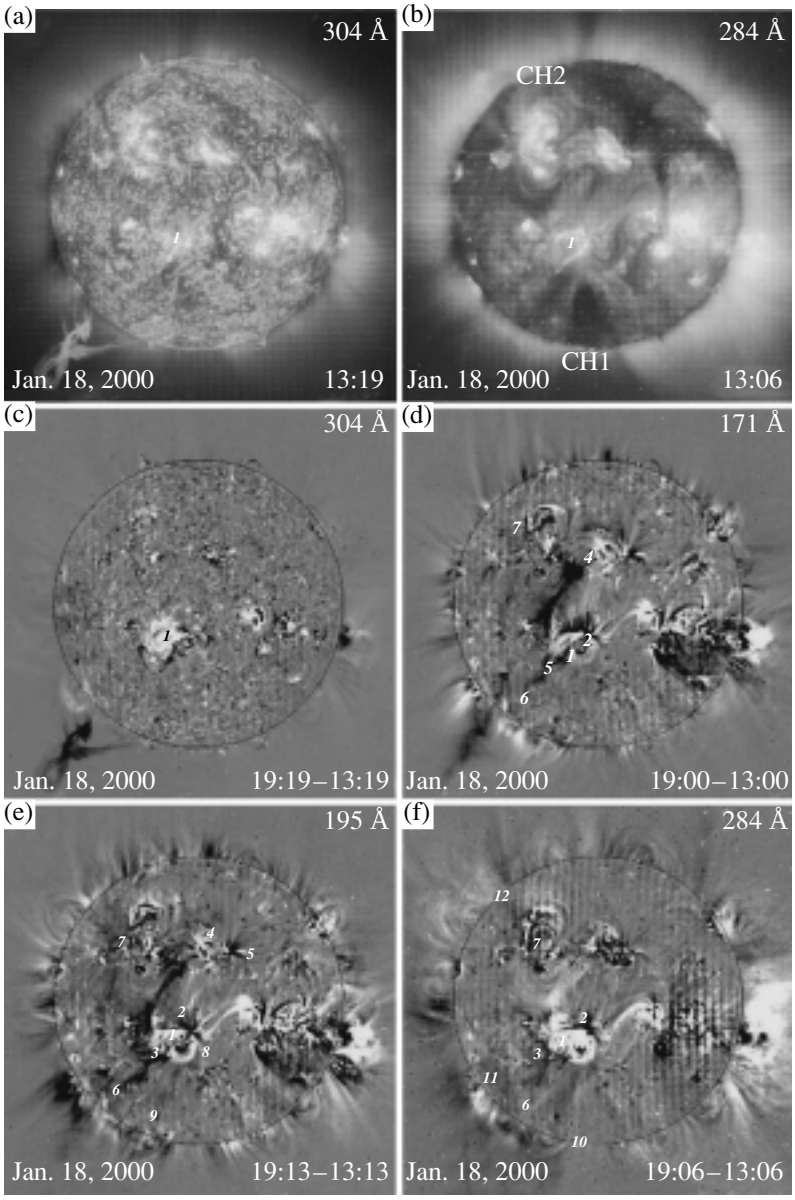

Fig. 2. Same as Fig. 1 for the event of January 18, 2000.

emerges from the southeastern dimming loop 3-6 and stretches toward the active region in the northeastern quadrant. Some of the other dimmings described in [26] are also visible in this difference image, in particular, the relatively faint dimming 8-9, which partially borders the edge of a southern polar coronal hole observed, e.g., in the 284 Å line (Fig. 2b). The relationship between this eruptive event and the dimmings observed in the southwestern quadrant remains unclear.

The main dimmings, including elements 2 and 3 joining to the eruption center, and both channeled transequatorial structures, 6-3-4 and 6-7, are also observed in the lower-temperature 171 Å line (Fig. 2d).

However, substantial differences are visible in the highest-temperature 284 Å line (Fig. 2d), which are important for our understanding of the origin of dimmings. In this case, they are mainly associated with the transequatorial dimmings noted above. The eastern dimming branch that goes toward the northeastern active complex 7 is almost indistinguishable, while an appreciable brightening is present in the transequatorial structure 1-4 between the eruption center and the northern active region, instead of a deep dimming. The brightening also affects the entire equator-facing loop system, which connects active

centers 4 and 7 of the northern hemisphere; i.e., a considerable portion of the space between the eastern and western transequatorial dimming branches seen at 171 and 195 Å. At the same time, in the 284 Å line, the deepest dimmings 2 and 3 adjoining to the eruption center and the southeastern dimming loop 3-6 are similar to those observed at 171 and 195 Å. In addition, at 284 Å, there is some intensity decrease (i.e., a weak dimming) localized generally within the southeastern sector 1-10-11-1, although the eastern edge of the sector, which coincides with the eastern boundary of the southern polar coronal hole CH1, seems to be slightly brightened, as are some other narrow structures inside the sector. The dimming 7-12, which stretches toward the southeastern limb along the narrow CH2, is more pronounced at 284 Å than in the other lines.

Transition-region manifestations of dimmings in the 304 Å line are not significant in this event, as can be seen from the corresponding difference heliogram (Fig. 2c). Only dimming fragments with relatively small areas can be distinguished in the coronal lines around the eruption center 1. In particular, there are well-defined dimming fragments and narrow loop structures east, south, and west of the eruption center.

3.3. Event of February 17, 2000

This event occurred one rotation after the preceding one (Section 3.2) at nearly the same site on the solar surface; i.e., in the presence of a similarly complex global solar magnetosphere. It developed against the background of two sympathetic flares of classes M2.5/1B and M1.3/2N that occurred in the nearby southern active regions AR 8869 and AR 8872 (S25 W16 and S29 E07; regions 1 and 2 in Fig. 3) and reached their maxima in soft X-rays at 18:52 and 20:35 UT, respectively [40]. According to the SOHO/LASCO coronagraph data, the first signs of a CME were observed over the southwestern limb at 19:31 UT, and an almost symmetric halo CME developed around the entire occulting disk after 20:06 UT.

Since the event developed most intensely near 19 UT, the frames taken at that time cannot be used to construct the difference images in the four lines. For this reason, the rotated difference heliograms shown in Figs. 3c–3f were obtained using a twelve-hour interval, as the difference between the frames of 13 UT on February 17 and 1 UT on February 18, after they were first rotated to 19 UT on February 17 at each wavelength. Analysis of these heliograms confirms the conclusion drawn earlier based on detailed data in the 195 Å line [26] that the dimmings that accompanied this event were global, mainly channeled, reflected the magnetic coupling between all the active regions situated in the central sector of the disk on both sides of the equator, and covered the entire space between these regions (see also [40]).

This dimming structure is also visible in the twelve-hour 195 Å difference heliogram (Fig. 3e). Some dimmings here abut on the western eruption center 1. One or two dimming strips 3-4 and the dimming 5 to the north of these emerge from this center toward the west. Two dimming arcs branch at point 3, curving around the eruption center 1 from the south and stretching toward the eastern dimming 6. In turn, this last dimming is connected with the eastern eruption center 2 by the narrow channel 6-7. This center is surrounded by the dimming 7-8 from the north and the well-defined dimming loop 9 from the south. The transequatorial dimming channel 7-10-11 goes from the eruption center 2 through the knot 10 toward the northern active region. The connection between these regions and those to the east of the central meridian is also indicated by a number of dimming channels, such as the southern horizontal strips 8-12 and 10-13-14-15, the northern horizontal strip 16-17-11, and the vertical dimmings 13-17 and 14-16 connecting these strips. The northern horizontal strip 16-17-11 extends in the west to the northwestern region 18—the neighborhood of a small coronal hole. Relatively faint but clearly visible transequatorial dimmings directly connect the western eruption center 1 with the northern region 11 and the northwestern region 18. On the whole, the global system of channeled dimmings described here reflects the fact that the magnetic structures of the entire circular central sector of the disk were involved in the eruption process. As is noted in [26], the positions and configurations of most channeled dimmings correspond to elements of the annular system of emitting UV chains that was visible in the 195 Å line before the event in the same central sector of the disk. It is probably the eruption of precisely these magnetic structures that determined the quasi-symmetric annular shape of the halo CME observed in this event.

A very similar dimming system can also be seen in the corresponding 171 Å heliogram (Fig. 3d). One small difference is the somewhat weaker manifestation of the dimmings between the western eruption center 1 and the northern regions 11 and 18. On the other hand, the knot 13 and the main eastern and northeastern channels aligned with the curve 15-14-16-17-11 are more pronounced.

A different situation with regard to dimmings is observed in the high-temperature 284 Å line. The corresponding difference heliogram (Fig. 3f) displays all the main dimmings adjoining to the eruption

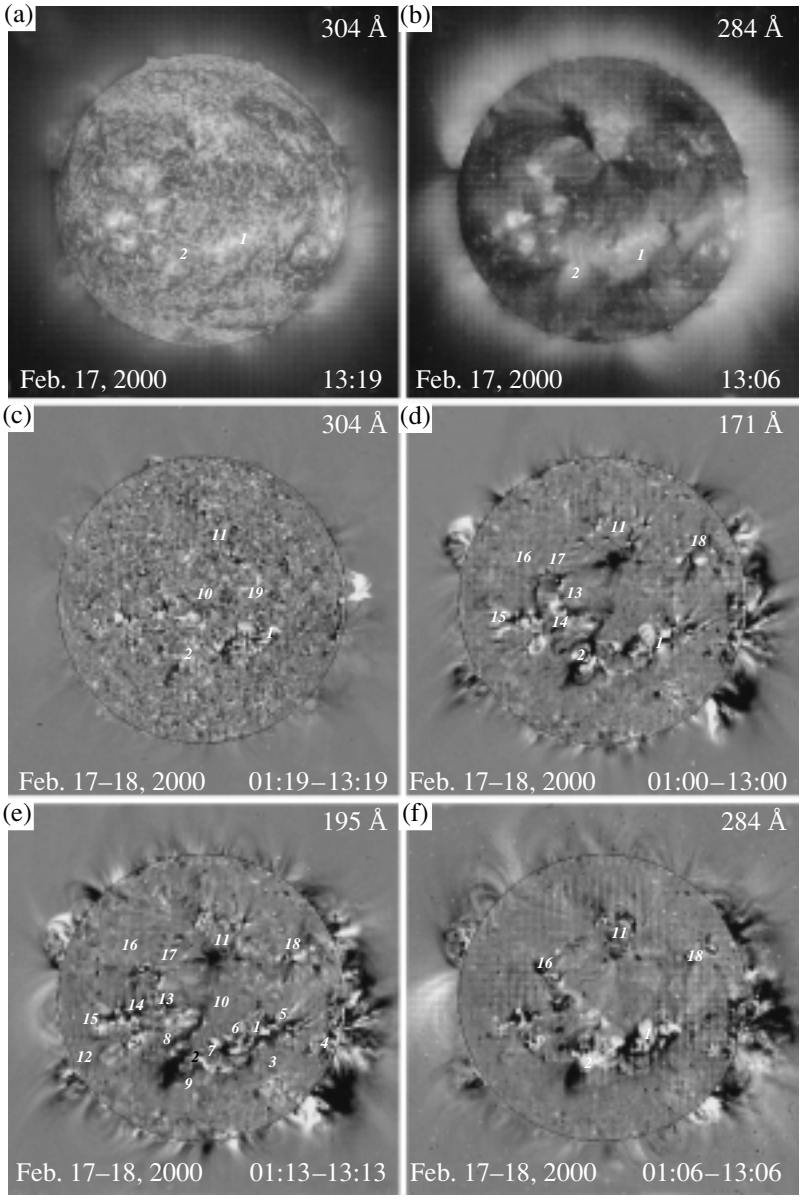

Fig. 3. Same as Fig. 1 for the event of February 17, 2000. The rotated difference images (c)–(f) were obtained over the twelve-hour interval between 13 UT and 1 UT on February 18, 2000; all the heliograms were rotated to 19 UT on February 17, 2000.

centers 1 and 2, and, as in the other two coronal lines, the pattern of the channeled dimmings generally reflects the connections between the active regions in the central sector of the disk. In this heliogram, some eastern transequatorial dimmings are less pronounced, while western dimmings, e.g., in the strip 1-18, are slightly more intense. However, no analogs of the dimming knot 10 and the entire transequatorial dimming between the eastern eruption center 2 and the northern active region 11 are observed at 284 Å (Figs. 3d, 3e). On the contrary, a weak brightening is present instead of the dimming, while a slight intensity decrease is observed west of this structure, up to the strip 1-18. Another brighter transequatorial structure stretches from the zone between the eruption centers 1 and 2 toward the northwestern dimming region 16.

As in the preceding events, fragments of the deepest dimmings are mainly visible in the 304 Å transition-region line (Fig. 3c), in particular, near the eruption centers 1 and 2. The dimming knot 10 on the line between regions 2 and 11 also stands out. It is noteworthy that a new dimming knot, 19, which is not observed in the coronal lines, appears slightly to the west and north of the knot 10. It is possible to see on a computer screen and in high-quality prints

of the heliogram (Fig. 3c) that the knot 19 is linked by a weak dimming with the northern active region.

3.4. Event of April 29, 1998

This eruptive event included a long-lived flare of class 3B/M6.8 in the southeastern AR 8210 (S21 E03), which had its soft-X-ray maximum at 16:37 UT. It is noteworthy that the corresponding CME appeared as a two-component halo with a white-light maximum over the northeastern and southwestern sectors of the limb. During the event, AR 8210 (region 1 in Fig. 4) was essentially the only active region on the disk, but it was adjacent to a large transequatorial coronal hole, CH1 (Fig. 4b), which occupied nearly the entire central zone of the disk, and two other, much smaller, coronal holes CH2 and CH3 located between the eruption source 1 and the eastern limb. The unusual appearance of the CME [35] and the peculiar manifestations of the dimmings in the 304 Å line (see below) were apparently related to these features of the global solar magnetosphere.

Analysis of running-difference and fixed-difference images and movies obtained using the available 195-Å data in intervals of 12-16 min shows additional features of this event (see the web site noted at the beginning of Section 3). A well-defined coronal wave propagated from the eruption center in two angular sectors—toward the northeastern limb and westward; i.e., through CH1. (Such propagation of a coronal wave within a coronal hole is apparently rarely observed, and deserves special consideration in a separate study.) Like the first sector of the coronal wave, the main dimmings in 195 Å developed in the direction of the northeastern limb. This can also be seen in the rotated six-hour difference heliogram (Fig. 4e). The deep transequatorial dimming 2-3, which formed in the immediate vicinity of CH1, extends toward northern region 3 and joins with the eastern meridional dimming 4-5. The narrow dimming 6-7 curves around the eruption center 1 from the south; this dimming also has a number of even narrower extensions in the eastern and northeastern directions, in particular, toward point 4. Weak manifestations of dimmings and fragmentary brightenings are also observed along the southern (8-9) and northern (3-10) boundaries of CH1. A distinct but small intensity reduction also covers the entire space of CH1, and several arc-like channels stand out against this background, in particular, between points 6 and 9. A similar slight intensity reduction is probably present to the east of the dimming 7-4-5.

As usual, an overall similar dimming structure is also observed in the 171 Å line (Fig. 4d). Minor differences from the 195 Å pattern include the fact that dimming 2 is somewhat broadened in the longitudinal direction in the immediate vicinity of the eruption center 1, and only individual fragments replace the eastern meridional dimming 4-5 (Fig. 4e), especially its northern part. In addition, the small-area dimmings 11 near the northern boundary of CH1 and some dimmings within CH1, including the arc-like structure 2–9, are enhanced.

In this event, the main dimming 2 and its transequatorial extension toward the northeastern region 3 are also present, although somewhat weaker, in the difference heliogram taken in the high-temperature 284 Å coronal line (Fig. 4f). The Yohkoh/SXT data confirm the tendency for the dimming 2-3 to weaken with increasing temperature: as shown in [35], dimming 2 is almost invisible in soft X-rays, i.e., at even higher plasma temperatures, $T_e > 2.5$ MK. The eastern meridional dimming 4-5, which is most pronounced in the 195 Å line (Fig. 4a) is likewise only weakly manifest at 284 Å. At the same time, the dimming 6-7-4 passing to the south and southeast of the edge of the posteruptive arcade 1 is visible as clearly as in the other two coronal lines. Intensity variations corresponding to CH1 can also be seen in 284 Å. The faint chains 3-10 are observed along the northern boundary of CH1, which adjoins on narrow, faint dimming structures. Similar structures are also localized within CH1 (6-9) and along its western boundary (9-10).

In contrast to the three events described above, in this case, the largest-area dimmings can be distinguished in the difference heliogram in the 304 Å transition-region line (Fig. 4c). Not only are analogs of the northern (2) and southern (6-7) coronal dimmings clearly visible near the eruption center 1, but an intensity decrease with appreciable amplitude and area (i.e., a transient coronal hole) can also be seen in the southern part of CH1. A deep dimming starts from the broadened eastern end (6) of the southern dimming, covers an extensive portion of CH1 stretching westward to point 12, then bends and extends toward the southeast to point 13, located near the southern boundary of CH1. This seems to imply that the CME-induced restructuring of the magnetic field (opening of the field lines) has also affected the underlying transition-region plasma in the southern portion of CH1 and in the space between the eruption center 1 and CH1. As can be seen in Fig. 4c, this process was also accompanied by the formation of a emitting chain along the southern boundary of dimming 6. A similar emitting structure is also present at the eastern edge of dimming 2.

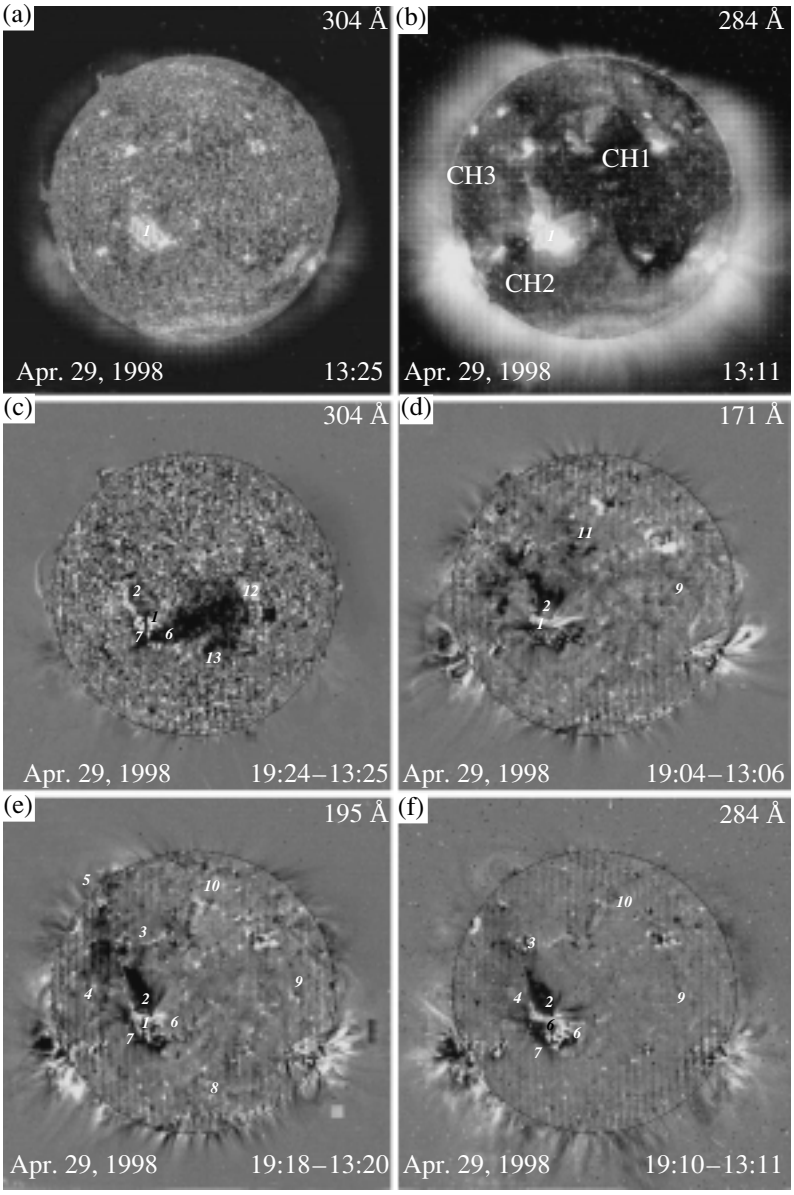

Fig. 4. Same as Fig. 1 for the event of April 29, 1998.

4. DISCUSSION AND CONCLUSION

The reduction of SOHO/EIT UV heliograms with compensation for the rotation of the solar disk has enabled us to construct rotated difference images over six-hour (and 12-hour, when required) intervals, and to study CME-produced dimmings (transient coronal holes) simultaneously in three coronal lines (171, 195, and 284 Å) and a transition-region line (304 Å). This was also possible because the lifetimes of major dimmings normally exceed these time intervals.

Our analysis provides additional evidence that, at least in the three coronal lines, the dimming disturbances and transient brightenings produced by major CMEs are essentially global and span the space between well-separated activity centers that may be located on both sides of the heliographic equator. This means that a large portion of the global solar magnetosphere is involved in the eruption process. Structures directly linked to the eruption center by magnetic fields appear to be primarily involved in this process [41].

The data obtained in all four lines also confirm the conclusion obtained earlier based on a detailed analysis of 195 Å data [26] that, when the global solar magnetosphere is complex and several active-region filaments, coronal holes, etc. are present on the solar disk, many dimmings are anisotropic, and can be observed as narrow, extended channels.

In many cases, the reduction of the EUV intensity of the dimmings in all four lines is comparable to that at 195 Å and, according to the estimates of [3, 9, 26], can reach several tens of percent.

Well-defined dimmings adjacent to the eruption center are recorded in all four lines. They have nearly the same shapes and locations in the three coronal lines, while only the deepest fragments of such dimmings can normally be seen in the transition-region line, rather than the entire coronal dimming area. This, firstly, supports the previous conclusion of [3, 8, 9] that such dimmings result from a CME-related partial or complete opening of field lines in the corresponding magnetic structures and the evacuation of plasma from them, and, secondly, suggests that this process involves not only the coronal plasma but, to some extent, the transition-region plasma as well. We should bear in mind that the fragmentary dimmings observed in the immediate neighborhood of eruption centers (or active regions) at 304 Å may be partially due to a reduction in the emission of the SiXI coronal line, which is also recorded by the 304 Å filter [13, 30].

Channeled dimmings, especially transequatorial ones, are manifested most clearly and in very similar ways in the 171 and 195 Å coronal lines, which have moderate excitation temperatures. The events of January 18 and February 17–18, 2000 (Sections 3.2 and 3.3) show that a brightening in pre-existing loop structures is observed at the locations of some transequatorial dimmings in the high-temperature 284 Å line. This suggests that the reduction in EUV intensity of such structures in some lines and its enhancement in others may be initiated by changes in the plasma temperature—for example, by heating due to CME-related thermal or MHD disturbances. The above considerations do not mean that transequatorial dimmings cannot be observed at all in the high-temperature UV line, or even in soft X-rays. Examples of transequatorial dimmings in the 284 Å line were indicated in our description of the events of February 17–18, 2000 (Section 3.3), and April 29, 1998 (Section 3.4), and examples of observations of such dimmings in soft X-rays are given in [21, 42] (based on Yohkoh/SXT data). The analysis of numerous transequatorial loops presented in [43] shows that CME-related variations in the EUV (195 Å) and soft-X-ray intensities can be manifested as dimmings in some cases and as loop brightenings in others.

Our analysis of heliograms in the 304 Å transition-region line reveals dimmings invisible in the coronal lines that appear far from the active region. In some cases (the event of February 17–18, 2000; Section 3.3), these are dimming elements with relatively small areas and intensity reductions, while in others (the event of April 29, 1998; Section 3.4), they are clearly defined dimmings occupying a considerable portion of a large coronal hole adjacent to the eruption center. Such dimmings can arise, in particular, as a result of the heating of the transition-region plasma (sinking of the corona) during a CME. Another possibility is the opening of field lines and outflow of material from originally closed, low-lying structures in the transition region. In the case of the event of April 29, 1998, this second possibility implies that the transition-region plasma is directly involved in the eruption process, so that a continuation of a substantial portion of the existing coronal hole is formed at transition-region altitudes; i.e., a transient "hole" develops in the transition region.

As noted in Section 3, brightenings in the form of elongated emitting chains are observed during some events after a CME, at the boundaries of some dimmings or coronal holes in all four lines [8, 39]. Their origin is most likely related to the energy released due to reconnection during the interaction of evolving open magnetic structures involved in the CME with neighboring closed magnetic fields.

In conclusion, we emphasize again that our analysis of dimmings was based on rotated six-hour and 12-hour SOHO/EIT difference heliograms, obviously, this should be considered as the first step only. A more detailed analysis of CME-related disturbances (especially coronal waves) in various spectral lines requires much higher time resolution, reaching tens of seconds or several minutes. Such observations—in the same four lines—are being conducted on the TRACE (which has a restricted field of view) [31, 33] and CORONAS F (which can obtain full-disk images) [44, 45] space observatories.

ACKNOWLEDGMENTS

The authors are grateful to the SOHO/EIT team for the data used in our analysis. This work was supported by the Russian Foundation for Basic Research (project nos. 03-02-16049 and 03-02-16591) and partially by the Ministry of Industry, Science, and Technology of the Russian Federation.

REFERENCES

1. D. F. Webb, J. Atmos. Sol.-Terr. Phys. **62**, 1415 (2000).
2. A. C. Sterling, J. Atmos. Sol.-Terr. Phys. **62**, 1427 (2000).
3. N. Gopalswamy and B. J. Thompson, J. Atmos. Sol.-Terr. Phys. **62**, 1458 (2000).
4. H. S. Hudson and E. W. Cliver, J. Geophys. Res. **106**, 25199 (2001).
5. D. M. Rust, Space Sci. Rev. **34**, 21 (1983).

6. A. C. Sterling and H. S. Hudson, Astrophys. J. Lett. **491**, L55 (1997).
7. H. S. Hudson and D. F. Webb, in *Coronal Mass Ejections,* Ed. by N. Crooker, J. Joselyn, and J. Feynman (1997), AGU Geophys. Monogr. Ser., No. 99, p. 27.
8. B. J. Thompson, S. P. Plunkett, J. B. Gurman, *et al.*, Geophys. Res. Lett. **25**, 2465 (1998).
9. D. M. Zarro, A. C. Sterling, B. J. Thompson, *et al.*, Astrophys. J. Lett. **520**, L139 (1999).
10. L. K. Harra and A. C. Sterling, Astrophys. J. Lett. **561**, L215 (2001).
11. S. W. Kahler and H. S. Hudson, J. Geophys. Res. **106**, 29239 (2001).
12. S. Tsuneta, L. Acton, M. Bruner, *et al.*, Solar Phys. **136**, 37 (1991).
13. J.-P. Delaboudiniere, G. E. Artzner, J. Brunaud, *et al.*, Solar Phys. **162**, 291 (1995).
14. R. A. Harrison, E. C. Sawyer, and M. K. Carter, Solar Phys. **162**, 233 (1995).
15. B. J. Thompson, J. B. Gurman, W. M. Neupert, *et al.*, Astrophys. J. Lett. **517**, L151 (1999).
16. A. Klassen, H. Aurass, G. Mann, *et al.*, Astron. Astrophys. **141**, 357 (2000).
17. B. J. Thompson, B. Reynolds, H. Aurass, *et al.*, Solar Phys. **193**, 161 (2000).
18. A. Warmuth, B. Vrsnak, H. Aurass, *et al.*, Astrophys. J. Lett. **560**, L105 (2001).
19. D. A. Biesecker, D. C. Myers, B. J. Thompson, *et al.*, Astrophys. J. **569**, 1009 (2002).
20. N. Narukage, H. S. Hudson, T. Morimoto, *et al.*, Astrophys. J. Lett. **572**, L109 (2002).
21. J. L. Khan and H. Aurass, Astron. Astrophys. **383**, 1018 (2002).
22. H. S. Hudson, J. L. Khan, J. R. Lemen, *et al.*, Solar Phys. **212**, 121 (2003).
23. G. E. Moreton and H. E. Ramsey, Publ. Astron. Soc. Pac. **72**, 357 (1960).
24. C. Delannee and G. Aulanier, Solar Phys. **190**, 107 (1999).
25. C. Delannee, Astrophys. J. **545**, 512 (2001).
26. I. M. Chertok and V. V. Grechnev, Astron. Zh. **80**, 162 (2003) [Astron. Rep. **47**, 139 (2003)].
27. D. F. Webb, R. P. Lepping, L. Burlaga, *et al.*, J. Geophys. Res. **105**, 27251 (2000).
28. I. M. Chertok, in *Proceedings of the Fifth SOHO Workshop*; ESA SP **404**, 269 (1997).
29. S. Pohjolainen, D. Maia, M. Pick, *et al.*, Astrophys. J. **556**, 421 (2001).
30. D. Moses, F. Clette, J.-P. Delaboudinire, *et al.*, Solar Phys. **175**, 571 (1997).
31. B. N. Handy, L. W. Acton, C. C. Kankelborg, *et al.*, Solar Phys. **187**, 229 (1999).
32. M. J. Wills-Davey and B. J. Thompson, Solar Phys. **190**, 467 (1999).
33. L. K. Harra and A. C. Sterling, Astrophys. J. **587**, 429 (2003).
34. A. C. Sterling, H. S. Hudson, B. J. Thompson, *et al.*, Astrophys. J. **532**, 628 (2000).
35. H. Wang, P. R. Goode, C. Denker, *et al.*, Astrophys. J. **536**, 971 (2000).
36. C. Delannee, J. Atmos. Sol.-Terr. Phys. **62**, 1471 (2000).
37. G. E. Brueckner, R. A. Howard, M. J. Koomen, *et al.*, Solar Phys. **162**, 357 (1995).
38. S. P. Plunkett, B. J. Thompson, R. A. Howard, *et al.*, Geophys. Res. Lett. **25**, 2477 (1998).
39. I. M. Chertok, Solar Phys. **198**, 367 (2001).
40. H. Wang, V. Yurchyshyn, J. Chae, *et al.*, Astrophys. J. **559**, 1171 (2001).
41. T. Wang, Y. Yan, J. Wang, *et al.*, Astrophys. J. **572**, 580 (2002).
42. J. I. Khan and H. S. Hudson, Geophys. Res. Lett. **27**, 1083 (2000).
43. A. Glover, L. K. Harra, S. A. Matthews, and C. A. Foley, Astron. Astrophys. **400**, 759 (2003).
44. V. N. Oraevsky and I. I. Sobelman, Pis'ma Astron. Zh. **28**, 457 (2002) [Astron. Lett. **28**, 401 (2002)].
45. I. A. Zhitnik, O. I. Bougaenko, J.-P. Delaboudinière, *et al.*, in *Proceedings of 10th European Solar Physics Meeting* (2002) ESA SP-506, p. 915.

Translated by A. Getling

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 946–952. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 1026–1033.

Quasi-Periodic Line-of-Sight Velocity Variations at the Bases of Polar Coronal Holes

N. I. Kobanov and D. V. Makarchik

Institute for Solar and Terrestrial Physics, P.O. Box 4026, Irkutsk, 664033 Russia

Received March 5, 2003; in final form, May 8, 2003

Abstract—Spatial and temporal variations in the line-of-sight velocities and brightnesses measured in the Hα and FeI 6564 Å Hβ and FeI 4864 Å NiI 4857 Å lines at the bases of polar coronal holes are analyzed. Time series with durations of 43–120 min were recorded using a CCD strip (3700 pixels 200×7 μm in size) and a CCD array (256×1024 pixels 24 μm in size). Quasi-stationary upward flows (with radial velocities reaching 3 km/s in the photosphere and 12–15 km/s in the chromosphere) were observed near dark points at the boundaries of the chromospheric network. The acoustic 3-min and 5-min oscillations are amplified in the coronal hole, and reach 1 km/s in the photosphere and 3–4 km/s in the chromosphere. The spectra of fluctuations of the line-of-sight velocity exhibit significant maxima at low frequencies, clustering near 0.4, 0.75, and 1 mHz. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

It is known that the solar wind "blows" from coronal holes, whose localization on the disk indicates the presence of open magnetic configurations [1–3]. The characteristics of motion in coronal holes have been studied in detail in the past several years [4, 5]. The SOHO satellite has enabled us to obtain the pattern of motions in the inner corona and transition region, including the first data on oscillatory and wave processes in these regions [6, 7]. On the other hand, motions and oscillations at the bases of coronal holes, at the lower chromospheric and photospheric levels, remain poorly understood. The main aim of the present work is to fill this gap by analyzing several polar coronal holes using modern equipment and new measurement methods [8, 9].

2. EQUIPMENT AND METHOD

The observations were carried out using the horizontal solar telescope of the Sayany Observatory, equipped with a photoelectric system capable of scanning and guiding the image with an accuracy up to 1″. We used a Toshiba CCD strip (3700 pixels 200×7 μm in size) and a Princeton Instruments RTE/CCD-256H CCD array (256×1024 pixels 24 μm in size). The array is equipped with a thermoelectric cooler and controller, providing automatic control of temperature in the range +30°C to −45°C. To reduce undesired illumination during recording, the light ray is interrupted by an electromechanical shutter supplied with the array. We used WinSpec 32 software to control the entire process of observing and recording the information from the CCD array. The operation of the CCD strip was controlled by a software package developed at the Institute for Solar and Terrestrial Physics. We developed an unmodulated measurement method in order to simultaneously measure the line-of-sight velocities at several heights and the longitudinal magnetic field using these two photodetectors [8, 9].

The observations using the CCD strip were carried out primarily in the Hβ 4861 Å, NiI 4857 Å, and FeI 4864 Å lines. The CCD array was able to detect motions in the lower chromosphere in the Hα line, while the photospheric velocities and the magnetic-field intensity were measured in the FeI 6569 Å line, where the Landé factor was ~ 1.4.

3. OBSERVATIONAL MATERIAL AND DATA REDUCTION

We shall analyze data for several polar coronal holes, whose characteristics are briefly summarized in the table. The observed region was usually located near the central meridian, 50″–100″ from the pole. In measurements using the CCD array, the image was usually rotated by a Dove prism so that the spectrograph slit was parallel to the limb.

The size of the observed region was determined by the entrance slit width and the height of the array along the slit, which were $1.5'' \times 60''$. Since the spatial resolution over one-hour intervals was usually not better than 1″ due to atmospheric effects, the excess spatial resolution of the array (with 256 pixels

1063-7729/03/4711-0946$24.00

Summary of observational data

No.	Position of the region on disk	Date and time of observations (UT)	Photodetector	Spectral lines	Exposure/storage time, s	Number of frames/minute
1	~ 50″ from S pole	August 2, 2002, 10:02–10:46	CCD array	Hα, FeI 6569 Å	0.3	24
2	~ 67°N	September 14, 2002, 02:08–02:52	CCD array	Hα, FeI 6569 Å	5	6
3	~ 270″ from S pole ~ 100″ to W	May 31, 2001, 02:51–04:16	CCD strip	Hβ, FeI 4864 Å	10	6
4	~ 100″ from S pole	May 31, 2001, 04:21–05:46	CCD strip	Hβ, FeI 4864 Å	10	6
5	~ 50″ from S-pole	June 1, 2001, 02:01–03:33	CCD strip	Hβ, NiI 4857 Å	10	6
6	~ 50″ from S-pole	June 1, 2001, 04:12–05:37	CCD strip	Hα, FeI 6569 Å	10	6
7	~ 70°N	June 2, 2001, 05:00–06:25	CCD strip	Hα, FeI 6569 Å	10	6

covering 60″) was reduced by averaging the signals over four neighboring pixels. As a result, the final spatial resolution along the slit became only 1″, in agreement with the actual resolution. Each observational series represents a sequence of frames with the same exposure time taken at equal time intervals. The special control software enabled us to record only the required parts of the spectrum, rather than the entire spectrum covered by the sensitive surface of the array. This helped us to reduce considerably the digital size of each frame and, consequently, the volume of an entire time series, which was recorded in a single file about 120–130 MB in size, on average.

4. RESULTS

Even a cursory visual examination of the obtained data reveals some structures in the polar coronal holes at the lower chromospheric level (Fig. 1). The sharp boundary between the left- and right-hand sides of the frame in Fig. 1a is due to the considerable extent of the Hα wings.

The line-of-sight velocities in these structures were quite large. An example of the velocity distributions along the spectrograph slit in the chromosphere and photosphere is given in Fig. 1b, which represents the data for frame No. 51 from time series No. 1, recorded on August 2, 2002. The line-of-sight velocity in the Hα line was measured at three height levels: 0.3, 0.5, and 0.7 Å, and reaches values of 3–4 km/s in the chromosphere and 1 km/s in the photosphere. If the motion is assumed to be purely radial, the radial velocity corrected for projection onto the line of sight will be 9–12 km/s. These values are close to the estimates of the upward flows in the vicinity of dark points presented by Kozlova and Somov [5], but the spatial dimensions of our structures are substantially smaller, ~ 4″–7″. These structures should probably be identified with the rosettes of fibrils forming the boundaries of the chromospheric network that turned out to be oriented along the slit. We use the following notation to label the axes in all the figures: v is the line-of-sight velocity, P the spectral power, t the time in minutes, f the frequency in mHz, and L the length of the entrance slit in arcsec. In all diagrams, positive and negative velocities correspond to the directions toward and away from the observer. Positive and negative velocities are represented in the gray-scale diagrams by light and dark regions.

A comparison of the patterns shown by the line-of-sight velocity and the corresponding brightness distributions indicates that upward flows are localized near dark structures observed in the Hα line (Fig. 2). The time variations of the line-of-sight velocity and brightness along the slit for series No. 1 and No. 2 are presented in Fig. 3. Short-period oscillations in the Hα emission that correspond to the 3-min chromospheric oscillations are clearly visible.

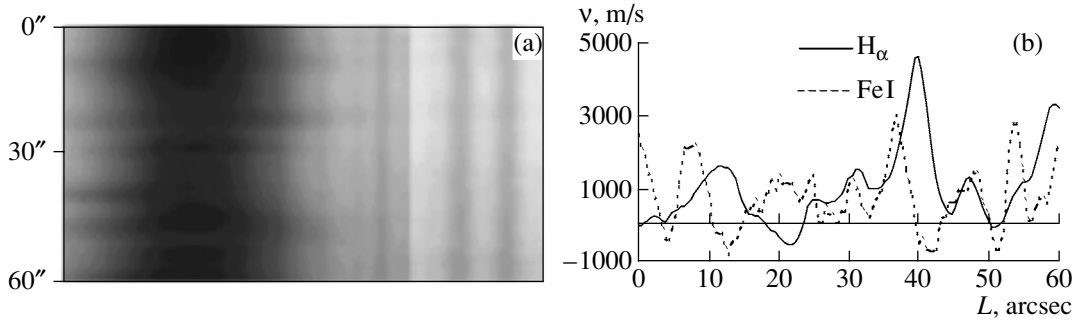

Fig. 1. Chromospheric structures in a polar coronal hole: (a) frame No. 51 from time series No. 1 and (b) line-of-sight velocity along the slit (for convenience, the photospheric velocities have been increased by a factor of four).

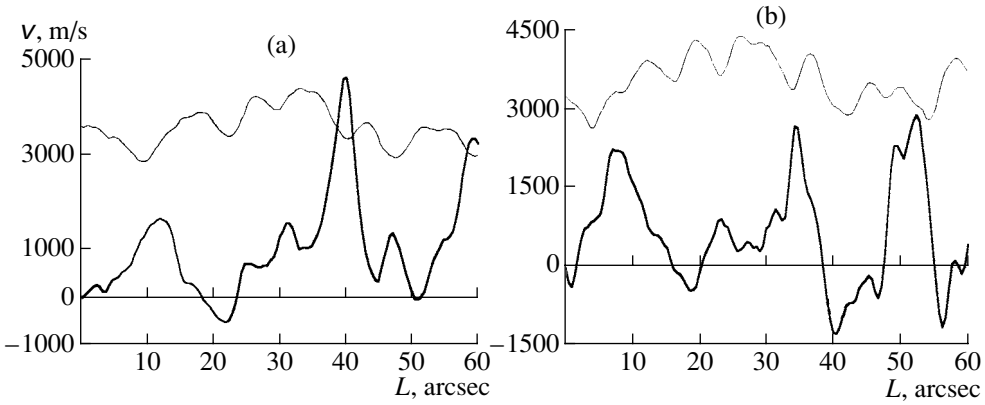

Fig. 2. Examples of correspondences between upward chromospheric flows and regions of decreased Hα brightness. The thick curve is the line-of-sight velocity and the thin curve is the brightness in arbitrary units.

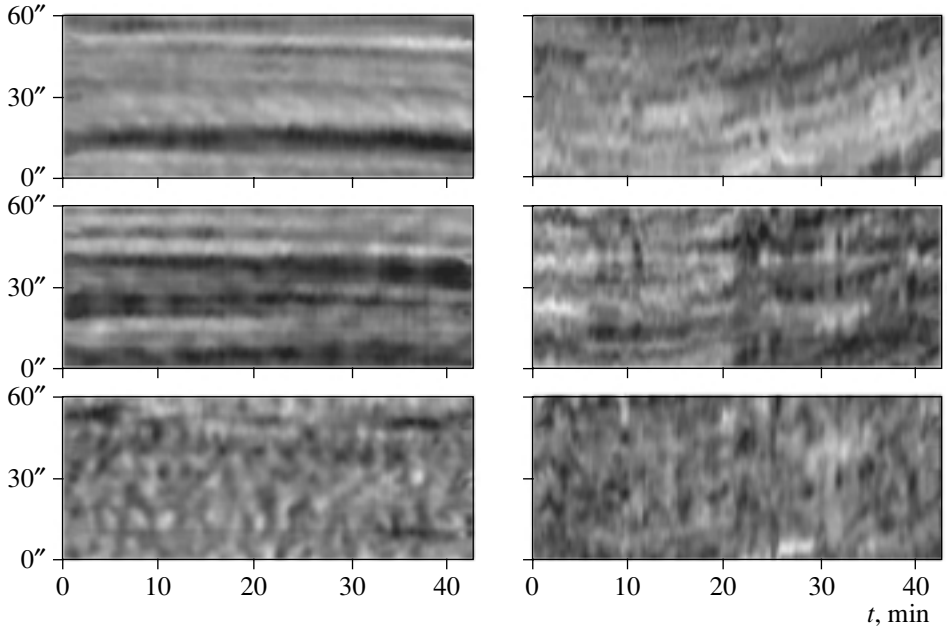

Fig. 3. Spatial and temporal variations in the Hα brightness (upper row) and the photospheric (middle row) and chromospheric (lower row) line-of-sight velocities for time series No. 1 (right column) and No. 2 (left column). The angular coordinate is plotted along the vertical axis and time is plotted along the horizontal axis.

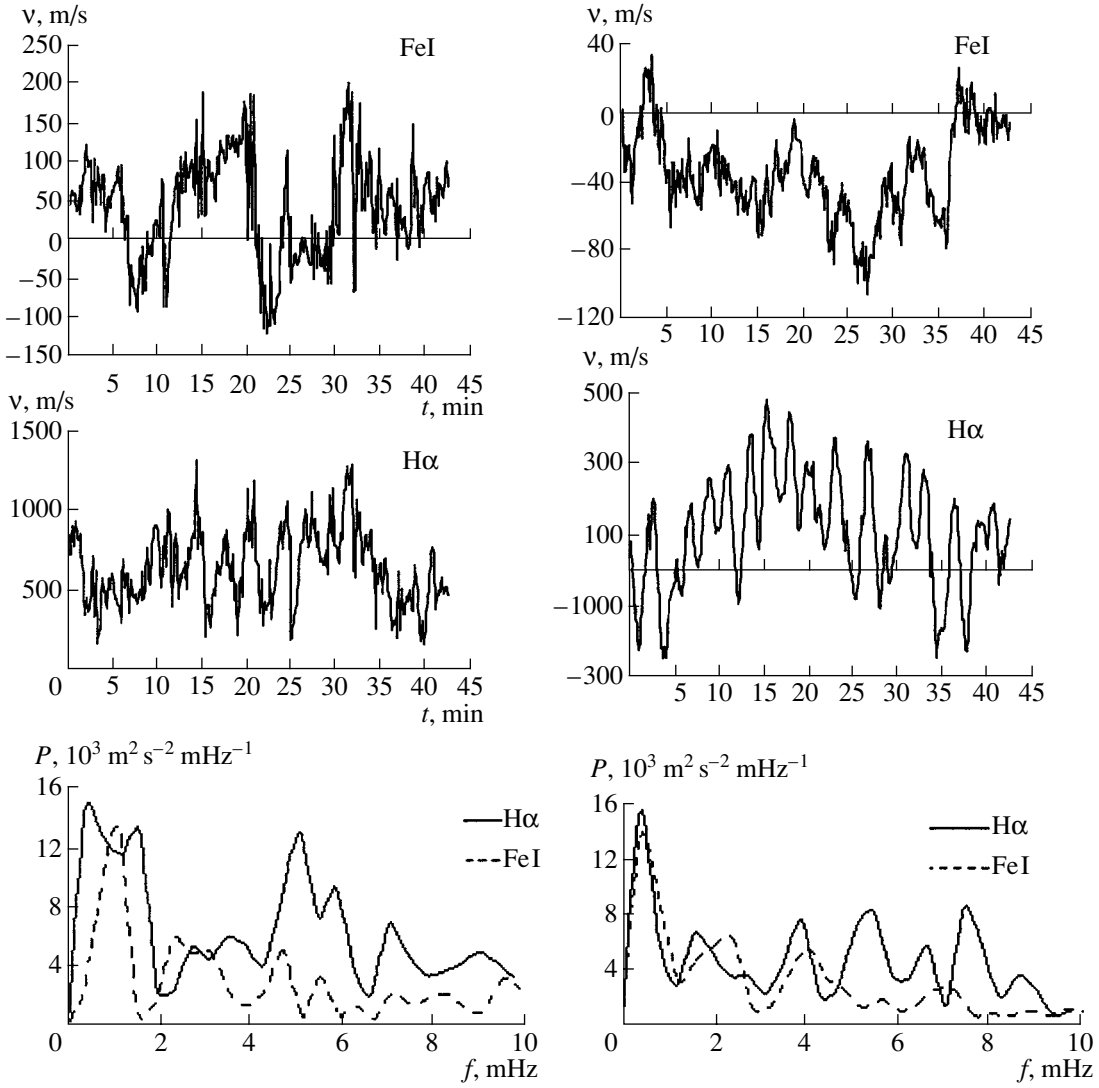

Fig. 4. Line-of-sight velocities averaged over the array height and the corresponding power spectra for time series No. 1 (left column, values of the photospheric spectrum have been increased by a factor of two) and No. 2 (right column, values of the photospheric spectrum have been increased by a factor of four).

It is interesting that the maximum at 0.39 mHz dominates in the spectra of time series No. 2 integrated over the entire slit width (right-hand plots in Fig. 4). A peak corresponding to 4 mHz is common. The period 10 min is also significant in the chromosphere, while a shorter period, 7 min, was detected in the photosphere. In general, the chromospheric spectrum is characterized by a finer structure at high frequencies than is the photospheric spectrum. The same spectra for time series No. 1 indicate periods of 20 min in the chromosphere and 15 and 7 min in the photosphere, along with the 3-min and 5-min oscillations (left-hand plots in Fig. 4). The time series in which each point represents a mean line-of-sight velocity for the entire slit shows clearly both low-frequency (15–20 min) and high-frequency (3–5 min) variations, which are very typical for the unperturbed atmosphere at the disk center. Despite the considerable averaging over the spatial coordinate ($\sim 60''$), the amplitude of these variations is about 100 m/s in the photosphere and 400 m/s in the chromosphere. After correcting for projection, the mean line-of-sight velocities should be 300 and 1200 m/s, respectively. Consequently, the corresponding oscillations cover large areas of the coronal holes.

There are no appreciable differences in the spectral components of the velocity oscillations at the height scales covered by the Hα contour. This is illustrated in Fig. 5a, which presents power spectra of the line-of-sight velocity variations measured in the Hα line at levels of ± 0.3 Å and ± 0.7 Å. The spectra of the velocity and brightness variations in different regions of the

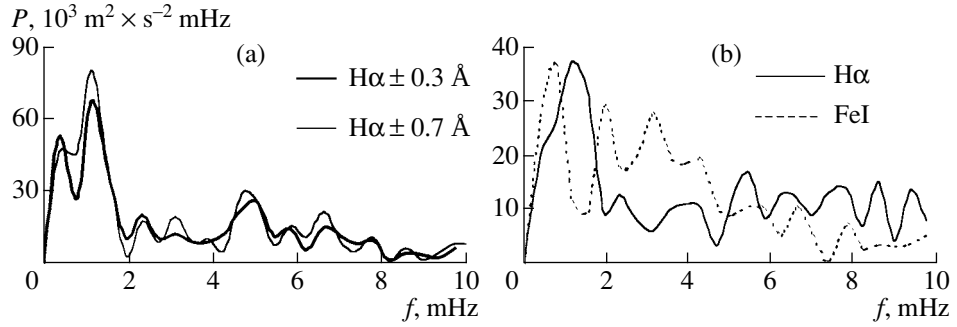

Fig. 5. Power spectra of the line-of-sight velocity for various spatial elements of time series No. 2: (a) spectra of chromospheric oscillations at two heights in a 54″ element, and (b) spectra of photospheric and chromospheric oscillations in a 38″ element. The values of the photospheric spectrum have been increased by a factor of five.

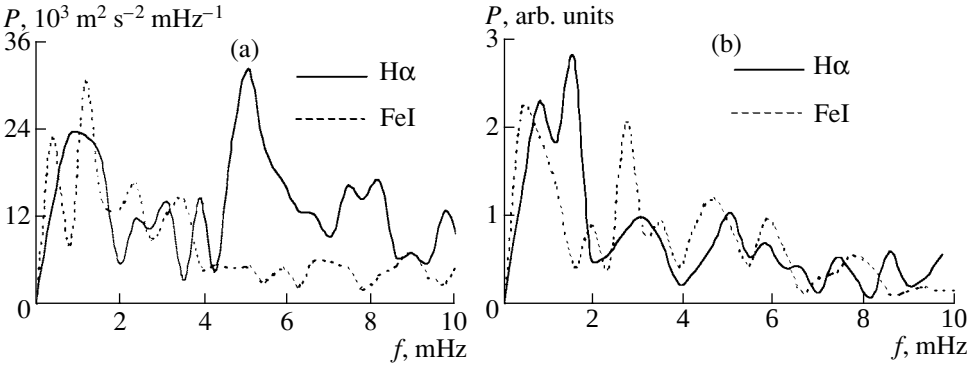

Fig. 6. Comparison of the spectra of the temporal variations in the (a) velocity and (b) intensity in a 18″ element from series No. 1. The values of the photospheric spectra have been increased by a factor of two.

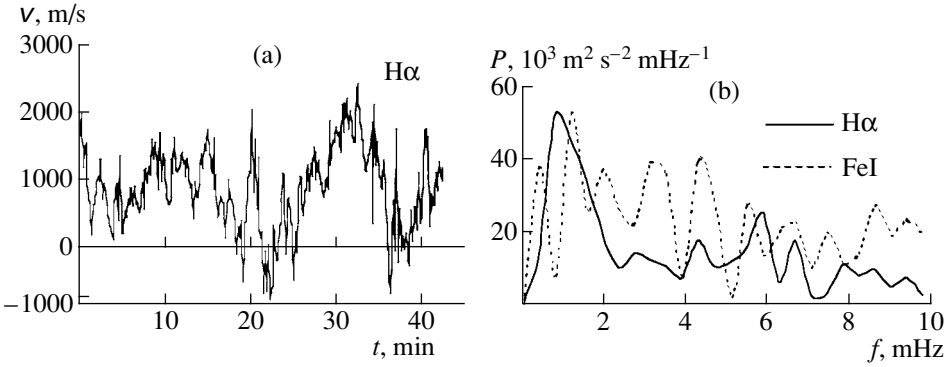

Fig. 7. Example of low-frequency velocity variations in a 40″ element from time series No. 1: (a) original variations in the chromosphere and (b) the corresponding spectra. The values of the photospheric spectrum have been increased by a factor of five.

image differ considerably. The region corresponding to the 80th pixel of the array along the slit (time series No. 1, 18″) is characterized by velocity oscillations with a period of about 14 min at both levels (Fig. 6). This period is detected in the $H\alpha \pm 0.3$ Å brightness, but is suppressed in the brightness spectrum of the photosphere. The spectrum of the photospheric velocity oscillations in the region of the 40th pixel (series No. 2, $\sim 9''$) contains maxima corresponding to the periods 20, 8.5, and 5 min (Fig. 5b). Since the duration of time series No. 1 and No. 2 is 43 min, the power near 0.4 mHz will be overestimated, so that the reliability of the low-frequency maxima is not very high. On the other hand, the reality of these maxima is supported by the fact that they are sometimes clearly

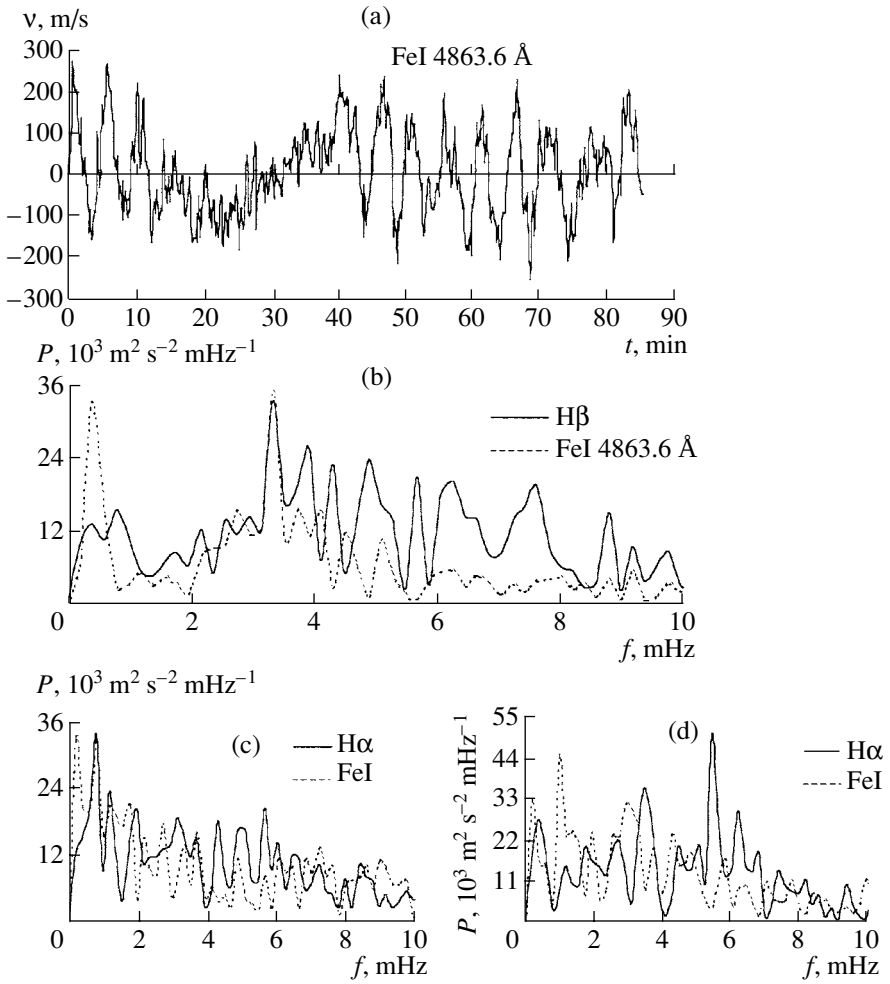

Fig. 8. Variations in the line-of-sight velocities measured by the CCD strip: (a) a fragment of time series No. 3, (b) the corresponding spectra in the Hβ and FeI 4863.6 Å lines (values of the photospheric spectrum have been increased by a factor of two), (c) and (d) spectra of oscillations of the line-of-sight velocity for series No. 6 and No. 7, respectively (values of the photospheric spectrum have been increased by a factor of three).

expressed in the photospheric spectrum, but not observed in the chromospheric spectrum from the same time series (Figs. 6 and 7), or vice versa (Fig. 4). In some time series, the variations of the line-of-sight velocities at these two levels of the solar atmosphere are in antiphase. All this behavior would be unlikely if these maxima were artifacts. In addition, similar oscillation periods are visible in some of the original time series even without the aid of spectral analysis (Figs. 4, 7, and 8).

Unfortunately, we were not able to derive reliable estimates of variations of the magnetic-field intensity in the recorded series (the Landé factor of the photospheric line used was below 1.4). We can only conclude that no temporal variations in this parameter were observed at the achieved sensitivity (a few tens of Gauss).

In general, the time series obtained using the CCD strip confirm the results of the CCD-array measurements. However, since the duration of the continuous measurements with the CCD strip exceeded 1 h 20 min in several cases, they enabled us to analyze longer periods in the velocity variations. The time series themselves and the corresponding spectra contain a fairly prominent period of about 40 min, as is illustrated in Fig. 8. The spectral resolution of these data is approximately double that of the array data, leading to the appearance of fine structure in the spectral maxima at medium and high frequencies.

5. COMPARISON OF RESULTS AND CONCLUSIONS

Let us compare our results with those obtained in studies of the dynamics of higher layers of the solar atmosphere. Banerjee *et al.* [10] investigated oscillations in a polar coronal hole using data from the coronal diagnostic spectrometer onboard the SOHO

satellite. The analysis was carried out for the OV 629 Å spectral line, which is formed in the lower layers of the transition region between the chromosphere and corona. The spectrograph slit was oriented along a plume. The main periods in the intensity variations were 10 and 25 min. However, these periods do not appear convincing in the line-of-sight velocity spectra presented in [10]. On this basis, those authors concluded that the detected oscillations are acoustic or magnetosonic, but not Alfvenic waves.

Almost all the periods we have detected from 0.4–2 mHz are present in Figs. 2 and 9 of [10]. The main difference is that our results concern variations in the line-of-sight velocity rather than in the intensity. When Banerjee *et al.* [10] studied an inter-plume region, they also found velocity variations with periods of about 40 min and more in the OIII 599 Å and OIV 554 Å lines, as well as variations with a period of 20 min in the HeI 580 Å line [11]. In connection with our period of ~ 8 min, we note that very similar oscillations of the brightness and degree of polarization have been observed in the inner corona [7, 12].

It follows from our observations that the 3-min and 5-min oscillations are amplified rather than attenuated in polar coronal holes (if we assume them to be radial). Their amplitude reaches 1 km/s in the photosphere and 3–4 km/s in the chromosphere (when projection onto the line of sight is taken into account). This result contradicts the work of Malanushenko *et al.* [13], who concluded that oscillatory processes were suppressed in coronal holes. This contradiction may be due to the fact that we observed only polar coronal holes. In addition, as can be seen in the figures presented here, we have reliably detected low-frequency variations with periods of 8, 10–15, 20, and 40 min. We suggest that quasi-stationary upward flows in the lower atmospheres of coronal holes coincide with boundaries of the chromospheric network. Kozlova and Somov [5] proposed that high-speed flows are formed at the boundaries of coronal holes rather than at their centers. If the boundaries of coronal holes follow boundaries of the chromospheric network, high-speed flows should also be detected inside coronal holes, since the chromospheric network is still observed in the coronal holes, though with lower contrast.

The velocity oscillations observed in the polar coronal holes at the photospheric and lower chromospheric levels are characterized by a wide range of frequencies, with 0.4, 0.75, and 1 mHz being encountered most frequently. The power spectrum of the line-of-sight velocity averaged over a $60'' \times 1.5''$ region contains significant periods of 40, 15, and 7 min, indicating that the oscillatory modes observed have a considerable extent. The amplitudes of acoustic oscillations (at 3 and 5 mHz) increase in the polar coronal holes. We suggest based on our observations that the velocity and brightness variations observed in the transition zone and inner corona above coronal holes are formed in the lower chromosphere and photosphere.

ACKNOWLEDGMENTS

This work was partially funded by the Program of Support of Leading Scientific Schools of the Russian Federation (project no. NSh-733.2003.2) and the State Science and Technology Program *Astronomy*.

REFERENCES

1. J. B. Zirker, *Coronal Holes and High-Speed Wind Streams* (Colorado Associated Univ. Press, 1977), p. 124.
2. V. N. Obridko and B. D. Shelting, Solar Phys. **124**, 73 (1989).
3. V. N. Obridko and B. D. Shel'ting, Astron. Zh. **67**, 890 (1990) [Sov. Astron. **34**, 449 (1990)].
4. N. N. Stepanyan, Izv. Ross. Akad. Nauk, Ser. Fiz. **59** (7), 63 (1995).
5. L. M. Kozlova and B. V. Somov, Astron. Zh. **77** (6), 63 (2000) [Astron. Rep. **44**, 401 (2000)].
6. D. Banerjee, E. O'Shea, J. G. Doyle, *et al.*, Astron. Astrophys. **380**, L39 (2001).
7. L. Ofman, M. Romoli, G. Poletto, *et al.*, Astrophys. J. **529**, 592 (2000).
8. N. I. Kobanov, Prib. Tekh. Éksp., No. 4, 110 (2001).
9. N. I. Kobanov and D. V. Makarchik, Prib. Tekh. Éksp., No. 4, 116 (2001).
10. D. Banerjee, E. O'Shea, J. G. Doyle, *et al.*, Solar Phys. **196**, 63 (2000).
11. D. Banerjee, E. O'Shea, J. G. Doyle, *et al.*, Astron. Astrophys. **377**, 691 (2001).
12. L. Ofman, M. Romoli, G. Poletto, *et al.*, Astrophys. J. Lett. **491**, L111 (1997).
13. E. V. Malanushenko, V. P. Malanushenko, and N. N. Stepanian, *Motions in the Solar Atmosphere,* Ed. by A. Hanslmeier and M. Messerotti (Kluwer Academic, Dordrecht, 1999), Vol. 239, p. 251.

Translated by Yu. Dumin

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 953–962. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 1034–1045.

Global Solar Magnetology and Reference Points of the Solar Cycle

V. N. Obridko and B. D. Shelting

Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Troitsk, Moscow oblast, 142190 Russia

Received March 5, 2003; in final form, May 8, 2003

Abstract—The solar cycle can be described as a complex interaction of large-scale/global and local magnetic fields. In general, this approach agrees with the traditional dynamo scheme, although there are numerous discrepancies in the details. Integrated magnetic indices introduced earlier are studied over long time intervals, and the epochs of the main reference points of the solar cycles are refined. A hypothesis proposed earlier concerning global magnetometry and the natural scale of the cycles is verified. Variations of the heliospheric magnetic field are determined by both the integrated photospheric $i(B_r)_{ph}$ and source surface $i(B_r)_{ss}$ indices, however, their roles are different. Local fields contribute significantly to the photospheric index determining the total increase in the heliospheric magnetic field. The $i(B_r)_{ss}$ index (especially the partial index ZO, which is related to the quasi-dipolar field) determines narrow extrema. These integrated indices supply us with a "passport" for reference points, making it possible to identify them precisely. A prominent dip in the integrated indices is clearly visible at the cycle maximum, resulting in the typical double-peak form (the Gnevyshev dip), with the succeeding maximum always being higher than the preceding maximum. At the source surface, this secondary maximum significantly exceeds the primary maximum. Using these index data, we can estimate the progression expected for the 23rd cycle and predict the dates of the ends of the 23rd and 24th cycles (the middle of 2007 and December 2018, respectively). © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The traditional view of the properties of the 11-year solar-activity cycle was formed at the end of the 20th century, taking into consideration numerous phenomena associated with solar activity. There are two types of intensity variations observed in these phenomena—smooth and sharp—which alternate during the cycle. Taking into account these features, the concept of the main phases of the solar cycle was introduced, with these phases being characterized by smooth changes in solar-activity phenomena and their interrelations during some time interval. Intervals of comparatively smooth changes in various parameters were named the "cycle phases," and their boundaries "reference points." Thus, the initial definition of the reference points was quite natural. The reference points t_{Dm} and t_{mA} were related to the start and end of the minimum of the 11-year cycle, while t_{AM} and t_{MD} were associated with the start and end of the maximum. The first detailed description of the solar-cycle phases was given in 1986 by Vitinskii *et al.* [1], who concentrated their studies on cyclic behavior of solar activity governed mainly by local magnetic fields, since the properties of the global magnetic fields remained virtually unknown due to the extremely limited experimental data available. The first synoptic magnetic-field maps were obtained at the Mount Wilson Observatory (MWO) from August 1959 to December 1978, for the 1417–1648th Carrington rotations (CRs), and at the Kitt Peak Observatory (KPO) from January 1975 to July 1984 (CR 1622–1751). Long-term observations were started at the Stanford Wilcox Solar Observatory (WSO) in May 1976, and are still ongoing. Although there were no detailed studies of the global magnetic fields or deep understanding of their properties at that time, Vitinskii *et al.* [1] concluded that obtaining a clear understanding of solar cyclicity requires analysis of features and processes related to the quiescent Sun (coronal holes, the total and background magnetic fields, the quiescent corona, the solar wind, etc). Interesting behavior of the large-scale fields had been revealed near the solar maxima: the reference point t_{AM} coincides with the onset of the polarity reversal of the magnetic field, which is manifest as a rapid poleward drift of the polarity boundaries of the background fields; long-lived coronal holes are observed during the decline phase after the reference point t_{MD}; there is a clear recurrence in the pattern of the interplanetary magnetic field, with the polarity boundaries of the background fields being stable in latitude.

As the amount of observational data on global fields increased, systematic studies of their properties

1063-7729/03/4711-0953$24.00

were stimulated—first and foremost being their cyclic behavior. In particular, comprehensive studies of the behavior of both local and global magnetic fields in descriptions of the 11-year cycle were first presented in [2–7]. The concept of global solar magnetology was first introduced in [6, 7], where the cycle is described using integrated indices of the magnetic field. It was discovered that a natural cycle duration and the concept of normal and anomalous field fluxes can be introduced. A new and very effective (as has been shown by subsequent long-term investigations) energy index $i(Br)$ for the global magnetic field was introduced. This index is defined as the square of the radial magnetic field averaged over a sphere of a given radius:

$$i(B_r) = \langle B_r^2 \rangle.$$

A description of the mathematical apparatus used in calculations of the integrated magnetic-field indices and the partial indices at the photosphere and source surface is presented in [3, 4], as well as in Section 2 below.

Further, an entire set of integrated magnetic-field indices was constructed as a function of time, and their cyclic behavior was studied in detail taking into account the cycle phases and reference points determined from the behavior of the local fields. It turns out that the integrated indices supply us with a "passport" for reference points, enabling more precise determination of their epochs.

In addition, numerous phenomena of solar and geophysical activity form a continuous series between local fields, with Wolf numbers being the most widely used index, and global fields, which are most accurately described by the index $i(B_r)$ at the source surface.

One of the characteristics of the behavior of $i(B_r)$ near the solar maximum is that the Wolf number maximum does not correspond to the cycle maximum in the integrated energy indices for the global magnetic field. On the contrary, the integrated index demonstrates a characteristic two-peak form at the maximum, with a local dip being observed near the epoch of the Wolf number maximum. A comparison between the photospheric and source surface indices indicates that the first maximum (preceding the Wolf number maximum) is related predominantly to fields on small and intermediate spatial scales, whereas the second maximum observed at the end of 1982 is completely determined by global fields.

The dip at the Wolf-number maximum was subsequently studied by numerous researchers using various indices, including the integrated magnetic-field indices introduced in [2–7]. This dip, whose nature remains unclear, was named the Gnevyshev dip [8–14]. It is more clearly expressed in high-power phenomena [9]. Since the occurrence of the dip coincides with the polarity reversal of the global magnetic field, it may simply indicate a decrease in the total magnitude of the global and local fields. However, as was noted above and will be shown based on more complete data below, the first and second maxima in the integrated indices are by no means equivalent, with the role of global fields being much more significant in the second maximum.

The fine structure of the geophysical index aa also demonstrates a double-peaked behavior. Ahluwalia [11] and Webb [15] concluded that the first maximum is due to coronal mass ejections, while the second is associated with coronal holes. On the other hand, Richardson *et al.* [12] have argued that this dip is related to a decrease in the solar-wind velocity, which, in turn, is due to a decrease in the magnitude of the global magnetic field.

The goal of our work is to analyze integrated magnetic-field indices over a time interval exceeding three cycles, to compare the main reference points with characteristic features in the time behavior of the indices, and to study some relations between the integrated magnetic-field indices and other heliospheric and geophysical indices.

2. DATA USED

In [2–7], we used only data on global fields for 11-year solar cycle 21, and it was impossible to make any general conclusions about the applicability of our findings to all solar-activity cycles. Today, we have measurements of large-scale magnetic fields obtained using low-resolution magnetographs over more than forty years (1960–2002). We have currently applied our technique to three entire cycles—20, 21, and 22—and to the first half of cycle 23.

The observational data for the 20th cycle (and therefore the corresponding results and conclusions obtained) are somewhat less reliable, since data obtained at different observatories with different spatial resolutions and calibrations were combined. Some mismatches may remain in the data, no matter how carefully we have tried to reduce them to a single system. In addition, there are quite a few gaps in the observations during this period. We believe that the WSO data on the 21st and 22nd cycles (entire magnetic cycles) are the most trustworthy.

The following integrated magnetic-field indices were introduced in [3–5]:

$$i(B_r)|_{R_o} = \sum_{lm} \frac{(l+1+l\zeta^{2l+1})^2}{2l+1}(g_{lm}^2 + h_{lm}^2), \quad (1)$$

$$i(B_r)|_{R_s} = \sum_{lm}(2l+1)\zeta^{2l+4}(g_{lm}^2 + h_{lm}^2). \quad (2)$$

These formulas were derived for a potential magnetic-field model with a source surface (see, for example, [16]). Here, g_{lm} and h_{lm} are the coefficients of the expansion of the photospheric magnetic field in Legendre polynomials with the indices l and m, R_o and R_s are the radii of the photosphere and source surface, respectively, and their ratio ζ is assumed to be 0.4. The summation over l and m was carried out from 0 to 9.

Along with the total integrated field index, the so-called partial integrated indices were obtained: the zonal-even index ZE ($m = 0$, $l = 2k$), zonal-odd index ZO ($m = 0$, $l = 2k + 1$), sectorial-even index SE ($m = l = 2k$), and sectorial-odd index SO ($m = l = 2k + 1$). The physical definition of the partial indices is the following.

The index ZO corresponds to the magnetic-field component with zonal-odd symmetry (similar to a vertical dipole). Its physical meaning is that it best describes the main global field, which is used as a basis for generating the local fields that are traditionally employed to describe solar activity. Therefore, as we will see below, this is among the main indices determining the reference points of the cycle.

The ZE index is low as a result of the Hale law. The sectorial-odd index SO characterizes a tilted dipole, and appears when the two-sector and four-sector structures are present. The even-sectorial index SE usually appears in the presence of four-sector structures.

We calculated the coefficients g_{lm} and h_{lm} using the MWO data for August 1959 to December 1978 (CR 1417–1648), the KPO for January 1975 to July 1984 (CR 1622–1751), and the WSO data for May 1976 to present. The calculation technique we used is described in [17].

3. INTEGRATED MAGNETIC-FIELD INDICES

Figures 1–3 show the time behavior of the integrated magnetic-field indices. Of course, the initial calculations were performed with a resolution of one Carrington rotation. The final results presented in all our plots except for Figs. 4 and 5 are running averages over six Carrington rotations; the results in Figs. 4 and 5 are running averages over two years. The solid curves in Figs. 1–3 show the indices calculated using the WSO data, while the dashed curves show the results of the calculations using the earlier observations. In spite of the overall qualitative agreement, we can see that there are some quantitative discrepancies, whose origin remains unclear. They may be associated with differences in sensitivity, spatial and temporal resolution, calibration technique, and polar correction. Our technique [17] has enabled us to reduce these data to a single system, although it, naturally, cannot completely remove the differences between the datasets.

The results presented in Figs. 1–3 enable us to verify and refine on the basis of four cycles some results obtained earlier [5] for the 21st cycle only.

We have already noted that the main index $i(B_r)$ at the photosphere (Fig. 1) displays a clear quasi-eleven-year cyclicity with a marked rise at the cycle maxima. However, right at the cycle maxima, there is a clear dip resulting in the typical two-peaked form. The maximum succeeding the dip is always higher than that preceding the dip. At the source surface (Fig. 2), this second maximum significantly exceeds the first maximum, and displays different shifts in phase with respect to the local field maximum in different cycles. Thus, at the source surface, the 11-year cycle is expressed more weakly than at the photosphere. We will return to this point in Section 4.

As expected, the index ZO determined by quasi-dipolar components—first and foremost, by the coaxial ("vertical") dipole—is in antiphase with $i(B_r)$. The maximum at the source surface somewhat precedes that at the photosphere. This may be associated with the fact that the polarity reversal occurs earlier at the source surface than at the photosphere [19]. The ZO curves at the photosphere and source surface are rather similar, with the values differing by a factor of ~ 300, which is close to the factor of 244 that would be observed for a pure, free (i.e., without a source surface) dipole (ζ^{-6}), but is significantly higher than the ratio 108 resulting from setting $l = 1$ in (1)–(2). This is probably associated with the influence of the source-surface model and the contributions of other quasi-dipolar components.

The behavior of the remaining indices is somewhat more complex. In [5], we did not consider ZE at all, assuming that its presence was prohibited by the Hale law for hemispheres. Indeed, this index was very low during cycles 20–22, and its magnitude was lower than those of the other partial indices by a factor of a few (even an order of magnitude at the photosphere during the 22nd cycle). In principle, this could be due to some calculational or observational errors. However, ZE rose sharply in the 23rd cycle (especially at the source surface). If this is not an artifact, it indicates a significant north–south asymmetry in the large-scale magnetic field at the maximum of the 23rd cycle.

The SE index describing predominantly four-sector structures is also comparatively low and increases at maxima, reaching especially high values at the start of the decline phase. SO nearly always

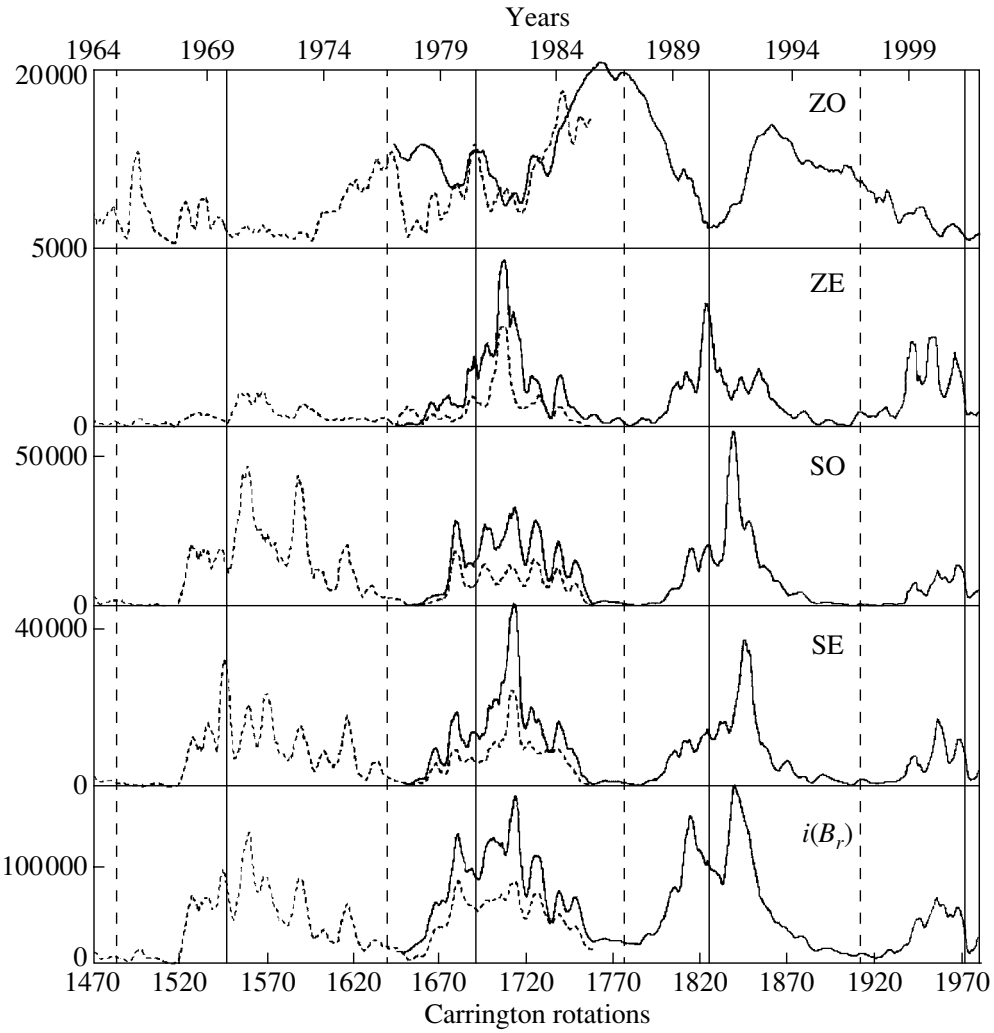

Fig. 1. Cyclic variations of $i(B_r)$ and the partial indices (in μT^2) at the photosphere. The dashed curves correspond to the Kitt Peak and Mount Wilson data, and the solid curve to the Wilcox Observatory data. The solid vertical lines mark the maxima of the local field activity cycles, while the dashed vertical lines mark the minima.

exceeds SE (sometimes substantially, especially at the source surface). This reflects the more frequent occurrence of two-sector than four-sector structures, as has been repeatedly noted, for example, in [20–22].

Since $i(B_r)$ and ZO are the most important integrated magnetic-field indices, we show their ratios as functions of time in Fig. 3. As expected, the $ZO/i(B_r)$ ratio increases during the decline phase both at the photosphere and at the source surface, reaching its maximum near the cycle minimum. The increase of this ratio at the source surface is sharper than at the photosphere, and occurs somewhat earlier. Therefore, the behavior of the ZO_{ss}/ZO_{ph} ratio is especially interesting. This ratio increases rather sharply right at the reference point t_{MD} marking the end of the maximum phase (see dates in the table below).

4. REFERENCE POINTS ACCORDING TO THE INTEGRATED MAGNETIC-FIELD INDICES

Let us generalize the behavior of the integrated magnetic-field indices at the reference points based on all four cycles considered (20–23).

(1) **Reference point t_{mA}.** The onset of a sharp increase in the total and partial indices in the photosphere is observed in this period, and they reach their first moderate local maximum. The exception is ZO, which continues to decrease after its maximum, observed during the preceding minimum of the local fields. The behavior at the source surface is similar, but with lower temporal gradients for all the indices.

(2) **Reference point t_{AM}.** The first significant maxima (we call them MAX1) are observed in the total index and SE and SO partial indices at the

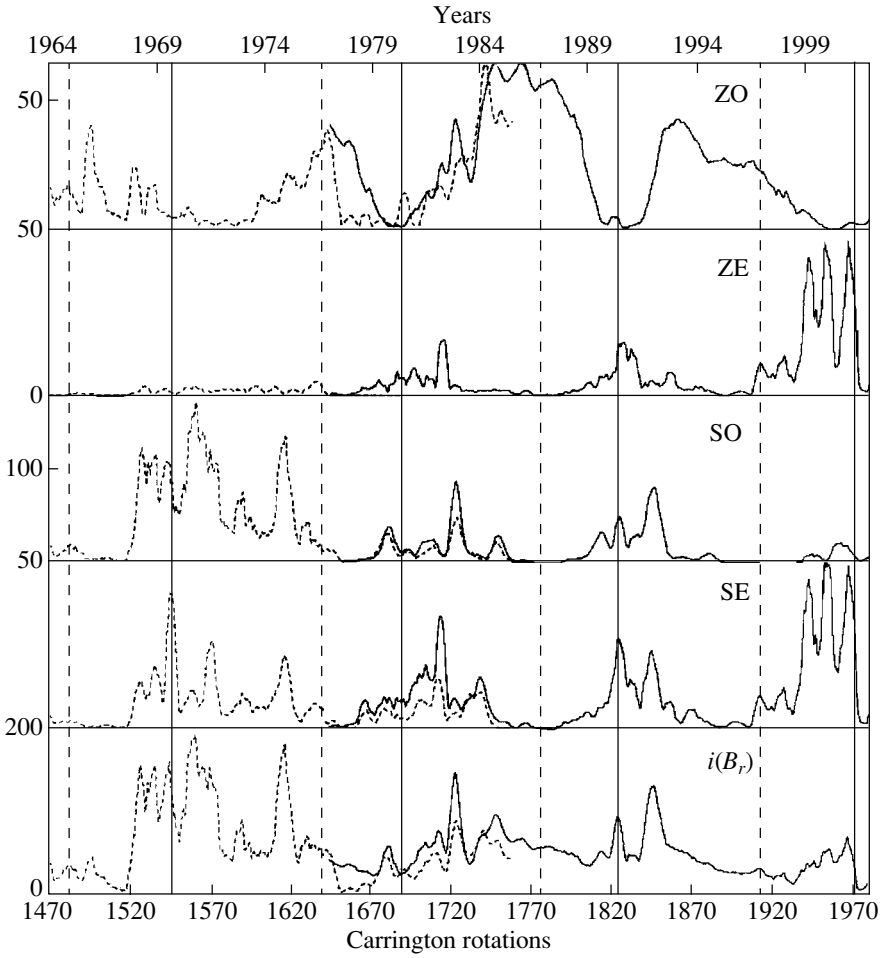

Fig. 2. Same as Fig. 1 for the source surface.

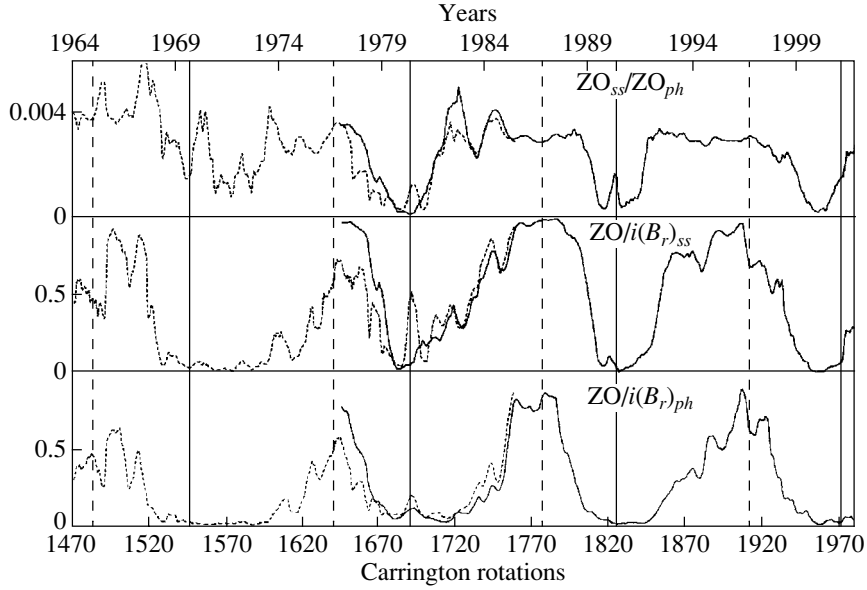

Fig. 3. Ratio of the indices ZO and $i(B_r)$ at the photosphere and the source surface. Notation is the same as in Fig. 1.

Main features of the reference points for cycles 20–23

Cycle	CR Date	$i(B_r)_{ph}$	$i(B_r)_{ss}$	ZO_{ph}	ZO_{ss}
			MIN		
20	1484 July, 1964	MIN	MIN	moderate local MAX	moderate local MAX
21	1640 Apr., 1976	MIN	MIN	main MAX	main MAX
22	1778 July, 1986	MIN	MIN	main MAX	main MAX
23	1916 Dec., 1996	MIN	MIN	decline of the main MAX	decline of the main MAX
			t_{mA}		
20	1499 Sept., 1965	MIN	MIN	local MAX	local MAX
21	1656 June, 1977	start of the rapid rise	MIN	main MAX	decline of the main MAX
22	1791 July, 1987	start of the rapid rise	MIN	decline of the main MAX	decline of the main MAX
23	1940 Aug., 1998	start of the rapid rise	MIN	decline of the main MAX	decline of the main MAX
			t_{AM}		
20	1520 May, 1967	start of the rise	start of the rise	local MIN	local MIN
21	1680 Apr., 1979	MAX1	MAX1	local MIN	MIN
22	1812 Apr., 1989	MAX1	local MAX	decline of the main MAX	MIN
23	1955 Oct., 1999	MAX1	MAX1	end of the decline of the main MAX	MIN
			MAX		
20	1540 Oct., 1968	before MAX1	dip between MAX	start of the MIN phase	start of the MIN phase
21	1692 Jan., 1980	dip between main MAX	dip between MAX	MAX2	MIN
22	1818 July, 1989	decline after MAX1	before MAX1	decline of the main MAX	MIN
23	1961 Apr., 2000	dip between main MAX	dip between MAX	MIN	MIN
			t_{MD}		
20	1558 Apr., 1970	MAX2I	main MAX	long-term MIN	short MIN
21	1715 Sept., 1981	MAX2	local MAX	MIN	start of the rise
22	1840 March, 1991	MAX2	dip between MAX	start of the rise	main MAX
			t_{Dm}		
20	1615 June, 1974	last local MAX before MIN	last local MAX before MIN	rise before MAX	rise before MAX
21	1748 March, 1984	last local MAX before MIN	last local MAX before MIN	rise before MAX	MAX
22	1875 Nov., 1993	MIN	MIN	decline after MAX	decline after MAX

photosphere. ZO approaches its minimum. Recall that there is a small decrease in the growth of the Wolf number at this time (a horizontal plateau) before the main maximum of the local fields. At the source surface, the indices demonstrate a similar behavior but with lower local maxima. ZO reaches its absolute minimum.

(3) The **maximum phase MAX** occurs between the reference points t_{AM} and t_{MD}. On average, all the indices acquire high values at the photosphere, although, as a rule, the maximum of the local fields occurs in the dip between the main maxima of the integrated indices. As always, the behavior of ZO is opposite, and this index reaches its minimum. At this phase, the polarity reversal occurs and the large-scale zonal fields disappear, which is equivalent to the vanishing of the vertical (axial) dipole. At the source surface, all the indices demonstrate a similar behavior.

(4) The **reference point t_{MD}** corresponds at the photosphere to the second local maximum MAX2 in almost all the indices. This is the highest maximum for the global field; it is interesting that it coincides in time with the second maximum in the Wolf number. ZO exhibits a moderate increase. At the source surface, only a few indices demonstrate local maxima at this reference point, and not in all cycles. More often, the second local maximum at the source surface occurs later, approximately ten Carrington rotations after the second maximum at the photosphere.

(5) **Reference point t_{Dm}.** As a rule, there is a moderate local peak in the intensity (moderate local maximum) of all the indices at the photosphere and source surface before the final onset of the minimum phase. ZO reaches its broad maximum. At this reference point, there is a local maximum in the $ZO/i(B_r)$ ratio at both the photosphere and source surface.

(6) At the **local field minimum MIN**, all the integrated global field indices fluctuate near zero, with the quasi-dipolar index ZO remaining nonzero and taking its maximum.

This discussion shows that the total indices and ZO play the main roles in determining the reference points. This is quite natural, since the cyclic variations in the total magnetic field are determined by cyclic variations in the local fields, which contribute to the total index, and by the global field, which mainly determines ZO, especially at the source surface. These results are summarized in the table.

Krainev *et al.* [14] have obtained a very interesting result. They have calculated the energy of the large-scale fields (i.e., essentially our indices) during the 21st, 22nd, and beginning of the 23rd cycles, though not for the entire disk as in (1)–(2), but individually for equatorial (10°–40°) and polar (40°–90°) zones. The cyclic behavior of these energies turns out to be quite different. This is quite natural, since the relative

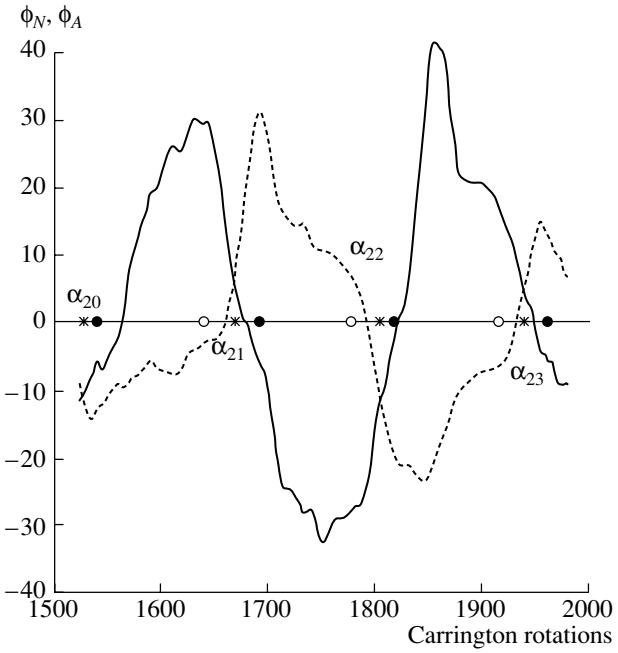

Fig. 4. The fluxes Φ_N (solid curve) and Φ_A (dashed curve) as functions of time. The asterisks indicate the equilibrium points α, while the filled and hollow circles show the maxima and minima of the 11-year cycles, respectively. The flux unit is 3×10^{20} Mx.

contributions of the local and large-scale fields, on the one hand, and of the global fields, on the other hand, vary during the cycle. At the source surface, both these variations intersect at the boundaries of the cycle maximum phases (t_{AM} and t_{MD}) indicated in [18]. Thus, the latitudinal gradient of the magnetic-field energy becomes zero near these points, and the contributions of the local and global fields to the total energy at the source surface become equal.

In [14], the last intersection point for the 23rd cycle was indicated as August 1999 (our estimates give October 1999 for the reference point t_{AM}), and the middle of the maximum was forecast for June 2000. The maximum determined by the smoothed Wolf number arrived in April 2000. This shows once more the utility of the reference points and integrated magnetic-field indices for refining our understanding of the 11-year cycles and ability to forecast them.

5. GLOBAL MAGNETOMETRY AND COMPLEMENTARY FLUXES

In [6], we introduced the concept of complementary fluxes of the global magnetic field, namely the normal Φ_N and anomalous Φ_A fluxes. These quantities determine the magnetic flux integrated over the surface of a solar hemisphere (we have used the northern hemisphere). During this averaging, all axially symmetric components and active-region

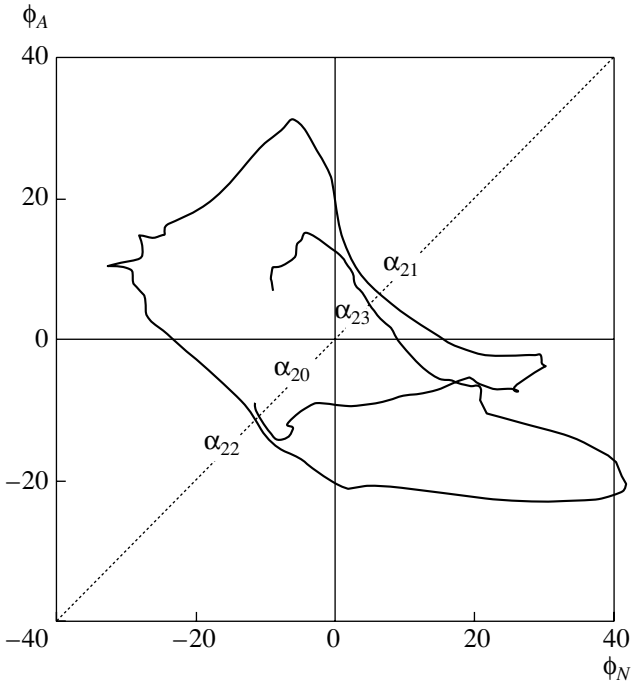

Fig. 5. Relation of the anomalous flux Φ_A to the normal flux Φ_N. The flux unit is 3×10^{20} Mx.

fields are eliminated, with only the zonal global fields remaining. These fluxes are completely determined by the coefficients g_{l0} introduced above, making them similar to ZO. The fundamental difference between Φ_N, Φ_A, and ZO is that the index ZO results from integrating the surface magnetic-field energy of the zonal component over an entire sphere, and can therefore, in principle, contain a contribution, albeit weaker, of smaller-scale fields as well as global fields. In contrast, the fluxes Φ_N and Φ_A contain only the global field flux through the hemisphere.

The complementary fluxes are calculated via the formula

$$\Phi = 2\pi R_0 \sum g_{l0} P_{l-1}(0). \qquad (3)$$

Here, $P_l = P_{l,0}$, $P_{l,m} = P_{l,m}(\cos\theta)$, and $\sqrt{(2l+1)} \times P_{l,m}$ are the associated Legendre polynomials of degree l and order m ($0 \le l \ge |m|$), which are orthonormal on l for fixed m. In this expression, the dipole portion ($l = 1$) gives the normal flux Φ_N, while the non-dipole portion ($l > 1$) gives the anomalous flux Φ_A. We showed in [6] that these fluxes are complementary. In each cycle, there is an equilibrium point α where these fluxes become equal. We showed in [6], based on limited statistics, that the total magnetic fluxes at the α points of the even (20th and 22nd) cycles were the same, while the time integral of this flux from α_{20} to α_{22} was zero. We suggested that one 22-year magnetic cycle finishes and another begins at even equilibrium points, thus determining a natural time scale for successive cycles.

Figure 4 shows the fluxes Φ_N and Φ_A as functions of time, while Fig. 5 relates Φ_A to Φ_N in a phase diagram. The flux unit in these plots is 3×10^{20} Mx.

Unfortunately, we do not have sufficient data to fully verify the main hypothesis concerning the existence of independent 22-year cycles, since we must experimentally determine the point α_{24} for this purpose. However, we can establish the following results, which are in overall agreement with the main hypothesis.

(1) The points α_{20} (CR 1528, November 1967) and α_{22} (CR 1805, July 1988) coincide in the phase diagram, and the corresponding fluxes are each -3.45×10^{21} Mx. The points α_{21} (CR 1669, June 1978) and α_{23} (CR 1939, July 1998) are also close to each other. The flux at α_{23} is 1.41×10^{21} Mx, and is close to the value 1.84×10^{21} Mx at α_{21} within the errors.

(2) The average total flux over the 22-year cycle α_{20}–α_{22} was 0.5×10^{20} Mx, and was 4.4×10^{20} Mx over the cycle α_{21}–α_{23}. Both fluxes are close to zero within the errors.

(3) The duration of the cycle α_{21}–α_{23} is 271 Carrington rotations, somewhat shorter than that of the cycle α_{20}–α_{22}, which was 278 Carrington rotations.

It is possible that the observed weak deviations from the equality of (i) the fluxes at the points α_{21} and α_{23} to each other, (ii) the average total flux to zero, and (iii) the durations of the 22-year cycles to each other are associated with the fact that the Hale cycle starts from an even cycle, according to the Gnevyshev–Ol' rule, and so we should compare the pairs (α_{20}–α_{22}) and (α_{22}–α_{24}). Nevertheless, taking 274 Carrington rotations for the average duration of the 22-year cycle, we can identify the epochs for α_{24} (CR 2079, January 2009) and α_{25} (CR 2212, December 2018).

Note that the reference points t_{mA} for the 21st (CR 1665), 22nd (CR 1804), and 23rd (CR 1940) cycles virtually coincide with the natural boundaries of the cycles α_{21}, α_{22}, and α_{23}, but these points were observed a half a year or a year earlier during the 20th cycle. Consequently, our forecast of the α_{24} and α_{25} points is essentially a forecast of the onsets of the rising phases of the 24th and 25th cycles.

6. GLOBAL ACTIVITY IN THE HELIOSPHERE

Figure 6 compares the main global-activity indices at the photosphere and source surface with the monthly averaged Wolf numbers W, magnetic field B, and velocity V in the heliosphere and with the

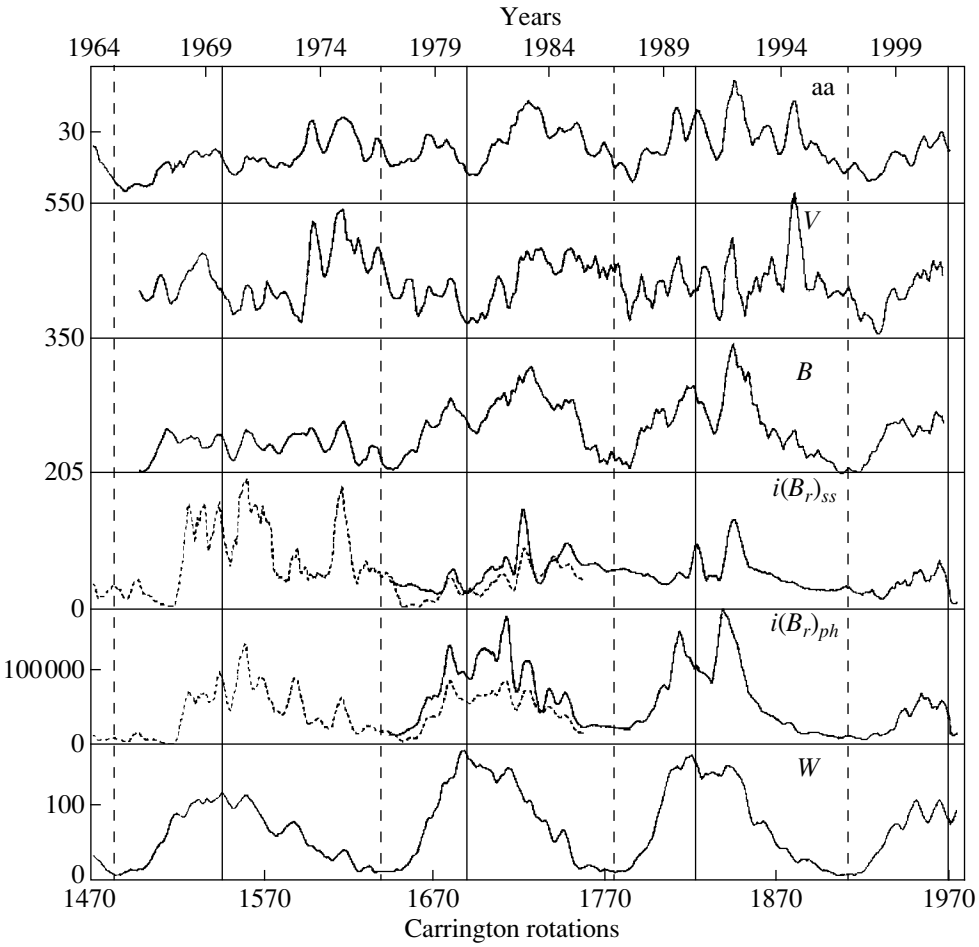

Fig. 6. Cyclic variations of some helio-geophysical indices: the Wolf number W, total global field indices at the photosphere $i(B_r)_{ph}$ and source surface $i(B_r)_{ss}$, heliospheric magnetic field B, solar-wind velocity V (in km/s), and geomagnetic aa index.

geomagnetic aa index. We can see that the similarity between variations of the heliospheric and geomagnetic parameters and variations of $i(B_r)_{ss}$, found in [5] for the 21st cycle, is verified by the broader observational material covering cycles 20–23. However, we can indicate some new features.

The larger data set shows that both $i(B_r)_{ph}$ and $i(B_r)_{ss}$ contribute to variations of the heliospheric magnetic field. However, their roles are different: the photospheric index determines the general increase in the heliospheric magnetic field, while $i(B_r)_{ss}$ determines narrow extrema. The relationship to V is significantly weaker. This is probably associated with the fact that the global indices describe only the integrated energy, while the generation of high-velocity flows is predominantly determined by structural features, in particular, the surface area occupied by the open magnetic field.

The peak in the velocity V during CR 1890 at the end of 1993 and in the first half of 1994 and the related peak in geomagnetic activity is a typical example. No appreciable activity in the integrated indices was seen at that time. However, the area occupied by open magnetic fields had grown strongly in this period [23], and a powerful polar coronal hole was observed [24, 25]. A similar sharp peak in the solar-wind velocity and geomagnetic activity was observed in 1974, 20 years earlier. Ahluwalia [11] and Tsurutani *et al.* [26] have attributed these effects to corotating high-velocity flows from two coronal holes. Ahluwalia has suggested that this is a property of even cycles within the Hale magnetic-field cycle.

The heliospheric parameters B and V together lead to the geomagnetic variations described by the aa index. Therefore, the relationship between the parameters of the solar magnetic field and the aa index is even more complex. Although we can trace virtually all the extrema detected in the solar indices in the aa index, in general, the 11-year cycle of geomagnetic variations is manifest appreciably less clearly than is the solar cycle.

7. DISCUSSION

Our analysis has shown that the integrated global indices introduced earlier enable us to determine more accurately the epochs of the cycle reference points, which, in turn, makes it possible to obtain a more accurate description of the solar cycle. There is no doubt that the entire process of the solar cycle results from interactions of magnetic fields on various scales, most notably local fields (scales of several hundredths of the solar radius), large-scale fields (scales of tenths of the solar radius), and global fields (scales of the order of the solar radius). The contributions of the fields on various scales differ considerably depending on the helio-geophysical processes considered.

The entire solar cycle can be described as a complex interaction of two main fields: global/large-scale and local. In general, this approach agrees with the traditional dynamo scheme, although there are numerous discrepancies in details.

Special attention must be drawn to the cycle reference points. Using our results and data on earlier cycles beginning with the 15th (sunspot reference points were used before the 20th cycle), we have found that the reference points within a cycle correspond to the following times (in fractions of the cycle duration): $t_{mA} = 0.116 \pm 0.08$, $t_{AM} = 0.230 \pm 0.10$, $\mathrm{MAX} = 0.370 \pm 0.05$, $t_{MD} = 0.515 \pm 0.15$, and $t_{Dm} = 0.71 \pm 0.1$ (the rms errors are also indicated). The cycle amplitudes, Wolf numbers, and integrated global indices at the reference points differ from cycle to cycle much more strongly than the positions of the points within the cycle. One has the impression that the periodicity and energetics of helio-physical cycles are formed independently to some extent.

Using the results obtained above and the position of the point α_{24}, we can estimate the duration of the 23rd cycle (CR 140–145) and indicate the epochs of the points t_{MD} (CR 1988–1990, March–June 2002), t_{Dm} (CR 2014–2020, April–June 2004), and MIN (CR 2056–2060, the middle of 2007), and, thus, determine the expected progression of the 23rd cycle.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16199) and by INTAS (grant 2000-840).

REFERENCES

1. Yu. I. Vitinskii, G. V. Kuklin, and V. N. Obridko, Sol. Dannye, No. 3, 53 (1986).
2. G. V. Kuklin, V. N. Obridko, and Yu. Vitinsky, in *Solar Terrestrial Predictions: Proceedings of Workshop at Leura, Australia, 1989* (1990), Vol. 1, p. 474.
3. V. N. Obridko and F. A. Ermakov, Astron. Tsirk. No. 1539 (1989).
4. V. N. Obridko, B. D. Shel'ting, and F. A. Ermakov, Astron. Tsirk., No. 1540, 23 (1989).
5. V. N. Obridko and B. D. Shelting, Solar Phys. **137**, 167 (1992).
6. F. A. Ermakov, V. N. Obridko, and B. D. Shel'ting, Astron. Zh. **72**, 98 (1995) [Astron. Rep. **39**, 86 (1995)].
7. F. A. Ermakov, V. N. Obridko, and B. D. Shel'ting, Astron. Zh. **72**, 753 (1995) [Astron. Rep. **39**, 672 (1995)].
8. E. S. Vernova, M. I. Tyasto, D. G. Baranov, *et al.*, Geomagn. Aeron. **37**, 105 (1997).
9. F. Feminella and M. Storini, Astron. Astrophys. **322**, 311 (1997).
10. J. G. Luhmann, J. T. Gosling, J. T. Hoeksema, *et al.*, J. Geophys. Res. **103**, 6585 (1998).
11. H. S. Ahluwalia, J. Geophys. Res. **105**, 27481 (2000).
12. J. G. Richardson, E. W. Cliver, and H. V. Cane, J. Geophys. Res. **105**, 18203 (2000).
13. G. A. Bazilevskaya, M. V. Krainev, V. S. Makhmutov, *et al.*, Solar Phys. **197**, 157 (2000).
14. M. V. Krainev, G. A. Bazilevskaya, and V. S. Makhmutov, Adv. Space Res. **29**, 331 (2000).
15. D. F. Webb, Adv. Space Res. **16**, 57 (1995).
16. J. T. Hoeksema and P. H. Scherrer, *Solar Magnetic Field—1976 through 1985* (WDCA, Boulder, USA, 1986).
17. V. N. Obridko and B. D. Shelting, Solar Phys. **184**, 187 (1999).
18. E. V. Ivanov, V. N. Obridko, and B. D. Shel'ting, Astron. Zh. **74**, 273 (1997) [Astron. Rep. **41**, 236 (1997)].
19. B. D. Shelting and V. N. Obridko, in *The Sun at an Epoch of a Magnetic-Field Polarity Reversal* [in Russian] (GAO RAN, 2001), p. 391.
20. G. V. Kuklin, V. N. Obridko, and Yu. Vitinsky, in *Solar Terrestrial Predictions: Proceedings of Workshop at Leura, Australia, 1989* (1990), Vol. 1, p. 474.
21. G. V. Kuklin and V. N. Obridko, Izv. Ross. Akad. Nauk, Ser. Fiz. **59** (7), 12 (1995).
22. G. V. Kuklin and V. N. Obridko, in *Physics of Solar Activity* [in Russian] (Nauka, Moscow, 1988), p. 146.
23. V. N. Obridko and B. D. Shelting, Solar Phys. **187**, 185 (1999).
24. V. N. Obridko, *Advances in Solar Connection with Interplanetary Phenomena*, Ed. by X. Feng, F. Wei, and M. Dryer (International Academic Publishers, Beijing, 1998), p. 41.
25. V. N. Obridko, V. V. Fomichev, A. F. Kharshiladze, *et al.*, Astron. Astrophys. Trans. **18** (6), 819 (2000).
26. V. T. Tsurutani, W. D. Gonzalez, A. L. Gonzalez, *et al.*, J. Geophys. Res. **100**, 21717 (1995).

Translated by V. Badin

Astronomy Reports, Vol. 47, No. 11, 2003, pp. 963–973. Translated from Astronomicheskiĭ Zhurnal, Vol. 80, No. 11, 2003, pp. 1046–1056.

Characteristics of Motions in the 2D Isosceles Three-Body Problem with Unequal Masses

V. V. Orlov[1] and A. I. Martynova[2]

[1] *Astronomical Institute, St. Petersburg State University, Universitetskiĭ pr. 28, St. Petersburg, Peterhof, 198504 Russia*

[2] *St. Petersburg Forest Technical Academy, Institutskiĭ per. 5, St. Petersburg, 194022 Russia*

Received March 25, 2003; in final form, May 8, 2003

Abstract—We analyze the general 2D isosceles three-body problem for various ratios ε of the mass of the central body to the mass of each of the other two bodies. We set the initial conditions using two parameters: the virial coefficient k and the parameter $\mu = \dot{r}/\sqrt{\dot{r}^2 + \dot{R}^2}$, where $\dot{r}$ is the relative velocity of the two outer bodies and $\dot{R}$ is the velocity of the central body relative to the center of mass of the outer bodies. We compare statistical dependences between evolutionary parameters of triple systems with various values of ε, and analyze the k and μ dependences of the number of crossings of the center of mass of the triple system by the central body and the lifetime of the system. We construct the functions $R_{\max}(r_{\max})$, where $r_{\max}$ and $R_{\max}$ are the maximum achievable distances between the outer bodies, and between the central body and the center of mass of the outer bodies in the triple system. The parameter ε proves to be the most important parameter of the problem, and determines the relationship between the measures of the regular and stochastic trajectories. However, there exist "seeds" of stochasticity, even at small $\varepsilon \sim 10^{-2}$. The measure of the stochastic orbits increases with ε; when $\varepsilon \geq 10$, virtually the entire region of the initial conditions corresponds to stochastic trajectories. © *2003 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

One well known special case of the three-body problem is the isosceles problem, in which the configuration of triple system has the form of an isosceles triangle at any time. A large number of papers have been dedicated to this problem (see, e.g., [1–9]).

Sitnikov [1] found a new class of motions—oscillating motions—in the restricted isosceles problem, in which a body with zero mass moves along a line perpendicular to the orbital plane of the two other bodies. If the binary has an elliptical orbit, the central body may experience arbitrarily distant ejections without escape (oscillations). Alekseev [2] suggested that the motions in the Sitnikov problem could be described using the methods of symbolic dynamics. He theoretically found a one-to-one correspondence between the set of all solutions to this problem and the set of all symbolic sequences of a certain form. Zare and Chesley [7] and Chesley [8] performed a similar analysis for the general isosceles three-body problem using numerical simulations.

Broucke [3] numerically found a number of periodic solutions and orbits with triple collisions in the 2D isosceles three-body problem with equal-mass bodies. Chesley and Zare [7], Chesley [8], and Orlov *et al.* [9] found a region of trajectories with finite motions in the neighborhood of the central stable periodic orbit. Chesley [8] determined the region of stability for the central orbit as a function of the mass ratio of the bodies and the angular momentum of the binary formed by the two outer bodies.

Devaney [4], Simó and Martinez [5], and Chesley and Zare [7] analyzed the manifolds of initial conditions corresponding to triple collisions. The behavior of the solutions on this manifold can be described in terms of specially chosen variables. The simplest examples of solutions with triple collisions are so-called homothetic solutions, which correspond to a linear configuration or an equilateral triangle.

Simó and Martinez [5] and Antonov and Chernin [6] determined the critical ratio of the mass of the central body to that of one of the outer bodies (which is equal to $55/4$) that separates regular and stochastic behavior of the trajectories for the special case of parabolic motion of the outer bodies and small oscillations of the central body.

In [9], we classified the types of motions occurring in the 2D, isosceles three-body problem. We analyzed the regions of escape after various numbers of triple collisions as a function of the initial conditions, which were defined using two parameters: the virial coefficient k and the parameter $\mu = \dot{r}/\sqrt{\dot{r}^2 + \dot{R}^2}$, where

1063-7729/03/4711-0963$24.00

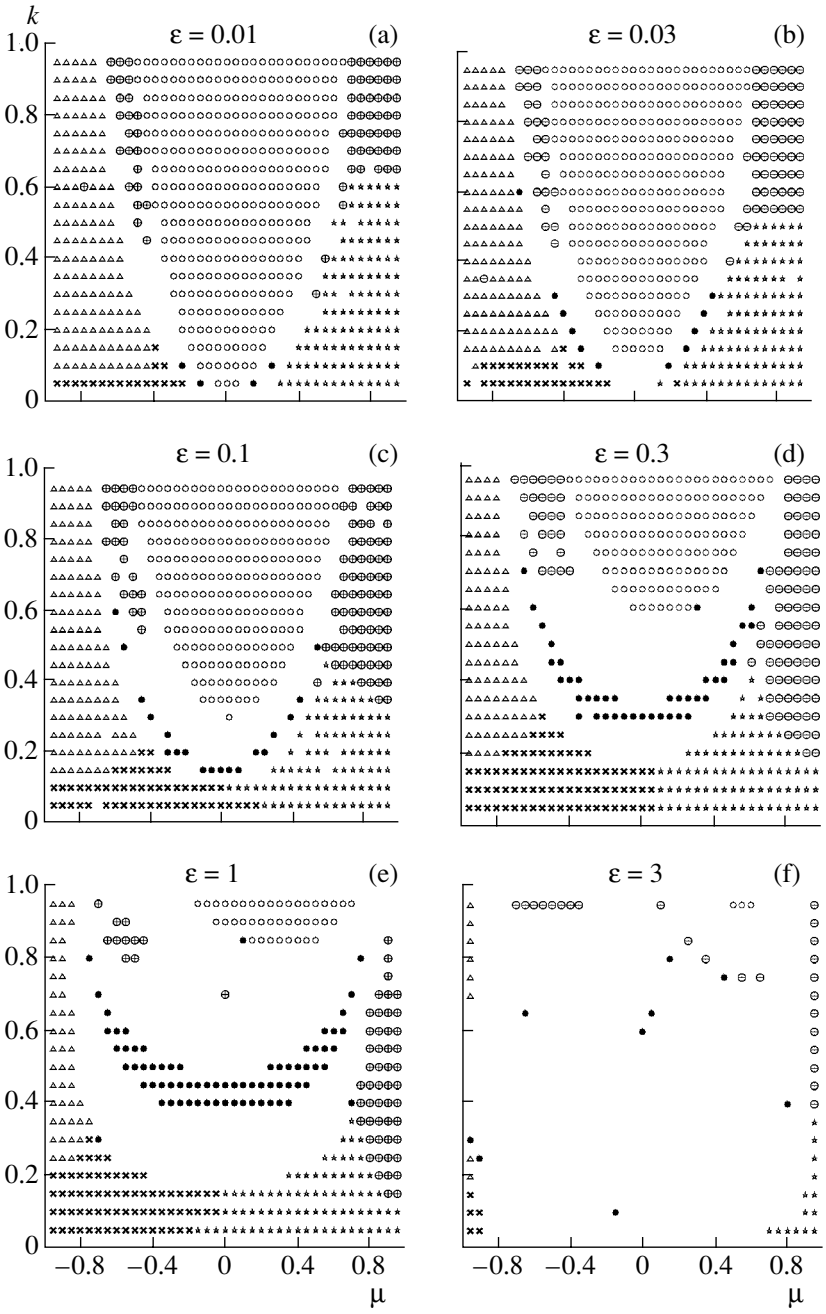

Fig. 1. Regions of initial conditions (k, μ) corresponding to trajectories for which the escape of the central body occurs within various time intervals t for various ε—the ratio of the mass of the central body to the mass of each of the outer bodies. The time intervals are denoted by circles ($t = 0$), triangles ($0 < t \leq 1\tau$), crosses ($1\tau < t \leq 2\tau$), asterisks ($2\tau < t \leq 3\tau$), circled crosses ($3\tau < t \leq 4\tau$), and dots ($t \geq 1000\tau$).

$\dot{r}$ is the relative velocity of the two outer bodies and $\dot{R}$ is the velocity of the central body relative to the center of mass of the outer bodies. The central periodic orbit generates regions of stochastic motions and regions of Lagrange-stable trajectories.

In the current paper, which is a continuation of [9], we analyze a number of other characteristics of motions in the 2D, isosceles three-body problem. Section 2 briefly describes the analysis method used, Section 3 presents our results, and Section 4 formulates our main conclusions and discusses our results.

2. ANALYSIS METHOD

In [9], we proposed a set of initial conditions such that all three bodies are collinear, the central body moves vertically upward, and the outer bodies located

at identical distances from the central body can move either toward or away from each other. Let R be the distance between the central body and the center of mass of the outer bodies, and r be the distance between the outer bodies. The masses of the outer bodies and the gravitational constant are taken to be equal to unity, and the total energy is $E = -1$. The initial conditions are determined using two parameters:

(1) the virial coefficient

$$k = \left(\frac{\dot{r}^2}{4} + \frac{\varepsilon \dot{R}^2}{\varepsilon + 2}\right) \Big/ \left(\frac{1 + 4\varepsilon}{r}\right) \qquad (1)$$

and

(2) the velocity ratio

$$\mu = \frac{\dot{r}}{\sqrt{\dot{r}^2 + \dot{R}^2}}, \qquad (2)$$

where $\dot{r}$ is the relative velocity of the outer bodies, $\dot{R} > 0$ is the velocity of the central body relative to the center of mass of the binary formed by the outer bodies, and ε is the ratio of the mass of the central body to the mass of one of the outer bodies.

We numerically integrated the equations of motion. In cases of binary collisions ($r/R < \delta = 10^{-7}$), we obtained an analytical approximation using a superposition of two two-body problems. We monitored the accuracy of this approximation by verifying that the integral of the energy was conserved. As a rule, the fractional error in the energy integral at the end of the computations did not exceed 10^{-4} (typically, $\sim 10^{-6}$).

We continued the computations until either the escape criterion [10, 11] was satisfied, or the central body reached a specified large critical distance R_c, or the computations reached a specified long critical time t_c (see below).

Let us consider various values of ε:

$$\varepsilon = \{0.01, 0.03, 0.1, 0.3, 0.5, 1, 2, 3, 10, 55/4, 30\}. \qquad (3)$$

For each ε, we scanned over a series of k and μ values in steps of 0.05:

$$k = 0.05(0.05)0.95, \qquad (4)$$

$$\mu = -0.95(0.05)0.95. \qquad (5)$$

In addition, we carried out a more detailed scanning of k and μ in the region of initial conditions in steps of 0.01 for the four values $\varepsilon = \{0.1, 0.5, 1, 2\}$:

$$k = 0.01(0.01)0.99, \qquad (6)$$

$$\mu = -0.99(0.01)0.99. \qquad (7)$$

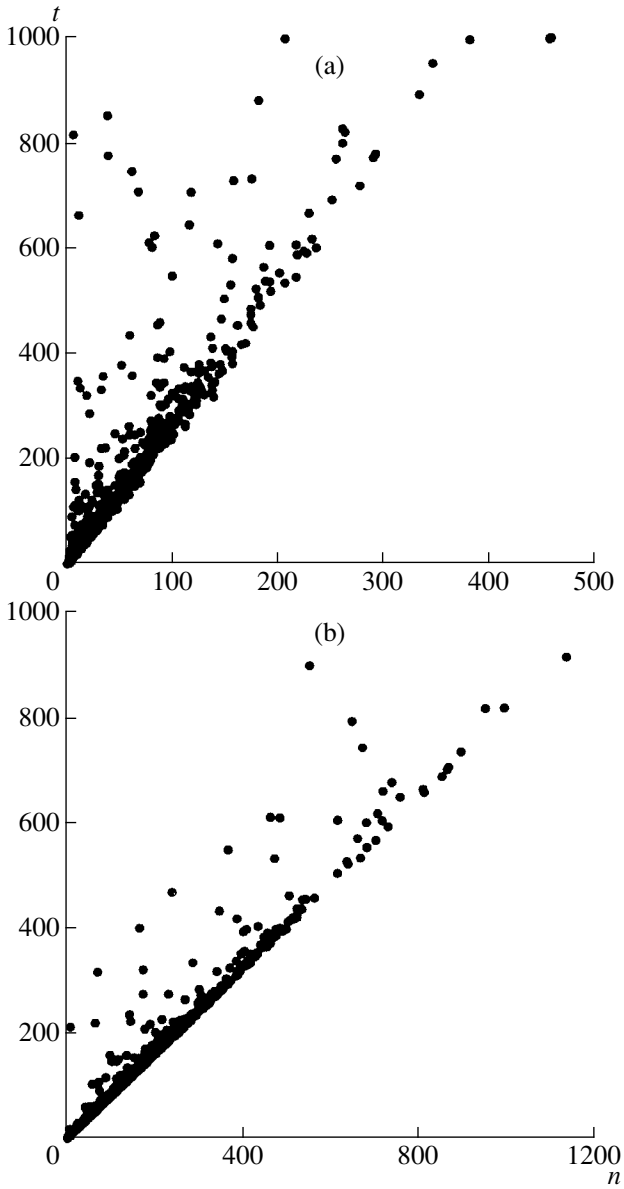

Fig. 2. Relations between the lifetime t and the number of crossings by the central body of the center of mass of the triple system n for (a) $\varepsilon = 2$ and (b) $\varepsilon = 30$.

To facilitate comparison of the results for different ε values, we used a dynamical system of units (see, e.g., [12]). The unit of distance in this system is taken to be the mean size of the system:

$$d = \frac{\sum_{i<j} m_i m_j}{2|E|} = 1/2 + \varepsilon, \qquad (8)$$

and the unit of time the mean crossing time:

$$\tau = \frac{\sum_{i<j} m_i m_j \sqrt{\sum_i m_i}}{2\sqrt{2}|E|^{3/2}} = (1/2 + \varepsilon)\sqrt{1 + \varepsilon/2}. \qquad (9)$$

Table 1. Parameters of the distributions of the number of crossings $f(n)$

ε	$n_{\max}$	$\nu_{\max}$	$\nu(n \le 10)$	$\nu(n > 200)$	$\nu(t > 1000\tau)$
0.01	1	0.42	0.98	0	0.005
0.03	1	0.39	0.95	0	0.015
0.1	1	0.30	0.94	0	0.026
0.3	3	0.23	0.85	0	0.057
0.5	3	0.25	0.82	0.001	0.082
1	3	0.23	0.64	0	0.096
2	3	0.13	0.35	0.02	0.036
3	3	0.05	0.16	0.08	0.012
10	12	0.02	0.06	0.04	0.001
30	18, 20	0.01	0.01	0.34	0

Table 2. Behavior of the maxima in the lifetime distributions $f(t)$

ε	$\Delta t_{\max}/\tau$	$\nu_{\max}$
0.01	[0, 5]	0.88
0.03	[0, 5]	0.85
0.1	[0, 5]	0.78
0.3	[0, 5]	0.66
0.5	[0, 5]	0.58
1	[0, 5]	0.43
2	[0, 5]	0.25
3	[0, 5]	0.08
10	[45, 50]	0.06
30	[45, 50]	0.03

We adopted the critical parameter values for distance and time $R_c = 100d$ and $t_c = 1000\tau$, and analyzed a total of 75 000 versions of the initial conditions.

3. RESULTS

Regions of Escape after a Specified Time

Zare and Chesley [7], Chesley [8], and Orlov *et al.* [9] identified in the regions of initial conditions zones corresponding to escape after a certain number of crossings of the line connecting the outer bodies by the central body. In addition, Zare and Chesley [7] analyzed binary collisions between the outer bodies and constructed symbolic sequences of binary collisions and crossings. Zare and Chesley [7], Chesley [8], and Orlov *et al.* [9] found continuous regions of initial conditions corresponding to small numbers n and n_2 of crossings and binary collisions, respectively. As n and n_2 increase, the pattern becomes stochastic—continuous regions are transformed into islands, broken chains, and sets of scattered points. The boundaries of the regions with $n = \text{const}$ are apparently determined by conditional disruptions and triple collisions.

Table 3. Inverse slopes k_{tn} of the boundary lines in $t(n)$ for various ε

ε	k_{tn}	$\bar{t}/\tau$	$\bar{t}'/\tau$	a
0.01	0.32	2.9	3.1	0.5
0.1	0.34	2.5	3.0	0.5
0.3	0.35	2.6	2.9	0.5
0.5	0.37	2.6	2.7	0.5
1	0.37	2.6	2.5	0.5
2	0.38	2.1	2.2	0.5
3	0.45	2.0	2.3	0.5
10	0.75	1.2	–	0.2
30	1.2	0.8	–	0.2

Table 4. Results of analysis of $r_{\max}(t)$

ε	$r^{th}_{\max}$	$r^{h}_{\max}$	$r^{st}_{\max}$	t_1	t_2	ν
0.01	1.04	1.02	1.02	1.0	1.8	0.24
0.03	1.12	1.09	1.06	1.0	1.8	0.23
0.1	1.4	1.3	1.25	1.2	2.3	0.24
0.3	2.2	1.3	1.2	1.2	2.3	0.13
0.5	3	2.4	2.0	2.4	4.6	0.24
1	5	4.1	3.1	4.0	7.5	0.19
2	9	8.0	5.0	7	14	0.13
3	13	13.0	12.5	11.5	20	0.05
10	41	41	–	–	–	–
30	121	121	–	–	–	–

Here, we analyze the regions of escape after a specified time t in units of the mean crossing

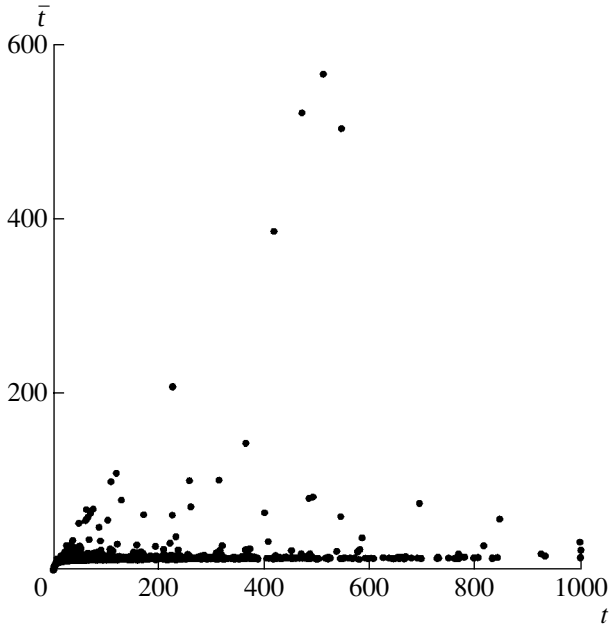

Fig. 3. Dependence of the mean time between consecutive crossings $\bar{t}$ on the lifetime t for $\varepsilon = 3$.

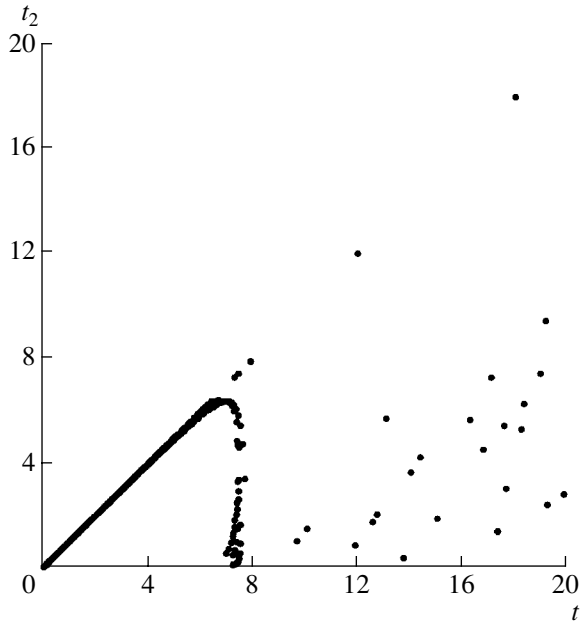

Fig. 5. Dependence of the time of the second crossing t_2 on the lifetime t for $\varepsilon = 1$.

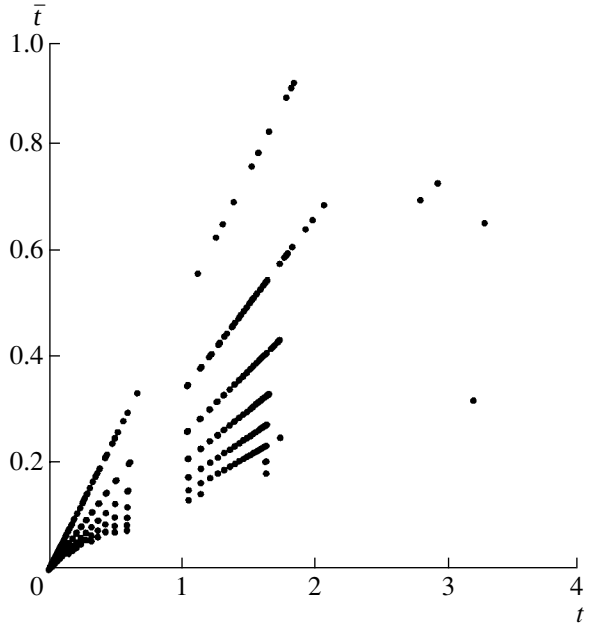

Fig. 4. Fragment of the $\bar{t}$ dependence on t for systems with short lifetimes for $\varepsilon = 0.01$.

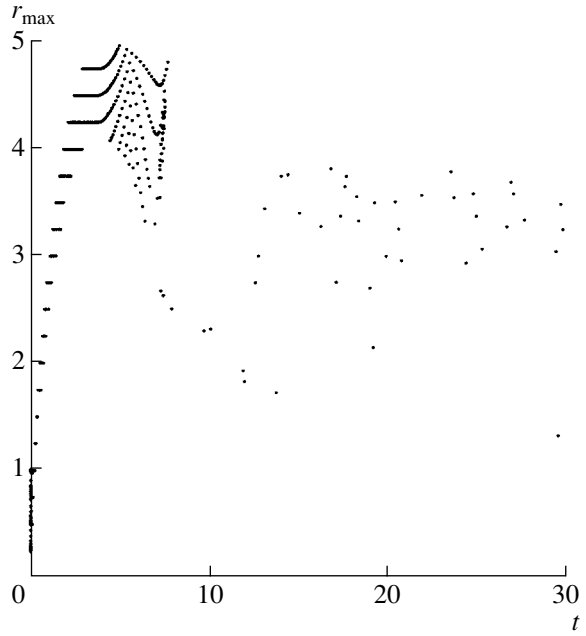

Fig. 6. Dependence of the maximum achieved distance between the outer bodies, $r_{\max}$, on the lifetime t for $\varepsilon = 1$.

time τ. As the time of disruption of the triple system, we adopted the time of the last crossing resulting in the escape of the central body along a hyperbolic orbit (true disruption) or a very distant ejection ($R > R_c = 100d$) (conditional disruption).

Figures 1a–1e show the results for various ε. The time step is 1τ. We can see the regions of escape after the first crossing, with $t = 0$. In addition, the figures also show a sequence of adjoining continuous regions running from the top left corner to the top right corner and bordering the region with $t = 0$. The boundaries between these regions are rather sharp. The lifetime t increases smoothly along this sequence.

Table 5. Results of analysis of $R_{\max}(t)$

ε	$R^{st}_{\max}$	$R^{h}_{\max}$	ν_e	ν_h
0.01	1.25	1.35	0.80	0.10
0.03	1.25	1.4	0.75	0.15
0.1	1.2–1.5	1.5	0.64	0.20
0.3	1.25–1.5	1.5	0.41	0.33
0.5	1.5–2.4	2.2	0.27	0.41
1	2.3–3.5	3.6	0.10	0.57
2	3.1–4.1	5.0	0.019	0.77
3	≈14	6.5	0.003	0.93
10	–	22	0	0.84
30	–	60	0	0.95

Distribution Functions of the Number of Crossings and Lifetime

Let us now consider the distribution function $f(n)$ of the number of crossings preceding the escape of the central body for various masses of this body (ε). Table 1 summarizes the main results. We considered a total of 741 versions of the initial conditions for each ε. The columns in Table 1 present (1) ε, (2) the number of crossings resulting in the highest peaks of the function $f(n)$, $n_{\max}$, (3) the corresponding fraction $\nu_{\max}$, (4) the fraction of disrupted systems with small numbers of crossings ($n \le 10$), (5) the fraction of systems with large numbers of crossings $n > 200$, and (6) the fraction of systems that were not disrupted during a time $t = 1000\tau$. It is clear from Table 1 that the maximum of $f(n)$ decreases and shifts toward higher n with increasing ε. When $\varepsilon \ge 0.5$, disrupting systems with $n > 200$ appear, and, on average, the fraction of such systems increases with ε. The fraction of undisrupted systems with lifetimes $t > 1000\tau$ first increases, reaches its maximum (about 10%) when $\varepsilon = 1$ (equal masses), then decreases to zero. The fraction of triple systems with large numbers of crossings increases, and the fraction of systems with small numbers of crossings ($n < 10$) decreases sharply with ε.

The lifetime distribution functions $f(t)$ are similar to the $f(n)$ distributions. These results are summarized in Table 2, which gives ε, the lifetime intervals Δt in which $f(t)$ reaches its maxima, and the fraction of systems in these intervals $\nu_{\max}$. When $\varepsilon \le 3$, the $f(t)$ functions have their maxima in the interval $t \in [0, 5]\tau$, and these maxima decrease with increasing ε. When $\varepsilon > 3$, the maxima shift toward longer t, and the maxima themselves become broader.

At the same time, the behavior of $f(n)$ differs somewhat from that of $f(t)$: after the main maximum, the distributions $f(n)$ are more uniform than $f(t)$, whose main maxima are more prominent. The main maximum of $f(t)$ remains visible at large $\varepsilon > 10$, whereas it is virtually absent for $f(n)$—the $f(n)$ distributions have the form of plateaus that slowly descend with increasing n.

Relation between the Number of Crossings and Lifetime

Let us now consider the dependences between n and t for various ε. Figures 2a and 2b show examples of such dependences for $\varepsilon = 2$ and 30, respectively. When $\varepsilon \le 0.3$, most of the points are concentrated near the origin (small n and t). There is a sequence of points along the t axis with $n < 25$ (triple systems with very distant ejections), a set of undisrupted trajectories with $t > 1000\tau$, and a few points along the interval connecting the origin and the points corresponding to stable trajectories. When $\varepsilon = 1$, the points are scattered inside some angle with its vertex at the coordinate origin; the set of points along the vertical axis and the set of stable trajectories with $t > 1000\tau$ and $n > 400$ are preserved. When $\varepsilon = 2$ (Fig. 2a), most of the points are located along a line that nearly passes through the origin and forms the right-hand boundary of the set in question. The region corresponding to stable orbits is located slightly to the right of this boundary. The number of points along the t axis decreases. When $\varepsilon = 30$ (Fig. 2b), the overwhelming majority of points are concentrated toward the line, with a small scatter of points above it. There are virtually no points along the t axis.

Note that every dependence $t(n)$ has a "zone of avoidance" of points. This zone is bounded from above by a straight line passing approximately through the coordinate origin. The set of stable trajectories with $t > 1000\tau$ is located slightly to the right of this boundary. The existence of this "angle of avoidance" confirms the absence of triple systems with large numbers of crossings ($n > 5$) and comparatively short lifetimes. Note that Zare and Chesley [7] argue that any number of crossings is possible in a comparatively short time, in contradiction with our numerical simulations. We apparently failed to detect these points near the boundary of the region of initial conditions because our step size was too large. That is, for a fixed number of crossings, there appears to be a fairly sharp lower boundary for the lifetime of the triple system. The concentration of points toward this line increases with increasing ε.

Table 3 gives the inverse slopes $k_{tn} = \Delta n/\Delta t$ of the boundary lines for various ε. No points were detected near the hypothetical boundary line for $\varepsilon =$

0.03, and we therefore were not able to determine the slope in this case. When $\varepsilon \leq 2$, the slope k_{tn} increases gradually with the mass of the central body. Further increase of ε results in a strong increase of k_{tn}. Physically, k_{tn} is the mean number of crossings during one mean crossing time τ. Thus, the mean number of crossings during τ increases with the mass of the central body.

Dependence of the Mean Time between Crossings on Lifetime

We will now construct the dependences of the mean time between consecutive crossings $\bar{t}$ on the lifetime t for various ε. Figure 3 shows a typical example. The points in this figure are located inside a sector bounded by two lines—one nearly horizontal and one inclined. Outside this sector (along the coordinate axes), points are found only near $t = 0$. We found a similar pattern for other ε values. The boundary $\bar{t}$ of the lower zone of avoidance (given in column 3 of Table 3) is close to the $\bar{t}'$ values for stable orbits (given in column 4 of Table 3).

The zone of avoidance along the horizontal axis is due to the well defined lower horizontal boundary of the mean time between triple encounters for systems with long lifetimes: $t > 10\tau$. The upper zone of avoidance indicates an absence of triple systems with $\bar{t} > t$, which is natural. The lower boundary of this zone of avoidance is the line $\bar{t} = at$. In most cases, the triple systems are concentrated near the lower, nearly horizontal, boundary. The last column of Table 3 gives the coefficients a corresponding to the lower boundary of the upper zone of avoidance. The coefficient a increases abruptly in the interval $\varepsilon \in (3, 10)$, possibly due to a qualitative change in the behavior of the interactions in the triple systems. This is also supported by the change in the lower boundary of the mean time $\bar{t}$ between consecutive crossings for the horizontal line bounding the lower zone of avoidance as a function of ε (see column 3 in Table 3).

Let us now analyze in more detail the relationship between $\bar{t}$ and t for short lifetimes of the systems. Figure 4 shows an example of such a dependence for $\varepsilon = 0.01$. The points are concentrated along a system of lines emerging from the coordinate origin. Each of these lines corresponds to some number of crossings n preceding the disruption of the triple system: the smaller n, the higher the corresponding line. As the central mass (i.e., ε) increases, the number of lines emerging from the coordinate origin decreases, and a system of parallel lines emerging from the upper boundary of the lower zone of avoidance appears, with a small number of points on each of these lines.

We can also analyze the dependence of the time of the second crossing t_2 on the lifetime t of the triple system. Figure 5 shows an example of the $t_2(t)$ dependence for $\varepsilon = 1$. At small values $\varepsilon \leq 1$, sequences of points can be identified that emerge from the coordinate origin, bend, and descend approximately perpendicular to the horizontal axis. The points outside these sequences and under the bisectrix $t_2 = t$ are scattered randomly. At large values $\varepsilon \geq 10$, these sequences disappear, and the points are scattered randomly within a horizontal band.

Analysis of the Maximum Achieved Distances between the Bodies

Consider now the dependences between the maximum achieved distance between the outer bodies $r_{\max}$ and the lifetime t of the triple system for various ε. The results are presented in Table 4 and Fig. 6.

We find from the equation for the zero-velocity curve (see Section 3.1 in [9]) the theoretical estimate $r^{th}_{\max} = 1 + 4\varepsilon$ for $R = 0$. The theoretical maximum distances $r^{th}_{\max}$ are given in the second column of Table 4.

It is evident from the $r_{\max}(t)$ dependences (Fig. 6 for $\varepsilon = 1$) that the triple systems break up into several populations:

(1) a group along the vertical axis with $t = 0$ (escapes after the first triple encounter),

(2) a band stretching to the right and upward from the vertical axis,

(3) a family of bent curves, and

(4) a horizontal band of scattered points corresponding to the zone of irregular motions.

This pattern is preserved when $\varepsilon \leq 3$. When $\varepsilon \geq 10$, the first three populations disappear and only the horizontal band remains. The third column of Table 4 gives the $r^{h}_{\max}$ values for the upper boundaries of these horizontal bands, which increase with ε. At large values of ε, some points lie above the theoretical boundary $r^{th}_{\max}$ due to the accumulation of computing errors. The fourth column gives the mean values for the stable triple systems, $r^{st}_{\max}$, which are located in the horizontal band.

The third population fits in with the classification of triple systems based on the number of crossings n described above. Each of the bent curves corresponds to some n. Let us estimate the times t_1 and t_2 corresponding to the beginning and end of this region. These times are given in the fifth and sixth columns of Table 4. The last column gives the fraction ν of triple systems in this region relative to the total number of triple systems. On average, this fraction decreases with increasing ε. There are no such systems when $\varepsilon \geq 10$ (dashes in the table).

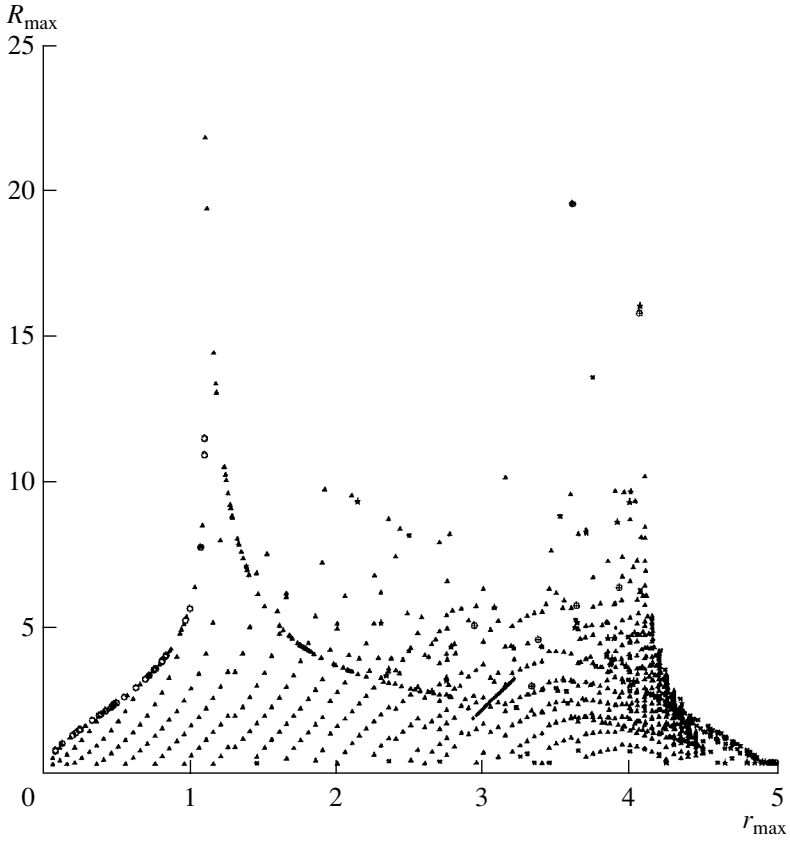

Fig. 7. Dependences of $R_{\max}$ on $r_{\max}$ for $\varepsilon = 1$ for triple systems that disrupt after $n = 1$ (circles), $n = 2$ (triangles), $n = 3$ (crosses), $n = 4$ (asterisks), and $n = 5$ (circled crosses) crossings. The points correspond to stable orbits.

The four populations we have pointed out here also show up in the $R_{\max}(t)$ dependences. Table 5 gives the intervals of $R^{st}_{\max}$ for stable triple systems, as well as the values $R^{h}_{\max}$ for the approximately horizontal line bounding from below the region in which the points are located. The region of points also has a well defined upper envelope whose shape is close to parabolic.

Table 5 also gives the fraction ν_e of the systems whose $R_{\max}$ values are close to the critical distance R_e used in the disruption criterion, and also the fraction ν_h of systems located in the fourth population, which is bounded from below by a nearly horizontal line and from above a parabola. The fraction ν_e decreases with increasing ε, and such systems are virtually absent at $\varepsilon \geq 10$. On average, the fraction ν_h increases with ε.

Figure 7 shows the relationship between $r_{\max}$ and $R_{\max}$ for $\varepsilon = 1$ for triple systems that disrupt after $n = \{1, 2, 3, 4, 5\}$ crossings and for stable triple systems. The points fill a region whose shape resembles the base of a vase and is somewhat asymmetric. Inside this region, we can see linear structures and isolated subsets of points corresponding to escapes after some number of crossings. Points with $n = 2$ form a family of regular lines contained between the left-hand boundary ($n = 1$) and the envelope ($2 \leq n \leq 5$) separating regular and stochastic sets of points. Only cases with $n = 2$ are realized in the envelope itself, in its upper part, while points with $2 \leq n \leq 5$ are located in its lower part. Stable orbits fill two sets of points: an almost rectangular region and a compact set at $r_{\max} \approx 4$ and $R_{\max} \approx 5$. Points with $n = 1$ form the left-hand boundary of the entire set. Points with $n = \{3, 4, 5\}$ form mutually penetrating sets in the right-hand side of the figure. The fairly well defined envelopes of these sets can be easily traced. There are regular linear structures and sets of scattered points for $n = 2, 3$.

Analysis of the Orbit Topology

We will now analyze the characteristic features of the trajectories of bodies in continuous regions of initial conditions with equal n and variations of the topological properties when going from a zone to a neighboring zone. Points corresponding to initial conditions leading to triple collisions are located at the zone boundaries.

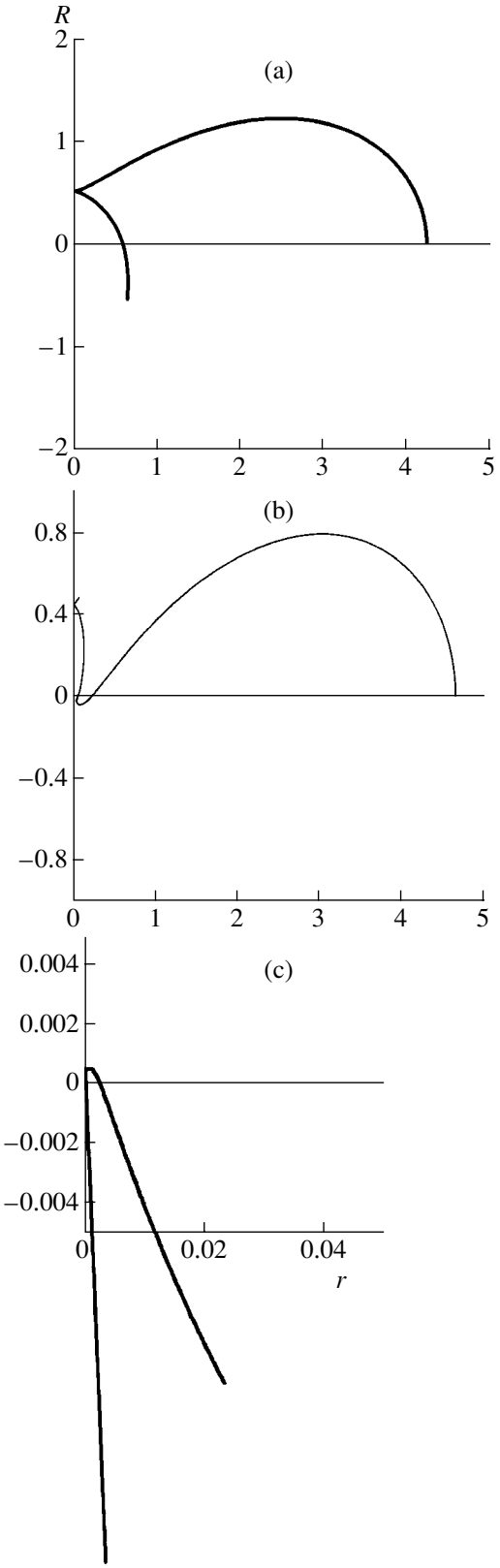

Fig. 8. Trajectories of motion for (a) $(k, \mu) = (0.15, 0)$ and $\varepsilon = 1$ (number of crossings $n = 2$), (b) $(k, \mu) = (0.06, 0)$ and $\varepsilon = 1$ (number of crossings $n = 3$), and (c) $(k, \mu) = (0.01, 0)$ and $\varepsilon = 1$ (number of crossings $n = 4$).

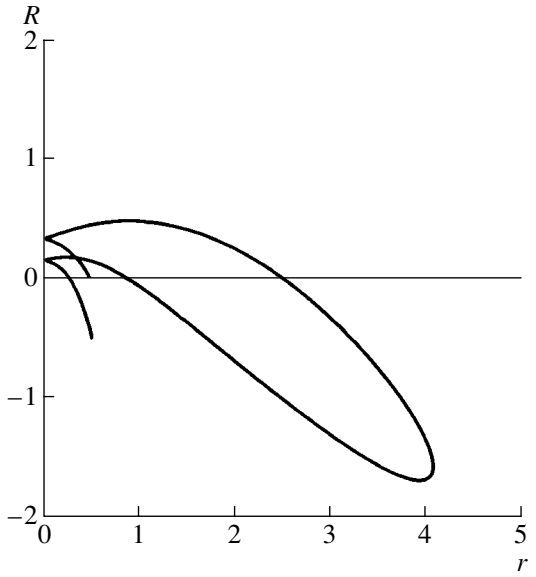

Fig. 9. Trajectories of motion for $(k, \mu) = (0.9, -0.5)$ and $\varepsilon = 1$. The number of crossings is $n = 4$.

Consider, for example, variations of the orbit topology for $\mu = 0$ $(\varepsilon = 1)$ and $k = 0.15$ $(n = 2)$, $k = 0.06$ $(n = 3)$, and $k = 0.01$ $(n = 4)$. Figures 8a–8c show the corresponding trajectories. The number of times the trajectory crosses the horizontal axis $R = 0$ before the first binary collision increases with n. A triple encounter resulting in the escape of the central body immediately follows the binary collision.

A somewhat different topology can be observed in the relatively small continuous zone with $n = 4$ for $\varepsilon = 1$ in the upper left-hand part of the region of initial conditions [9, Fig. 7]. Figure 9 shows an example of a trajectory with initial conditions in this zone ($k = 0.9$, $\mu = -0.5$). In this case two rather than one binary collisions of the outer bodies precede the escape.

We can suppose that the topological structure of the trajectories remains unchanged within each continuous region with a particular n. This structure is reorganized after the system undergoes a triple collision (see also [7]).

4. CONCLUSIONS AND DISCUSSION

The results of this work lead us to the following conclusions.

(1) The distribution functions $f(n)$ depend on ε: when $\varepsilon \leq 0.1$ (small central masses), the distribution function reaches its maximum at $n = 1$; when $0.3 \leq \varepsilon \leq 3$ (comparable masses for all the bodies), the distribution reaches its maximum at $n = 3$; and when $\varepsilon \geq 10$ (large central masses), $f(n)$ has the

form of a plateau. The maximum broadens with increasing ε.

(2) The dependences of the lifetime on the number of crossings $t(n)$ differ substantially for small ($\varepsilon \leq 0.3$) and large ($\varepsilon \geq 3$) masses of the central body. In the case of small ε, most of the points are concentrated along a system of lines near the coordinate origin, whereas, when ε is large, the overwhelming majority of points are located along a straight line emerging approximately from the coordinate origin. All the $t(n)$ dependences have angles of avoidance located below this line and along the vertical axis. The mean number of crossings per mean crossing time increases with the central mass.

(3) In the dependences of the mean time between triple encounters on the mean lifetime $\bar{t}(t)$, points are located near the coordinate origin and inside a sector bounded by two lines, one nearly horizontal and one inclined. We have determined the parameters of these lines as functions of ε. When ε is small, the set of points near the coordinate origin dominates, while the population of points inside the sector dominates when ε is large. Points corresponding to stable orbits are located at the end of the horizontal line.

(4) Our dependences (i) $r_{\max}(t)$ and (ii) $R_{\max}(t)$ of the maximum distances (i) between the outer bodies and (ii) between the central body and the center of mass of the outer bodies achieved in the course of the system's evolution for $\varepsilon \leq 3$ show several populations of points corresponding to escapes after some number of crossings. Only one population located in a horizontal band remains when $\varepsilon \geq 10$.

(5) In the $R_{\max}(r_{\max})$ dependences, the points fill a region with well-defined boundaries whose shape resembles the base of a vase. This region contains both regular structures and sets of isolated scattered points, sometimes bearing signs of regularity. Stable orbits form compact sets of points.

Let us now give a physical interpretation of the results obtained. Inside the band of continuous variations of the lifetime of the triple system as a function of the initial conditions (Fig. 1), the lifetime increases with μ. This is due to the direction of the relative motion of the outer bodies at the initial time. When μ is negative, the outer bodies approach each other, and undergo their binary collision earlier than in the case of positive μ. Therefore, the disruption of the triple system that takes place near the binary collision occurs somewhat earlier in the case of negative μ. In this case, for the same value of the virial coefficient k within the given band, the system can experience different numbers of crossings of the $R = 0$ axis by the central body. Note that this number increases with μ, although the lifetime itself remains virtually unchanged. This is due to the change in the orbit topology—additional crossings of the $R = 0$ axis by the central body accompanied by a very short ejection appear or disappear.

The change in the behavior of the distributions of the numbers of crossings n and lifetime t with increasing mass of the central body (ε) is due to the decrease in the regularity and increase in the stochasticity of the trajectories. When $\varepsilon \geq 10$, the regions of regularity disappear almost completely, and the evolution of the triple systems acquires a quasi-random nature: the probability of escape is nearly constant within rather wide intervals of n and t. The distribution functions have the shape of plateaus.

The quantity ε is the dominant parameter of the problem, since it is ε that determines the fractions of regular and stochastic trajectories. In a whole series of dependences between the evolutionary parameters, increasing ε results in a redistribution of the corresponding points from a system of lines near the coordinate origin to a band corresponding to stochastic orbits with long lifetimes. The manifold of points in these dependences is transformed from a one-dimensional to a two-dimensional structure, and disruption is transformed from a nearly deterministic to a quasi-random process. However, "seeds" of stochasticity are present even when ε is small. They are manifest in the region of the initial conditions (k, μ) in the form of a narrow "slit" near the set of stable orbits. The areas of both this slit and of the region of stable trajectories increase with ε up to $\varepsilon = 1$ (the case of equal masses). With further increase of the mass of the central body, the area of the region of stable orbits decreases and tends to zero, and the area of the set of initial conditions corresponding to stochastic trajectories increases strongly and tends to the area of the entire initial-condition region. Traces of regularity are preserved only near the boundary, and with further increase of ε vanish entirely.

At larger ε values, a large number of trajectories appear, virtually filling the region of possible motions. The evolution of these trajectories is close to ergodic.

ACKNOWLEDGMENTS

This work was supported by the Program of Support for Leading Scientific Schools of Russia (grant no. 00-15-96775), the Russian Foundation for Basic Research (project no. 02-02-17516) and the Universities of Russia Program of the Ministry of Education of the Russian Federation (grant no. UR.02.01.027). We are grateful to V.G. Surdin for a number of valuable comments on the draft of the paper.

REFERENCES

1. K. A. Sitnikov, Dokl. Akad. Nauk SSSR **133**, 303 (1960).
2. V. M. Alekseev, *Lectures on Celestial Mechanics* [in Russian] (NITs Regul. Khaotich. Dinamika, Izhevsk, 2001), p. 14.
3. R. Broucke, Astron. Astrophys. **73**, 303 (1979).
4. R. L. Devaney, Invent. Math. **60**, 249 (1980).
5. C. Simó and R. Martinez, Celest. Mech. **41**, 179 (1988).
6. V. A. Antonov and A. D. Chernin, Pis'ma Astron. Zh. **19**, 768 (1993) [Astron. Lett. **19**, 312 (1993)].
7. K. Zare and S. Chesley, Chaos **8**, 475 (1998).
8. S. Chesley, Celest. Mech. Dyn. Astron. **73**, 291 (1999).
9. V. V. Orlov, A. V. Petrova, and A. I. Martynova, Mon. Not. R. Astron. Soc. **333**, 495 (2002).
10. J. Yoshida, Publ. Astron. Soc. Jpn. **24**, 391 (1972).
11. C. Marchal, Celest. Mech. **9**, 381 (1974).
12. Zh. P. Anosova and V. V. Orlov, Tr. AO LGU **40**, 65 (1985).

Translated by A. Dambis