Astronomy Letters, Vol. 31, No. 10, 2005, pp. 645–662. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 723–742.

Nature of the Optical and Ultraviolet Variability of the Nucleus of NGC 4151: A New Concept

V. M. Lyuty*

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

Received April 27, 2005

Abstract—We have analyzed the optical (UBV) and ultraviolet (λ1000–2700 Å) observations of the nuclear variability of the Seyfert galaxy NGC 4151 in the period 1987–2001 (the second cycle of activity). The fast (tens of days) and slow ($\sim$10 years) components of the nuclear variability, F and S, respectively, are shown to be completely different, but thermal in nature. We associate the S component with the formation and evolution of an accretion disk and the F component (flares) with instabilities in the accretion disk and their propagation over the disk in the form of a shock wave. The S component is present not only in the optical, but also in the ultraviolet range, with its amplitude being comparable over the entire range λ1000–5500 Å under study. The amplitude of the average flare (the F component) doubles as the wavelength decreases from 5500 to 1000 Å, while the rise time of the brightness to its maximum Δt (the variability time scale) decreases from $27\overset{\mathrm{d}}{.}6 \pm 0\overset{\mathrm{d}}{.}3$ to $6^{\mathrm{d}} \pm 2^{\mathrm{d}}$. The brightness decline (flare decay) time decreases by a factor of 16. The extinction in the ultraviolet is shown to have been grossly underestimated: beginning from the first *IUE* data, only the extinction in our Galaxy, $A_V = 0\overset{m}{.}12$, has been taken into account. A proper allowance for the total extinction, i.e., for the extinction in the nucleus of NGC 4151 as well ($A_V = 1^m$, $A_{1450} = 3\overset{m}{.}8$) leads to a large increase in the luminosity of the variable source in the nucleus of NGC 4151: $L = (6–8) \times 10^{46}$ erg s^{-1}. The spectral energy distribution for the variable source (λ950–5500 Å) agrees well with two Planck distributions: $T_e = 65\,000$ ($\lambda_{\max} = 450$ Å) and 8000 K. The radiation with $T_e = 8000$ K is the reprocessing of the bulk of the ultraviolet radiation by the accretion disk with a lag of 0.5–0.6 days in the V band. The lag in the $U–B$ variability of the slow component revealed the existence of an extended broad line region (EBLR) at an effective distance of 1.5 lt-years, as confirmed by spectroscopic data obtained at the Crimean Astrophysical Observatory. This yields the following mass of the central object in NGC 4151: $M_{\mathrm{c}} = (1–3) \times 10^9 M_\odot$. The luminosity of the variable source then accounts for 50–60% of L_{Edd} rather than 1–2%, as has been thought previously. In general, the pattern of ultraviolet and optical variability in NGC 4151 agrees excellently with the theory of disk accretion instability for a supermassive black hole suggested by N. Shakura and R. Sunyaev 30 years ago: the energy release is at a maximum in the ultraviolet (in the case under consideration, at λ450 Å), the luminosity is $\sim 10^{47}$ erg s^{-1} for $M_{\mathrm{c}} \sim 10^9 M_\odot$ (several tens of percent of L_{Edd}), and the variability time scale ranges from several days to many years. © *2005 Pleiades Publishing, Inc.*

Key words: *active galactic nuclei, NGC 4151, UBV photometry, variability, ultraviolet variability, color analysis, accretion disk, accretion disk instabilities.*

1. INTRODUCTION

The discovery of quasars in the early 1960s marked the beginning of a totally new trend in astrophysics—research on active galactic nuclei (AGNs): quasars, Seyfert galactic nuclei, and Lacertids (BL Lac objects). Seyfert galaxies (SyG), whose active investigation began in the late 1960s, are of particular interest as objects that are brighter and, hence, more accessible to observation. At present, the universally accepted viewpoint on the nature of the variability of AGNs is disk accretion onto a supermassive compact object (a black hole?), with all features of the variability manifestations being attributable to an accretion disk (AD).

One of the most remarkable features in the light curves of AGNs is the presence of two variability components with different time scales: slow (several thousand days) brightness variations and fast (several tens of days) flares superimposed on them. In what follows, these are referred to as the S (slow) and F (flare) components, respectively, for short. Lyuty and Pronik (1975) were the first to point to the presence of two variability components in the light curves of SyG nuclei. Analyzing the photographic light curves

*E-mail: lyuty@sai.crimea.ua

for 20 AGNs, McGimsey *et al.* (1975) independently not only pointed out the presence of two components, but also identified four types of light curves: (1) the F component dominates, and the S component is very weak or completely absent; (2) the amplitude of the S component is much larger than that of the F component; (3) the two components have approximately equal amplitudes; and (4) occasional flares. Subsequently, based on the photographic light curves for 144 objects, Pica *et al.* (1988) refined this classification by excluding occasional flares.

At present, there is no doubt that the two variability components of AGNs actually exist, since a correlation between the flare amplitude and the slow variability amplitude has been found at least for several AGNs: NGC 4151 (Lyuty and Oknyanskij 1987; Doroshenko *et al.* 2001), 3C 120, and 3C 390.3 (Hagen-Thorn 1987a). This is a strong argument against the attempts to represent the light curves of AGNs as a superposition of random flares of unknown origin (Terebizh *et al.* 1989) or random supernova explosions in the galactic nucleus (Aretxaga and Terlevich 1994). However, another question that arose immediately with the assumption about the existence of two variability components in AGNs, the question about the nature of the components, in particular, whether the S and F components are identical or different in nature, remains unanswered. One-color (Pica *et al.* 1988) or even three-color (McGimsey *et al.* 1975) photographic observations are no longer enough to solve this question; high-accuracy photoelectric observations whose analysis can reveal the physical nature of the radiation are required.

Attempts to elucidate the nature of the S and F components have been made previously. Thus, for example, analyzing the nuclear variability of the Seyfert galaxies NGC 1275, 3516, 4151, 3C 120, and others by the method of flux–flux diagrams, Hagen-Thorn (1987b) concluded that the color indices of the S and F components are equal, i.e., the two components are identical in nature, while the variability mechanism is nonthermal. Interpreting the light curve for the nucleus of NGC 4151 as a superposition of random flares, Terebizh *et al.* (1989) also performed a color analysis of the nuclear variability in NGC 4151 by the method of flux–flux diagrams and concluded that the components have identical color characteristics. However, this result is a corollary of the component separation technique: when averaged by a moving polynomial, the S component already includes the F component, while when subtracted from the observed light curve, the F component is partially subtracted; i.e., the difference between the components is minimized.

A color analysis by the method of flux–flux diagrams leaves something to be desired. Of course, the variable part of the flux can be separated by this method independently of the contribution from the constant part (the surrounding galaxy). In this case, if the color indices of the variable source do not vary with brightness, then there will be a straight line in the diagram, but the reverse is not necessarily true (Doroshenko and Lyuty 1994). In addition, neither Hagen-Thorn (1987b) nor Terebizh *et al.* (1989) noticed that the scatter of data points (measurements) in the flux–flux diagrams exceeded the mean error of a single measurement by several tens of times! Incidentally, this was pointed out by Doroshenko and Lyuty (1994). Since my high-accuracy (the error of a single measurement is 1–3%) photoelectric observations (Lyuty 1972, 1977, 1979) were analyzed in both papers, the scatter by several tens of percent can in no way be the result of errors.

NGC 4151 is the brightest Seyfert galaxy with a large variability amplitude. At present, a large number of not only optical, but also infrared, ultraviolet (UV), and X-ray observations of its nuclear variability as well as spectroscopic variability observations are available. Therefore, NGC 4151 serves as a kind of a test in investigating the variable sources in AGNs. Suffice it to say that more than 300 papers with NGC 4151 in their titles have been published in the past fifteen years. Regular photoelectric three-color (UBV) and five-color ($UBVRI$) observations of the nuclear variability in NGC 4151 were begun, respectively, in 1968 (see Lyuty and Doroshenko (1999) and references therein) and 1989 (Doroshenko *et al.* 2001; Merkulova *et al.* 2001). However, most of the papers still pertain to spectroscopic studies even after it has been realized that the spectroscopic variations in AGNs are the secondary phenomena determined by continuum variability. The most comprehensive spectroscopic study of NGC 4151 has been (and is currently being) carried out at the Crimean Astrophysical Observatory (CrAO) since 1988 (Malkov *et al.* 1997; Sergeev *et al.* 2001).

UV (the dedicated IUE satellite) observations of NGC 4151 were begun in 1978 mostly simultaneously (or almost simultaneously) with optical and, in certain cases, X-ray observations (Ulrich *et al.* (1991) and references therein). Some of these observations were performed at the minimum and the second (B) cycle of activity. The last observations from IUE were made in 1993 (Crenshaw *et al.* 1996). Individual observations were performed with the Hopkins Ultraviolet Telescope (HUT) and the Orbiting and Retrievable Far and Extreme Ultraviolet Spectrometer (ORFEUS) during flights of the Shuttles *Columbia* and *Endeavour* (Kriss *et al.* 1992, 1995; Espey *et al.* 1998; Espey 1999) as well as with the Hubble Space Telescope (HST–SPIS) (Crenshaw

et al. 2000) in the period 1990–1999. Remarkably, the two components are also present in the UV light curve (Mittaz *et al.* 1990; Ulrich *et al.* 1991; Lyutyi 2005).

The X-ray observations of NGC 4151 were begun much earlier, in 1970, from the first dedicated UHURU X-ray satellite (Gursky *et al.* 1971). The first long series of observations was obtained from the Ariel V satellite (Lawrence 1980) and yielded two main conclusions: the slow variability component is completely absent, and the flare duration is very short, an order of magnitude shorter (0.5 day) than that in the optical range. The 2–10 keV X-ray luminosity was comparable to the optical one, while the 1–2000 keV luminosity was an order of magnitude higher than the combined optical and infrared luminosity. These results led to the subsequently universally accepted idea that the main energy release in the disk accretion model takes place in the X-ray range in a small ($<10^{15}$ cm) region.

A good correlation between the UV and optical and, in certain cases, X-ray variabilities without a noticeable lag between the long-wavelength and short-wavelength variabilities seems to confirm this conclusion. Whereas some of the authors consider two possibilities—variability due to local instabilities in the AD or the AD irradiation by a central variable X-ray source (Ulrich *et al.* 1991)—others definitely favor the second possibility (Edelson *et al.* 1996). However, the mechanism of energy release in the X-ray range itself remains unclear (the thermal mechanism requires a very high temperature, $\sim 5 \times 10^6$ K, for the range 1–10 keV).

By far the most remarkable thing in the variability history of NGC 4151 in the past decades was a five-year-long (1984–1989) minimum, when not only did the brightness of the variable source in the nucleus fall to the level of the galaxy through a 27″ aperture, but also the broad Hβ line wings disappeared (Lyuty *et al.* 1984; Penston and Perez 1984). This is most likely attributable to the total dissipation of the accretion disk in the 1968–1984 cycle of activity (cycle A). A new cycle of activity of NGC 4151 (cycle B) began in 1989–1990; its maximum was observed in 1995–1996 (Lyuty *et al.* 1998; Lyuty and Doroshenko 1999; Oknyanskij *et al.* 1999; Doroshenko *et al.* 2001). The mean color indices of the variable source in the second cycle of activity were approximately the same as those in the first cycle, but the luminosity at the maximum doubled; the luminosity of the variable source in cycle B increased by several tens of times in less than 10 years (Lyuty and Doroshenko 1999). This behavior was interpreted as the appearance and evolution of a new accretion disk.

The large increase in the variability amplitude of both the F and S components in cycle B makes it possible to separate these components with a much smaller relative error than in cycle A. Separating the fast and slow variability components in the nucleus of NGC 4151, color analysis of three-color photometry, analysis of UV observations, including those performed simultaneously with optical observations (1993), and elucidating the nature of the optical and UV variabilities based on this analysis are the main objectives of this study.

2. OBSERVATIONS

In this paper, we used the UBV observations performed over the period 1984–2004 at the Crimean Station (CS) of the Sternberg Astronomical Institute (SAI) (the main data set) and CrAO (Nauchnyi, Crimea, Ukraine; 80 nights (Merkulova and Metik 1993, 1995, 1996)) and the observations by V. Rakhimov (Sanglok, Tajikistan; 70 nights (Rakhimov 1989)), a total of more than 600 nights (dates) of observations, to analyze the nature of the variable source in the nucleus of NGC 4151. Incidentally, what is also important for our analysis is that half as many measurements were made over the period 1968–1983. Since the bulk of the data (about 90%) have been published (Lyuty and Doroshenko 1999; Doroshenko *et al.* 2001; and references therein), the table of additional observations is not given here, but will be published later in another paper. All of the observations were reduced to a 27″.5 aperture and to the photometric system of CS SAI. The observing and reduction techniques were described in detail by Lyuty and Doroshenko (1999) and Doroshenko *et al.* (2001).

The Maidanak U magnitudes (see Doroshenko *et al.* 2001) showed a fairly large scatter in the check color–magnitude diagrams and turned out to be unsuitable for our color analysis. The same is also true for the CCD observations at CS SAI performed with an ST-6UV camera (Doroshenko *et al.* 2001). Whereas the standard deviation for $B-V$ is approximately the same for all data series, the photoelectric Maidanak and CCD observations showed a factor of 1.5–2 larger scatter for $U-B$ than the remaining series. Naturally, this is because the measurement accuracy is lower: the quantum efficiency of ST-6UV in the U band is a factor of 4 to 5 lower than that at the sensitivity maximum, while a FEU-79 photomultiplier tube (PMT) with a glass entrance window is used at Maidanak Observatory. The UV sensitivity of this instrument is almost an order of magnitude lower than that of an instrument with a PMT with a quartz window (e.g., EMI 9789 used at CS SAI). The B and V magnitudes of all data series have a high accuracy and were subsequently

used to analyze the light curves (time–amplitude characteristics).

For this goal, we also used the BV CCD observations performed at the Special Astrophysical Observatory, Russian Academy of Sciences (Nizhnii Arkhyz, Karachai-Cherkessia, Russia), the Japanese BV CCD observations (Honda *et al.* 1998), and the CrAO spectroscopic data (Malkov *et al.* 1997; Sergeev *et al.* 2001). The continuum measurements at λ5100 Å made in the last two papers were transformed to the B and V magnitudes with calibration using 49 common nights and yielded good agreement with the photoelectric measurements. Incidentally, although the wavelength of the spectroscopic continuum measurements is closer to the effective wavelength of the V band, the dispersion of the V magnitudes was larger than that of the B magnitudes. The Japanese observations (Honda *et al.* 1998) were performed in the period 1995–1996 with a 9″ aperture. Unfortunately, these have two significant shortcomings: (1) the sky background was measured in a 24″–18″ ring, i.e., in fact, the background of the surrounding galaxy rather than the sky background was measured; and (2) there were errors in the magnitudes of the comparison stars. Although the authors used star "C3" = "2" with $B = 12\overset{m}{.}47$ and $12\overset{m}{.}49$, $V = 11\overset{m}{.}44$ and $11\overset{m}{.}47$ (Lyuty (1971) and Penston *et al.* (1971), respectively), whose UBV magnitudes were known, as a comparison star, they determined the V magnitude themselves, but clearly erroneously ($V = 11\overset{m}{.}56$); Honda *et al.* (1998) took the B magnitude from Penston *et al.* (1971). Nevertheless, based on 19 common dates and with allowance made for the aperture and the background of the surrounding galaxy (Lyuty and Doroshenko 1999), we were able to reduce the data by Honda *et al.* (1998) to the CS SAI photometric system. All these additional data were used to better represent the light curve and to analyze the composite flare.

The subsequent analysis pertains only to the variable source in the nucleus; the constant starlight from the galaxy through a 27″.5 aperture was taken into account as prescribed by Lyuty and Doroshenko (1999). Figure 1 shows the light curve for the variable source in the nucleus of NGC 4151 spanning 20 years (1984–2004). The light curve is presented on a linear scale (the flux density is in millijanskys, below referred to as the flux for simplicity) in the B band. We clearly see that the flux from the variable source fell almost to zero during the 1984–1989 minimum of activity. At the short minima of 2001 and 2004, the flux was slightly higher. In the first cycle of activity (1968–1983), the peak flux was half as high as that in the period 1995–1996 (see Fig. 2 from Lyuty and Doroshenko (1999)) and was comparable to the 2003 maximum. The large amplitude of the second cycle of activity (1990–2001) and the presence of a long (1984–1989) minimum give a unique opportunity to separate the S and F variability components with an adequate accuracy and to try to elucidate their nature on the basis of a color analysis. However, before turning to such an analysis, it is desirable to estimate the contribution from emission lines to the UBV bands.

Three fairly strong lines are known to fall within the B band: Hγ, Hβ, and He II λ4686. Among these lines, only Hγ, which is approximately half as strong as Hβ, lies at the maximum of the B passband. The position of Hβ corresponds to 30% of the maximum, while the position of the He II line, which is a factor of 2–3 weaker than Hβ, corresponds to 50%. According to Malkov *et al.* (1997), the integrated flux at the 1988–1989 minimum was $(2–3) \times 10^{-12}$ for Hβ and $<10^{-12}$ erg s^{-1} cm^{-2} for He II, while at the 1995 maximum, it was $(8–9) \times 10^{-12}$ and 3×10^{-12} erg s^{-1} cm^{-2}, respectively. The B-band continuum flux in the same units was 30 and 100. Thus, the total contribution from the strongest emission lines to the B band does not exceed 7–8%. There are no strong lines within the V band, while in the U band, the main contribution from the gas emission is the Balmer continuum; it can reach 30%.

Figure 1 also shows the UV observations at λ1450 Å. The vertical dashes represent the IUE observations, and the open circles represent the UV observations performed after IUE—Astro-1, Astro-2, ORPHEUS–SPAS, and HST–STIS (see Section 1 for references). To compensate for the decrease in flux density F_ν with frequency for the so-called power-law spectrum, the UV data were increased by a factor of 3, which corresponds to an increase in frequency compared to the B band. Thus, Fig. 1 shows quantities proportional to νF_ν: for B, V, U, and λ1450, we took $\nu = 1$, 0.8, 1.22, and 3.03, respectively. All of the data in Fig. 1 are given without any correction for the extinction. We see that both the F and S components are present in the UV variability, with the amplitude of the latter being comparable to the optical variability amplitude. It should be noted that the contribution of the starlight emission from the surrounding galaxy at λ1450 is negligible. Indeed, extrapolating the contribution of the galaxy in the $UBVR$ bands to λ2690 and λ2500 yields, respectively, no more than 3–4% and less than 1% in the 15″ aperture with which the UV observations in 1993 were performed.

3. EXTINCTION

The results of our color analysis also depend strongly on the adopted extinction both in the Galaxy and in NGC 4151. Since the object is near the

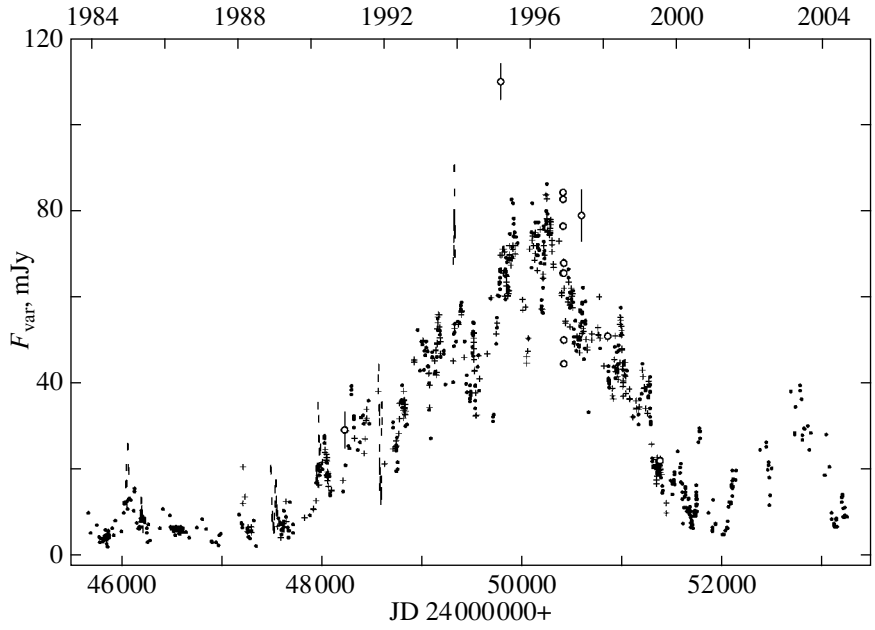


Fig. 1. Light curve for the variable source in the nucleus of NGC 4151, the minimum and the second cycle of activity, on a linear scale (the B flux is in millijanskys). The dots, pluses, and vertical dashes represent the UBV observations, the additional BV measurements, and the IUE UV observations, respectively; the open circles represent the HUT, HST, and ORPHEUS II observations (see the text).

Galactic pole ($b \sim 75°$), the Galactic extinction is low and corresponds to $E_{B-V} = 0.02-0.04$. A value of 0.04 has been taken in all studies of the UV variability in NGC 4151 since 1980 (the first IUE data). Moreover, some of the authors attempted to prove that a larger reddening was in conflict with observations, since there is no dip at $\lambda 2200$ in the spectra of NGC 4151 (Penston *et al.* 1981). At the same time, these authors drew attention to the discrepancy between the UV and X-ray extinctions: in the range 1–10 keV, the neutral hydrogen column density is estimated to be no less than $(2-3) \times 10^{22}$ cm^{-2}, which corresponds to $E_{B-V} = 3.5-4$ at a normal gas-to-dust ratio. The maximum estimate of the gas column density is an order of magnitude larger (Perola *et al.* 1982). The color excess $E_{B-V} = 0.04$ corresponds to a Galactic HI column density toward NGC 4151 of $\sim 2 \times 10^{20}$ cm^{-2}, i.e., two to three orders of magnitude lower than that in the X-ray range.

On the other hand, there are studies in which the intrinsic extinction in NGC 4151 was determined (Ward *et al.* 1987; and references therein). The intrinsic extinction in the nucleus of NGC 4151 was variously estimated to be between $A_V = 0.3$ (from the [S II] $\lambda 6724/4072$ ratio) and $A_V = 1.4$ (from Paβ/Hγ). $A_V = 0.9$ was found from the [S II] lines, but in a wider wavelength range ($\lambda 10\,400/4072$). The most accurate estimate is probably obtained from the He II $\lambda 10123/3203$ lines: $A_V = 1.0 \pm 0.1$ (Ward *et al.* 1987). At $R = A_V/E_{B-V} = 3.1$, this corresponds to a color excess of $E_{B-V} = 0.32$. It is this value that was adopted by Lyuty and Doroshenko (1999) and Doroshenko *et al.* (2001). In this paper, we also took this reddening. Figure 2 shows the positions of the color indices for the variable source in NGC 4151 with zero reddening (the observed colors after the subtraction of the constant emission from the surrounding galaxy) and with a reddening of $E_{B-V} = 0.32$. We clearly see that the optical (UBV) extinction does exist and, since the Galactic extinction toward NGC 4151 is very low, it should be attributed to the intrinsic extinction in the nucleus of NGC 4151. The comet-shaped distribution of points in Fig. 2 will be discussed below.

A paradoxical situation arises: the extinction is fairly high in the optical range and corresponds to $E_{B-V} = 0.3$, is very high (>3.5) in the X-ray range, which is taken into account when calculating the X-ray luminosity, and is virtually absent in the UV! This conclusion (low extinction) was drawn by Penston *et al.* (1981) from the extinction curve obtained by Seaton (1979) for Galactic stars: a Lorentz function was superimposed on the curve of the normal extinction law $1/\lambda$ to obtain an absorption band near 2200 Å. Other authors used this conclusion in the ensuing 20 years. The paradoxicality of the situation can be emphasized even further by noting that the extinction at $\lambda 3600$ (the U band) is $A_U = 1^m.6$ (see Fig. 2), but falls sharply, according to Penston *et al.*

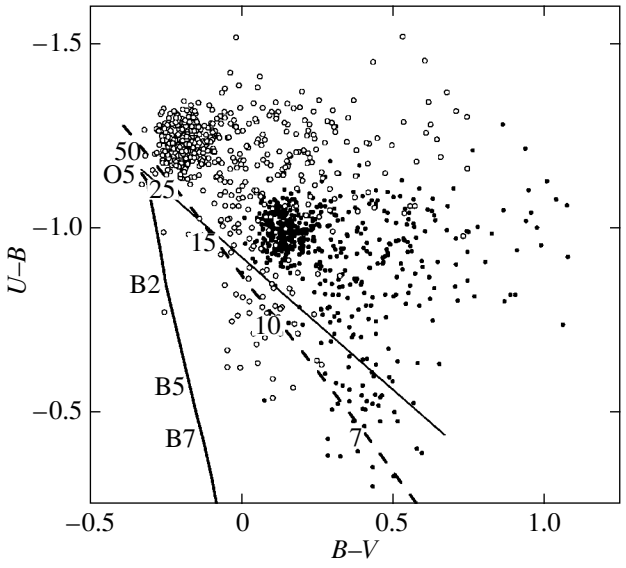


Fig. 2. Positions of the color indices of the variable source in NGC 4151 in the $(U-B)-(B-V)$ diagram without any correction for the reddening (dots) and with $E_{B-V} = 0.32$ (circles). The main sequence and a blackbody with an indication of its temperature in kilokelvins are shown in the diagram. The thin solid line corresponds to the normal reddening law.

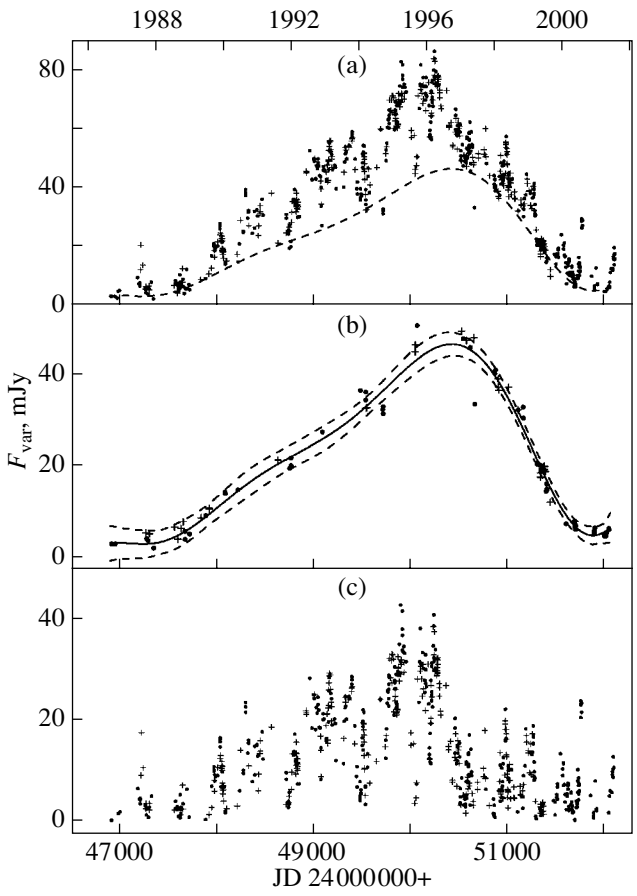


Fig. 3. Separation of the components: (a) the part of the light curve (see Fig. 1) for which our analysis was performed (the dashed line represents a fitting polynomial); (b) the S component (the filled circles indicate the individual values, the solid line represents the fitting polynomial, and the dashed lines indicate the 95% confidence intervals); and (c) the F component obtained by subtracting the polynomial from the light curve of the variable source. The pluses represent the additional BV observations.

(1981), by no more than 0.2–0.3 at λ2700–3000. In addition, as was pointed out by Mushotzky *et al.* (1978), if one assumes the optical extinction to be $A_V = 0.2$ and the HI column density in the X-ray range to be 3×10^{22} cm^{-2}, which corresponds to $A_V \sim 13^m$, then the gas-to-dust ratio in the nucleus of NGC 4151 will be a factor of 200 smaller than its normal value. This is unlikely, since the presence of a large amount of dust in NGC 4151 has long been proven: the bulk of the infrared flux is attributable to dust reradiation.

Is it not easier to assume that the absence of a dip at 2200 Å in NGC 4151 is attributable to low UV extinction, not to the absence of such a band (different dust properties)? The normal law $1/\lambda$ that holds in the optical range should then be also extended to the UV. In this case, the extinction at λ1450 Å (most of the published data are given for this wavelength) will be $A_{1450} = 3^m.8$, which should be taken into account when determining the UV luminosity.

4. THE SLOW COMPONENT

4.1. Separating the S Component

We see from Fig. 1 that no point falls within the region bounded above by the lower envelope of the light curve. Therefore, the observations near the lower envelope were chosen to separate the S component, a total of 60 points (dates). This method for separating the components makes the difference between them, if it exists, largest. For our subsequent analysis, we used not all of the data presented in Fig. 1, but only the observations in the period 1986–2001, i.e., from the time the activity of the old accretion disk ceased completely (Lyuty and Doroshenko 1999) until the first deep minimum in 2000/2001 (JD 2 451 850–2 452 050; below, for convenience, we use truncated Julian days: TJD = JD − 2 400 000). The second wave (2001–2004) was not analyzed due to the small amplitude and, accordingly, the large relative errors.

Figure 3 show the observed B fluxes from (a) the variable source (in millijanskys), (b) the S component, and (c) the F component after the subtraction of a fitting polynomial from the observations, the S component was fitted by a ninth-order polynomial using

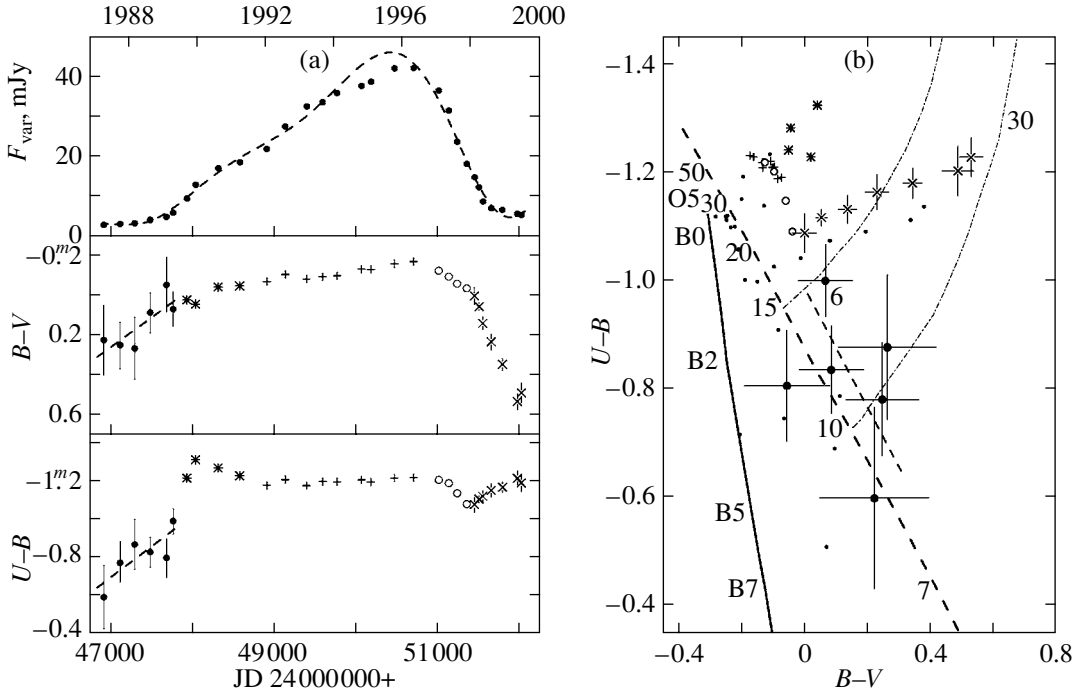


Fig. 4. (a) Averaged light and color curves for the S component and (b) positions of the color indices in the $(U-B)-(B-V)$ diagram. Different symbols indicate the distinguished segments of the color curves (see the text). Panel (b) also shows the color indices of main-sequence stars (solid line), a blackbody (heavy dashed line) with an indication of the temperature in kilokelvins, and radiation of a hot gas (dash–dotted line) that is optically thick in the Balmer continuum. The dots indicate the shift of the entire picture at a different reddening (see the text).

only the UBV data; the fitting polynomial is indicated by the solid line, and the 95% confidence intervals are indicated by the dashed lines. The additional B magnitudes are indicated in Fig. 3 by the pluses. We see from Fig. 3b that these are in satisfactory agreement with the S component separated using UBV, which may be considered to be the confirmation that the S component was separated correctly. It should be noted that the S component is clearly asymmetric: a slow (10 years) rise in brightness and a fast (3 years!) decline. The F component (Fig. 3c) shows a clear correlation (which, incidentally, is also seen from Fig. 3a) between the flare amplitude and the amplitude of the S component (Lyuty and Oknyanskij 1987; Doroshenko *et al.* 2001). Another peculiarity of the F component is a sharp (approximately twofold) decrease in the flare amplitude after the maximum.

In certain cases, negative values are obtained when a polynomial is subtracted from the observational data. All the dates of observations when the flux for the F component was negative or zero at least in one band were excluded. We also excluded the dates when the flux from the F component was less than 1σ and 3σ for the analysis of the light curves and the color indices, respectively. The error is determined by the significance of the polynomial fit and is $\sigma F_\nu(U) = 0.41$, $\sigma F_\nu(B) = 0.34$, $\sigma F_\nu(V) = 0.41$, i.e., approximately the same as the error in the contribution of the surrounding galaxy in a $27''$ aperture (Lyuty and Doroshenko 1999).

The Color Indices

Since the color indices of the variable source have a fairly large scatter for individual measurements (Fig. 2), it makes sense to perform a color analysis for smoothed light curves. This does not change the result qualitatively, but makes it possible to obtain quantitative estimates. We smoothed the light curves for the S component by a moving average over one or two seasons (depending on the number of observations in the season), which halved the number of points. Subsequently, we calculated the color indices for each mean date. Figure 4a shows the light curve smoothed in this way (flux $F_\nu(B)$) and the color indices. For comparison, the dashed line indicate the fitting polynomial drawn through all 60 points. The two curves agree satisfactorily; the small difference at the maximum plays no significant role. Figure 4b shows the color variations with brightness.

Five different segments of the light curve can be distinguished by the pattern of behavior of the color indices; these are indicated by different symbols both

in the color curves (Fig. 4a) and in the two-color diagram. Since even the smoothed color indices have fairly large errors at the minimum, the figure shows a linear regression line drawn through the first six points (the numbers near the symbols in Fig. 4b correspond to the point number in the smoothed light curve).

From the minimum light (1987–1988) until mid-1989, when the formation of a new accretion disk appears to have just begun, the increase in luminosity was attributable mainly to the rise in temperature from 7–8 to 14–15 kilokelvins. Although a factor of 16 increase in luminosity must correspond to a factor of 2 increase in temperature, this is true for the stellar emission (spherical symmetry). The luminosity of a disk as a plane figure, first, is lower, and, second, the observed flux also depends on the disk orientation relative to the observer.

In 1992–1996, the color indices (the pluses in Fig. 4) were almost constant, although the luminosity doubled over this period. The $B-V$ and $U-B$ color indices decreased by $0^m.1$ and $0^m.15$, respectively. We clearly see an UV excess of $\approx 0^m.1-0^m.15$, which is hard to explain by an underestimation of the contribution from the surrounding galaxy in the U band: this would have a stronger effect at the minimum than at the maximum. However, the UV excess is approximately the same even at the minimum (filled circles). Since the flux from the minimum (1987) to the maximum (1995–1996) increased by a factor of 10, the UV excess can be said to increase with luminosity.

The UV excess increased sharply late in 1989 (the asterisks in Fig. 4). At the same time, the total luminosity increased insignificantly (points 7 and 8 in the light curve). The $U-B$ color index decreased by more than $0^m.3$, while $B-V$ decreased by only $0^m.1$. If we recognize the existence of an overall luminosity-dependent UV excess and take into account a slightly larger reddening than that adopted above, then the color indices in 1992–1996 will correspond to a blackbody temperature of 25 000–30 000 K or B0–O5 stars. This shift (the overall UV excess is $0^m.2$ and $E_{B-V} = 0^m.43$) is indicated in Fig. 4 by the dots. For a hot star, the increase in temperature from 25 000 to 30 000 K exactly corresponds to a twofold increase in luminosity. However, this is no longer enough for the disk, which, in addition, may be tilted to the line of sight, and an increase in the size of the emitting surface is required. It is even possible that an increase in disk size plays a greater role than a rise in temperature.

Thus, two main segments can be distinguished on the ascending branch of the light curve for the S component: the temperature increase in luminosity (1987–1989) by 10–15% from the maximum value (filled circles) and the main increase in luminosity by more than a factor of 4 (1989–1995, the asterisks and pluses), which is probably attributable to an increase in the AD size. The nature of the UV excess in 1989–1990 will be discussed below. Two different segments can also be distinguished on the descending branch (1997–2000): the temperature decrease (open circles) and the decrease with the appearance of a strong UV excess (crosses). The increase in $B-V$ in 1996–1997 (the open circles and the last plus) corresponds to the decrease in blackbody temperature approximately from 30 000 to 18 000 K. At the same time, the luminosity must decrease by almost a factor of 4, but it fell by only a factor of 2. Hence, about half of the observed fall in luminosity is the result of a decrease in AD size. The further fall in luminosity and the increase in $B-V$ were accompanied by the appearance of an UV excess and its increase (the crosses in Fig. 4). The temperature gradient (the change in $B-V$) became steeper, but the radiation of a hot gas (hydrogen plasma) that is optically thick in the Balmer continuum showed up quite clearly in the U band (Chalenko 1999).

This radiation cannot be produced by gas clouds; the latter are generally believed to be responsible for the fast (of the order of a month) flux variability in the Hα and Hβ lines, which lags behind the continuum variability by 1–3 weeks (the broad line region, BLR). First, this radiation appeared only in mid-1989, i.e., approximately one and a half years after the temperature began to rise (see Fig. 4), and, second, after the luminosity fell to its minimum in mid-1998, the UV excess remained at least for another year (the gas responsible for this UV excess did not recombine in 1 year), while the gas in BLR clouds recombines approximately in a month. Since the recombination time is determined by the gas density and since the recombination time in our case is more than a year, the gas density in the region under consideration must be an order of magnitude lower than that in the BLR. It thus follows that an extended broad line region (EBLR) with a density that is an order of magnitude lower than that in the BLR exists at an effective distance of 1.5 lt-years ($\approx 1.4 \times 10^{18}$ cm) from the central source. The existence of this region has not yet been discussed. However, without the assumption about the presence of an extended broad line region, it is hard or even impossible to explain not only the appearance of an UV excess in the S variability component, but also some of the spectral features, which are discussed in more detail in Section 7.

5. THE FAST COMPONENT

The individual color measurements of the S component are difficult to analyze (see above); this is

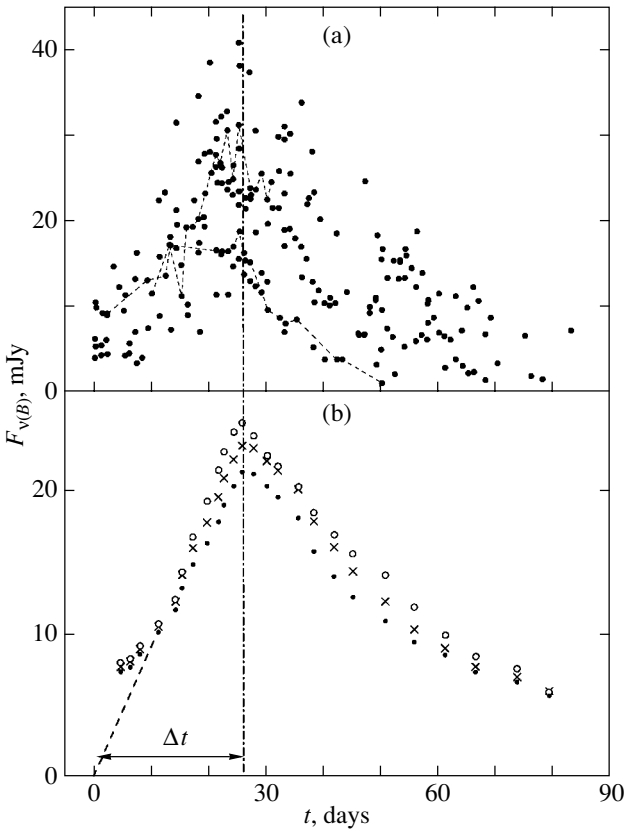


Fig. 5. Average composite flare of the F component: (a) individual B-band measurements (the dashed line connect the points belonging to one flare); (b) the shape of the average flare in U (circles), B (crosses), and V (dots); the dashed line indicates a linear extension of the ascending branch (B) to the flare onset. The vertical dash–dotted line indicates the position of the maximum.

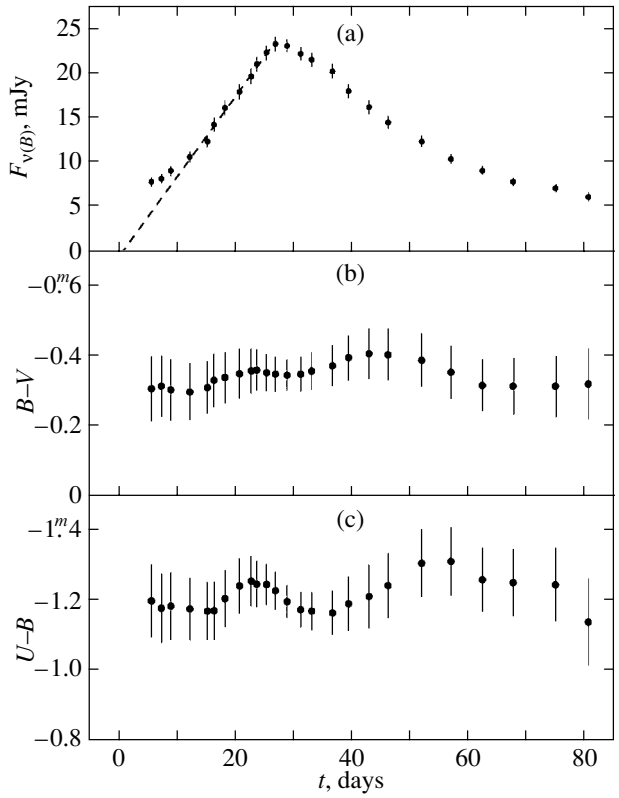


Fig. 6. (a) Average flare in the B band. The dashed straight line indicates an extrapolation to the flare onset. Panels (b) and (c) show the mean color indices of the composite flare.

especially true for the F component, since the relative error increases appreciably when the S component is subtracted. Therefore, we analyzed the color characteristics of not the individual flares, but the so-called composite flare (Lyuty 1977). In 1977, it was established for nine Seyfert nuclei that the time of the rise in brightness by the mean amplitude (the variability time scale Δt) does not depend on the flare amplitude, but it is different for different objects. In particular, the 1968–1976 observations yielded $\Delta t = 10$ days for NGC 4151. It would be natural to associate this time with the size of the active region, the accretion disk (Dibaĭ and Lyuty 1984).

In this paper, we also constructed the composite flare from data on the variability of the F component (Fig. 3): the individual flares were aligned by the flux maximum if it was recorded reliably. There were 16 such events (Fig. 5a). Subsequently, the composite flare was averaged by a moving average and smoothed over five neighboring points for each of the UBV bands (Fig. 5b). The shape of the average flare turned out to be the same as that in the period 1968–1976—an almost linear rise in flux to its maximum followed by an exponential decline, but the rise time Δt increased by almost a factor of 3, to 26–27 days. The position of zero (zero flux) was determined from linear regression for the ascending branch, except the first three points (dispersion minimum). A certain deviation of the first three points from the regression line is most likely attributable to the completion of the preceding flare: a superposition of flares is commonly observed when the succeeding flare begins before the complete decay of the preceding flare (more than 100 days from the maximum). Only in rare cases was almost complete decay observed. Given the time for the zero flux, $\Delta t_B = 26\overset{\text{d}}{.}0 \pm 0\overset{\text{d}}{.}5$. Approximately the same values were obtained for V and U: $\Delta t_V = 27.6 \pm 0.3$ and $\Delta t_U = 25.0 \pm 0.9$, although there is still a tendency for Δt to increase with wavelength.

The color indices (Fig. 6) are constant, within the error limits, throughout the flare and are, on average, $(B-V)_0 = -0\overset{m}{.}34 \pm 0\overset{m}{.}02$ and $(U-B)_0 = -1\overset{m}{.}21 \pm 0\overset{m}{.}02$, corresponding to blackbody radiation with a temperature of ~60 000 K. It thus follows that the evolution of the flare correlates weakly with the temperature variations. Undoubtedly, such a correlation

must exist, since the flare amplitudes differ by a factor of 4 or 5 for the same brightness rise time. However, a 30–40% change in temperature would suffice for this. This change could hardly be noticed in the two-color diagram: at high temperatures, the color indices depend weakly on the temperature. An increase in the size of the emitting region (AD) must then be mainly responsible for the rise in brightness during the flare. Since the luminosity during the flare is comparable to the mean luminosity of the variable source, the flare must affect the entire disk.

If the disk size is estimated as $c\Delta t$, where c is the speed of light, then the disk radius must be $\sim$26 lt-days ($\sim 7 \times 10^{16}$ cm). Previously, this was assumed to be the case. However, evidence that the radius of the optical disk, i.e., the AD region emitting efficiently in the optical range, is at least an order of magnitude smaller has been obtained in recent years. Thus, for example, based on quasi-simultaneous optical UBV and infrared $JHKL$ observations of the nuclear variability in NGC 4151, Lyuty *et al.* (1998) found that the radius of the optical disk is 7×10^{14} cm. Oknyanskij *et al.* (2003) found the variability in the V, R, and I bands to lag behind that in the B band by 0.2, 1.0, and 1.4 days, respectively (with an error of $\pm$0.4 day). Sergeev (2005) obtained almost the same results: 0.1, 1.0, and 1.5 days with a slightly smaller error ($\pm$0.3 day). This lag implies a temperature stratification over the disk radius, while its value determines the radius of the region efficiently emitting in this band. Thus, the radius of the optical (UBV) disk does not exceed 0.5–0.6 lt-day ($\sim 1.5 \times 10^{15}$ cm), in good agreement with the above value determined from the luminosity and the spectral energy distribution. This value is a factor of 50 lower than that derived from the formula $c\Delta t$; i.e., the velocity of propagation of the flare is much lower than the speed of light. This velocity is estimated from the relation $\Delta t = 26^{\rm d}$ and the disk radius in light days to be $\sim$6000 km s^{-1}. This is discussed in more detail below.

We can now return to Fig. 2 and explain the comet-shaped distribution of points: the head of the comet is mainly the F component whose color indices correspond to a blackbody with $T \sim 60\,000$ K, while the two tails (parallel to the blackbody line and the $B-V$ axis) correspond to the S component (see Fig. 4).

6. THE UV CONTINUUM VARIABILITY

In November–December 1993, an international campaign to study the nuclear variability of NGC 4151 in the X-ray, UV, and optical spectral ranges as well as in the C IV λ1549, He II λ1640, C IIIλ1909, Hα, and Hβ lines was conducted (Crenshaw *et al.* 1996; Kaspi *et al.* 1996; Warwick *et al.* 1996). The X-ray (ROSAT, ASCA and CGRO), UV (IUE), and optical observations were performed for 14 days, 18 days (most intensively for 9 days), and 3 months, respectively. The main goal of this campaign was to investigate the fast (days–hours) variability of NGC 4151 over a wide wavelength range and to find the time lag, if it exists, between the variabilities in different ranges. The main shortcoming of the campaign is an overly short duration of the X-ray and UV observations. However, the campaign yielded quite definite results that were analyzed in the fourth paper of this series (Edelson *et al.* 1996). The authors concluded that the results of the campaign are consistent with the reprocessing hypothesis (the reprocessing model), in which the X-ray range 1–10 keV plays a primary role, then the UV–optical–infrared and, in the gamma-ray range (50–200 keV), the inverse Compton effect, although they point out that the UV luminosity is an order of magnitude higher than the 1–10-keV luminosity. This is even despite the fact that a very low extinction (corresponding to $E_{B-V} = 0.04$) was adopted in the UV. However, if the gamma-ray range is added, then the 1–200-keV luminosity is comparable to the UV luminosity (a factor of 2 lower).

These conclusions were drawn primarily from the strong (in the opinion of the authors) wavelength dependence of the variability amplitude: at λ1275, the normalized (i.e., related to the mean flux) variability amplitude is a factor of 10 larger than that in the optical range (Edelson *et al.* 1996). However, these authors forgot that the dependence $F_\lambda(\lambda)$ is nonlinear. Figure 7a (a modification of Fig. 1 from Edelson *et al.* (1996)) shows the variability of the flux density F_λ at various wavelengths and F_ν for the same wavelengths (Fig. 7b). We see that the UV and optical variability amplitudes in linear units (F_ν) are comparable, although the normalized amplitude increases with decreasing wavelength, but not so catastrophically as Edelson *et al.* (1996) imagined: the normalized amplitude at λ1275 is only twice that at λ4600. Of course, as was correctly pointed out by Kaspi *et al.* (1996), the decrease in optical amplitude is partly attributable to the effect of constant starlight from the surrounding galaxy. However, allowance for the contribution of the galaxy ($\sim$25% at λ4600) leads to an increase in amplitude by less than a factor of 1.5 (in linear units, the constant component does not affect the absolute variability amplitude, but affects only the relative (normalized) amplitude). Figure 7b shows both the observed fluxes at λ4600 and λ2688 and the flux after the subtraction of the starlight from the surrounding galaxy—the flux from the variable source at λ4600 is in excellent agreement with that in the B band (λ4400). Thus, scale nonlinearity is

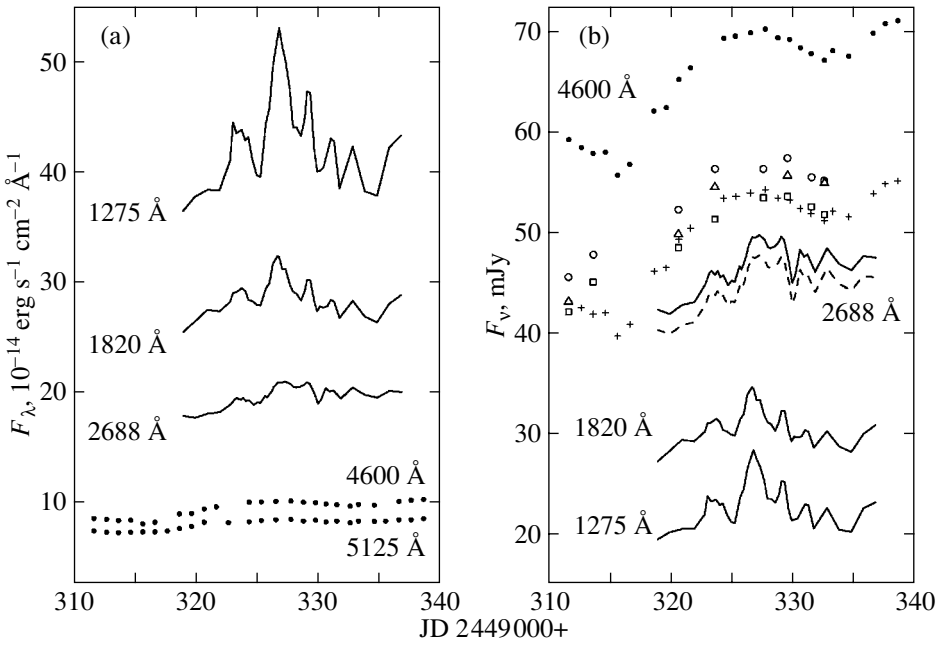


Fig. 7. Simultaneous UV and optical observations: (a) a modified analogue of Fig. 1 from Edelson *et al.* (1996), the flux is in units of 10^{-14} erg s^{-1} cm^{-2} Å^{-1}; (b) the flux in millijanskys. Our data: the triangles for U, the squares for B, the circles for V, and the pluses and the dashed lines for λ4600 and λ2688, respectively, with allowance made for the emission from the surrounding galaxy.

mainly responsible for the large difference in the variability amplitude in units of F_λ (Fig. 7a). In addition, we see that the UV flare was not recorded from the minimum: according to the optical data, the minimum was observed at TJD 49315, while the flux at the minimum was almost a factor of 2 lower relative to the maximum than that at TJD 49320.

The 1993 observations are the last data obtained from the IUE satellite. Apart from these data, the 1988/1989, 1990, and 1991 observations (Ulrich *et al.* 1991; Ulrich and Horne 1996) as well as observations with a different aperture, including those in the far UV, wavelength shorter than Lyα, were performed during the second cycle of activity in NGC 4151. All these data are shown in Fig. 1 and reveal the presence of F and S variability components in the UV. The data were obtained from Astro-1 and Astro-2 in the range 912–1840 Å (Kriss *et al.* 1992, 1995), from ORPHEUS-II in the range 912–1220 Å (Espey *et al.* 1998; Espey 1999), and from HST–STIS in various ranges from 1150 to 3060 Å (Crenshaw *et al.* 2000). A common wavelength for all of the UV data, except ORPHEUS-II, is λ1450 Å. The ORPHEUS-II data are given for λ1000 Å, but these can be extrapolated to λ1450 using the Astro-2 spectrum to be presented in the light curve (Fig. 1). The 1991 UV observations were performed in a period when there were no optical observations at all (TJD 48570–48606). The 1993 flare is one of the strongest in amplitude. In Fig. 1, the UV data were superimposed on the B-band light curve. The UV data (IUE, 1993 and Astro-2, 1995) stand out clearly at the maximum of the light curve, but the same data are in excellent agreement with the U-band light curve (also with allowance made for the frequency). Espey (1999) pointed out that the Astro-2 observations (March 1995) revealed the strongest UV flux ever observed. However, comparison with the U-band light curve indicates that an UV flux higher by 20–30% could be observed in the second half of 1995 and the first half of 1996.

Further, it is of interest to separate the S and F components in the UV as well. The UV observations alone are not enough for this to be done, but in combination with the U-band data, these allow the S component to be separated in the UV as the lower envelope of the light curve. The fitting polynomial slightly differs from the optical one: a slightly lower flux in the periods 1988–1990 and 1994–1996. Allowance for these differences would lead to an increase in the amplitude of the F component in the U band. It is difficult to trace the evolution of the S component as was done for the UBV bands due to the scarcity of data, but it is quite possible to separate the F component. Just as in the optical range, we then constructed the composite flare for λ1450 and λ1720 Å.

The UV composite flare for λ1450 Å is shown in

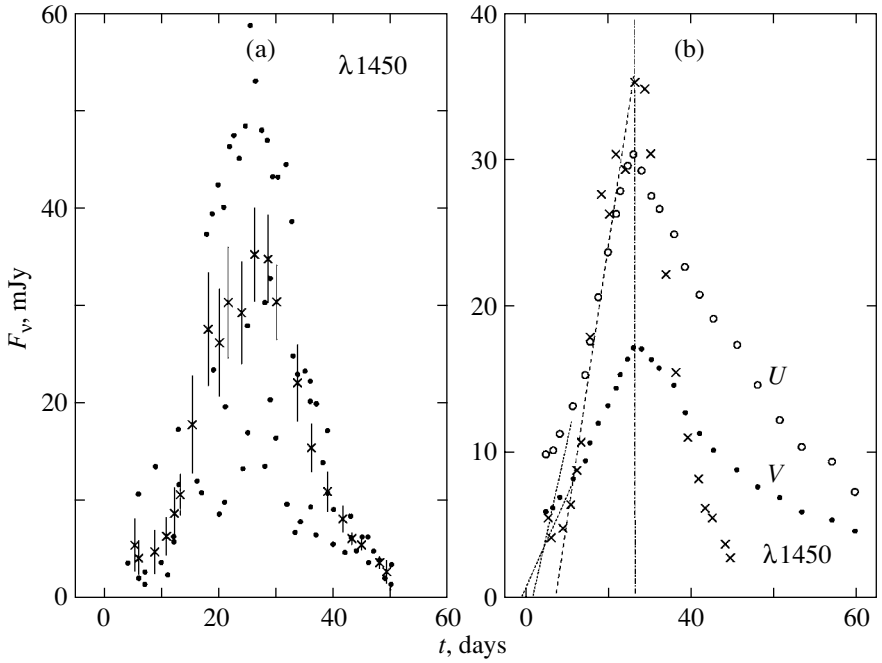


Fig. 8. (a) UV composite flare (λ1450 Å) from *IUE* data (1988/1989, 1990, 1991, and 1993): the dots represent the individual measurements, the crosses with error bars represent the mean values; (b) the shape of the average flare at λ1450 Å (crosses), in U (circles), and in V (dots), the dotted lines indicate a linear extrapolation to the flare onset in U and V, the dashed line is a linear representation of the ascending branch of the UV flare.

Fig. 8a. The average composite flare (averaging by a moving average) is obtained with a slightly lower accuracy than that in the optical range, but this accuracy is quite sufficient for the subsequent analysis. Figure 8b compares the parameters of the average flare in V and U and at λ1450 Å (the frequency difference was taken into account; see above). First of all, we clearly see that the brightness rise time of the UV flare at almost the same mean amplitude is appreciably shorter, $\Delta t_{1450} = 19 \pm 3$ days. The decay time of the flare is severalfold shorter, about 30 days, while it is ∼100 days in U and >160 in V (exponential decay). The UV flare confirms the wavelength dependence of Δt suspected above: Δt decreases from 27.5 days in V (λ5500 Å) to 19 days in the UV (λ1450 Å). This is in good agreement with the universally accepted assumption that the size of the active region decreases with wavelength in the disk accretion model: at λ1450 Å, the size of the active region is almost a factor of 1.5 smaller. At λ1720 Å, the average flare is approximately the same, but the amplitude is slightly smaller, while $\Delta t = 20^{\rm d} \pm 3^{\rm d}$ is longer (within the error limits).

There is also evidence that the flare rise and decay times at λ1000 Å are event shorter than those at λ1450 Å. Whereas six events with different amplitudes were included in the composite flare at λ1450 Å (Fig. 8), only one flare (in 1996) was observed at λ1000 Å (ORPHEUS-II), but its reduced ($\propto \nu F_\nu$) amplitude is of the order of the mean amplitude. This makes it possible to estimate the brightness rise time, $\Delta t_{1000} \sim 6$–8, while the decay time is 8–10 days.

7. DISCUSSION

7.1. The Extended Broad Line Region

Analysis of the UBV observations of the nuclear variability in the Seyfert galaxy NGC 4151 showed that the S and F components are completely different in nature. Moreover, to explain the peculiarities of the color behavior of the S component required assuming the existence of an extended broad line region (EBLR) with a density that is one or two orders of magnitudes lower than that in the BLR at an effective distance of 1.5 lt-years from the central source. Is there spectroscopic confirmation of the existence of this region?

There is such confirmation. NGC 4151 has been regularly observed in the Hα, Hβ, and He IIλ4686 lines at CrAO since 1987 (Malkov *et al.* 1997; Sergeev *et al.* 2001). In the former paper, the authors pointed out a puzzling effect: the splitting of the line–continuum diagrams into several sequences. This showed up most clearly for the Hα and Hβ lines. The authors also pointed out that the main feature in the line variability is an overall increase in flux from 1988 until 1995 with short (50–100 days) events (flares) superimposed on it. However, they did not

pay particular attention to the origin of this slow line variability component, although the amplitude of the S component is a factor of 3 or 4 larger than that of the F component, while these are roughly identical in the continuum (see Fig. 5 in Malkov *et al.* 1997). Thus, the diagram splitting is just a manifestation of the S component.

The origin of the F component is clear—this is the response of gas clouds in the BLR to ionizing continuum variability with a lag of 1–3 weeks. However, if only the region at a distance of $\sim 10^{17}$ cm and with a density of $10^{10}-10^{12}$ cm^{-3} that is responsible for the fast line variability with a lag relative to the continuum existed, then no slow line variability component would be observed at all. If, however, we also assumed the existence of an EBLR at a distance of 1.5 lt-years and with a density that is one or two orders of magnitudes lower than that in the BLR, as follows from our analysis of the UBV variability, then the presence of a slow component in the line variability could be easily explained. In the U band, an excess appears through the radiation of the same extended gaseous region if the gas is optically thick in the Balmer continuum. There is also other evidence for the existence of an extended gaseous region: adding the second peak to the transfer function at a distance of ~600 days improves greatly the continuum–line correlation (Sergeev *et al.* 2001).

7.2. The Continuum

Parameters of the F component. Analysis of the optical and UV observations of the variable source in NGC 4151 showed that the S and F components are clearly completely different, but thermal in nature. There is no need to discuss the synchrotron radiation, as was the case in the first years of AGN studies, or the inverse Compton scattering (the early 1980s), especially since these models gave no answer to the question about the energy source. Therefore, the disk accretion model was universally accepted after 1990. However, the question of which spectral range is the primary one, i.e., the range in which the main energy release takes place, has not yet been solved. In the past 10 years, based on both the flattening of the spectrum ($\log\nu-\log F_\nu$) in the range 1 keV–2 MeV and on the variability rate (small region), most authors have assumed it to be the X-ray range.

Table 1 gives basic parameters of the fast variability component in NGC 4151 in the range 1000–5500 Å. As has been said above, the S component is present not only in the optical range, but also in the UV, probably up to the Lyman limit (Astro-1, Astro-2, and ORPHEUS data). The S component is completely absent in the X-ray range 1–10 keV, suggesting that the UV/optical and X-ray emissions are different in nature.

Table 1. Parameters of the F component of the variable source

λ, Å	$F_{\max}$, mJy	Δt, days	$t_{\rm d}$	F/S
5500	16.8 ± 0.4	27.6 ± 0.3	>160	0.40
4400	22.9 ± 0.4	26.0 ± 0.5	~115	0.50
3600	29.5 ± 0.6	25.0 ± 0.9	~100	0.55
1720	30 ± 2	19 ± 2	~30	0.65
1450	35 ± 3	19 ± 3	~30	0.75
1000	37 ± 5	6 ± 2	~10	0.90

Column 2 in Table 1 gives the reduced (see above) amplitude of the average composite flare, columns 3 and 4 give the rise and decay times (see Figs. 5 and 8), and column 5 gives the amplitude ratio of the F and S components. The F/S ratio increases with decreasing wavelength approximately by the same amount as the amplitude of the F component (column 2); i.e., the reduced amplitude of the S component is virtually independent of the wavelength. The color peculiarities of the S component in the optical range were discussed above, while the UV data are not enough for this to be done. Analysis of the F component (Table 1) yields important parameters of the variability in a wide spectral range.

First of all, the rise time of the flare Δt does not depend on its amplitude over the entire wavelength range under study, from $\lambda 5500$ (V) to $\lambda 1000$ Å. However, Δt decreases with wavelength by a factor of 4 or 5 from V to $\lambda 1000$ Å, which directly suggests that the size of the emitting region decreases or, what is less likely, that the velocity of propagation of the flare over the disk increases. The decrease in flare decay time from the optical range to the UV is even larger, a factor of 10–15. If the decrease in Δt is attributable to a decrease in the AD size, then the AD radius in the UV is <0.1 lt-days.

The Luminosity of the variable source. To compare the parameters of the variability in different ranges, Edelson *et al.* (1996) used the monochromatic luminosity $L_\nu = 4\pi D^2 \nu F_\nu$ by multiplying it by the variability fraction in the corresponding range (from 25% at 1–2 keV to 1% in the optical range), the variable luminosity $L_{\rm var}$. Incidentally, even for this definition, the 1–2-keV luminosity was a factor of 35 lower than L_{1275}. Given the aforesaid, within the limits of 3 and 6 (the extinction and the variability amplitude), the spectral energy (monochromatic luminosity) distribution was be constructed for both the total (observed) flux and the average composite flare (Fig. 9). The observed distribution is given for the

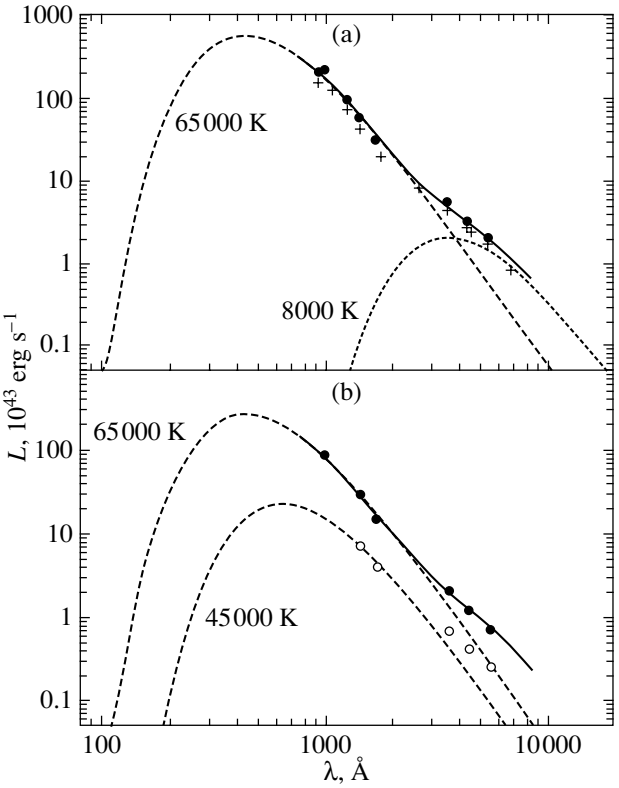


Fig. 9. Spectral energy distribution for the variable source in NGC 4151: (a) the observations near maximum light (March 1995, filled circles) and the instantaneous spectrum (TJD 49329, crosses); (b) the spectral energy distribution at the maximum of the average flare (filled circles) and on the eighth day after the flare onset (open circles). Also shown are the Planck distributions for different temperatures (dashed lines) and the total distribution for 65 000 and 8000 K (solid lines).

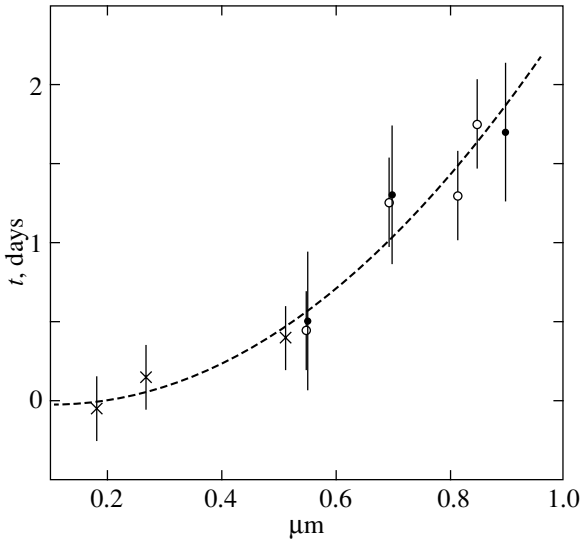


Fig. 10. Lag τ of the variability in different spectral ranges relative to λ1275 Å; the dots, circles, and crosses represent the data by Oknyanskij *et al.* (2003), Sergeev (2005), and Edelson *et al.* (1996), respectively; the dashed line indicates a parabolic fit with allowance made for the weights of the individual measurements.

flux near the maximum (March 1995, UV data from Astro-2), and the instantaneous spectrum is given at TJD 49 329, to within 0.1–0.2 day.

First, Fig. 9 shows that the UV energy distribution for both the observed flux (Fig. 9a) and the average flare is in excellent agreement with the Planck distribution at $T_e = 65\,000$ K. For the average flare near the minimum, $T_e = 45\,000$ K (open circles). It was noted in Section 5 that analysis of only the UBV data yields a temperature at the flare maximum of $\sim$60 000 K, while the temperature variations with flare amplitude can be 30–40%. The flux from the average flare on $t = 8^{\rm d}$ from its onset roughly corresponds to the flare maximum with a minimum amplitude, and 45 000 K is 30% lower than 65 000 K. Thus, the UV data completely confirm the conclusions of Section 5. Second, the optical luminosity is seen to be higher than the long-wavelength tail of the Planck distribution for the source with $T_e = 65\,000$ K, while the energy distribution in the range 1000–5500 Å agrees well with the total distribution for the two sources, with $T_e = 65\,000$ and 8000 K (solid line); i.e., the radiation of the low-temperature source is added to the long-wavelength tail of the source with $T = 65\,000$ K. The same is also true for the average flare (Fig. 9b)—the solid line represents the sum of the two sources, 65 000 and 8000 K. Third, even the monochromatic luminosity of the variable source proved to be very high: the luminosity[1] observed at λ1000 Å is $L_{1000} = 2 \times 10^{45}$ erg s^{-1}. The luminosity in the V band (λ5500) is two orders of magnitude lower, while at the maximum (λ450 Å), it is a factor of 3 higher, $L_{450} = 6 \times 10^{45}$ erg s^{-1}. The total luminosity of the UV source is an order of magnitude higher: $L_{\rm tot} = (6-8) \times 10^{46}$ erg s^{-1}.

The low-temperature source. Before we turn to the nature of the additional low-temperature source, let us return to the AD size determined from the lag of the emission in different spectral ranges. This lag in the optical range was briefly discussed in Section 5. However, one of the objectives of the 1993 campaign (Crehshaw *et al.* 1996) was exactly the determination of the possible lag between the long-wavelength and short-wavelength emissions. Analysis of the variability in the range λ1275–5125 Å showed that there is a lag relative to λ1275 Å, which increases from

[1] Edelson *et al.* (1996) assumed the distance to NGC 4151 to be $D = 20$ Mpc, which corresponds to $H_0 = 50$; in this paper, as previously (Lyuty *et al.* 1998; Lyuty and Doroshenko 1999), we assumed that $D = 14$ Mpc ($H_0 = 75$).

Table 2. Determination of the SBH mass in NGC 4151

M_c, $10^8 M_\odot$	$\log M_c$	L_{Edd}, 10^{46} erg s^{-1}	R_g, 10^{13} cm	Method*	Reference
0.6	7.8	0.8	1.9	D (Hβ)	Dibaĭ (1977)
0.5	7.7	0.6	1.5	D (Hβ)	Dibaĭ (1980)
0.4	7.6	0.4	1.1	D (C IV)	Clavel *et al.* (1987)
0.5	7.7	0.6	1.5	D (C IV, Mg II)	Gaskell (1988)
0.6	7.8	0.8	1.9	D (Hβ)	Wandel (1989)
0.2	7.3	0.2	0.6	D (C IV)	Ulrich and Horne (1996)
0.5	7.7	0.6	1.5	D (C IV)	Ulrich *et al.* (1984)
0.6	7.8	0.8	1.9	D (Mg II)	Ulrich *et al.* (1984)
1	8.0	1.2	3.0	D (Hβ)	Ulrich *et al.* (1984)
20	9.3	24	59	D ([O III])	Sanders (1970)
1.1	8.0	1.3	3.2	X	Wandel and Mushotzky (1986)
4.9	8.7	5.9	14	D ([O III])	Wandel and Mushotzky (1986)
2.8	8.5	3.4	8.3	AD, $L(\lambda 1450)$	Wandel (1989)
2.0	8.3	2.4	5.9	S	Bochkarev (1998)
25	9.4	30	74	D (Hα)	This paper

* Abbreviated names of the methods: D—dynamical, X—from the X-ray variability rate, AD—from the UV luminosity, S—from Hβ satellites.

0.15 ± 0.2 days at $\lambda 2688$ Å to 0.4 ± 0.2 day at $\lambda 5125$ Å (Edelson *et al.* 1996).

Individually, all these results have a low statistical significance. If, however, we combine all of the data on the lag determination (Oknyanskij *et al.* 2003; Sergeev 2005; Edelson *et al.* 1996) by reducing them to the lag relative to $\lambda 1275$ Å, then we will obtain a clear wavelength dependence of the time lag τ. This dependence is shown in Fig. 10. The dashed line represents a parabolic fit with allowance made for the weight of each point. The lag is $\tau < 0\overset{d}{.}09 \pm 0\overset{d}{.}18$ in the range 1000–3000 Å and $\tau_{4500} = 0\overset{d}{.}33 \pm 0\overset{d}{.}18$ at $\lambda 4500$ Å, i.e., almost twice the error of the parabolic fit, and $\tau_{5500} = 0\overset{d}{.}56 \pm 0\overset{d}{.}18$ at $\lambda 5500$ (the V band) relative to $\lambda 1275$ Å. The lag of the optical ($\lambda > 4000$ Å) variability is determined quite reliably, which yields an effective AD radius of $(1.5 \pm 0.5) \times 10^{15}$ cm in the optical range and no more than 2.3×10^{14} cm (<2.5 lt-hours) in the UV ($\lambda < 3000$ Å). It follows from Figs. 9 and 10 that there is no lag for $\lambda <$ 2000 Å—the energy distribution in this range is well represented by a Planck curve with $T = 65\,000$ K.

It thus follows that the low-temperature ($T =$ 8000 K) source is the reprocessing of energy of the UV source by the accretion disk in the long-wavelength range. Whereas the U luminosities of the two sources are approximately equal, the luminosity of the source with $T = 8000$ K is almost a factor of 3 and 4 higher, respectively, in the V band and at $\lambda 6925$ (close to λ_{eff} of the R band). For the average flare (Fig. 9b), the relationship is different: the luminosity of the UV source at the flare maximum is a factor of 2 higher in the U band and lower in the V band. Incidentally, it is possibly for this reason that our analysis of only the UBV data yielded a temperature at the flare maximum of ~60 000 K. Thus, the reprocessing of the flux from the UV source by the accretion disk with a lag from 0.3 in B to 1 day in R dominates in the optical range (BVR). The gas radiation is also added in the U band.

7.3. The Nature of the Variability

The mass of the central object. A proper allowance for the extinction in the UV (see Section 3) leads to a large, more than two orders of magnitude, increase in the luminosity of the variable source in NGC 4151: $L_{tot} = (6–8) \times 10^{46}$ erg s^{-1}. This value is an order of magnitude higher than the Eddington luminosity for the mass of the central object M_c (below referred to as a supermassive black hole (SBH)) in NGC 4151 determined by the dynamical method from broad lines (C IV, Mg II, Hβ) and is equal to L_{Edd} for M_c determined by the dynamical method

from [O III] lines (Wandel and Mushotzky 1986). Since, in general, the spread in mass estimates for the SBH in NGC 4151 is two orders of magnitude, this question should be considered separately.

Table 2 lists all of the available M_c estimates that have been obtained by various methods since 1970. This table also gives the Eddington luminosity and the gravitational radius for the corresponding mass.

All of the SBH mass estimates can be divided into low ($<M_8$) (in what follows, the mass is given in units of $10^8 M_\odot$), and high ($>M_8$) ones. All of the low mass estimates were obtained by the dynamical method ($M_c = v^2 R/kG$) from broad lines. Whereas the velocity is determined from the line width satisfactorily, the radius is usually estimated by an indirect method that depends on the line luminosity, density, and temperature. The radius of the broad line region seems to be grossly underestimated; hence the low mass. Even the estimation of the radius from the lag of the fast line variability does not yield the true size of the entire BLR. A strange result was obtained by Ulrich *et al.* (1984): M_c depends on the ionization potential of the corresponding line. The minimum and maximum masses are derived from the C IV and Hβ lines, respectively, although Gaskell (1988) derived the same mass from the C IV and Mg II lines. Subsequently, Ulrich and Horne (1996) get the lowest mass estimate, $M_8 = 0.2$, from the C IV line.

The masses estimated by other methods (from the X-ray variability rate, from the UV luminosity, etc.) proved to be close to the mass estimated by Wandel and Mushotzky (1986) by the dynamical method from the [O III] line. However, these estimates are not adequate either. For example, the mass estimate from the X-ray variability rate was made by assuming that the size of the hard X-ray region was $\sim 5R_g$. Since the indirect NLR radius estimates seem to be less dependent on the assumed physical conditions (great stability—no variability), M_c estimated from the [O III] lines is an order of magnitude larger. Unfortunately, what was known in the late 1960s was forgotten in the late 1970s: for nearby Seyfert galaxies, the NLR diameter can be measured directly. In particular, the NLR diameter estimated for NGC 4151 from the [O III] lines is $\sim$50 pc ($0''.8$ at $D = 14$ Mpc). Then, $M_c = 10$–$30 M_8$ (Sanders 1970). This estimate is probably closest to the true mass of the central object.

It was shown above (see Sections 4 and 7.1) that, apart from the well-known broad line region (BLR) responsible for the fast broad line variability with a lag relative to the continuum, there is an extended broad line region (EBLR) in the nucleus of NGC 4151 with an effective radius of $\sim$1.5 lt-years (0.45 pc) that determines the slow component of the Hα variability (Malkov *et al.* 1997). However, the gas velocities in this region are the same as those in the BLR, $\pm$5000 km s^{-1} (Sergeev *et al.* 2001). Taking this velocity and $R = 0.45$ pc, we obtain $M_c = 25 M_8$ from the above formula (the coefficient k, which is equal to 0.5–2, depends on the kinematic model and, in our case, was set equal to unity), which is in excellent agreement with the estimate by Sanders (1970) obtained from [O III] (the mean value is given in Table 2). An Eddington luminosity of $L_{Edd} = 3 \times 10^{47}$ erg s^{-1} and a gravitational radius of $R_g \approx 7 \times 10^{14}$ cm correspond to a mass of $2.5 \times 10^9 M_\odot$. This mass of the central object is in better agreement with the huge bulge of the giant spiral NGC 4151 (Arp 1977; Doroshenko *et al.* 1998) than the low mass estimates. The total luminosity of $(6$–$8) \times 10^{46}$ erg s^{-1} then corresponds not to 1–2% of L_{Edd}, as is generally believed, but to 20–40%.

Disk accretion onto a SBH. Thus, our analysis of the optical and UV variability leads to the following picture. After a prolonged minimum (1984–1989), the formation of a new AD with different parameters and a factor of 2 higher luminosity began. The rate of mass inflow into the disk determines the amplitude of the S component, which depends on both the temperature and the size of the disk. The S component is observed not only in the optical range, but also in the UV, at least up to 1000 Å; its amplitude depends weakly on the wavelength. In contrast to the UV/optical range, the S component is almost completely absent in the X-ray (1–10 keV) range (Lawrence 1980; Papadakis and McHardy 1995). This is also clearly demonstrated by the RXTE data: from 1996 until 2001, the amplitude of the S component in the range 1–10 keV did not change, within the error limits, while in the optical range, it fell by more than an order of magnitude (Fig. 1).

However, the mean flux in the second cycle of activity (1989–2003) increased by a factor of 2 compared to the first cycle (1968–1983), from 10 to 20 mCrab, in agreement with the same increase in optical luminosity (Lyuty and Doroshenko 1999).

Aperiodic (or possibly quasi-periodic on short, 2–3 year, time scales) flares are superimposed on the S component. The main property of the average flare is that its rise time Δt does not depend on the flare amplitude in both the optical and UV ranges. However, Δt decreases with wavelength from 27 in V to 6 days at λ1000 Å. This dependence (Table 1) is satisfactorily, within the error limits, represented by a parabola whose extrapolation to $\Delta t = 0$ yields $\lambda = 680$ Å, which is close to λ_{max} for $T_e = 65\,000$ K. This is direct evidence, first, for the decrease in the

size of the emitting region with decreasing wavelength and, second, for the main energy release precisely in the UV spectral range. In addition, the lag between the long-wavelength and short-wavelength emissions (Fig. 10) yields the AD radius: $R = c\tau$. In the optical range, $\lambda 5500$, $R_V = 15 \times 10^{14}$ cm or $2R_g$ for $M = 2.5 \times 10^9 M_\odot$. In combination with $\Delta t_V = 27^d$, this gives $\sim$6000 km s^{-1} for the velocity of propagation of the flare over the disk. If this velocity does not depend on the wavelength, then the disk radius at $\lambda = 1000$ Å will be a factor of 4 or 5 smaller, i.e., $0.5R_g$. In general, this is exactly the radius of the marginally stable orbit for a Kerr black hole.

The mass of the central object is at best estimated to within a factor of 2 or 3. Sanders (1970) gave $(11-35)M_8$. In this paper, we also obtained $25M_8$ to within a factor of at least 2 ($k = 0.5-2$, R could be slightly smaller than 0.45 pc). On the other hand, the luminosity of the variable source, $(6-8) \times 10^{46}$ erg s^{-1}, requires no less than $5M_8$; in this case, the luminosity will be equal to the critical (Eddington) luminosity. The most plausible mass estimates appears to be $M_c = 1 \times 10^9 M\odot$. The AD luminosity then accounts for 50–60% of the critical luminosity $L_{Edd} = 1.2 \times 10^{47}$, the gravitational radius is $R_g = 3 \times 10^{14}$ cm, and the disk radius in the UV is $R_{1000} = 1.3R_g$.

The observations in December 1993 showed that the flare maximum is reached almost simultaneously (the lag does not exceed $0\overset{d}{.}5$) not only in the UV/optical range ($\lambda 1275-6990$ Å), but also in the X-ray range (1–2 keV). A linear extrapolation of the ascending branch of the X-ray flare similar to that for the UV/optical flare (Fig. 8) yields $\Delta t_{XR} = 6^d$, as at $\lambda 1000$ Å. Hence, the X-ray flare cannot precede the UV flare, as Edelson *et al.* (1996) believed; i.e., the UV flux is not the reprocessed X-ray flux. In addition, the UV luminosity is two orders of magnitude higher than the X-ray (1–10 keV) luminosity.

This picture is in excellent agreement with the theory of disk accretion instability by Shakura and Sunyaev (1976) when applied to supermassive black holes: for $M \sim 10^9 M_\odot$, disk accretion can provide a luminosity of $\sim 10^{47}$ erg s^{-1}, the radiation temperature at $L/L_{Edd} \simeq 0.1-1$ is $\sim 10^5$ K, the minimum variability time scale is 2 or 3 days. The inner disk region where the radiation pressure dominates is unstable against small perturbations. The growing short-wave perturbations propagate in the form of concentric waves. This closely agrees with the observed fast (flare) component with Δt decreasing from the optical range to the UV and with the fact that the velocity of propagation of the flare over the disk is much lower than the speed of light.

This paper by Shakura and Sunyaev has been almost forgotten, although the theory of stationary accretion (Shakura and Sunyaev 1973) has still been alluded to. The reason is clear: the X-ray observations began much earlier than the UV ones and yielded an X-ray luminosity that is an order of magnitude higher than the optical luminosity. It was established that the bulk of the energy is released in the X-ray range and is subsequently reprocessed into low-frequency ranges. The UV observations were consistent with this view, because the extinction in the UV was grossly underestimated. Incidentally, note an interesting feature: the observed flux (and, hence, the luminosity) in the optical and UV ranges in 1995 tripled compared to that in 1978–1980, while the X-ray (1–10 keV) flux remained approximately the same. Our analysis showed that when the UV extinction is properly taken into account, the observed pattern of variability in NGC 4151 is in good agreement with the theory by Shakura and Sunyaev (1976) not only qualitatively, but also quantitatively.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 03-02-17123) and the Federal Science and Technology Program (contract no. 40.022.1.1.1101 of the Ministry of Science).

REFERENCES

1. I. Aretxaga and R. Terlevich, Mon. Not. R. Astron. Soc. **269**, 462 (1994).
2. H. Arp, Astrophys. J. **218**, 70 (1977).
3. N. G. Bochkarev, *IAU Symposium No. 188: The Hot Universe*, Ed. by K. Koyama, Sh. Kitamoto, and M. Itoh (Kluwer, Dordrecht, 1998), p. 421.
4. N. N. Chalenko, Astron. Zh. **76**, 529 (1999) [Astron. Rep. **43**, 459 (1999)].
5. J. Clavel, A. Altamore, A. Boksenberg, *et al.*, Astrophys. J. **321**, 251 (1987).
6. D. M. Crenshaw, P. M. Rodrigues-Pascual, S. V. Penton, *et al.*, Astrophys. J. **470**, 322 (1996).
7. D. M. Crenshaw, S. B. Kraemer, J. B. Hutchings, *et al.*, Astrophys. J. Lett. **545**, L27 (2000).
8. É. A. Dibaĭ, Pis'ma Astron. Zh. **3**, 1 (1977) [Sov. Astron. Lett. **3**, 1 (1977)].
9. É. A. Dibaĭ, Astron. Zh. **57**, 677 (1980) [Sov. Astron. **24**, 389 (1980)].
10. É. A. Dibaĭ and V. M. Lyuty, Astron. Zh. **61**, 10 (1984) [Sov. Astron. **28**, 7 (1984)].
11. V. T. Doroshenko and V. M. Lyuty, Pis'ma Astron. Zh. **20**, 703 (1994) [Astron. Lett. **20**, 606 (1994)].
12. V. T. Doroshenko, V. M. Lyuty, and N. G. Bochkarev, Pis'ma Astron. Zh. **27**, 83 (2001) [Astron. Lett. **27**, 65 (2001)].

13. V. T. Doroshenko, V. M. Lyuty, and V. I. Shenavrin, Pis'ma Astron. Zh. **24**, 197 (1998) [Astron. Lett. **24**, 160 (1998)].
14. R. A. Edelson, T. Alexander, D. M. Crenshaw, *et al.*, Astrophys. J. **470**, 364 (1996).
15. B. R. Espey, G. A. Kriss, J. H. Krolik, *et al.*, Astrophys. J. Lett. **500**, L13 (1998).
16. B. R. Espey, Astron. Soc. Pac. Conf. Ser. **175**, 379 (1999).
17. S. Honda, K. Sadakane, and K. Tanabe, Mem. Osaka Kyoiku Univ., Ser. 3 **46**, 219 (1998).
18. V. A. Hagen-Thorn, Astrofizika **26**, 415 (1987a) [Astrophys. **26**, 252 (1987a)].
19. V. A. Hagen-Thorn, *Active Nuclei and Stellar Cosmogony* (Mosk. Gos. Univ., Moscow, 1987b) [in Russian].
20. C. M. Gaskell, Astrophys. J. **325**, 114 (1998).
21. H. Gursky, E. M. Kellogg, C. Leong, *et al.*, Astrophys. J. Lett. **165**, L43 (1971).
22. S. Kaspi, D. Maoz, H. Netzer, *et al.*, Astrophys. J. **470**, 336 (1996).
23. G. A. Kriss, A. F. Davidsen, W. P. Blair, *et al.*, Astrophys. J. **392**, 485 (1992).
24. G. A. Kriss, A. F. Davidsen, W. Zheng, *et al.*, Astrophys. J. Lett. **454**, L7 (1995).
25. A. Lawrence, Mon. Not. R. Astron. Soc. **192**, 83 (1980).
26. V. M. Lyutyi, Astron. Tsirk **619**, 1 (1971) [Astron. Circ. **5**, 626 (1971)].
27. V. M. Lyutyi, Astron. Zh. **49**, 930 (1972) [Sov. Astron. **16**, 763 (1972)].
28. V. M. Lyutyi, Astron. Zh. **54**, 1153 (1977) [Sov. Astron. **21**, 655 (1977)].
29. V. M. Lyutyi, Astron. Zh. **56**, 918 (1979) [Sov. Astron. **23**, 518 (1979)].
30. V. M. Lyutyi, Astron. Soc. Pac. Conf. Ser. (2005) (in press).
31. V. M. Lyutyi and V. T. Doroshenko, Pis'ma Astron. Zh. **25**, 403 (1999) [Astron. Lett. **25**, 341 (1999)].
32. V. M. Lyutyi and V. L. Oknyanskij, Astron. Zh. **64**, 465 (1987) [Sov. Astron. **31**, 245 (1987)].
33. V. M. Lyutyi, V. L. Oknyanskij, and K. K. Chuvaev, Pis'ma Astron. Zh. **10**, 803 (1984) [Sov. Astron. Lett. **10**, 335 (1984)].
34. V. M. Lyutyi and V. I. Pronik, *IAU Symposium No. 67: Variable Stars and Stellar Evolution*, Ed. by Sherwood Plaut (Reidel, Dordrecht, 1975), p. 591.
35. V. M. Lyutyi, O. G. Taranova, and V. I. Shenavrin, Pis'ma Astron. Zh. **24**, 243 (1998) [Astron. Lett. **24**, 199 (1998)].
36. Yu. F. Malkov, V. I. Pronik, and S. G. Sergeev, Astron. Astrophys. **324**, 904 (1997).
37. B. Q. McGimsey, A. G. Smith, R. L. Scott, *et al.*, Astron. J. **80**, 895 (1975).
38. N. I. Merkulova and L. P. Metik, Izv. Krym. Astrofiz. Obs. **87**, 135 (1993).
39. N. I. Merkulova and L. P. Metik, Izv. Krym. Astrofiz. Obs. **90**, 178 (1995).
40. N. I. Merkulova and L. P. Metik, Astron. Astrophys. Trans. **10**, 295 (1996).
41. N. I. Merkulova, L. P. Metik, and I. I. Pronin, Astron. Astrophys. **374**, 770 (2001).
42. J. P. D. Mittaz, M. V. Penston, and M. A. J. Snijders, Mon. Not. R. Astron. Soc. **242**, 370 (1990).
43. R. F. Mushotzky, S. S. Holt, and P. J. Serlemitsos, Astrophys. J. Lett. **225**, L115 (1978).
44. V. L. Oknyanskiĭ, V. M. Lyutyi, O. G. Taranova, and V. I. Shenavrin, Pis'ma Astron. Zh. **25**, 563 (1999) [Astron. Lett. **25**, 483 (1999)].
45. V. L. Oknyanskij, K. Horne, V. M. Lyutyi, *et al.*, Astron. Soc. Pac. Conf. Ser. **290**, 119 (2003).
46. I. E. Papadakis and I. M. McHardy, Mon. Not. R. Astron. Soc. **273**, 923 (1995).
47. M. V. Penston and E. Perez, Mon. Not. R. Astron. Soc. **211**, 33P (1984).
48. M. V. Penston, M. J. Penston, and A. Sandage, Publ. Astron. Soc. Pac. **83**, 783 (1971).
49. M. V. Penston, A. Boksenberg, G. E. Bromage, *et al.*, Mon. Not. R. Astron. Soc. **196**, 857 (1981).
50. G. C. Perola, A. Boksenberg, G. E. Bromage, *et al.*, Mon. Not. R. Astron. Soc. **200**, 293 (1982).
51. A. J. Pica, A. G. Smith, J. R. Webb, *et al.*, Astron. J. **96**, 1215 (1988).
52. V. Yu. Rakhimov, Byull. Astrofiz. Inst. Tadzhik. Akad. Nauk **78**, 50 (1989).
53. R. H. Sanders, Astrophys. J. **159**, 1115 (1970).
54. M. J. Seaton, Mon. Not. R. Astron. Soc. **187**, 73P (1979).
55. S. G. Sergeev, Astron. Soc. Pac. Conf. Ser. (2005) (in press).
56. S. G. Sergeev, V. I. Pronik, and E. A. Sergeeva, Astrophys. J. **554**, 245 (2001).
57. N. I. Shakura and R. A. Sunyaev, Astron. Astrophys. **24**, 337 (1973).
58. N. I. Shakura and R. A. Sunyaev, Mon. Not. R. Astron. Soc. **175**, 613 (1976).
59. V. Yu. Terebizh, A. V. Terebizh, and V. V. Biryukov, Astrofizika **31**, 76 (1989).
60. M.-H. Ulrich and K. Horne, Mon. Not. R. Astron. Soc. **283**, 748 (1996).
61. M. H. Ulrich, A. Boksenberg, G. E. Bromage, *et al.*, Mon. Not. R. Astron. Soc. **206**, 221 (1984).
62. M. H. Ulrich, A. Boksenberg, G. E. Bromage, *et al.*, Mon. Not. R. Astron. Soc. **382**, 483 (1991).
63. A. Wandel, *Active Galactic Nuclei*, Ed. by D. E. Osterbrock and J. S. Miller (Kluwer, Dordrecht, 1989), p. 62.
64. A. Wandel and R. F. Mushotzky, Astrophys. J. Lett. **306**, L61 (1986).
65. M. J. Ward, T. Geballe, M. Smith, *et al.*, Astrophys. J. **316**, 138 (1987).
66. R. S. Warwick, D. A. Smith, T. Yaqoob, *et al.*, Astrophys. J. **470**, 349 (1996).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 663–667. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 743–748.

Kinematic Parameters of the Galactic Spiral Pattern from Data on Open Star Clusters and OB Stars

M. É. Popova and A. V. Loktin*

Astronomical Observatory, Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia

Received December 24, 2004

Abstract—Data on the positions, radial velocities, and proper motions of open star clusters and OB stars are used to obtain the rotation curve of the Galaxy fitted by a polynomial in inverse powers of the distances from the Galactic rotation axis. We determine the locations of the corotation region and the inner and outer Lindblad resonances using a previously estimated pattern speed. Based on data for objects of the Carina–Sagittarius and Orion arms, we have determined the distortion amplitudes of the velocity field of the Galactic disk, $f_R = -3.97 \pm 4.79$ km s^{-1} and $f_\theta = +13.27 \pm 2.57$ km s^{-1}. © *2005 Pleiades Publishing, Inc.*

Key words: *Galactic spiral pattern, Galactic rotation curve, open star clusters, OB stars.*

INTRODUCTION

Previously (Popova and Loktin 2005), we determined some of the parameters of the Galactic spiral pattern using the positions and ages of open star clusters (OSCs). In particular, we determined the pattern speed, $\Omega_P = 20.4 \pm 2.5$ km s^{-1} kpc^{-1}; the initial phase of the spiral pattern; and the pitch angle of the spiral arms, $i = 21°.5$. Additional information about the properties of the spiral pattern can be obtained from kinematic data on objects of the Galactic disk. With the publication of catalogs of stellar proper motions based on Hipparcos observations, the proper motions of objects, in particular, those from the Tycho-2 catalog, can be used along with their radial velocities. The proper motions for a large number of OSCs based on Tycho-2 data are given in Beshenov and Loktin (2004).

THE GALACTIC ROTATION CURVE

To analyze the global distortions of the circular velocity field produced by a spiral density wave in the Galactic disk, the Galactic rotation curve must be determined from the data set used here. In this case, we can perform our analysis in terms of a single distance scale. Therefore, we decided to construct the Galactic rotation curves from the radial velocities V_r and proper motions μ of OSCs. To determine both the rotation curve and its errors more reliably, we also decided to use the radial velocities and proper motions of a sample of OB stars. As a result, we obtained four samples (two for inferring the rotation curve from radial velocities and two for inferring the rotation curve from proper motions); their parameters are listed in the first three columns of Table 1, where N is the sample size. The radial velocities of OSCs were chosen from separate publications, and this list was supplemented by as yet unpublished mean values kindly provided by N. Kharchenko and A. Piskunov. The proper motions of OSCs were taken from Beshenov and Loktin (2004). The kinematic and photometric data for OB stars were chosen from a card catalog compiled by one of the authors, and the proper motions were taken from the Tycho-2 catalog. We determined the heliocentric distances of OB stars using the luminosity calibration related to the OSC distance scale adopted here (Beshenov and Loktin 2001) that was constructed from Hipparcos trigonometric parallaxes.

We determined the angular velocity of the Galaxy $\omega = \omega(R)$, where R is the distance from the Galactic rotation axis, using Bottlinger's formulas

$$V_r = R_\odot(\omega - \omega_0)\sin l \cos b, \qquad (1)$$
$$V_l = R_\odot(\omega - \omega_0)\cos l \cos b - \omega r \cos b.$$

In these formulas, we used the Galactocentric distance of the Sun, $R_\odot = 8.3$ kpc, from Gerasimenko (2003), who determined this parameter from OSC data on the same distance scale as here. We took $\omega_0 = (24.5 \pm 0.8)$ km s^{-1} kpc^{-1} from Loktin and Beshenov (2003). The radial, V_r, and tangential, V_l, velocities were corrected for the solar motion with its components $U_\odot = -8.5$ km s^{-1}, $V_\odot = 14.8$ km s^{-1}, and $W_\odot = 7.0$ km s^{-1} taken from Loktin (1978).

For the convenience of subsequent calculations, we fitted the derived dependences of the angular ve-

*E-mail: alexhander.loktin@usu.ru

Table 1. Coefficients of the polynomial fit to the Galactic rotation curve

Sample	Data	N	A	B	C
OSCs	V_r	196	140.4	220.4	2641.1
OSCs	μ	98	152.5	184.4	1724.1
OB stars	V_r	1671	145.3	227.5	2108.5
OB stars	μ	795	209.6	−23.5	218.0
Mean			162.0 ± 32.2	152.2 ± 118.6	1672.9 ± 1040.3

locity ω on the Galactocentric distance R by polynomials of the form

$$\omega = \frac{A}{R} + \frac{B}{R^2} + \frac{C}{R^3}, \qquad (2)$$

where A, B, and C are the unknown coefficients that were determined separately for each of the four samples by the least-squares method. We did not use higher powers of R, because the corresponding coefficients differed insignificantly from zero for all our samples. The derived coefficients A, B, and C are listed in columns 4–6 of Table 1. The last row of the table gives the coefficients averaged over the four samples. We used no weights in the averaging procedure, because a justified system of weights was difficult to choose for the four different samples and because several systems of weights that we tested yielded virtually indistinguishable results.

Figure 1 shows the angular velocities of the Galaxy $\omega(R)$ as a function of the distance from the Galactic rotation axis calculated from the proper motions (filled circles) and radial velocities (open circles) of OSCs. The solid and dotted lines indicate polynomial fit (2) with the coefficients corresponding to the sample under consideration for the proper motions and radial velocities, respectively. A characteristic feature of this plot is that the scatter of data points based on proper motions is smaller than the scatter of data points based on radial velocities. This suggests that the OSC proper motions from Beshenov and Loktin (2004) are highly reliable. It is also clear that the discrepancies between the curves obtained from proper motions and radial velocities can be explained in terms of random errors in the data and cosmic dispersion, so the Tycho-2 proper motions require no appreciable systematic corrections.

Figure 2 shows the angular velocities $\omega(R)$ calculated from the proper motions (filled circles) and radial velocities (open circles) of OB stars, and the curves indicate the corresponding polynomial fits. In this case, the data points determined from proper motions also exhibit a smaller scatter than those corresponding to radial velocities. The significantly larger scatter of data points in Fig. 2 than that in Fig. 1 can be explained by much larger errors in the heliocentric distances of OB stars than those in the heliocentric distances of OSCs. That is why the slopes of the rotation curves differ markedly. We see from Figs. 1 and 2 that OSC data make it possible to construct the rotation curve in the range of Galactocentric distances R approximately from 6.5 to 12 kpc, while data on OB stars are useful in the range from 5.2 to 11.5 kpc (with significant errors near the ends of this range).

In what follows, we use polynomial (2) with the coefficients given in column 5 of Table 1 as the rotation curve.

Figure 3 compares our rotation curve (solid line) with the rotation curves from Fich *et al.* (1989) (dashed line) and Brandt and Blitz (1993) (dotted line). These three curves constructed from different initial data for the R range in which we determined our rotation curve are similar. The slightly larger slope of our curve is probably attributable to a "compression" of our distance scale and to small differences in the adopted values of $R_\odot$ and ω_0. It should also be pointed out that the slope of the rotation curve is actually determined by a few data points with the smallest Galactocentric distances, for which the random errors of the initial data are largest. Note, in particular, that the curve obtained from the proper motions of OB stars with the largest deviation from the mean curve (see Table 1) has the smallest slope.

THE LOCATIONS OF RESONANCES IN THE GALACTIC DISK

The rotation curve derived above and the pattern speed Ω_P estimated previously (Popova and Loktin 2005) allow an attempt to be made to determine the locations of the regions that characterize the properties of the spiral pattern and the Galactic disk: the corotation region and the Lindblad resonances. Figure 4 shows the adopted Galactic rotation curve $\omega(R)$ and the pattern speed Ω_P. The hatching indicates the error corridors for our rotation curve and Ω_P. Also shown are the lines corresponding to the rotation curve shifted upward and downward by half

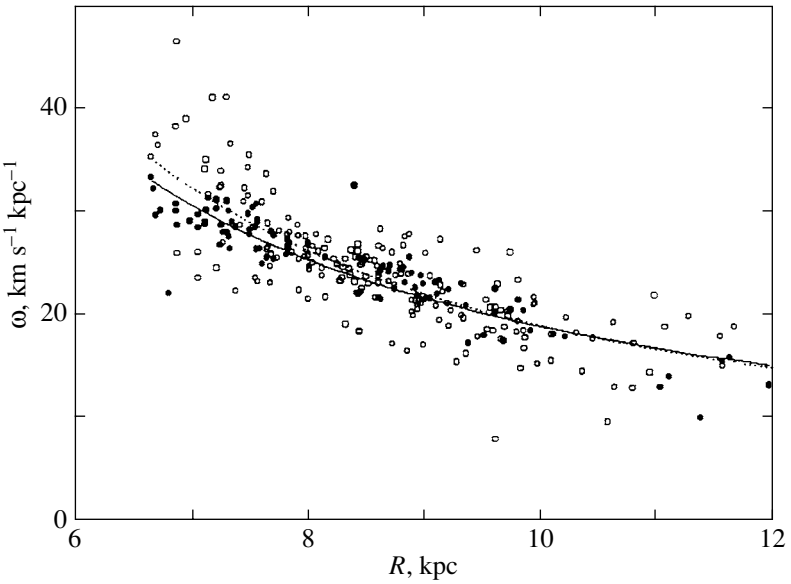


Fig. 1. Galactic rotation curve calculated from the proper motions (filled circles) and radial velocities (open circles) of OSCs. The solid and dotted lines indicate the corresponding polynomial fits.

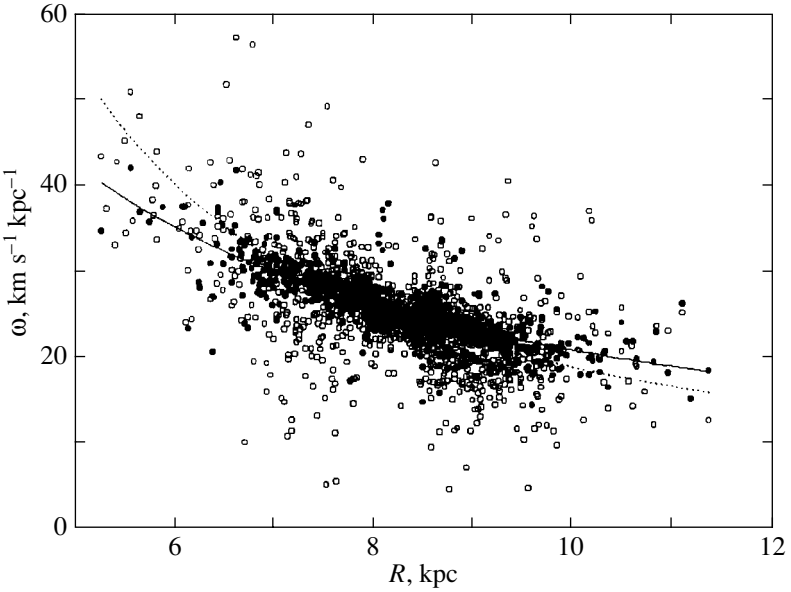


Fig. 2. Same as Fig. 1 for OB-stars.

the epicyclic frequency for our rotation curve and shifted downward for the rotation curve from Fich *et al.* (1989). We determined the epicyclic frequency using the standard formula

$$\kappa^2 = 4\omega^2 \left(1 + \frac{R}{2\omega}\frac{\partial\omega}{\partial R}\right). \qquad (3)$$

It is clear from the figure that the corotation region, the Galactocentric distance at which the angular velocity of the Galactic disk is equal to the pattern speed, is located approximately between 8.9 and 10.8 kpc, and the accuracy of its location is determined mainly by the estimation error of Ω_P. Previously (Popova and Loktin 2005), we estimated the current location of the Perseus arm to be $R =$ 10.5 kpc (along the Sun–Galactic center line); the OSC age gradient points in the same direction as that in the Orion and Carina–Sagittarius arms. It can thus be concluded that the corotation region in our Galaxy lies at a small distance from the Perseus arm (beyond it), i.e., beyond $R = 10.5$ kpc. The corotation

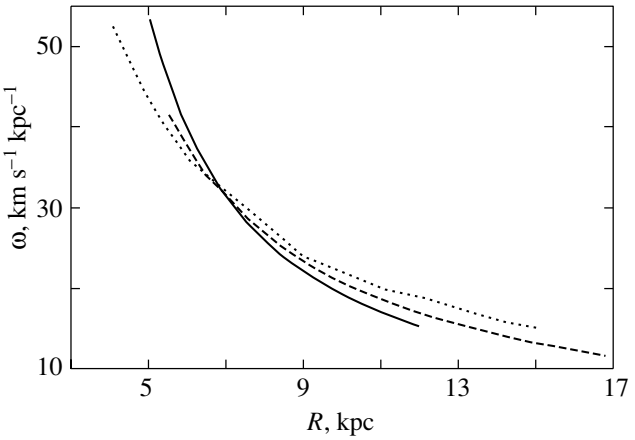


Fig. 3. Comparison of our rotation curve (solid line) with the rotation curves from Fich *et al.* (1989) (dashed line) and Brandt and Blitz (1993) (dotted line).

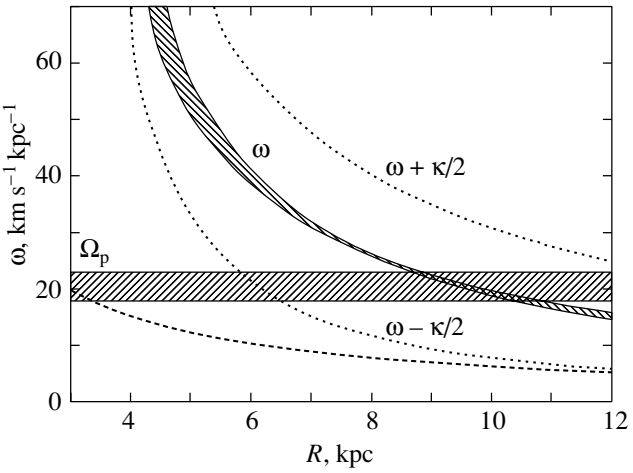


Fig. 4. Dependence of the adopted Galactic rotation curve, pattern speed, and epicyclic frequency on the Galactocentric distance. The hatching indicates the error corridors for the rotation curve $\omega(R)$ and the pattern speed Ω_P. The dashed lines indicate the R dependence of $\omega \pm \kappa/2$ for the rotation curve from Fich *et al.* (1989).

region is unlikely to lie in the immediate vicinity of the solar orbit in the Galaxy.

The curves $\omega \pm \kappa/m$ (m is the number of arms in the spiral pattern) can be used to determine the locations of the outer and inner Lindblad resonances for the given $\Omega_P = 20.4$ km s^{-1} kpc^{-1}, between which, according to the theory (see, e.g., the monograph by Marochnik and Suchkov (1984)), a spiral density wave can propagate. It is clear from Fig. 4 that, according to our data, the inner Lindblad resonance at $m = 2$ lies at a distance of $R \approx 6$ kpc. However, since this value is at the boundary of the range of Galactocentric distances used to construct our rotation curve, it is determined with a large error, especially if it is considered that the epicyclic frequency is calculated via the derivative of the rotation curve. If our Galaxy has a four-armed spiral pattern, then the inner Lindblad resonance must be located no farther than $R = 7.2$ kpc, which disagrees with the locations of the Carina–Sagittarius and inner arms. Thus, although many authors (for references, see Popova and Loktin 2005) have determined the spiral pattern of our Galaxy as a four-armed one, our data favor a two-armed pattern.

If we used the rotation curve from Fich *et al.* (1989) that is less steep than our rotation curve, then the inner Lindblad resonance would displace toward the Galactic center to a Galactocentric distance of about $R = 3$ kpc. This leads us to conclude that the rotation curve inferred from stellar objects is not reliable enough for such works. Note also that we have now obtained the rotation curve from data on classical Cepheids, which is also steeper than that from Fitch *et al.* (1989) in the region of interest. Clearly, at present, we cannot determine the location of the inner Lindblad resonance with a high accuracy from the data used here. In conclusion, note that the rotation curves determined from the motions of gas and stellar objects do not necessarily coincide closely, but this question requires a separate analysis.

The data used here do not make it possible to draw any definitive conclusions about the location of the outer Lindblad resonance either (see Fig. 4); we can only assert that it is definitely beyond $R = 12$ kpc.

THE RESIDUAL VELOCITY FIELD

To complete the picture of peculiarities of the motions of Galactic disk objects associated with the effect of the spiral pattern, let us determine the distortion amplitudes of the velocity field of the Galactic disk. The formulas describing the velocity field of circular motions in the Galactic disk with allowance made for the density wave effects are given in a monograph by Marochnik and Suchkov (1984). Since we determined many parameters that appear in these formulas previously, the formulas simplify to

$$\begin{aligned} V_r &= R_\rho(w - w_0)\sin l\cos b \\ &- f_R\cos(\chi_0 + \epsilon)\cos(l + \theta)\cos b \\ &+ f_\theta\sin(\chi_0 + \epsilon)\sin(l + \theta)\cos(b), \\ V_l &= -r\cos b\omega_0 - f_R\cos(\chi_0 + \epsilon)\sin(l + \theta) \\ &+ f_\theta\sin(\chi_0 + \epsilon)\cos(l + \theta), \end{aligned} \tag{4}$$

where V_r and V_l are the radial and tangential velocities of objects corrected for the solar motion in space,

$$\epsilon = m[(\theta - \theta_0) + \cot(i)\ln(R/R_0)], \tag{5}$$

and χ_0 was calculated from the formula

$$R_2 - R_0 = \Delta R(2\pi - \chi_0)/2\pi, \qquad (6)$$

$$R_8 - R_1 = \Delta R\chi_0/2\pi,$$

where $\Delta R = R_2 - R_1$ and R_1 and R_2 are the distances from the center to the inner and outer arms, respectively. The quantities with the subscript zero refer to the location of the Sun. We set θ_0 equal to zero and took $i = 21^\circ\!.5$ from our previous paper (Popova and Loktin 2005). For the reasons given above, we used $m = 2$ in our calculations, i.e., we adopted a two-armed model of the Galaxy.

We took the Carina–Sagittarius and Orion arms as the inner and outer arms, respectively. The Galactocentric distances of these arms along the Galactic center–Sun line were determined previously and were assumed to be 7.21 and 8.80 kpc, respectively. We did not use the Perseus arm, because our sample contains a small number of objects in this region. The amplitudes f_R and f_θ were estimated from formulas (4) by the least-squares method. To obtain more robust estimates, we excluded objects with large residual velocities and repeated the calculations. Unfortunately, this process reduced only slightly the scatter of estimates obtained from different samples, although it slightly reduced the dispersions of these estimates. Therefore, we decided to use the weighted mean estimates for all samples where the inverse squares of the rms errors of the estimates obtained for each sample were used as weights. Table 2 presents the results of our calculations for the samples described above. Averaging without weights yielded $f_R = (-7.9 \pm 5.4)$ km s^{-1} and $f_\theta = (+15.0 \pm 2.4)$ km s^{-1}. We see that using weights in this case slightly reduces the errors of the estimates.

In this case, the solutions based on radial velocities resulted in smaller estimation errors of the amplitudes than the solutions based on proper motions. Our weighted mean amplitudes proved to be close to those determined previously by Pavlovskaya and Suchkov (1980): $f_R = (-3.1 \pm 2.3)$ km s^{-1} and $f_\theta = (+13.6 \pm 2.3)$ km s^{-1}.

Table 2. Estimates of the distortion amplitudes of the velocity field of the Galactic disk

Sample	Date	f_R, km s^{-1}	f_θ, km s^{-1}
OSC	V_r	-3.03 ± 4.84	18.26 ± 2.91
OSC	μ	-21.26 ± 8.92	18.05 ± 3.64
OB stars	V_r	3.79 ± 1.64	7.92 ± 1.10
OB stars	μ	-11.25 ± 2.44	15.68 ± 1.32
Weighted mean		-3.97 ± 4.79	13.27 ± 2.57

CONCLUSIONS

Data on the motions and spatial locations of OSCs and OB stars yielded an almost complete set of parameters characterizing the properties of the Galactic spiral pattern. At the same time, we failed to determine the number of arms in the spiral pattern. So far, it is also hard to hope to infer the dependence of the distortion amplitudes of the velocity field on the Galactocentric distance and the locations of the inner and outer Lindblad resonances from the available data. This requires more data on the OSCs closest to the Galactic center.

In the future, we plan to improve the inferred values based on the locations and motions of Cepheids. Cepheids, which are objects that are seen far from the Sun and whose ages can be estimated from the period–age relation, will help to obtain another independent series of estimates for the parameters of interest. This is especially true for the estimates of the distortion amplitudes of the velocity field that were determined from our data with an insufficient accuracy. The unique choice of the number of arms in the spiral pattern remains the main question. Analysis of the motions of gaseous molecular clouds of the Galactic disk may also help solve this question.

REFERENCES

1. G. V. Beshenov and A. V. Loktin, Astron. Zh. **78**, 819 (2001) [Astron. Rep. **45**, 711 (2001)].
2. G. V. Beshenov and A. V. Loktin, Astron. Astrophys. Trans. **23**, 103 (2004).
3. J. Brand and L. Blitz, Astron. Astrophys. **67**, 275 (1993).
4. M. Fich, L. Blitz, and A. A. Stark, Astrophys. J. **342**, 272 (1989).
5. T. P. Gerasimenko, Astron. Rep. **48**, 103 (2003).
6. A. V. Loktin, Astron. Tsirk, No. 1021, 1 (1978).
7. A. V. Loktin and G. V. Beshenov, Astron. Zh. **80**, 8 (2003) [Astron. Rep. **47**, 6 (2003)].
8. L. S. Marochnik and A. A. Suchkov, *The Galaxy* (Nauka, Moscow, 1984) [in Russian].
9. E. D. Pavlovskaya and A. A. Suchkov, Astron. Zh. **57**, 280 (1980) [Sov. Astron. **24**, 164 (1980)].
10. M. É. Popova and A. V. Loktin, Pis'ma Astron. Zh. **31**, 190 (2005) [Astron. Lett. **31**, 171 (2005)].

Translated by A. Dambis

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 668–671. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 749–753.

Breakup of Moving Clusters in the Galactic Disk

A. A. Mülläri[1], V. V. Orlov[2*], and A. V. Petrova[2]

[1] *Department of Information Technologies, Turku University, Turku, Finland*

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

Received March 18, 2005

Abstract—Numerical simulations are used to analyze the breakup of moving groups in the Galactic disk through the internal velocity dispersion of the group components and the tidal effect of the external regular gravitational Galactic field and giant molecular clouds. The initial locations of the group centroids correspond to well-known moving streams: the Hyades, the Pleiades, the Ursa Major cluster, and the group HR 1614. The mean group breakup times have been found as a function of the adopted limiting group size. The interactions of stream stars with giant molecular clouds reduce significantly the group lifetime.

Key words: *star clusters and associations, stellar dynamics, Galaxy (Milky Way)*.

INTRODUCTION

Analysis of the space velocities of stars in the solar neighborhood shows that this distribution contains groups with various numbers of stars (see, e.g., Mülläri and Orlov 1997; Chereul *et al*. 1999; and references therein). The origins of these groups of stars with similar space motions may be different (see, e.g., Orlov *et al*. (1995) for a discussion). Some of the groups may be the products of the breakup of open star clusters, such as the Hyades, the Pleiades, or the Ursa Major cluster, with the breakup of clusters taking place before our eyes (Brown 2001). It is of interest to estimate the breakup time scales of such groups under the effect of the external Galactic gravitational field, through the internal velocity dispersion of the stars that have escaped from the cluster, and under the effect of giant molecular clouds (GMCs) populating the Galactic disk.

In this paper, we attempted to obtain such estimates.

THE METHOD OF ANALYSIS

Stars dissipate from open clusters because of encounters between stars within the cluster, the tidal effect of the Galactic gravitational field on the cluster periphery, and encounters of the cluster with other Galactic objects (e.g., GMCs). This results in the formation of a structure that consists of a physically bound star cluster and a stream of stars ejected from the cluster. We are interested in the subsequent dynamical evolution of the stream.

We will consider the stream as a set of physically unbound stars with an internal one-dimensional velocity dispersion σ (the superposition of three Gaussians). At the initial time, the stream stars are uniformly and randomly distributed within a cube with a 40-pc edge. The cube center coincides with the current location of the Sun in the Galaxy. The stream centroid velocities correspond to the heliocentric motions of four well-known moving groups: the Hyades, the Pleiades, the Ursa Major cluster, and the group HR 1614. We took data from Orlov *et al*. (1995) for the first three groups and from Feltzing and Holmberg (2000) for the fourth group. Table 1 lists the heliocentric centroid velocities of these groups. The U, V, and W axes are directed toward the Galactic center, along the Galactic rotation, and toward the North Galactic Pole, respectively. We took $\sigma = 1$ km s^{-1} for the bulk of our statistical experiments, although we also considered other values of σ in certain specified cases.

Table 1. Initial velocities of group centroids

Group	U_0, km s^{-1}	V_0, km s^{-1}	W_0, km s^{-1}
Hyades	−40.5	−17.5	−3.4
Pleiades	−14.2	−21.7	−10.2
UMa	+13.5	+2.8	−8.5
HR 1614	−25	−63	−14

*E-mail: vor@astro.spbu.ru

1063-7737/05/3110-0668$26.00

We analyzed the motion of each of the $N = 1000$ stars in the moving group relative to the group centroid; the location of the group centroid at the initial time coincided with the current location of the Sun in the Galaxy. The gravitational field of the Galaxy was represented by the potentials suggested by Kutuzov and Osipkov (1989) and Flynn *et al.* (1996). The times t the star escaped from the spheres with radii of $r_c = 50$, 100, 200, and 400 pc relative to the group centroid were recorded. We determined the mean times $\bar{t}$ and their rms deviations.

We used the following approach to allow for the effect of GMCs on the evolution of the stream in the Galactic gravitational field. We numerically simulated the interaction of stream stars with GMCs by assuming that the GMCs move in circular orbits in the Galactic plane. We took into account the interaction of stream stars only with the GMCs located within a square with a 2-kpc side; the center of the square coincided with the projection of the group centroid onto the Galactic plane. The number of GMCs within this square was fixed at 16, as in Hänninen and Flynn (2001). If one of the clouds escaped from the square, then a new cloud was introduced on the opposite side of the square.

The GMC masses were taken to be $m = 10^6 M_\odot$. The cloud potential was represented by the formula

$$\varphi(r) \sim \frac{m}{\sqrt{r^2 + \varepsilon^2}},$$

where r is the distance from the star to the cloud center and $\varepsilon = 50$ pc is the softening length. This approach makes it possible to allow for close encounters between group stars and GMCs.

Note that the adopted characteristic GMC mass may be slightly overestimated (see, e.g., Sanders *et al.* 1985). At the same time, as was shown by Wielen (1985), the effect of GMCs on the dynamics of star clusters is virtually independent of their masses and depends only on their total density. Following Wielen (1985), we took the characteristic GMC density to be $\sim 0.02 M_\odot$ pc^{-3}.

SIMULATION RESULTS

Table 2 lists our estimates of the mean breakup times $\bar{t}$ of the moving groups and their rms deviations at $\sigma = 1$ km s^{-1} for the two models of the regular Galactic gravitational field under consideration. The observed pattern of increase in the group lifetime with increasing critical distance r_c is approximately the same for all four groups.

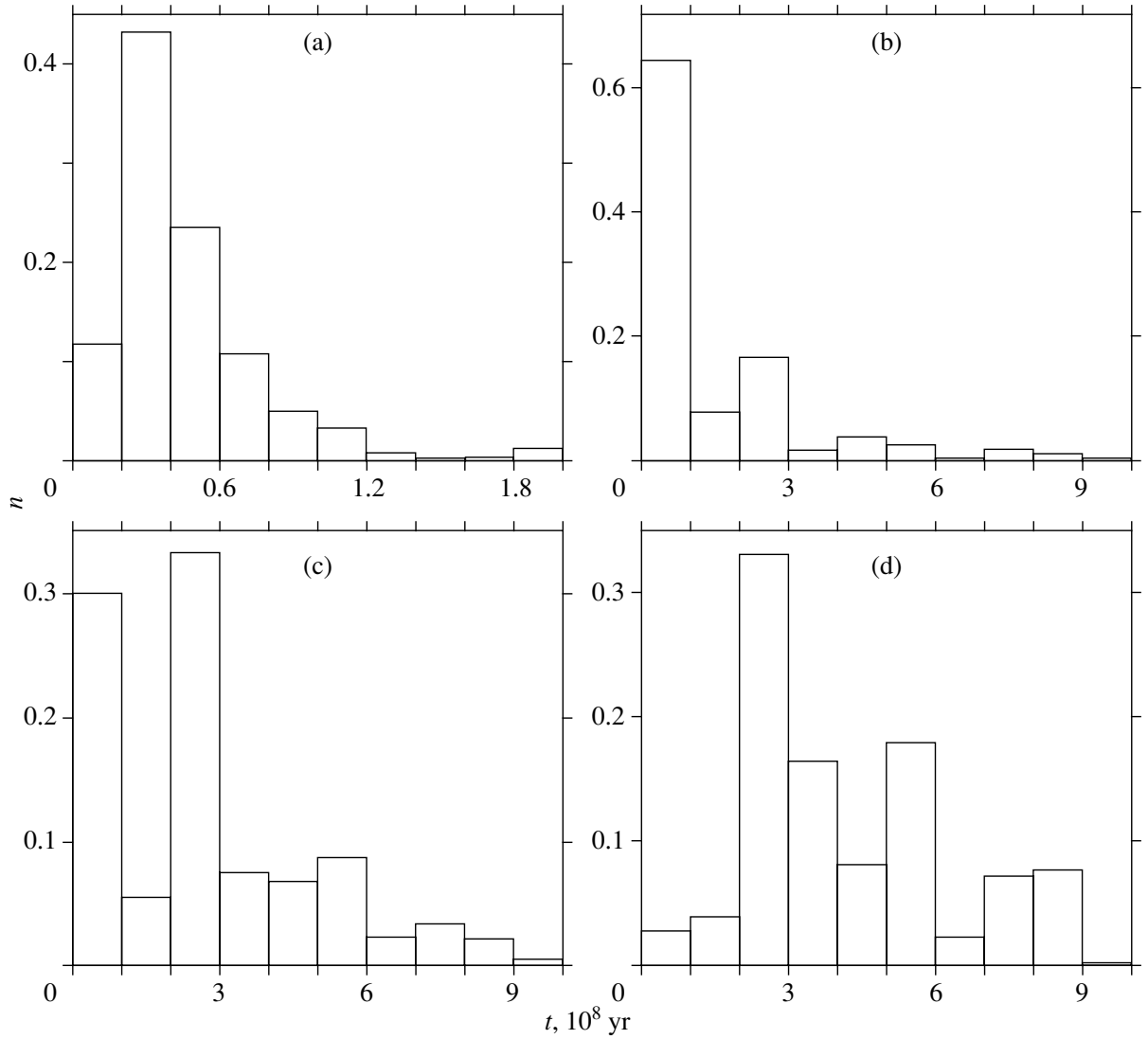
The figure shows the distributions $n(t)$ for the Hyades group at the chosen values of r_c. The distributions for other groups are similar. The distributions are asymmetric: the maxima are reached at comparatively short times. At the same time, secondary maxima are seen at long t, which may suggest that part of the group survives on a significant time interval, $\sim 10^8 - 10^9$ yr.

Table 2. Mean group breakup times and their rms deviations in units of 10^8 yr

r_c, pc	Group			
	Hyades	Pleiades	UMa	HR 1614
Model by Kutuzov and Osipkov (1989)				
50	0.98	1.05	1.16	0.97
	±11	±11	±24	±07
100	2.7	3.7	4.9	3.1
	2	4	1.8	2
200	7.0	10.4	8.1	8.8
	7	1.9	1.6	6
400	14.4	21.1	15.6	16.4
	1.3	3.4	3.0	1.1
Model by Flynn *et al.* (1996)				
50	1.27	1.42	1.24	1.22
	16	19	15	11
100	3.9	5.0	8.9	3.7
	4	5	4.2	2
200	12.2	19.3	12.5	9.5
	1.2	4.2	1.9	6
400	18.0	29.1	24.2	21.5
	1.4	3.4	3.8	1.7

The distribution $n(t)$ depends on the initial velocity dispersion σ in the group. Table 3 illustrates this dependence for the Hyades cluster at $r_c = 200$ pc. The mean breakup time $\bar{t}$ is seen to decrease with increasing σ. Thus, for example, if we increase the characteristic initial velocity spread in the group by a factor of 5, then the mean lifetime will decrease by a factor of about 8–10; this decrease is slightly faster than the linear law. The accelerated decrease is probably attributable to the additional effect of the tidal Galactic field.

As an example, Table 4 illustrates the effect of GMCs on the breakup of groups for the Hyades cluster in the model by Kutuzov and Osipkov (1989). The results are similar for the other groups. As we see from the table, the interaction with GMCs decreases significantly the lifetimes of moving groups.



Distribution of escape times of the Hyades moving group stars beyond the radii $r_c = 50$ pc (a), $r_c = 100$ pc (b), $r_c = 200$ pc (c), and $r_c = 400$ pc (d).

DISCUSSION

In this paper, we consider the evolution of a moving stellar group in the regular gravitational field of the Galaxy. The group stars are not bound to one another and are affected by GMCs moving in circular orbits in the Galactic plane around the Galactic center.

The group is gradually blurred within the main body of the Galaxy. The main cause of this blurring is the difference in the space velocities of the stream stars, which is produced, for example, by the dynamical processes that lead to the breakup of the open clusters—the progenitors of moving groups. The second cause is the external gravitational field of the Galaxy, which acts as an additional tidal force. The interaction of stream stars with GMCs plays a secondary role and only at the late evolutionary stages of the group when it becomes very wide (several hundred parsecs).

Let us now compare our estimates of the breakup times for the simulated moving groups with the ages of five observed young kinematic groups of the Galactic disk population (Montes *et al.* 1999): the Pleiades, IC 2391, the Castor group, the Ursa Major group, and the Hyades. The ages of these groups do not exceed 6×10^8 yr, with four groups being no older than 3×10^8 yr. Comparison of these estimates with the data from Table 4 shows that the mean escape time of group members beyond the sphere of radius $r_c = 400$ pc from the group center is less than 3×10^8 yr. Therefore, a considerable fraction of the stars of the streams under consideration may be located at greater distances from the group centers.

Table 3. Dependence of the mean breakup time of the Hyades group (in units of 10^8 yr) on the initial velocity dispersion σ of the group stars. $r_c = 200$ pc

σ, km s^{-1}	1	2	3	4	5
Model by Kutuzov and Osipkov (1989)					
Mean time $\bar{t}$	7.0 ±7	3.2 ±3	1.6 ±2	1.2 ±2	0.9 ±2
Model by Flynn *et al.* (1996)					
Mean time $\bar{t}$	12.2 ±1.2	3.9 ±4	2.3 ±3	1.5 ±2	1.2 ±2

Table 4. Comparison of the results obtained with and without allowance for the interaction between group stars and GMCs for the Hyades group at $\sigma = 1$ km s^{-1}

r_c, pc	Without GMCs	With GMCs
50	0.98 ±11	0.53 ±1
100	2.74 ±22	1.00 ±2
200	7.0 ±7	1.68 ±3
400	14.4 ±1.3	2.61 ±5

The physical sizes of young (no older than 0.5×10^8 yr) OB associations in the solar neighborhood are ~10–100 pc (Brown 2001). It can be assumed that the observed kinematic groups were formed from such weakly bound systems. Therefore, we can apply our estimates to these groups as well.

Numerical N-body simulations with allowance made for the sweeping-out of gas from the cluster and the external Galactic gravitational field show (Kroupa *et al.* 2001) that a stellar association with a radius of ~100–150 pc is formed around an open cluster in $\sim 0.5 \times 10^8$ yr. The observed kinematic groups may be parts of such associations surrounding some of the nearby open clusters, such as the Hyades, Praesepe, the Pleiades, and others. The results obtained can also be applied to such groups, although the continuous replenishment of the association with stars ejected from the cluster should be taken into account here.

Thus, the halos of open clusters will gradually break up due to the internal velocity dispersion, the external gravitational field of the Galaxy, and encounters with GMCs. The breakup time scales are several hundred million years for a critical group radius of several hundred parsecs.

ACKNOWLEDGMENTS

V.V. Orlov thanks the Foundation for Support of Leading Scientific Schools (project no. 1078.2003.02) and the Russian Foundation for Basic Research (project no. 04-02-17447).

REFERENCES

1. A. G. A. Brown, astro-ph/0101207 (2001).
2. E. Chereul, M. Creze, and O. Bienayme, Astron. Astrophys., Suppl. Ser. **135**, 5 (1999).
3. S. Feltzing and J. Holmberg, Astron. Astrophys. **357**, 153 (2000).
4. C. Flynn, J. Sommer-Larsen, and P. R. Christensen, Mon. Not. R. Astron. Soc. **281**, 1027 (1996).
5. J. Hänninen and C. Flynn, Astron. Soc. Pac. Conf. Ser. **228**, 143 (2001).
6. P. Kroupa, S. Aarseth, and J. Hurley, Mon. Not. R. Astron. Soc. **321**, 699 (2001).
7. S. A. Kutuzov and L. P. Osipkov, Astron. Zh. **66**, 965 (1989) [Sov. Astron. **33**, 498 (1989)].
8. D. Montes, A. Latorre, and M. J. Fernández-Figueroa, astro-ph/9906069 (1999).
9. A. A. Mülläri and V. V. Orlov, *Structure and Evolution of Stellar Systems*, Ed. by T. A. Agekian, A. A. Mülläri, and V. V. Orlov (St. Petersburg Univ. Press, St. Petersburg, 1997), p. 97.
10. V. V. Orlov, I. E. Panchenko, A. S. Rastorguev, and A. V. Yatsevich, Astron. Zh. **72**, 495 (1995) [Astron. Rep. **39**, 437 (1995)].
11. D. B. Sanders, N. Z. Scoville, and P. M. Solomon, Astrophys. J. **289**, 373 (1985).
12. R. Wielen, *Dynamics of Star Clusters*, Ed. by J. Goodman and P. Hut (Reidel, Dordrecht, 1985), p. 449.

Translated by A. Dambis

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 672–680. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 754–763.

Outburst of the X-ray Transient SAX J1818.6−1703 Detected by the INTEGRAL Observatory in September 2003

S. A. Grebenev[1*] and R. A. Sunyaev[1,2]

[1] *Space Research Institute, Russian Academy of Sciences, Profsoyuznaya ul. 84/32, Moscow, 117997 Russia*

[2] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, Postfach 1317, D-85741 Garching, Germany*

Received May 16, 2005

Abstract—During the observation of the Galactic-center field by the INTEGRAL observatory on September 9, 2003, the IBIS/ISGRI gamma-ray telescope detected a short (several-hours-long) intense (∼380 mCrab at the peak) outburst of hard radiation from the X-ray transient SAX J1818.6−1703. Previously, this source was observed only once in 1998 during a similar short outburst. We present the results of our localization, spectral and timing analyses of the object and briefly discuss the possible causes of the outburst. The release time of the bulk of the energy in such an outburst is appreciably shorter than the accretion (viscous) time that characterizes the flow of matter through a standard accretion disk.

Key words: *X-ray sources, transients, accretion.*

INTRODUCTION

The source SAX J1818.6−1703 was discovered by the BeppoSAX observatory on March 11, 1998, during an X-ray outburst that lasted only a few hours (in't Zand *et al.* 1998; in't Zand 2001). The appearance of a new transient near the well-known burster GX 13+1 (at an angular distance of ∼1°) was recorded by the WFC-2 wide-field X-ray camera at 19^h10^m (UT). By 20^h40^m, the photon flux from SAX J1818.6−1703 reached its maximum: ∼100 mCrab in the range 2–9 keV and ∼400 mCrab in the range 9–25 keV. The observation was interrupted 3 h later, but the flux was almost halved by this time, indicative of a fast decay of the transient. The source's position, R.A. $= 18^h18^m39^s$ and Decl. $= -17°03'.1$ (epoch 2000.0), was determined with an accuracy of 3′. Only one catalogued B3 III star, HD 168078, with $V = 10^m.7$ is within the error circle (in't Zand *et al.* 1998), but there is no additional reason to believe it to be a real candidate for optical identification with SAX J1818.6−1703.

All that has been known about SAX J1818.6−1703 until recently is listed above. Its nature and, primarily, the mechanism that produced such a short outburst with a duration much shorter than the characteristic time scale for the propagation of disturbances in a standard accretion disk ($t_{\rm vis} \gtrsim 1.4$ days) have remained unclear. In this paper, we present the results of our observations of the second outburst of hard radiation from this source that has allowed us to investigate it in more detail. This flare was detected on September 9, 2003, by the INTEGRAL observatory.

OBSERVATIONS

The INTEGRAL international gamma-ray observatory (Winkler *et al.* 2003) was placed in a high-apogee orbit by a PROTON launcher on October 17, 2002 (Eismont *et al.* 2003). There are four telescopes onboard the observatory that allow concurrent X-ray, gamma-ray, and optical observations of cosmic sources. This work is based on the data obtained with the IBIS gamma-ray telescope at energies above 18 keV. Unfortunately, no concurrent observations of SAX J1818.6−1703 were performed in the standard X-ray range (by the JEM-X telescope). Since the source is fairly far from the Galactic center toward which the observatory was pointed, it was not within the field of view of the JEM-X telescope, which is narrower than that of the IBIS telescope.

The IBIS telescope (Ubertini *et al.* 2003) uses the principle of a coded aperture to image the sky in hard X rays and gamma rays in a 30° × 30° field of view (the fully coded field is 9° × 9°) with an angular resolution of 12′ (FWHM). It is equipped with two position-sensitive detectors: ISGRI (Lebrun *et al.* 2003) composed of 128 × 128 CdTe semiconductor elements with a high sensitivity in the range 18–200 keV and PICsIT located under it (Labanti *et al.* 2003) and composed of 64 × 64 CsI(Tl) scintillators

*E-mail: sergei@hea.iki.rssi.ru

with an optimal sensitivity in the range 175 keV–10 MeV. In this paper, we use only the ISGRI data. The total area of the sensitive elements of this detector is 2620 cm^2; the effective area for sources at the center of the field of view is ~1100 cm^2 (half of the detector is shadowed by opaque aperture elements). The detector provides fairly good energy, $\Delta E/E \sim 7\%$ (FWHM), and high time, $\Delta t \simeq 61$ μs, resolutions.

Although the outburst of SAX J1818.6–1703 was initially detected using a standard software package for analyzing the INTEGRAL data (at that time, OSA-3.1), all of our results presented here were obtained using the data processing codes developed for the IBIS/ISGRI telescope at the Space Research Institute of the Russian Academy of Sciences. A general description of the procedures used can be found in a paper by Revnivtsev *et al.* (2004). Application of the latest version of these codes to the observations of the Crab Nebula has shown that the systematic measurement error of the absolute photon flux from the source does not exceed 7%, while the measurement error of the relative fluxes in various spectral channels does not exceed 3%. In our spectral analysis, we used the response matrix of the OSA-4.2 standard package (the rmf file of version 12 and the arf file of version 6), which proved to be good at fitting the spectra of the Crab Nebula, in particular, the spectra measured in August 2003 immediately before the observations under consideration. The spectrum of the Nebula was assumed to be $dN(E)/dE = 10E^{-2.1}$ phot. $cm^{-2} \times s^{-1}$ keV^{-1}, where the energy E is given in keV. For our study within the framework of the same general approach to analyzing the IBIS/ISGRI data, we developed codes for reconstructing the source's light curves with a good time resolution.

When the outburst of the transient SAX J1818.6–1703 occurred, the INTEGRAL observatory was performing a long (with a total exposure time of 2 Ms) series of observations of the Galactic-center region. During this series, the IBIS telescope detected 60 hard radiation sources of various intensities (Revnivtsev *et al.* 2004). Curiously enough, another poorly studied source, AX J1749.1–2733, flared up in this region on September 9, 2003, almost simultaneously with SAX J1818.6–1703. The results of its study are presented elsewhere (Grebenev and Sunayev 2005). The INTEGRAL observations in this period were performed by successive pointings at points of the Galactic-center field spaced ~2° apart according to the 5×5 scheme. Depending on the pointing, the exposure efficiency of SAX J1818.6–1703 changed greatly. The duration of each pointing was ~3450 s.

RESULTS

The appearance of the transient source in the field of view was first recorded at a statistically significant level by the IBIS/ISGRI telescope (at a signal-to-noise ratio of $S/N = 9.4$) during the pointing that began on September 9, 2003, at 00^h01^m (UT). Analysis of its X-ray image showed that SAX J1818.6–1703 flared up. The measured photon flux from it was 69 ± 7 mCrab in the range 18–45 keV and 43 ± 16 mCrab in the range 45–70 keV (the flux of 1 mCrab in these ranges corresponds to radiation fluxes of 1.1×10^{-11} and 4.7×10^{-12} erg cm^{-2} s^{-1}, respectively). The brightness of the source remained the same during the next pointing; subsequently, it faded for 2–3 h,[1] but flared up again. During the pointings that began at 10^h42^m and 12^h41^m, two superintense bursts were detected from the source, during which it became the brightest among all of the sources in the field of view. During the first (stronger) burst, the measured photon flux in the above ranges reached 230 ± 5 and 172 ± 10 mCrab, respectively. The source remained moderately bright (~50–70 mCrab) until 20^h, we failed to detect it in several subsequent pointings, and the observations were then interrupted, because the satellite entered the Earth's radiation belts in the final segment of its orbit. During the next orbital cycle (September 10–13), the source was detected only at the telescope's sensitivity limit ($S/N \simeq 6.0$) with a mean 18–45-keV flux of 6.6 ± 1.1 mCrab.

The described picture is illustrated by Fig. 1, which shows the source's light curves constructed from the observations of September 7–13 in two energy ranges. Each point of these curves is the measurement of the photon flux from the source in the corresponding sky image obtained during an individual pointing.

The vertical dashed lines in Fig. 1 indicate the interval of the source's statistically significant detection. We used this interval to accumulate the integrated images of the sky near SAX J1818.6–1703 (signal-to-noise maps) in the energy ranges 18–45 and 45–70 keV shown in Figs. 2 and 3, respectively. The total exposure time was 75 400 s. Apart from SAX J1818.6–1703, four more X-ray sources are seen in Fig. 2: the bursters GX 17+2 and GX 13+1, the atoll source GX 9+1, and the X-ray pulsar IGR J18027–2016 (also known as SAX J1802.7–2017). SAX J1818.6–1703 is the brightest of these sources—it was detected at a

[1] Since the pointings at this time were particularly unfavorable for the source's observation (it was at the very edge of the field of view), the flux was measured with large errors; nevertheless, the fall in flux appears statistically significant (see Fig. 1).

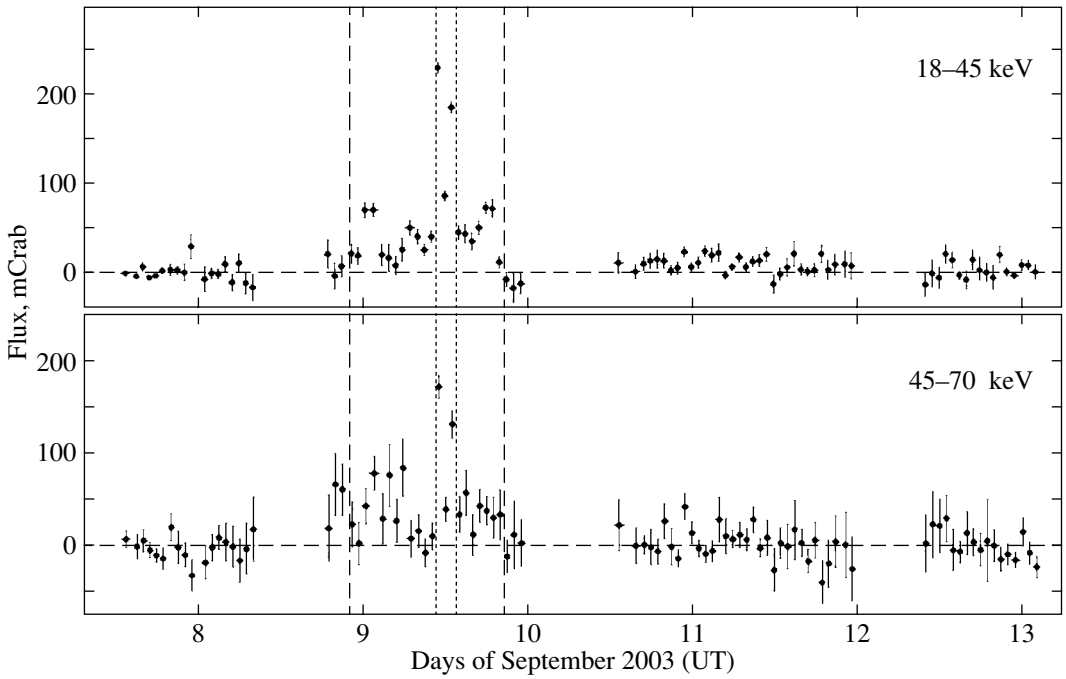


Fig. 1. IBIS/ISGRI light curve for SAX J1818.6–1703 in the energy ranges 18–45 and 45–70 keV obtained in the period September 7–13, 2003. Each point of this curve corresponds to an individual ~3450-s-long pointing of the INTEGRAL observatory. The dashed and dotted vertical lines indicate the source's activity period and the main outburst event (two intense bursts of hard radiation).

signal-to-noise ratio of S/N ≃ 48, while the next brightest source GX 9+1 has S/N ≃ 20. In Fig. 3, SAX J1818.6–1703 is the only source (it is seen at S/N ≃ 15). Estimates indicate that this is attributable not just to the natural decrease in the hard X-ray photon flux typical of all sources,—the flux from SAX J1818.6–1703 falls in this case more slowly; i.e., the source has a much harder (with the possible exception of the X-ray pulsar IGR J18027–2016) spectrum.

The image shown in Fig. 2 was used to improve the localization of the source. The position found, R.A. = $18^h18^m38\overset{s}{.}2$ and Decl. = −17°03′11″ (epoch 2000.0, 1′.5 uncertainty), coincided with its position measured by BeppoSAX in 1998 to within 12″.

The Outburst Time Profile

To elucidate the nature of the source's outburst, it is crucially important to analyze the structure of the two intense short bursts occurred at about 11 and 13 h. Since the light curve in Fig. 1 does not allow this to be done, we reconstructed more detailed light curves. Figure 4a shows the 18–45-keV light curve with a time resolution (bin size) of 500 s. It spans only the source's activity period. The actually measured count rate, i.e., the count rate corrected for the dead time and other instrumental effects, but uncorrected for the variations in the effective (source-irradiated) area of the detector due to the change in the INTEGRAL pointing, is along the Y axis. This effect is important, since the source was outside the fully coded field of view of the telescope. We do not make the corresponding correction in order not to overload the figure. At the time resolution used, statistically insignificant spikes appear in the corrected light curve due to the Poissonian fluctuations of the count rate when the source approaches the edge of the field of view. Instead, Fig. 4b shows the curve of variations in the effective area of the detector. We clearly see its correlation with the count rate, which, however, does not distort severely the main event. The effective area for the observation of SAX J1818.6–1703 was only 620 cm^2 even at maximum; i.e., it was almost a factor of 2 smaller than the area typical of the sources in the fully coded field of view. For such an effective area, 1 count/s corresponded to a flux of ~16.4 mCrab in the energy range under consideration, so, according to this figure, the maximum flux from the source reached ~380 mCrab.

The figure suggests that the two bursts of the main outburst event have a fairly complex time profile. A narrow (10–20 min) precursor peak and a broader (1.5–2 h) main peak can be distinguished in each of them. The amplitude of the precursor peak in the first burst is almost twice that of the main peak,

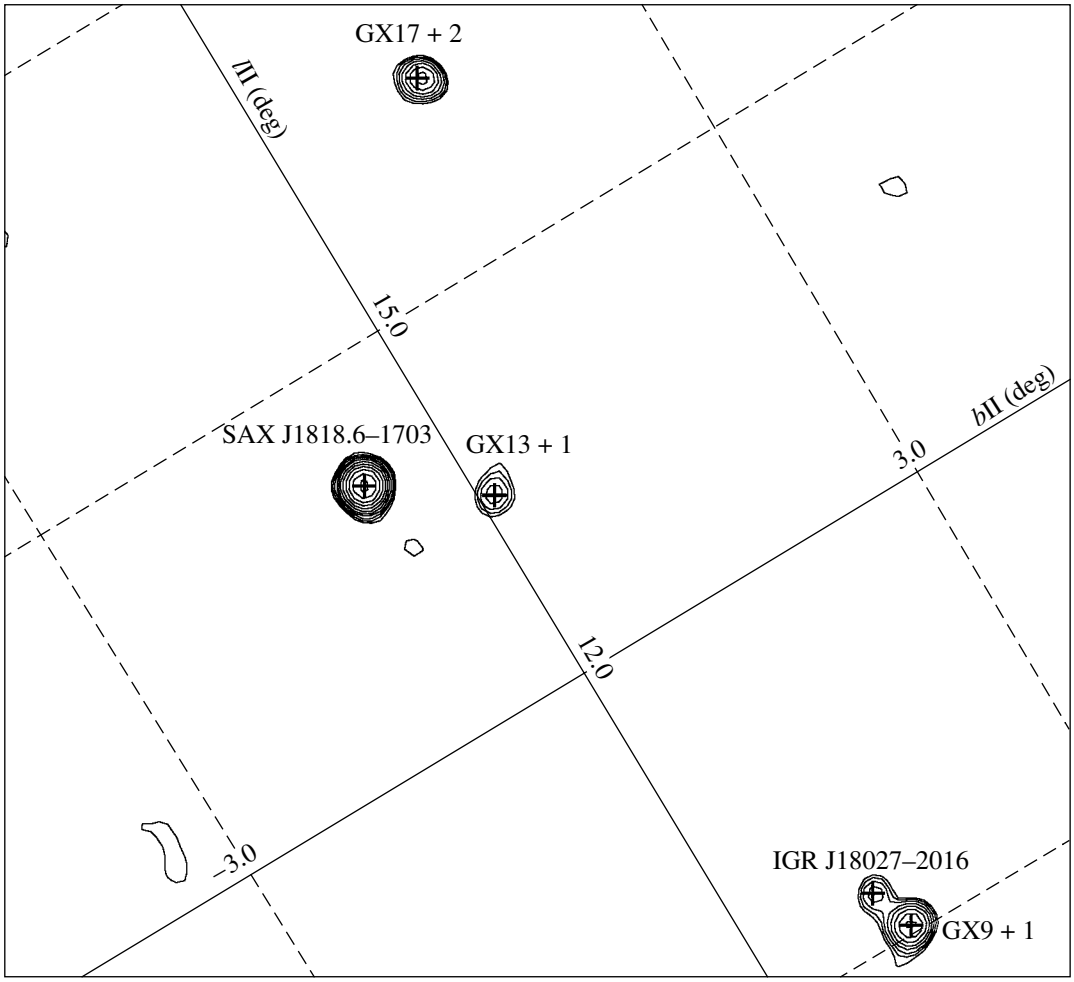


Fig. 2. IBIS/ISGRI X-ray image of the region around SAX J1818.6–1703 obtained on September 9, 2003, during its outburst. The contours indicate the regions of confident detection of sources in the energy range 18–45 keV and are given at signal-to-noise ratios of 3, 3.9, 5.2, 6.8, 8.9, 11.6, ...(on a logarithmic scale). The image size is approximately equal to $9^\circ \times 9^\circ$.

while the precursor peak in the second burst is appreciably smaller. In general, the two bursts resemble ordinary type-I X-ray bursts with photospheric expansion, i.e., bursts produced by a thermonuclear explosion on the neutron star surface at which the photospheric luminosity reached the Eddington limit (see Lewin *et al.* 1993). However, the duration of the bursts from SAX J1818.6–1703 was 2–3 h, which is much longer than the duration of ordinary X-ray bursts. The recently discovered superbursts (Kuulkers *et al.* 2002; in't Zand *et al.* 2004) have comparable durations, but exhibit completely different time profiles—a very fast rise and a long exponential decay. As we will see below, there are also more fundamental differences between these bursts.

Figure 5 shows the time profiles of the main event in several energy ranges after their reduction to the effective area of 620 cm^2. In general, they are similar in structure, suggesting that the spectral shape of the source changed little during the outburst. However, note a clear decrease with energy in the relative amplitude of the precursor peak in the first burst and a probable decrease in the amplitude of the precursor peak in the second burst. These changes were confirmed during a detailed spectral analysis.

The Radiation Spectrum

Figure 6 shows the average spectrum of SAX J1818.6–1703 obtained in the period of its activity (during the interval bounded by the vertical dashed lines in Figs. 1 and 4). The source's radiation is recorded up to ~200 keV, with an exponential cutoff being observed at energies above 70 keV. Note that an additional soft ($h\nu < 30$ keV) radiation component is present in the spectrum. The energy characteristics of the source's radiation in this period, its mean luminosity and energy release calculated by assuming that SAX J1818.6–1703 is actually near the Galactic

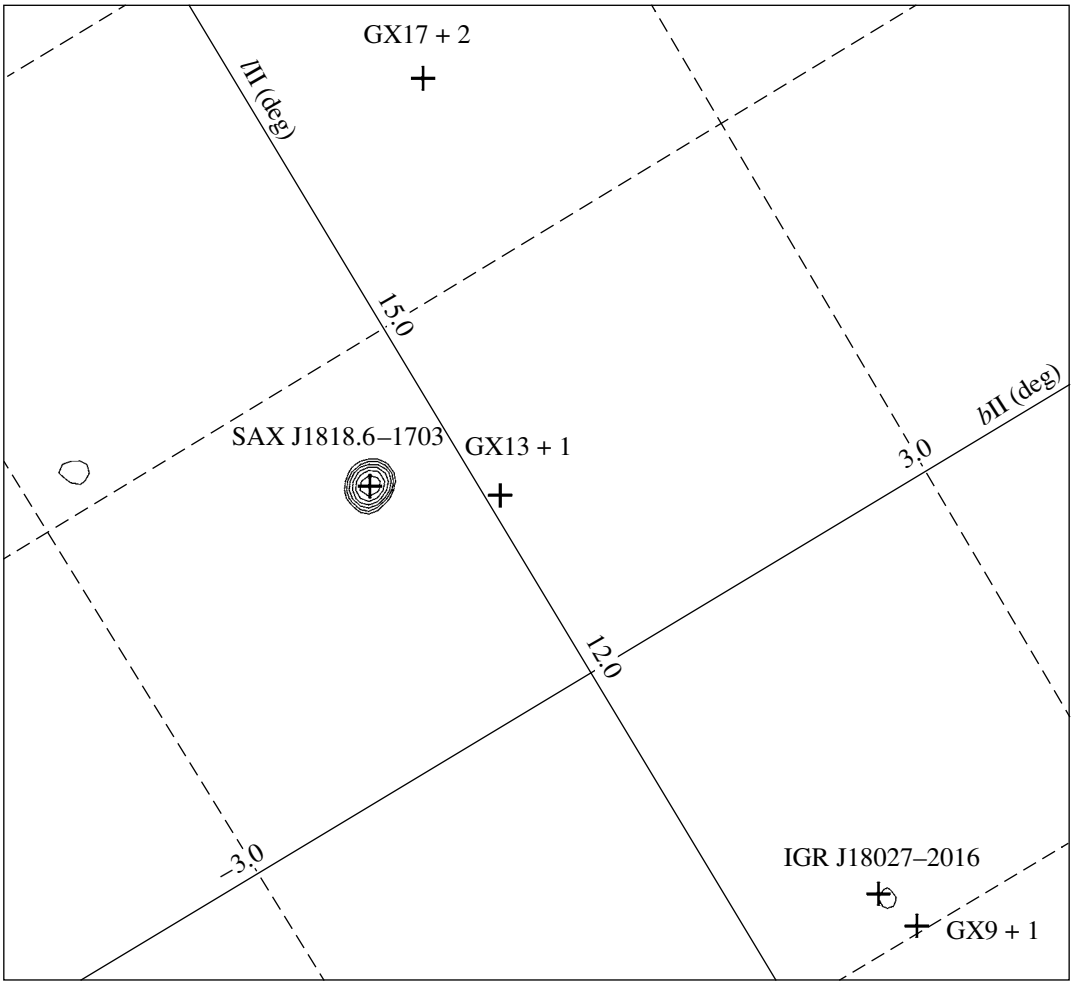


Fig. 3. The same image as that in Fig. 2, but in the energy range 45–70 keV. Of the five sources observed in this field at low energies, only SAX J1818.6–1703 is still bright (seen at S/N $\simeq 14.8\sigma$).

center, at a distance of $d \simeq 8$ kpc,[2] are given in Table 1.

Table 2 summarizes the results of fitting the spectrum by simple analytical models: a power law (PL), a power law with an exponential high-energy cutoff (PE), the radiation formed through Comptonization of low-energy photons in a cloud of high-temperature plasma (ST, Sunyaev and Titarchuk 1980), the bremsstrahlung of an optically thin thermal plasma (TB), the bremsstrahlung with an additional soft blackbody component (TB+BB), and the Comptonization radiation with an additional blackbody component (ST+BB). We see that even the single-component PL, PE, and ST models describe satisfactorily the radiation spectrum. Although the introduction of an additional soft radiation component affects only the first two or three points of the spectrum, this leads to its further significant improvement. Extrapolating the soft component to the X-ray energy range leads to a luminosity that is an order of magnitude higher than that in the hard energy range. Thus, the overall energetics of the source could be maintained at a level approaching the critical Eddington level for accretion onto a neutron star (or, given the uncertainty in the spectral shape of the soft component, even onto a black hole of a low $\sim 3\ M_\odot$ mass).

Figure 7 shows the spectral evolution of SAX J1818.6–1703 during the outburst under discussion. Spectra A, B, and C were measured by the IBIS/ISGRI telescope at different evolutionary phases of the main event (the time intervals corresponding to these phases are indicated in the upper panel of Fig. 5). Spectrum D was measured during the remaining activity period of the source. The solid lines indicate the best fit to spectrum A by the bremsstrahlung law of an optically thin plasma ($kT \simeq 29$ keV). The soft component, which is absent in the

[2] For the source GX 13+1 closest to SAX J1818.6–1703, $d \simeq 7 \pm 1$ kpc (Bandyopadhyay *et al.* 1999).

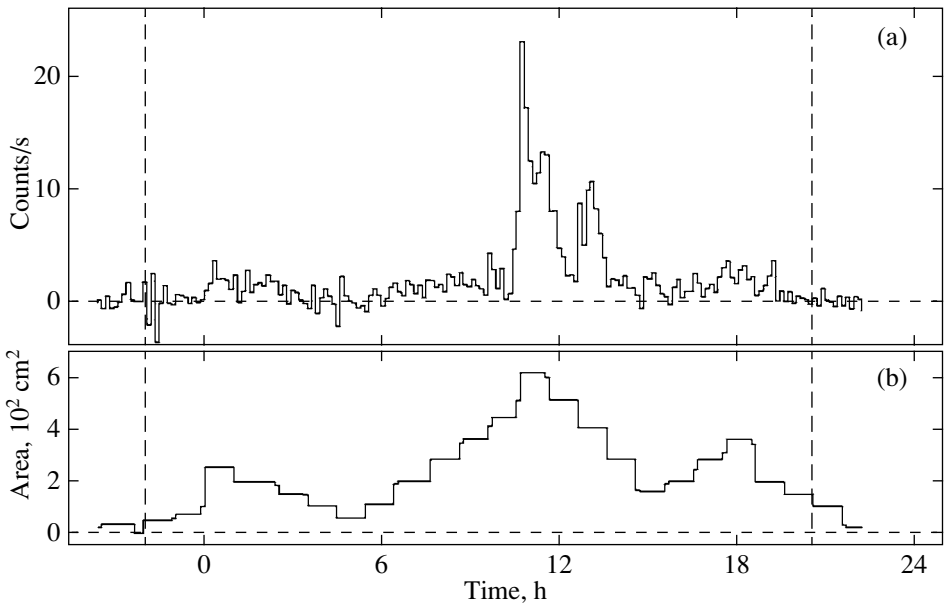


Fig. 4. (a) Detailed IBIS/ISGRI 18–45-keV light curve for SAX J1818.6–1703 obtained on September 9, 2003, in the period of its activity. The time in hours from the beginning of the day (UT) is along the X axis. The time resolution is 500 s. The count rate along the Y axis was corrected for the dead time of the detector and other instrumental effects, except the variations in the effective area for this source related to the change in the INTEGRAL pointing (the effect of partial coding). The corresponding change in the effective area is shown in panel b.

main radiation of the bursts (spectra B and C), is clearly seen at energies below ~30 keV in the spectra of the precursor peak in the first burst (spectrum A) and the period of moderate activity of the source (spectrum D), as well as in the average spectrum of the source. The results of fitting the presented spectra by a power law (PL) and the bremsstrahlung law of an optically thin thermal plasma (TB) are given in Table 3. We see that in the period of moderate activity, the source had a fairly soft radiation spectrum (with a photon index of $\alpha \sim 3$); at the onset of the main event, the spectral hardness increased ($\alpha \sim 2.7$) and continued to increase, reaching $\alpha \sim 2.4$ in the second burst.

DISCUSSION

Transients like SAX J1818.6–1703 form a special, fairly representative population among the X-ray sources discovered or recorded during outbursts by the INTEGRAL observatory (e.g., IGR J17544–2619, XTE J1739–302, and others). Their distinctive feature is a very short (several hours) lifetime and a long recurrence period (several years). Only a few such transients (including SAX J1818.6–1703) were observed in previous experiments.

The activity of these sources could in principle be caused by the following: (1) thermonuclear explosions on the neutron star surface, (2) magnetic energy release in the case of a neutron star with a very strong magnetic field, and (3) unsteady accretion onto a neutron star or a black hole in a binary system. The first two possibilities seem unlikely, since the lifetime of such transients is much longer than the duration of both soft gamma-ray bursts from gamma repeaters (magnetars) and ordinary X-ray bursts from neutron stars with a weak magnetic field (bursters). Our results show that the outbursts of SAX J1818.6–1703 also differ greatly from the superbursts discovered recently from bursters (Kuulkers *et al.* 2002; in't Zand *et al.* 2004) primarily by an increase in the hardness during the burst and by the burst profile. Note that the energy released during the main outburst event of SAX J1818.6–1703 accounted for only 1/3 of the

Table 1. Parameters of the outburst of SAX J1818.6–1703 observed on September 9, 2003

Interval	Parameter[a]	Value
Entire activity period [b]	ΔT	22.5 h
	L_X	7.5×10^{36} erg s^{-1}
	F_X	6.1×10^{41} erg
Main event[c]	ΔT	2.7 h
	L_X	1.9×10^{37} erg s^{-1}
	F_X	1.8×10^{41} erg

[a] Duration ΔT, luminosity L_X, and energy release F_X in the range 20–200 keV for an assumed distance of 8 kpc.
[b] Bounded by the vertical dashed lines in Fig. 1.
[c] The sum of intervals A, B, C marked in Fig. 5.

Table 2. Results of the best-fit approximation of the spectrum of SAX J1818.6−1703 averaged over the entire period of its activity

Model[a]	kT, keV	α[b]	kT_{bb},[c] keV	L_{bb},[c] 10^{38} erg s^{-1}	$\chi^2(N)$[d]
PL		2.75 ± 0.08			1.02(22)
PE	136 ± 4	2.46 ± 0.31			1.02(21)
ST	27.9 ± 2.1	2.51 ± 0.10			1.02(21)
TB	28.8 ± 1.9				1.37(22)
TB+BB	36.3 ± 3.5		1.7 ± 0.7	1.6 ± 0.4	0.75(20)
ST+BB	16.8 ± 3.7	2.08 ± 0.57	1.7[e]	1.4 ± 0.5	0.78(20)

[a] The notation of the models is given in the text.
[b] The photon index.
[c] Parameters of the soft radiation component: the blackbody temperature and bolometric luminosity (for $d = 8$ kpc).
[d] The χ^2 value of the best fit normalized to N (N is the number of degrees of freedom).
[e] A fixed parameter.

Table 3. Results of the best-fit approximation of the spectrum of SAX J1818.6−1703 at various evolutionary phases of the outburst

Spectrum	Model	kT, keV	α[a]	L_X,[b] 10^{37} erg s^{-1}	$\chi^2(N)$[c]
A	PL		2.74 ± 0.10	2.37 ± 0.77	1.03 (23)
	TB	28.6 ± 2.4		2.05 ± 0.18	0.83 (23)
B	PL		2.55 ± 0.10	1.98 ± 0.66	1.25 (23)
	TB	34.3 ± 3.4		1.70 ± 0.15	0.96 (23)
C	PL		2.41 ± 0.11	2.21 ± 0.79	1.76 (23)
	TB	39.0 ± 4.4		1.88 ± 0.17	1.04 (23)
D	PL		3.05 ± 0.18	0.42 ± 0.21	0.83 (23)
	TB	20.0 ± 2.4		0.36 ± 0.06	1.23 (23)

[a] The photon index.
[b] The 20–200-keV luminosity for an assumed distance of $d = 8$ kpc.
[c] The χ^2 value of the best fit normalized to N (N is the number of degrees of freedom).

total energy released in the period of its activity (Table 1). This is also difficult to explain in terms of a thermonuclear explosion on the neutron star surface without invoking unsteady accretion processes.

On the other hand, the lifetime of such outbursts is much shorter than the accretion (viscous) time that characterizes the propagation of disturbances in a standard accretion disk,

$$t_{vis} \sim \frac{2}{3\alpha}\frac{1}{\Omega_K(R)}\left(\frac{R}{H}\right)^2$$
$$\sim 1.4\left(\frac{R}{10^{10}\ \mathrm{cm}}\right)^{3/2}\left(\frac{M}{1.4M_\odot}\right)^{-1/2}\ \mathrm{days}$$

(Shakura and Sunyaev 1973). Here, R is the outer radius of the disk, H is the disk half-thickness

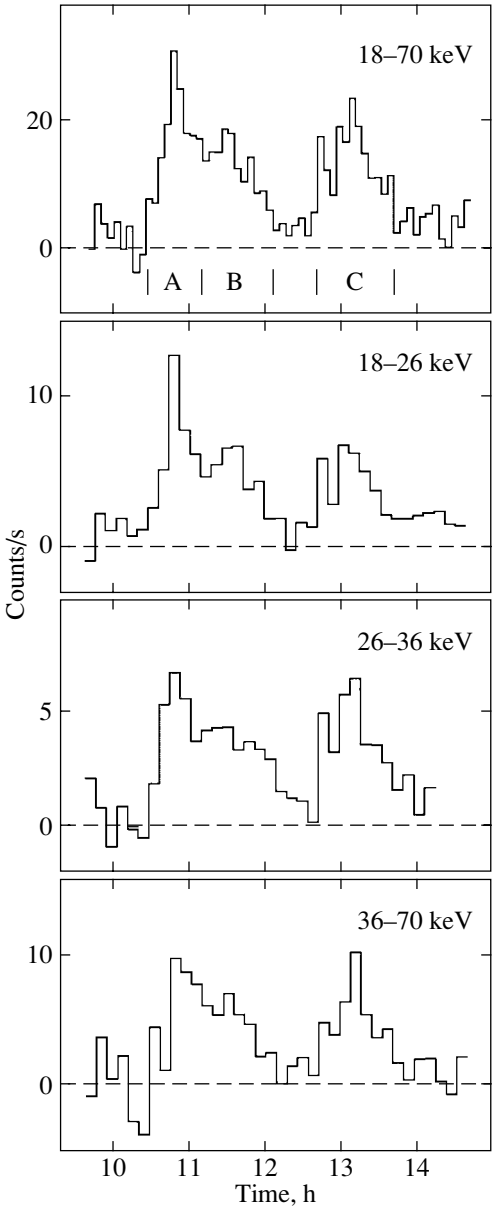


Fig. 5. Change of the time profile for the main outburst event of SAX J1818.6−1703 recorded by the IBIS/ISGRI telescope on September 9, 2003, with energy. The resolution is 500 s everywhere, except the profile measured in a wide energy range, 18−70 keV (the bin size for it is 300 s). The count rate along the Y axis was corrected for all instrumental effects and reduced to the same area of 620 cm^2, which corresponds to the maximum achieved efficiency of the source's observations (1 count/s is approximately equal to 36, 48, 31, and 12 mCrab in the ranges 18−26, 26−36, 36−70, and 18−70 keV, respectively). The time in hours from the beginning of the day (UT) is along the X axis.

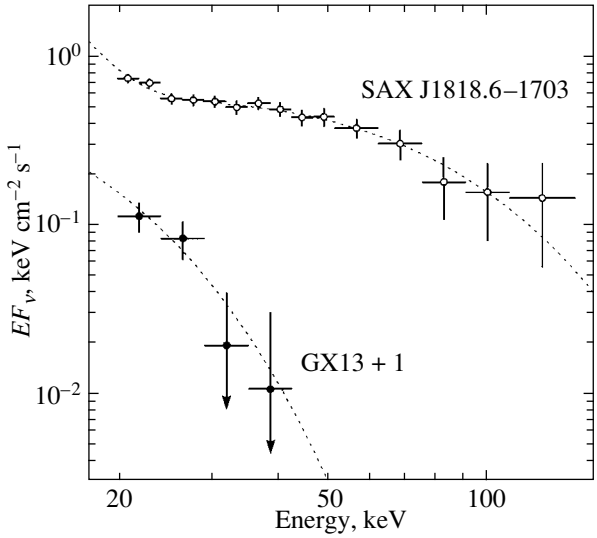


Fig. 6. Average IBIS/ISGRI spectrum of SAX J1818.6−1703 obtained on September 9, 2003, in the period of its activity (open circles). The spectrum is very hard; the characteristic temperature when fitting the spectrum by the bremsstrahlung law of an optically thin plasma is $kT \simeq 36$ keV. For comparison, the filled circles indicate the X-ray spectrum of the source GX13+1 closest to SAX J1818.6−1703 measured at the same time, which is typical of accreting neutron stars with a weak magnetic field ($kT \simeq 5$ keV). At low energies, an additional soft radiation component is apparently present in the spectrum of SAX J1818.6−1703. The dotted lines indicate the best fits to the spectra (see the text).

at this radius, $\Omega_K = (GM/R^3)^{1/2}$ is the Keplerian frequency, M is the mass of the compact object, and $\alpha \sim 1$ is the viscosity parameter. We disregarded the weak R dependence of H/R on the right-hand side of this expression (in the standard disk accretion model, $H/R \sim R^{1/8}$) and set $H/R \simeq 0.02$, which corresponds to the most compact binaries with $R \sim 10^{10}$ cm (for a discussion, see Gilfanov and Arefiev 2005). We can decrease t_{vis} by increasing H/R (advection-dominated regime), but this will cause the accretion efficiency to decrease appreciably. At an accretion efficiency of $\xi = 1/12$ typical of black holes or compact neutron stars ($R_{ns} \sim 3R_g$, where R_{ns} is the radius of the neutron star surface, and R_g is its gravitational radius), a compact object would accrete $M_{acc} \sim F_X/(\xi c^2) \simeq 8 \times 10^{21}$ g of matter for the energy release to be $F_X \simeq 6 \times 10^{41}$ ergs (Table 1). The other possibility to decrease t_{vis} appears in the case of unsteady accretion from the stellar wind. In this case the radius of the forming disk can be much smaller than the size of the binary.

The mechanisms responsible for the outbursts of such transients can be considered in detail only after their reliable optical identification. At present, this has not yet been done.

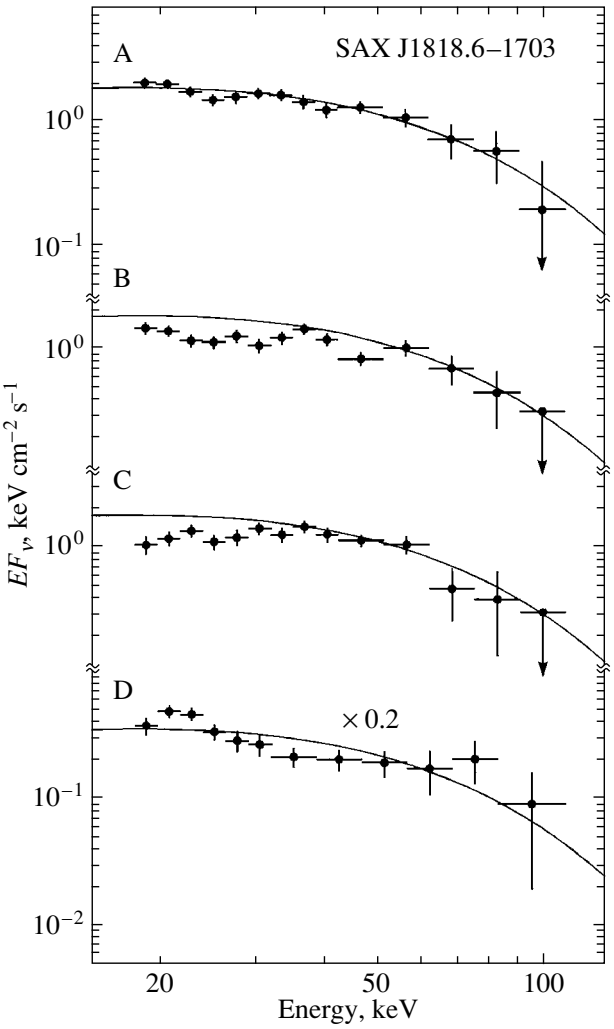


Fig. 7. Spectral evolution of SAX J1818.6–1703 during the hard X-ray outburst observed on September 9, 2003. Spectra A, B, and C were measured by the IBIS/ISGRI telescope at different evolutionary phases of the main event (see Fig. 5); spectrum D was measured during the remaining activity period of the source. The solid lines indicate the best fit to spectrum A by the bremsstrahlung law of an optically thin plasma ($kT \simeq 29$ keV; for spectrum D, the normalization of the fit was decreased by a factor of 5). The soft component, which is absent in the main burst radiation (spectrum B and particularly spectrum C), is clearly seen at energies below ~30 keV in the radiation spectra of the precursor peak of the first burst (spectrum A) and the period of moderate activity of the source (spectrum D).

ACKNOWLEDGMENTS

This work is based on the observational data obtained by the INTEGRAL observatory, an ESA satellite with instruments provided by ESA member states (especially France, Italy, Germany, Switzerland, Denmark, and Spain), the Czech Republic, and Poland, placed in orbit by Russia and operated by ESA with the participation of the United States, and provided via the Russian and European INTEGRAL Science Data Centers. We used some of the codes developed by E.M. Churazov to analyze the data. This study was supported by the Russian Foundation for Basic Research (project no. 05-02-17454), the Presidium of the Russian Academy of Sciences (the "Nonstationary Phenomena in Astronomy Program"), and the "Program of the Russian President for Support of Leading Scientific Schools" (project no. NSh.-2083.2003.2).

REFERENCES

1. R. M. Bandyopadhyay, T. Shahbaz, P. A. Charles, and T. Naylor, Mon. Not. R. Astron. Soc. **306**, 417 (1999).
2. N. A. Eismont, A. V. Ditrikh, G. Janin, *et al.*, Astron. Astrophys. **411**, L37 (2003).
3. S. A. Grebenev and R. A. Sunyaev, Pis'ma Astron. Zh. **31** (2005) (in press).
4. M. Gilfanov and V. Arefiev, Mon. Not. R. Astron. Soc. (2005) (in press).
5. J. in't Zand, in *Proceedings of the 4th INTEGRAL Workshop "Exploring the Gamma-Ray Universe"*, Ed. by B. Battrick, A. Gimenez, V. Reglero, and C. Winkler (Noordwijk, 2001), ESA SP-459, p. 463.
6. J. in't Zand, J. Heise, M. Smith, *et al.*, IAU Circ., No. 6840 (1998).
7. J. J. M. in't Zand, R. Cornelisse, and A. Cumming, Astron. Astrophys. **426**, 257 (2004).
8. E. Kuulkers, J. J. M. in't Zand, M. H. van Kerkwijk, *et al.*, Astron. Astrophys. **382**, 503 (2002).
9. C. Labanti, G. Di Cocco, G. Ferro, *et al.*, Astron. Astrophys. **411**, L149 (2003).
10. F. Lebrun, J. P. Leray, P. Lavocat, *et al.*, Astron. Astrophys. **411**, L141 (2003).
11. W. H. G. Lewin, J. van Paradijs, and R. E. Taam, Space Sci. Rev. **62**, 223 (1993).
12. M. G. Revnivtsev, R. A. Sunyaev, D. A. Varshalovich, *et al.*, Pis'ma Astron. Zh. **30**, 430 (2004) [Astron. Lett. **30**, 382 (2004)].
13. N. I. Shakura and R. A. Sunyaev, Astron. Astrophys. **24**, 337 (1973).
14. R. A. Sunyaev and L. G. Titarchuk, Astron. Astrophys. **86**, 121 (1980).
15. P. Ubertini, F. Lebrun, G. Di Cocco, *et al.*, Astron. Astrophys. **411**, L131 (2003).
16. C. Winkler, T. J.-L. Courvoisier, G. Di Cocco, *et al.*, Astron. Astrophys. **411**, L1 (2003).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 681–694. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 764–779.

Observations of the X-ray Burster MX 0836–42 by the INTEGRAL and RXTE Orbiting Observatories

I. V. Chelovekov[1*], A. A. Lutovinov[1], S. A. Grebenev[1], and R. A. Sunyaev[1,2]

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

[2] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, Postfach 1317, D-85741 Garching, Germany*

Received May 16, 2005

Abstract—We present the results of our study of the emission from the transient burster MX 0836–42 using its observations by the INTEGRAL and RXTE X-ray and gamma-ray observatories in the period 2003–2004. The source's broadband X-ray spectrum in the energy range 3–120 keV has been obtained and investigated for the first time. We have detected 39 X-ray bursts from this source. Their analysis shows that the maximum 3–20-keV flux varies significantly from burst to burst, $F \sim (0.5–1.5) \times 10^{-8}$ erg cm^{-2} s^{-1}. Using the flux at the maximum of the brightest detected burst, we determined an upper limit for the distance to the source, $D \simeq 8$ kpc. © *2005 Pleiades Publishing, Inc.*

Key words: *neutron stars—bursters, transients; X-ray sources—MX 0836–42.*

INTRODUCTION

The transient X-ray source MX 0836–42 was discovered in 1971 by the OSO-7 satellite (Markert *et al.* 1975). Its position almost coincided with the probable position of the point source detected in December 1970 and February 1971 by the UHURU observatory (Kellogg *et al.* 1971). Since the intensity of the latter was too low, it was not included in the official UHURU catalog of sources (Markert *et al.* 1977; Cominsky *et al.* 1978). In 1990, the WATCH all-sky X-ray monitor aboard the GRANAT orbiting observatory detected a bright transient source near MX 0836–42 (Sunyaev *et al.* 1990, 1991; Lapshov *et al.* 1992). Its 5–15-keV flux during these observations reached a level comparable to the flux from the Crab Nebula. The localization accuracy of the source was only 1°. Analysis of the ROSAT data revealed two point sources in this region spaced 24 arcmin apart (Hasinger *et al.* 1990). Subsequently, the presence of these two sources was confirmed by data from the ART-P telescope of the GRANAT observatory (Sunyaev 1991). Type-I X-ray bursts were detected from the northern source, which allows MX 0836–42 to be classified as a low-mass X-ray binary containing a neutron star with a weak magnetic field, while X-ray pulsations with a period of ~12 s were detected from the southern source GRS 0834–430, which characterize it as an X-ray pulsar (Makino 1990; Grebenev and Sunyaev 1991). Studies of the emission from MX 0836–42 showed that its energy spectrum could be described by a power law with a photon index of ~1.5 (Aoki *et al.* 1992). In this paper, based on the data obtained in 2003–2004 by the instruments of the INTEGRAL and RXTE orbiting observatories, we have constructed and analyzed the source's persistent broadband spectrum in the energy range 3–120 keV and its spectrum during X-ray bursts for the first time. We discuss the properties of the X-ray bursts detected from the source.

OBSERVATIONS AND DATA ANALYSIS

The INTEGRAL international orbiting gamma-ray observatory (Winkler *et al.* 2003) was placed in orbit by a Russian PROTON launcher on October 17, 2002 (Eismont *et al.* 2003). There are four instruments aboard the observatory: the SPI gamma-ray spectrometer, the IBIS gamma-ray telescope, the JEM-X X-ray monitor, and the OMC optical monitor. Here, we use the data obtained by the ISGRI detector, one (upper) of the two detectors of the IBIS gamma-ray telescope (Ubertini *et al.* 2003), and by the second module of the JEM-X X-ray monitor (Lund *et al.* 2003). The ISGRI/IBIS detector is sensitive to photons in the energy range 15–200 keV and has an energy resolution of ~7% at 100 keV. The IBIS telescope includes a coded mask that allows it to be used not only for spectral and timing analyses of the emission, but also for reconstructing the image of the sky in the 29° × 29° field of view of the instrument (the fully coded field of view is 9° × 9°) with an angular

*E-mail: chelovekov@hea.iki.rssi.ru

resolution of 12 arcmin (FWHM) and localizing X-ray and gamma-ray sources to within 1–2 arcmin. The JEM-X telescope is also based on the principle of a coded aperture. It is sensitive to photons in the energy range 3–35 keV, and its fully coded field of view is 4°.8 in diameter.

We analyzed the JEM-X and IBIS data using the OSA 4.1 data processing software package distributed by the INTEGRAL Science Data Center (ISDC). To construct the light curves and spectra for MX 0836–42, we used the fluxes that were obtained by reconstructing the image of the sky in the field of view of the instrument and identifying the observed sources. The photon spectrum of the source was reconstructed using a 50-channel response matrix of the instrument that was constructed from observations of the source in the Crab Nebula and that allows us to restore the spectral shape of the source to within 4% and the normalization to within 7%.

The RXTE observatory carries three main instruments: the PCA spectrometer based on five xenon proportional counters sensitive to photons in the energy range 2–60 keV, the HEXTE spectrometer sensitive to photons up to 200 keV, and the ASM all-sky monitor sensitive to photons in the energy range 2–12 keV. The RXTE observational data for the source under study were provided by the NASA archive (HEASARC). We used the LHEASOFT 5.3.1 software package and the XSPEC 11.3.1 code to process the PCA/RXTE and HEXTE/RXTE data and to analyze the source's spectra.

The X-ray transient MX 0836–42 was within the IBIS/ISGRI field of view several times in the period from March 2003 through May 2004 (Table 1), within the framework of both the Core and Open observing programs (Winkler *et al.* 2003). The total exposure time for this source was more than 2.5 Ms. We used the data obtained when scanning the Galactic plane and during deep observations of a region near the source Vela X-1.

The most recent observations of the above-mentioned Vela X-1 region were performed from June 12 through July 6 and from November 27 through December 11, 2003. The total JEM-X and IBIS exposure times for MX 0836–42 were ~0.73 and ~1.13 Ms for the former period and ~0.45 and ~1.0 Ms for the latter period, respectively. We used only the pointings during which the source under study was within the fully coded field of view of the instrument. This was done to avoid the inaccuracies in restoring the energy flux from the source under study as much as possible. The difference between the exposure times for the two instruments stems from the fact that the fully coded field of view of the JEM-X monitor is smaller than that of the IBIS telescope (see above).

Figure 1 shows the light curves of the source for the period 52650–53400 MJD constructed from the data of the ISGRI/IBIS (20–60 keV) and JEM-X (3–20 keV) telescopes aboard the INTEGRAL observatory and the ASM (2–12 keV) and PCA (3–20 keV) instruments aboard the RXTE observatory. Figure 2 presents the source's light curve constructed from all of the available ASM/RXTE data. The fluxes shown are the ratios of the fluxes from MX 0836–42 to the flux from the Crab Nebula in the corresponding energy range.

During two of the three groups of measurements consisting of nine (52773–52825 MJD) and six (53186–53261 MJD) observations, respectively, present in Fig. 1a, the flux from the source under study was below the sensitivity threshold of the ISGRI detector. The upper limits on the flux from MX 0836–42 for each of these pointings are given at the 3σ level.

Figures 1b and 2 show the light curve for MX 0836–42 in the energy range 2–12 keV constructed from the ASM/RXTE data. Each point in the figure corresponds to the flux from the source averaged over a 36-ks period.

Table 2 presents information about the observations by the RXTE orbiting observatory in 2003–2004 during which MX 0836–42 was within the PCA and HEXTE fields of view. Figure 3b shows the 3–120-keV fluxes from this source determined by processing the RXTE (PCA+HEXTE) data. Here, we use only the sessions of stable pointings of the instruments at the source.

THE PERSISTENT SPECTRUM

Based on the INTEGRAL data, we were able to construct the 3–120-keV spectrum of the source under study (Fig. 4, dashes). For this purpose, we used the observational data obtained by the JEM-X X-ray monitor (3–20 keV; Fig. 1c, region II) and the ISGRI/IBIS detector (20–120 keV; Fig. 1a, region II) during 137–141 orbital cycles (Table 1, November 27–December 9, 2003), when the source was in its high state and its flux in these energy ranges was ~50–70 mCrab. During the fitting, the JEM-X spectrum was renormalized to correspond to the normalization of the ISGRI/IBIS spectrum; the normalization factor was 1.15. We now attribute the spectral features in the regions 6–8, 12–15, and 20–25 keV to systematic measurement errors.

A spectral analysis of the emission from MX 0836–42 using the data averaged over each of the INTEGRAL orbital cycles mentioned above showed that the spectral shape of the source did not change significantly over this period. Therefore, we

Table 1. Observations of MX 0836–42 by the JEM-X and IBIS telescopes of the INTEGRAL orbiting observatory

Beginning of observation, UTC	End of observation, UTC	IBIS exposure time, s	JEM-X exposure time, s	ISGRI/IBIS flux (20–120 keV), mCrab
2003				
March 28	March 28	6571	–	53.94 ± 2.16
April 5	April 5	6600	–	59.95 ± 1.77
April 20	April 20	6600	–	27.85 ± 1.76
April 29	April 29	6713	4513	6.45 ± 1.68
May 14	May 14	8802	4399	10.60[a]
May 29	May 29	6600	2200	16.30[a]
June 5	June 5	8917	2201	9.84[a]
June 12	June 15	202 879	93 575	1.95[a]
June 16	June 18	142 958	94 976	3.13 ± 0.39
June 19	June 21	193 103	88 973	2.24[a]
June 26	June 27	106 271	85 499	2.66[a]
June 28	June 30	192 907	122 629	2.05[a]
July 1	July 3	175 646	145 142	1.95[a]
July 4	July 6	113 051	98 261	2.60[a]
Nov. 27	Nov. 29	206 101	54 131	80.56 ± 0.38
Nov. 30	Dec. 2	206 942	113 931	81.12 ± 0.32
Dec. 3	May 12	186 043	97 123	77.59 ± 0.35
Dec. 6	Dec. 8	207 723	88 880	74.39 ± 0.33
Dec. 9	Dec. 11	193 980	98 270	69.01 ± 0.32
Dec. 24	Dec. 24	6714	4514	69.37 ± 2.22
2004				
Jan. 2	Jan. 2	4463	2263	57.17 ± 2.39
Jan. 17	Jan. 17	5713	3513	46.81 ± 2.13
Jan. 26	Jan. 26	6602	2201	50.69 ± 1.88
Feb. 7	Feb. 7	6634	–	58.36 ± 2.21
March 7	March 7	4388	–	61.62 ± 2.55
March 26	March 26	6626	–	53.14 ± 1.88
April 24	April 24	6712	4512	45.69 ± 1.79
May 16	May 16	2314	–	46.95 ± 3.83
June 6	June 6	4400	–	19.78 ± 2.34
June 11	June 11	6602	2201	11.51 ± 1.82
June 30	June 30	8123	2200	4.62[a]
July 11	July 11	6602	2200	5.17[a]
July 23	July 23	4401	–	8.49[a]
Aug. 21	Aug. 21	5470	–	7.95[a]
Aug. 28	Aug. 28	2221	–	10.83[a]
Sept. 13	Sept. 13	8064	2171	5.14[a]

Note. Here, the exposure time is the total duration of all pointings during which the source was resolved.

[a] 3σ is an upper limit on the energy flux from the source.

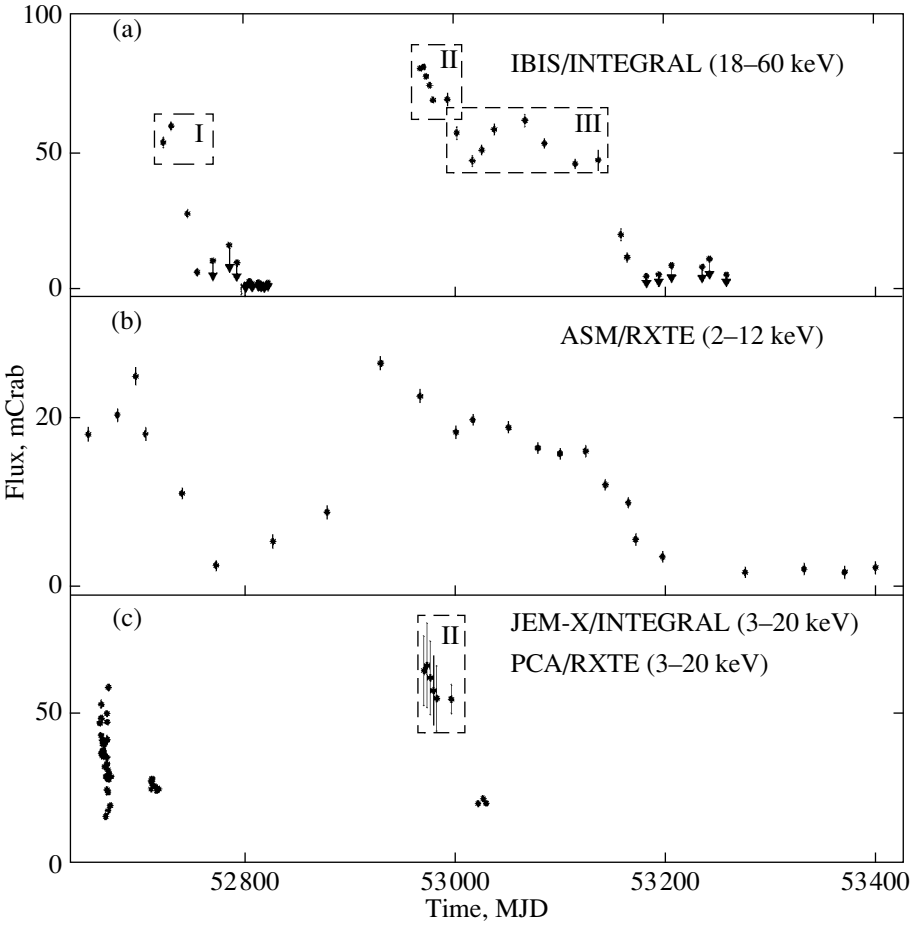


Fig. 1. Light curves for MX 0836–42 constructed from (a) ISGRI/IBIS/INTEGRAL data, (b) ASM/RXTE data, and (c) PCA/RXTE and JEM-X/INTEGRAL data (region II). The upper limits on the flux are given at the 3σ level.

used the averaged data of these six orbital cycles (Fig. 1a, region II) to construct and analyze the broadband spectrum. Fitting the constructed spectrum of the source by a power law with a high-energy exponential cutoff yields a photon index of $\alpha = 1.46 \pm 0.08$ and a cutoff energy of $E_{\rm cut} = 51.1 \pm 1.4$ keV; the 3–120-keV flux from the source was $F = (2.29 \pm 0.10) \times 10^{-9}$ erg cm^{-2} s^{-1}.

Based on the ISGRI/IBIS data averaged over 55–58 (Fig. 1a, region I) and 149–194 (Fig. 1a, region III) orbital cycles, we constructed the source's spectra only in the energy range 20–60 keV, since the total exposure times in these regions were only ~20 and ~50 ks, respectively, which are much shorter than the exposure time in region II (~1 Ms). The statistical significance of the source's detection at energies above 60 keV during these observations was insufficient to construct a qualitative spectrum. Fitting these spectra by a power law with a high-energy exponential cutoff yields a photon index of $\alpha = 1.29 \pm 0.17$ for the former and $\alpha = 1.31 \pm 0.14$ for the latter ($E_{\rm cut}$ during these fitting was fixed at 50 keV).

Figure 4 (solid lines) shows examples of the spectra for MX 0836–42 that were obtained from the PCA (3–20 keV) and HEXTE (20–60 keV) data averaged over several successive pointings (the total exposure time is ~2.9–10.5 ks). All of the spectra obtained were fitted in the energy range 3–60 keV by a power law with a high-energy exponential cutoff. We also added the reflection of radiation from the accretion disk, the photoabsorption under the assumption of solar heavy-element abundances in the interstellar medium, and the fluorescence iron line at $E_{\rm Fe} = 6.4$ keV to this model, which allowed the quality of the fit to be improved considerably (the χ^2 value per degree of freedom decreased from ~8–10 to ~1–2). In view of the uncertainty in the normalization of the HEXTE spectra, all of them were multiplied by a constant to be renormalized to the level of the PCA spectra obtained during the same observation.

Table 3 presents the fitting results and the model fluxes corrected for the dead time of the detector and the HEXTE spectral normalization constants mentioned above. Figure 3 shows the time dependences

Table 2. Observations of MX 0836–42 by the PCA and HEXTE instruments of the RXTE orbiting observatory

Date, UTC	PCA exposure time[a]	HEXTE exposure time[a]	$\dot{M}$[b] $\times 10^{-9}$, $M_\odot$ yr^{-1}	τ_1,[c] s	τ_2,[d] s
2003					
Jan. 23	11 296	5539	2.08 ± 0.03	3230	9680
Jan. 24	5088	2268	2.09 ± 0.04	3180	9810
Jan. 25	10 592	5689	1.82 ± 0.03	3640	10 930
Jan. 26	14 144	7506	1.75 ± 0.02	3810	11 420
Jan. 27	9504	5189	1.65 ± 0.03	4020	12 060
Jan. 28_1	1872	721	2.08 ± 0.12	3200	9580
Jan. 28_2	864	294	1.18 ± 0.08	5620	16 870
Jan. 29_1	1744	602	0.88 ± 0.12	7560	22 680
Jan. 29_2	944	393	1.31 ± 0.05	5070	15 200
Jan. 30_1	896	367	1.74 ± 0.05	3820	11 450
Jan. 30_2	736	327	1.23 ± 0.05	5410	16 220
Jan. 30_3	720	318	1.74 ± 0.06	3820	11 460
Jan. 31_1	5520	2253	1.28 ± 0.05	5170	15 510
Jan. 31_2	4032	2128	1.05 ± 0.04	6340	19 020
Feb. 1	8992	5085	1.48 ± 0.07	4710	13 460
Feb. 2	2944	1413	1.04 ± 0.06	6400	19 200
Feb. 3	2464	1007	1.39 ± 0.06	4780	14 340
Mar. 13	4672	2211	1.18 ± 0.04	5620	16 870
Mar. 14	10 496	5029	1.22 ± 0.02	5300	15 900
Mar. 17	3360	1627	1.23 ± 0.05	5400	16 200
Mar. 18	3392	1583	1.12 ± 0.05	5930	17 190
Mar. 20	3472	1652	1.19 ± 0.05	5590	16 750
2004					
Jan. 18	1120	464	0.90 ± 0.47	4990	14 960
Jan. 23	1216	520	0.99 ± 0.03	5040	15 110
Jan. 26	1552	723	0.78 ± 0.02	6390	19 160

Note. The subscript in the dates indicate the pointing number during the corresponding day.

[a] The total exposure time in seconds.
[b] The accretion rate.
[c] The burst recurrence period calculated in the case where 1/3 of the neutron star surface is covered with accreted matter.
[d] The burst recurrence period calculated in the case where the neutron star surface is completely covered with accreted matter.

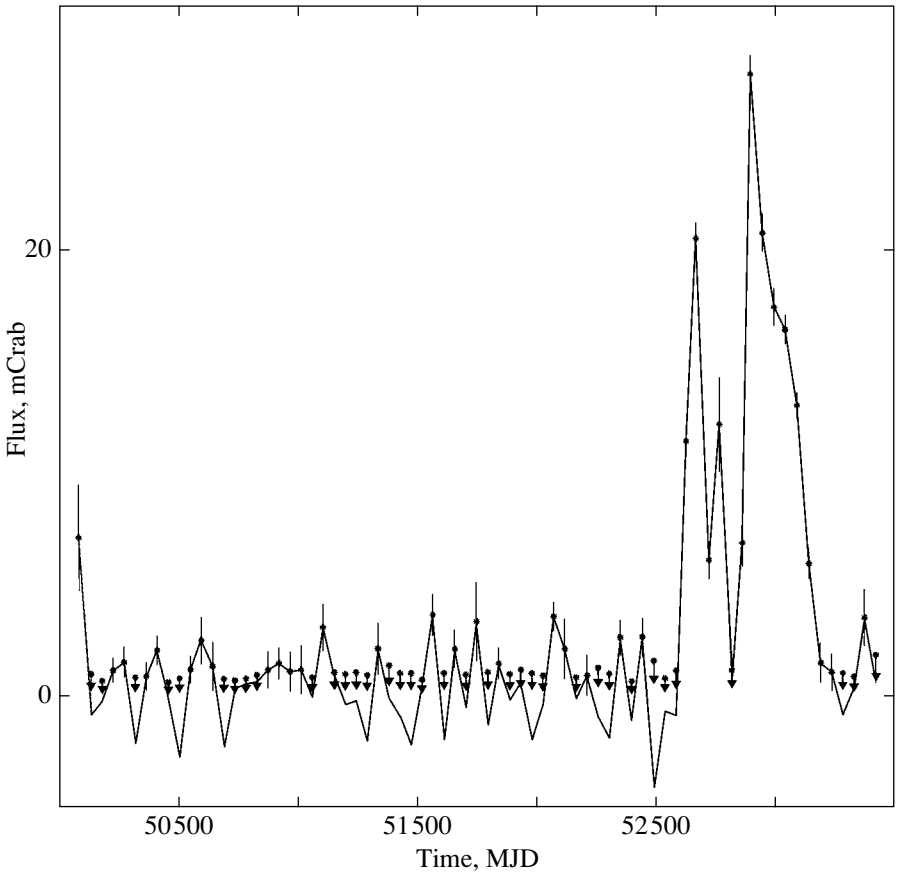


Fig. 2. Light curve for MX 0836–42 in the energy range 2–12 keV constructed from ASM/RXTE data.

of the photon index, the source's persistent model flux, and the hydrogen column density ($N_{\rm H}$) derived when fitting the spectra. The dotted line in this figure highlights the parameters determined from the data of January 29_1 and 31_2 (the subscript indicates the number of a given pointing among the pointings at the source on this day) and February 2, 2003. The relatively high values of $N_{\rm H}$ and the low values of the photon index and the flux obtained when fitting these data allow them to be separated into a group of observations with strong absorption. The derived mean photon index of the source's power-law spectrum, $\sim$(1.4–1.5), and the interstellar absorption, $\sim 3 \times 10^{22}$ atoms cm^{-2} (Fig. 3, without including the observations from the group with strong absorption mentioned above), are comparable to those obtained previously when studying this source (Aoki *et al.* 1992; Belloni *et al.* 1993).

The iron emission line at $E_{\rm Fe} = 6.4$ keV was detected in all PCA spectra. Since the PCA spectral resolution is too low for the line profile to be studied in detail, we fixed the line parameters at $E_{\rm Fe} = 6.4$ keV and $\delta E_{\rm Fe} = 0.1$ keV when fitting the spectra. The line equivalent width in the spectra under consideration was 100–310 eV.

X-RAY BURSTS

When analyzing the JEM-X 3–20-keV light curves for MX 0836–42, we found 24 X-ray bursts (Table 4). The light curves of this source were constructed only for the period of its reliable detection above the background level from the data obtained during 137–141 and 146 orbital cycles (Fig. 1, region III; Table 1, November 27–December 24, 2003).

There were no pointings containing more than one burst. The separation between the nearest of the neighboring bursts was $\sim$2 h, in good agreement with the burst recurrence period for this source estimated previously (Aoki *et al.* 1992).

To carry out a detailed analysis of the bursts from the source under study in the energy range 3–20 keV, we used the 25 observations of MX 0836–42 performed by the PCA detector of the RXTE orbiting observatory from January 24 through March 20, 2003, and from January 18 through January 26, 2004 (Table 2). We found 15 X-ray bursts in the light curves constructed from these data. The source's radiation temperature during the decay of these bursts decreased (Fig. 5b), which is characteristic of type-I X-ray bursts (Lewin and Joss 1981). Analysis of

Table 3. Results of fitting the RXTE (PCA+HEXTE) spectra of MX 0836−42 by a power law with a high-energy exponential cutoff with allowance made for the interstellar absorption and the iron emission line at $E_{Fe} = 6.4$ keV and the reflection of emission from the accretion disk

Date, UTC	α[a]	E_{cut},[b] keV	Flux,[c] erg cm^{-2} s^{-1}	N_H,[d] 10^{22} at cm^{-2}	K[e]	$\chi^2(N)$[f]
2003						
Jan. 23	1.45 ± 0.04	53.9 ± 4.9	3.18 ± 0.05	2.71 ± 0.08	0.75	1.27(60)
Jan. 24	1.44 ± 0.04	45.9 ± 4.6	3.20 ± 0.06	2.73 ± 0.09	1.00	0.87(66)
Jan. 25	1.47 ± 0.03	53.3 ± 3.8	2.79 ± 0.04	3.07 ± 0.06	0.74	1.16(63)
Jan. 26	1.50 ± 0.09	67.4 ± 3.8	2.67 ± 0.04	3.31 ± 0.07	0.79	2.08(50)
Jan. 27	1.47 ± 0.03	57.9 ± 3.1	2.53 ± 0.04	3.12 ± 0.08	0.76	1.09(59)
Jan. 28_1	1.38 ± 0.07	50.6 ± 10.1	3.18 ± 0.20	2.71 ± 0.28	1.03	0.87(51)
Jan. 28_2	1.43 ± 0.08	38.2 ± 7.7	1.81 ± 0.11	4.25 ± 0.26	0.57	0.97(52)
Jan. 29_1	0.81 ± 0.16	29.4 ± 4.6	1.34 ± 0.20	6.97 ± 0.79	0.77	1.99(52)
Jan. 29_2	1.45 ± 0.10	43.7 ± 14.7	2.01 ± 0.15	2.80 ± 0.29	0.93	0.88(44)
Jan. 30_1	1.41 ± 0.08	45.8 ± 8.2	2.66 ± 0.17	3.56 ± 0.28	0.97	0.91(52)
Jan. 30_2	1.52 ± 0.10	50.0(fixed)	1.88 ± 0.11	2.94 ± 0.30	1.12	1.05(53)
Jan. 30_3	1.40 ± 0.09	56.1 ± 14.4	2.66 ± 0.19	3.24 ± 0.32	0.98	0.93(49)
Jan. 31_1	1.44 ± 0.04	50.3 ± 5.3	1.95 ± 0.07	3.20 ± 0.11	0.75	1.25(52)
Jan. 31_2	1.32 ± 0.03	43.5 ± 4.1	1.61 ± 0.06	6.18 ± 0.19	0.70	1.88(52)
Feb. 1	1.44 ± 0.02	61.0 ± 3.8	2.27 ± 0.03	3.35 ± 0.07	0.54	1.27(66)
Feb. 2	1.16 ± 0.08	35.3 ± 3.4	1.59 ± 0.10	9.86 ± 0.32	0.87	1.29(50)
Feb. 3	1.46 ± 0.05	53.4 ± 7.9	2.13 ± 0.09	3.23 ± 0.19	0.90	1.03(52)
Mar. 13	1.51 ± 0.05	65.8 ± 8.3	1.80 ± 0.06	2.27 ± 0.16	0.71	1.27(55)
Mar. 14	1.54 ± 0.04	64.9 ± 5.5	1.87 ± 0.04	2.73 ± 0.09	0.76	1.08(50)
Mar. 17	1.54 ± 0.06	76.2 ± 12.6	1.88 ± 0.08	2.33 ± 0.20	0.86	0.88(55)
Mar. 18	1.51 ± 0.06	67.5 ± 12.4	1.71 ± 0.08	2.46 ± 0.21	0.81	0.98(52)
Mar. 20	1.58 ± 0.06	71.4 ± 14.1	1.82 ± 0.07	2.58 ± 0.17	0.92	0.87(48)
2004						
Jan. 18	1.49 ± 0.07	50.0(fixed)	1.37 ± 0.08	1.83 ± 0.28	1.01	0.96(54)
Jan. 23	1.54 ± 0.07	50.0(fixed)	1.51 ± 0.08	1.88 ± 0.25	0.98	0.93(54)
Jan. 26	1.59 ± 0.07	53.5 ± 15.2	1.19 ± 0.08	1.90 ± 0.26	0.65	1.27(53)

[a] The photon index.
[b] The exponential cutoff energy.
[c] The persistent (3–120 keV) flux from the source ($\times 10^{-9}$ erg cm^{-2} s^{-1}).
[d] The hydrogen column density obtained when fitting the spectrum.
[e] The scaling factor of the HEXTE spectrum.
[f] The χ^2 value of the best fit to the spectrum normalized to the number of degrees of freedom N.

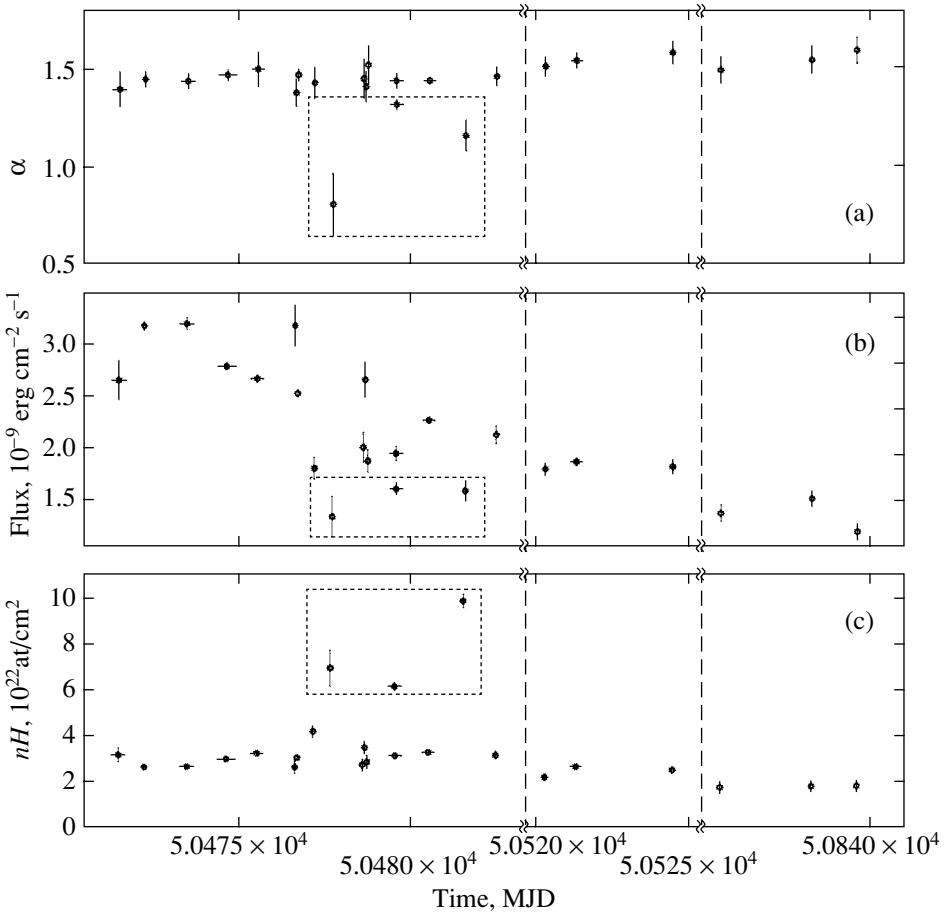


Fig. 3. (a) Time dependences of the photon index, (b) the 3–120-keV flux, and (c) the interstellar absorption derived from the RXTE (PCA+HEXTE) observations of MX 0836–42 in 2003–2004.

the exposures during which the observatory was repointed revealed no new burst.

The burst flux from the source reached its maximum, on average, in 6–8 s (Fig. 5b) and then remained at the same level for 3–4 s during some of the bursts (Fig. 5c). Table 5 gives the burst durations and the exponential burst decay times. The burst duration was defined as the ratio of the total energy released in the burst component of the emission from the burst onset time to the time the flux decreased to 10% of its maximum to the mean burst energy flux over this period. To calculate the exponential burst decay time, we fitted the burst profile by an exponential time dependence of the flux.

A characteristic feature of 80% of the X-ray bursts detected by the PCA spectrometer from MX 0836–42 is a more or less distinct double-peaked structure. An example of such a burst is shown in Fig. 5a. It is believed that a multipeaked burst shape can result from the following: (i) expansion of the photosphere under the pressure of a near-Eddington flux and (ii) peculiarities of the thermonuclear burning. Since we found no statistically significant increase in the color radius of the emitting object at the time of the dip between the peaks (the 3–20-keV flux decreased by ~15%), we can assume that the double-peaked structure of the burst in this case is not related to photospheric expansion, but could result from peculiarities of the thermonuclear burning in the source during the burst.

Interestingly, the maximum 3–20-keV flux from the source during the X-ray bursts detected by the PCA spectrometer ranged from $F \sim 1.5 \times 10^{-8}$ erg cm^{-2} s^{-1} (Fig. 5c) to $F \sim 0.5 \times 10^{-8}$ erg cm^{-2} s^{-1} (Fig. 5d). This allows us to trace the dependence of this quantity on the total energy released during the burst (Fig. 6a). We see that the maximum burst flux rises with increasing total energy released during the burst. A similar dependence was found for other sources of X-ray bursts, such as 1608–522 (Murakami *et al.* 1980), 1728–337 (Basinska *et al.* 1984), 1735–44 (Lewin *et al.* 1980), and 1837+049 (Sztajno *et al.* 1983).

Since no photospheric expansion of the neutron

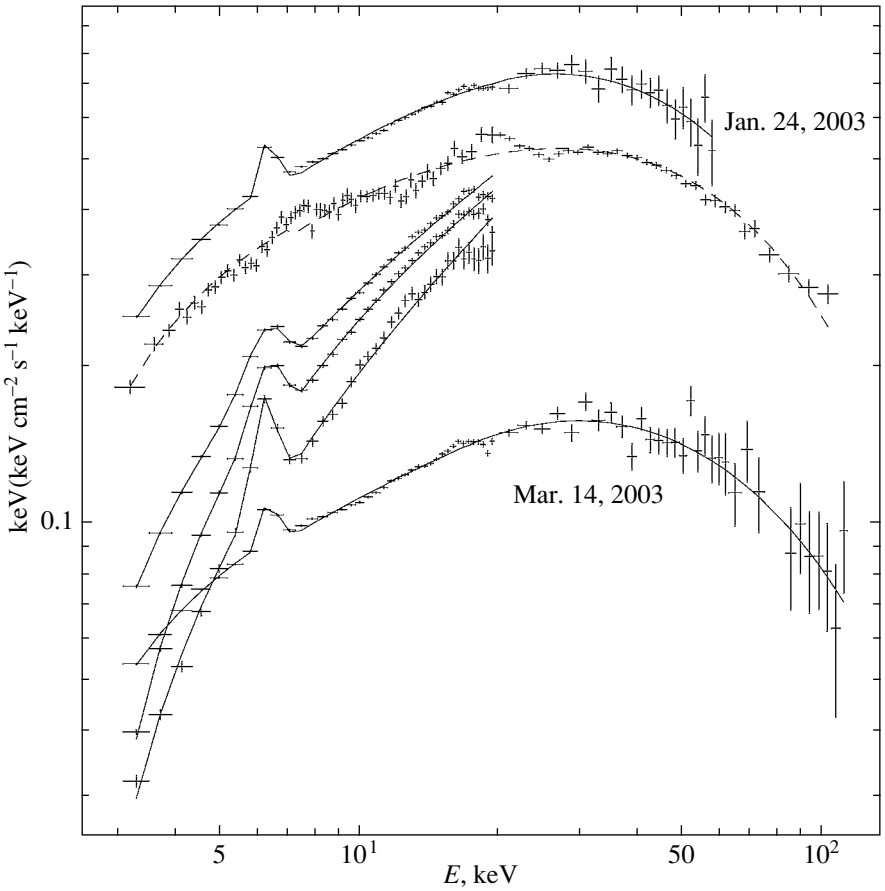


Fig. 4. Persistent spectra for MX 0836–42 constructed from the data of the JEM-X and ISGRI/IBIS telescopes aboard the INTEGRAL observatory (dashed line) and the PCA (3–20 keV) and HEXTE (20–60 keV) detectors aboard the RXTE observatory (solid lines). For convenience of perception, the values of these spectra and the model in the spectrum for March 14, 2003, were multiplied by a factor of 1/3.

star was reliably detected in any of the bursts studied, we used the flux at the maximum of the brightest of the X-ray bursts mentioned above (1.12 ± 0.24 Crab) to estimate the distance to MX 0836–42 by assuming that the source's luminosity at this time was close to the Eddington limit for a neutron star with a mass of 1.4 $M_\odot$. The derived upper limit for the distance to the source is $D \sim 8$ kpc. This value is close to the lower limit for the distance to the source $D \sim$ 10–20 kpc estimated previously by Aoki *et al.* (1992) by assuming that the color radius of the emitting object during an X-ray burst would correspond to the neutron-star radius, 10 km.

Figures 5c and 5d show the time dependences of the model flux and the color temperature and radius of the emitting object obtained when fitting the source's spectra by a blackbody during the brightest and weakest bursts detected by the PCA/RXTE spectrometer. All of the spectra studied were corrected for the background count rate of the detector and the persistent emission from the source under study. The color temperature of the emitting region, on average, rose at the burst onset to 2–2.5 keV in 1–4 s and gradually fell to 1.5–2.0 keV during the burst. The color radius R_c of the emitting region, on average, rose at the burst onset from 1–3 to 4–6 km in 3–5 s and decreased insignificantly by the burst end (by ~10–15%). This behavior of R_c may suggest that the size of the region affected by the explosion changes and that Comptonization plays a prominent role in shaping the spectrum. The mean values of the maximum temperature, $kT_{\rm bb} \sim 2.5$ keV, and the radius, $R \sim (4–6) \times (D/8\,{\rm kpc})$ km, are in good agreement with the values of these parameters obtained by Aoki *et al.* (1992).

Figure 6b shows the dependence of the persistent flux from the source on the total energy released during the burst constructed from the 15 X-ray bursts detected by the PCA/RXTE spectrometer. We see from this figure that there is a direct correlation between these quantities. This can serve as direct evidence for the current understanding of the burster phenomenon. Accreting matter falls to the neutron

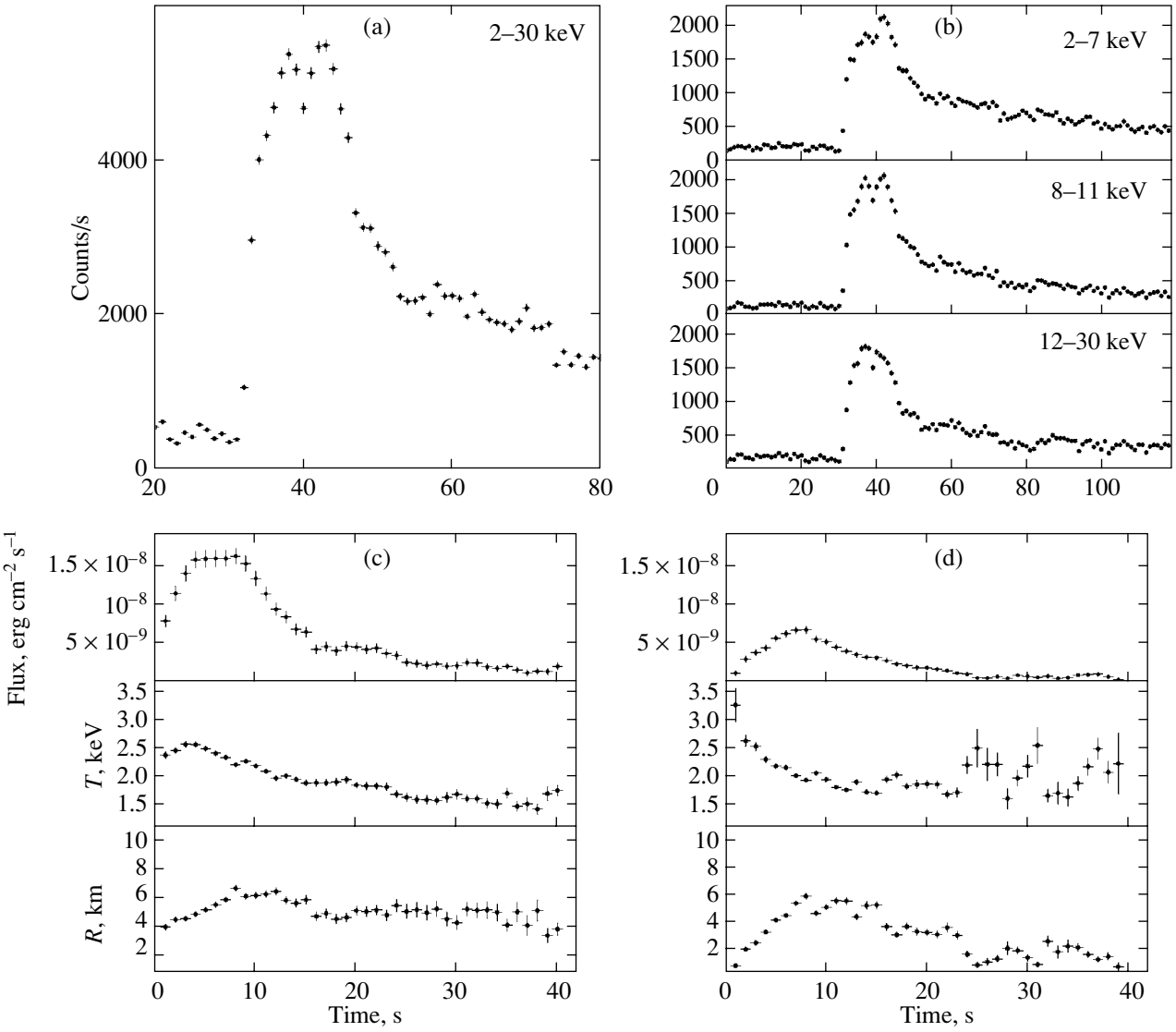


Fig. 5. Panels (a) and (b) show the time histories of the burst detected on January 24, 2003, from MX 0835–42 by the PCA/RXTE spectrometer in various energy ranges. A double-peaked structure of the burst near the maximum is clearly seen. Panels (c) and (d) display the time dependences of the 3–20-keV flux and the temperature and radius of the emitting object determined by fitting the spectra of MX 0836–42 by a blackbody during the bursts detected in 2003–2004 by PCA/RXTE. The time in panels (c) and (d) is measured from the burst onset; in panels (a) and (b), the burst onset corresponds to 30 s on the time axis.

star surface in the time between bursts, releasing part of its gravitational energy in the form of radiation that we observe as the system's persistent emission. Subsequently, this matter becomes a fuel for stable and explosive thermonuclear reactions, the latter of which are observed as an X-ray burst. In this case, if we assume that the entire accumulated store of fuel is used up during a burst, then the total energy released during the burst increases with persistent flux from the system and, hence, with accretion rate.

Since the PCA spectrometer is not a telescope, i.e., the sky in the field of view of the instrument cannot be imaged, we cannot assert with confidence that precisely MX 0836–42 is the source of the detected bursts. Note, however, that only one known burster, MX 0836–42, was within the PCA field of view when each of the bursts was detected.

The RXTE observatory is in a low near-Earth orbit; therefore, its instruments can continuously monitor the source only during 65% of its 90-min orbit. Since more than one X-ray burst occurred in none of the PCA observing sessions that we used, we cannot reliably determine the burst recurrence time τ_R for MX 0836–42 from these data. However, in the two successive sessions on January 31, 2003, the PCA spectrometer detected two X-ray bursts

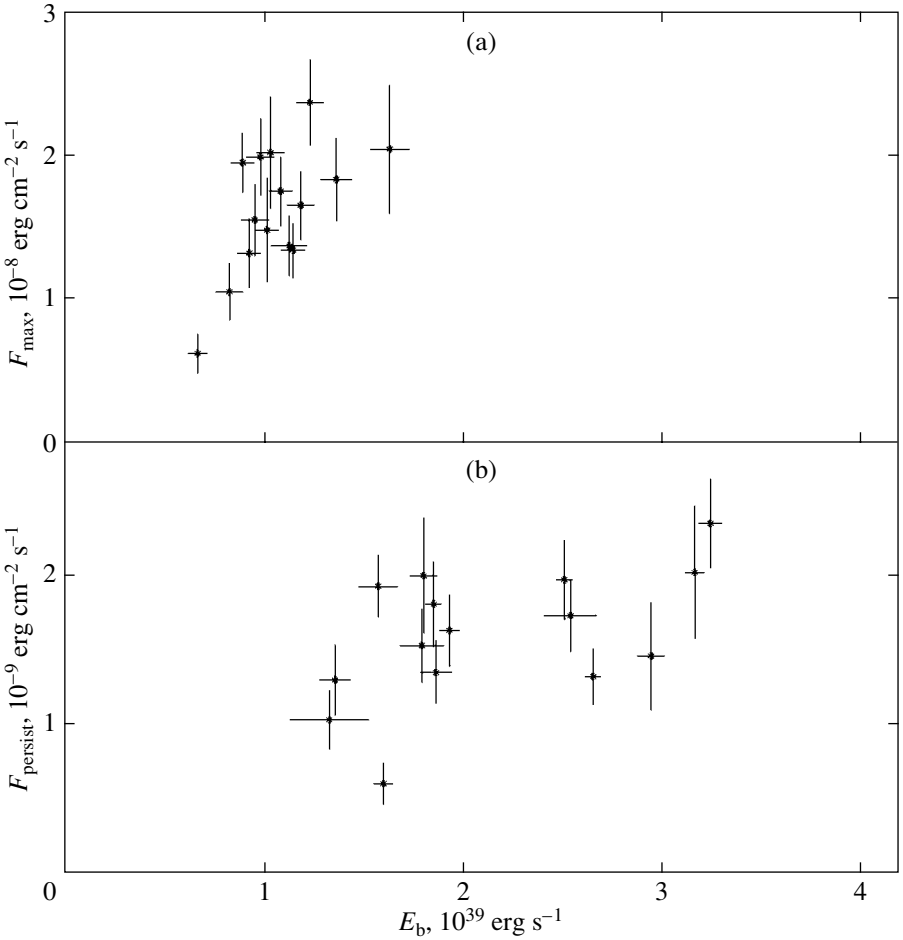


Fig. 6. (a) Maximum burst flux versus total energy released during the burst and (b) persistent flux from the source vs. total energy released during the burst. The dependences were constructed for the 15 X-ray bursts detected from MX 0836–42 using PCA/RXTE data.

separated by an interval of ~7205 s, which corresponds to the burst recurrence period determined for this source by Aoki *et al.* (1992). The source was continuously observed for ~2850 s after the first of these bursts; subsequently, the observations were interrupted for ~2610 s and then resumed. The 3–20-keV flux averaged over the second of the bursts was $F = (3.50 \pm 0.81) \times 10^{-9}$ erg cm^{-2} s^{-1}, and the mean integrated flux from the source in the energy range 3–120 keV between these bursts was $F = (1.61 \pm 0.06) \times 10^{-9}$ erg cm^{-2} s^{-1}. We assume that the bulk of the radiative energy during the burst is released in the range 3–20 keV.

In accordance with the current understanding of the burster phenomenon, the gravitational energy E_g of the matter accreted by the neutron star is released between X-ray bursts from such a system and the energy E_b of its thermonuclear burning is released during bursts. If we assume that no other bursts occurred over the interval between the pointings containing the first and second bursts during which the system was not monitored, then we can determine the thermonuclear burning parameter, the ratio $\alpha = E_g/E_b$, from the relationship

$$\alpha \approx \tau_R L_P / \tau_B L_B, \quad (1)$$

where $L_P = 1.24 \times 10^{37}$ $(D/8$ kpc$)^2$ erg s^{-1} and $L_B = 5.04 \times 10^{37}$ $(D/8$ kpc$)^2$ erg s^{-1} are the mean persistent and burst luminosities of the system and $\tau_B = 12.3$ s and $\tau_R = 7205$ s are the second burst duration and the period between bursts, respectively. In our case, formula (1) yields $\alpha \approx 144$, typical of a burst occurring through helium detonation (Bildsten 2000).

If we assume that the regime of thermonuclear burning did not change from burst to burst, then we can estimate the burst recurrence periods in other sessions using the above value of α. Table 5 gives the burst recurrence times for the source under study determined using formula (1) by assuming that $\alpha =$

Table 4. X-ray bursts detected from MX 0836−42 by the JEM-X instrument of the INTEGRAL orbiting observatory

Date,[a] MJD	F_m,[b] Crab	Date,[a] MJD	F_m,[b] Crab	Date,[a] MJD	F_m,[b] Crab
52971.954857	0.83 ± 0.20	52975.559510	0.75 ± 0.20	52980.954001	0.72 ± 0.18
52972.054063	0.62 ± 0.15	52977.293263	0.85 ± 0.21	52981.048922	0.74 ± 0.19
52972.158769	0.93 ± 0.22	52977.465705	0.69 ± 0.17	52981.146508	0.63 ± 0.16
52973.403827	0.87 ± 0.21	52978.798908	0.85 ± 0.26	52981.242242	0.56 ± 0.17
52974.639704	0.76 ± 0.21	52979.840899	1.12 ± 0.24	52982.428908	0.51 ± 0.16
52974.733813	0.69 ± 0.19	52979.930894	0.77 ± 0.17	52983.507589	0.57 ± 0.18
52975.372256	0.72 ± 0.24	52980.037844	0.84 ± 0.21	52983.710369	0.70 ± 0.17
52975.386177	0.55 ± 0.23	52980.124557	0.73 ± 0.18	52984.442369	0.85 ± 0.25

[a] The time the flux reaches its maximum.
[b] The maximum burst flux averaged over 1 s.

Table 5. X-ray bursts detected from MX 0836−42 by PCA aboard the RXTE observatory

Date, UTC	T,[a] MJD	T_{exp},[b] s	T_{eff},[c] s	F_{max},[d] erg cm^{-2} s^{-1}	F_p,[e] erg cm^{-2} s^{-1}	E_b,[f] erg	τ_R,[g] s
2003							
Jan. 23	52662.593715	8.3	16.3	1.63 ± 0.10	3.18 ± 0.05	2.04 ± 0.45	11 490
Jan. 24	52663.305984	13.0	25.2	1.23 ± 0.07	3.26 ± 0.06	2.37 ± 0.30	13 540
Jan. 25_1	52664.444097	17.3	21.2	1.08 ± 0.06	2.56 ± 0.13	1.75 ± 0.24	12 500
Jan. 25_2	52664.630428	19.4	19.1	1.01 ± 0.06	2.96 ± 0.07	1.48 ± 0.36	9140
Jan. 26	52665.702338	15.9	15.4	1.14 ± 0.06	2.67 ± 0.04	1.34 ± 0.19	9180
Jan. 27	52666.560185	20.4	26.5	0.98 ± 0.07	2.53 ± 0.04	1.99 ± 0.27	14 380
Jan. 28	52667.668495	14.7	21.3	0.95 ± 0.07	1.81 ± 0.11	1.55 ± 0.25	15 660
Jan. 29	52668.191574	12.4	16.7	0.82 ± 0.07	1.34 ± 0.20	1.05 ± 0.20	14 320
Jan. 31_1	52670.553704	12.6	18.3	1.18 ± 0.07	1.95 ± 0.05	1.65 ± 0.24	15 470
Jan. 31_2	52670.637095	10.9	12.3	0.66 ± 0.05	1.61 ± 0.05	0.62 ± 0.14	7205
Feb. 2	52672.624456	16.9	28.6	0.89 ± 0.06	1.59 ± 0.10	1.95 ± 0.21	22 430
Mar. 14	52712.408229	16.0	17.6	1.36 ± 0.08	1.87 ± 0.04	1.83 ± 0.29	17 890
Mar. 17	52715.453229	16.6	16.0	1.12 ± 0.09	1.88 ± 0.08	1.37 ± 0.21	13 320
Mar. 20	52718.536470	16.3	25.6	1.03 ± 0.07	1.82 ± 0.07	2.02 ± 0.39	20 290
2004							
Jan. 18	53022.663252	16.8	18.7	0.92 ± 0.06	1.37 ± 0.08	1.32 ± 0.24	17 620

Note. The subscript in the dates indicates the burst number during the corresponding day.

[a] The burst onset time.
[b] The exponential burst decay time.
[c] The effective burst duration.
[d] The 3–20-keV maximum burst flux ($\times 10^{-8}$).
[e] The persistent 3–120-keV flux ($\times 10^{-9}$).
[f] The energy released during the burst ($\times (D/8\ \mathrm{kpc})^2 \times 10^{39}$).
[g] The burst recurrence period (at $\alpha = 140$).

140. Given the relationship between the recurrence periods and the mean duration of the source's continuous monitoring by the observatory, the observation of only one burst in each of the sessions can be easily explained. It is worth mentioning that the estimates obtained by this method are a factor of 2–3 larger than the burst recurrent period for MX 0836–42 determined by Aoki *et al.* (1992): $\tau_R \sim 2$ h. If we assume that $\tau_R \sim 2$ h, then we can determine the parameter $\alpha \sim 80$ averaged over all of the observed bursts using formula (1); this value is typical of mixed hydrogen–helium bursts (Bildsten 2000). This model describes better the shape of the observed bursts, in particular, the relatively long (6–8 s) period of the burst rise to its maximum level.

Table 2 gives the accretion rate corresponding to the detected persistent integrated flux from the source under study and the corresponding recurrence periods of hydrogen–helium bursts from it calculated by assuming that accreted matter occupies 1/3 of the surface (τ_1) and the entire surface (τ_2) of the neutron star. We see from this table that if the bursts in MX 0836–42 are hydrogen–helium ones, then the accreted matter involved in the explosion during the burst occupies only part of the neutron star surface.

DISCUSSION

Figure 2 shows the light curve for MX 0836–42 constructed from all of the available ASM/RXTE data, in which we clearly see a rise in the flux from the source on a time scale of ~600 days. Clearly, even if this phenomenon is periodic, the period of such bursts is more than 7 years. For example, nonuniformity in the accretion process can be responsible for such bursts.

It follows from Fig. 3 that there is a group of three observations with an anomalously high absorption level among the RXTE observations of the source (Table 3). It can be assumed that additional absorption possibly associated with the outer regions of the accretion disk appears in the system during these observations. The following fact argues for this explanation: the harder (more absorbed) the spectrum, the stronger the iron emission line in it. For example, the line equivalent width in the spectrum constructed from the observations on January 26 was ~120 eV, while this parameter for the 29_1 observations was ~310 eV. This explanation suggests a large inclination of the accretion disk in the system under study. This process can be periodic and related to the orbital motion in the binary system; i.e., the source can be a dipper. To test this assumption, we analyzed the light curve of the source in the energy range 1.3–3.0 keV, which is subject to the strongest absorption, constructed from the ASM/RXTE data for the period from January 5, 1996, through April 7, 2005, for the presence of periodic variations in the frequency range $(5–300) \times 10^{-6}$ Hz, which is typical of the orbital motion of a low-mass binary. This analysis failed to reveal any significant period of the signal variations.

The transient X-ray pulsar GRS 0834–43 is in the immediate vicinity of the source under study (24 arcmin). Since the PCA and HEXTE spectrometers are incapable of spatially resolving the sources of the detected emission, the above pulsar could introduce distortions in the observed spectrum of MX 0836–42 when falling within the fields of view of these instruments. To test this hypothesis, we searched for the pulsations with a period of ~12 s that are typical of the pulsar GRS 0834–43 using all the available PCA observations of MX 0836–42. These studies failed to reveal any pulsating component whose confidence level would exceed 3σ in any of the observations. It is worth mentioning that the ASM/RXTE data revealed no flare activity in the emission from GRS 0834–43 during the period under consideration, while according to the ISGRI/IBIS data, the confidence level of the source's detection in the range 20–60 keV from January 2003 through March 2004 did not exceed 3σ; the upper limit on the flux from the source was 1 mCrab throughout this period and ~10 mCrab for the period of a single observation (~2 ks). Thus, the persistent spectra constructed from the available PCA and HEXTE data actually pertain to MX 0836–42.

In the course of some of the 15 X-ray bursts detected by the PCA spectrometer during the pointings at the source under study, the mean color radius of the emitting object was 4–7 km (Figs. 5c and 5d), which is smaller than the value obtained in terms of the standard models for the structure of a neutron star for its radius. A modification of the source's spectrum through scattering, which leads to an overestimation of the color temperature and an underestimation of the color radius (London *et al.* 1986; Sunyaev and Titarchuk 1986; Babul and Paczynski 1987), could be responsible for this discrepancy.

We see from Table 5 that the total energy released during the burst on January 31_2, 2003, is a factor of 2–3 lower than that for the remaining bursts, and that the burst itself occurred a factor of 2 earlier than the accumulation of a matter column density enough for detonation could occur (if detonation affects the entire surface of the neutron star) (Table 2). A similar situation was observed in the emission from this source on February 18, 1991 (Aoki *et al.* 1992), except that the separation between the bursts in that case was only ~10 min.

These phenomena can be explained as follows.

(1) Another burster is the burst source. Although this burst is morphologically similar to other bursts

from the source under consideration, given that the PCA spectrometer is incapable of imaging the observed sky region, this assumption cannot be refuted using the available data. This assumption can also explain the different maximum burst energy fluxes from the source.

(2) The thermonuclear burning during the burst proceeds in a regime different from the remaining cases. This is possible if, for example, some amount of fuel was not used up during a previous burst. In cases 1 and 2, we can say nothing about the parameter α and the burst recurrent period in other sessions.

(3) Accretion affected a smaller part of the neutron star surface, which led to a faster accumulation of matter to the column density required for detonation. This could be evidenced by the larger increase in the color radius at the burst onset and its larger decrease at the burst end than those for other bursts (Fig. 5d).

ACKNOWLEDGMENTS

We thank M.G. Revnivtsev for valuable suggestions and remarks made by him when preparing this paper. This work is based on the observational data obtained by the INTEGRAL observatory and provided through the Russian and European INTEGRAL Science Data Centers as well as the data obtained by the RXTE observatory and provided through the NASA's HEASARC Web page on the Internet (http://legacy.gsfc.nasa.gov). This study was supported by the Russian Foundation for Basic Research (project no. 05-02-17454), the Presidium of the Russian Academy of Sciences (the "Nonstationary Astronomical Phenomena Program"), and the Program for Support of Leading Scientific Schools (project no. NSh-2083.2003.2).

REFERENCES

1. T. Aoki, T. Dotani, K. Ebisawa, *et al.*, Publ. Astron. Soc. Jpn. **44**, 641 (1992).
2. A. Babul and B. Paczynski, Astrophys. J. **323**, 582 (1987).
3. E. M. Basinska, W. H. G. Lewin, M. Sztajno, *et al.*, Astrophys. J. **281**, 337 (1984).
4. T. Belloni, G. Hasinger, W. Pietcch, *et al.*, Astron. Astrophys. **271**, 487 (1993).
5. L. Bildsten, 2000arxt.confE..65B.
6. L. Cominsky, C. Jones, W. Forman, and H. Tananbaum, Astrophys. J. **224**, 46 (1978).
7. N. A. Eismont, A. V. Ditrikh, G. Janin, *et al.*, Astron. Astrophys. **411**, L37 (2003).
8. S. Grebenev and R. Syunyaev, IAU Circ., No. 5294 (1991).
9. G. Hasinger, W. Pietsch, and T. Belloni, IAU Circ., No. 5142 (1990).
10. E. Kellogg, H. Gursky, S. Murray, *et al.*, Astrophys. J. Lett. **169**, L99 (1971).
11. I. Yu. Lapshov, V. V. Dremin, R. A. Syunyaev, *et al.*, Pis'ma Astron. Zh. **18**, 30 (1992) [Sov. Astron. Lett. **18**, 12 (1992)].
12. W. H. G. Lewin and P. C. Joss, Space Sci. Rev. **28**, 3 (1981).
13. W. H. G. Lewin, J. van Paradijs, L. Cominsky, and S. Holzner, Mon. Not. R. Astron. Soc. **193**, 15 (1980).
14. R. A. London, R. E. Taam, and W. M. Howard, Astrophys. J. **306**, 170 (1980).
15. N. Lund, S. Brandt, C. Budtz-Joergesen, *et al.*, Astron. Astrophys. **411**, 231 (2003).
16. P. Magdziarz and A. Zdziarski, Mon. Not. R. Astron. Soc. **273**, 837 (1995).
17. F. Makino, IAU Circ., No. 5139 (1990).
18. T. Markert, H. V. Bradt, G. W. Clark, *et al.*, IAU Circ., No. 2765 (1975).
19. T. Markert, C. R. Canizares, G. W. Clark, *et al.*, Astrophys. J. **218**, 801 (1977).
20. T. Murakami, H. Inoue, K. Koyama, *et al.*, Astrophys. J. **240**, 143 (1980).
21. R. A. Sunyaev, IAU Circ., No. 5122 (1990).
22. R. A. Sunyaev, IAU Circ., No. 5180 (1991).
23. R. A. Sunyaev and L. G. Titarchuk, Pis'ma Astron. Zh. **12**, 857 (1986) [Sov. Astron. Lett. **12**, 359 (1986)].
24. M. Sztajno, E. M. Basinska, L. R. Cominsky, *et al.*, Astrophys. J. **267**, 713 (1983).
25. P. Ubertini, F. Lebrun, G. Di Cocco, *et al.*, Astron. Astrophys. **411**, 131 (2003).
26. C. Winkler, T. J.-L. Courvoisier, G. Di Cocco, *et al.*, Astron. Astrophys. **411**, L1 (2003).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 695–705. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 780–792.

Similarity Theory of Stellar Models and the Structure of Very Massive Stars

D. K. Nadyozhin[1,2*] and T. L. Razinkova[2]

[1] *Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, Postfach 1317, D-85741 Garching, Germany*

[2] *Institute for Theoretical and Experimental Physics, ul. Bol'shaya Cheremushkinskaya 25, Moscow, 117218 Russia*

Received May 13, 2005

Abstract—The similarity theory of stellar models is used to study the properties of very massive stars when all the opacity sources, except the Thomson scattering, can be disregarded. The dimensionless internal structure of such stars is essentially independent of the energy generation law. It is shown that the mass–luminosity relation can be fitted by an analytical expression that is virtually universal with regard to the chemical composition and the energy generation law. A detailed comparison with the Eddington standard model is made. The application of the results obtained to the observations of massive stars is briefly discussed. © *2005 Pleiades Publishing, Inc.*

Key words: *similarity theory, stellar structure, massive stars.*

INTRODUCTION

The similarity theory of stellar models was a major tool for investigating the properties of stellar structures even in the precomputer era of their study. It will suffice to mention a theoretical explanation of the mass–luminosity and mass–radius relations (Biermann 1931; Strömgren 1936; Sedov 1959). The similarity theory remains useful for interpreting the various aspects of the stellar structure (Schwarzschild 1961; Chiu 1968; Cox and Guili 1968; Dibai and Kaplan 1976; Kippenhahn and Weigert 1990).

Here, we first describe the similarity theory of stellar models as a boundary-value problem formulated by Imshennik and Nadyozhin (1968). We then discuss the structure of chemically homogeneous stars of such a large mass that the opacity may be considered to be due to the Thomson scattering alone. This approximation appears to be valid for still hypothetical Population-III stars and a number of observed luminous stars, e.g., massive O stars and Wolf–Rayet stars, as well as for some of the specific stars, such as η Car and the Pistol star. Particular attention is given to the comparison with the Eddington standard model.

THE SIMILARITY THEORY OF STELLAR MODELS

Let us consider a spherical star in hydrostatic and thermal equilibrium. If the pressure P is the sum of the perfect gas pressure $P_{\rm g}$ and the blackbody radiation pressure $P_{\rm r}$ and if the energy generation rate ε and the opacity κ are power functions of the temperature T and density ρ, then the structure of the star is described by the following system of differential equations:

$$\frac{dP}{dr} = -\rho\frac{Gm}{r^2}, \tag{1}$$

$$\frac{dm}{dr} = 4\pi r^2\rho, \tag{2}$$

$$\frac{dL}{dr} = 4\pi r^2\rho\varepsilon, \tag{3}$$

$$\frac{dT}{dr} = \begin{cases} -\frac{\rho T}{P}\frac{Gm}{r^2}\nabla_A, & \nabla_r \geq \nabla_A \\ \nabla_A = \left(\frac{\partial\log T}{\partial\log P}\right)_S & \\ -\frac{\rho T}{P}\frac{Gm}{r^2}\nabla_r, & \nabla_r \leq \nabla_A \\ \nabla_r = \frac{3\kappa LP}{16\pi caT^4Gm}, & \end{cases} \tag{4}$$

$$P = P_{\rm g} + P_{\rm r} = \frac{k}{m_{\rm u}}\frac{1}{\mu}\rho T + \frac{1}{3}aT^4, \tag{5}$$

$$\beta = \frac{P_{\rm g}}{P}, \quad \beta^{-1} = 1 + \mu\frac{am_{\rm u}}{3k}\frac{T^3}{\rho}, \tag{6}$$

$$\nabla_A = \frac{2(4-3\beta)}{32-24\beta-3\beta^2},$$

*E-mail: nadezhin@mail.itep.ru

$$\kappa = \kappa_0 \rho^{\alpha} T^{-\nu}, \quad \varepsilon = \varepsilon_0 \rho^{\delta} T^{\eta}, \tag{7}$$

where β is the ratio of the gas pressure to the total pressure; G, k, and a are the gravitational, Boltzmann, and radiation density constants, respectively; $m_{\rm u}$ is the atomic mass unit; and c is the speed of light. The mean molecular mass μ and the coefficients κ_0 and ε_0 depend on the chemical composition of the star.

The above equations must be solved under the boundary conditions

$$\text{Center} \quad r = 0: \quad m = 0, L = 0, \tag{8}$$

$$\text{Surface} \quad r = R: \qquad P = 0, \rho = 0, m = M. \tag{9}$$

Conditions (8) imply that there is neither a point mass nor a point energy source at the center of the star. Conditions (9) at the stellar surface suggest that the pressure and the density become zero there, and the mass m must be equal to the total mass M of the entire star. The stellar radius R must be obtained as a result of the solution of Eqs. (1)–(7); i.e., R is an eigenvalue of the problem. Simultaneously, the solution yields the luminosity of the star $L_0 \equiv L(R)$.

Assuming that the chemical composition and, consequently, μ, κ_0, and ε_0 are constant throughout the star, we can represent the above equations in dimensionless form if the physical quantities are measured in the following system of units (Schwarzschild 1961):

$$r \to R, \quad m \to M, \quad L \to L_0, \tag{10}$$
$$T \to \mu \frac{m_{\rm u}}{k} \frac{GM}{R} \equiv T_0, \quad P \to \frac{GM^2}{4\pi R^4} \equiv P_0,$$
$$\rho \to \frac{M}{4\pi R^3} \equiv \rho_0.$$

Introducing the dimensionless variables

$$x = r/R, \quad q = m/M, \quad l = L/L_0, \tag{11}$$
$$p = P/P_0, \quad \sigma = \rho/\rho_0, \quad t = T/T_0,$$

we can rewrite Eqs. (1)–(9) as (Imshennik and Nadyozhin 1968)

$$\frac{dp}{dx} = -\frac{\sigma q}{x^2}, \tag{12}$$

$$\frac{dq}{dx} = x^2 \sigma, \tag{13}$$

$$\frac{dl}{dx} = C_1 x^2 \sigma^{1+\delta} t^{\eta}, \tag{14}$$

$$\frac{dt}{dx} = \begin{cases} -\dfrac{q\sigma t}{x^2 p} \nabla_A, & \nabla_r \geq \nabla_A \\ \nabla_A(\beta) = \dfrac{2(4-3\beta)}{32-24\beta-3\beta^2} & \\ -\dfrac{q\sigma t}{x^2 p} \nabla_r, & \nabla_r \leq \nabla_A \\ \nabla_r = C_2 \dfrac{pl\sigma^{\alpha}}{qt^{4+\nu}}, & \end{cases} \tag{15}$$

$$p = \sigma t + B t^4, \quad \beta^{-1} = 1 + B\frac{t^3}{\sigma}, \tag{16}$$

where all of the parameters are gathered in three dimensionless constants:

$$C_1 = \frac{1}{(4\pi)^{\delta}} \left(\frac{Gm_{\rm u}}{k}\right)^{\eta} \mu^{\eta} M^{1+\delta+\eta} \frac{\varepsilon_0}{L_0 R^{3\delta+\eta}}, \tag{17}$$

$$C_2 = \frac{3(4\pi)^{-\alpha}}{64\pi^2 ac} \left(\frac{k}{Gm_{\rm u}}\right)^{4+\nu} \mu^{-4-\nu} M^{\alpha-3-\nu} \frac{\kappa_0 L_0}{R^{3\alpha-\nu}}, \tag{18}$$

$$B = \frac{4\pi a}{3G} \left(\frac{Gm_{\rm u}}{k}\right)^4 (\mu^2 M)^2 \tag{19}$$
$$= 0.78096 (\mu^2 M/M_{\odot})^2.$$

The dimensionless boundary conditions take the form

$$\text{Center} \quad x = 0: \quad q = 0, \quad l = 0, \tag{20}$$

$$\text{Surface} \quad x = 1: \quad p = 0, \quad t = 0, \quad l = 1, \quad q = 1. \tag{21}$$

Thus, we now have six boundary conditions for four first-order differential equations (12)–(15). This means that for each given B (or $\mu^2 M$), the constants C_1 and C_2 can have only certain values (eigenvalues) if the solution of the *four* differential equations is to satisfy the *six* boundary conditions. The solution itself, as well as C_1 and C_2, depend only on $\mu^2 M$ and the exponents α, ν, δ, η: $C_{1,2} = C_{1,2}(\mu^2 M, \alpha, \nu, \delta, \eta)$.

Equations (12)–(16) can be solved by standard methods. For the calculations described in the next section, we used the following algorithm. Having chosen trial values of C_1 and C_2 and numerically integrating Eqs. (12)–(16) from the surface ($x = 1$, boundary conditions (21)) inward to a certain point $x = x_{\rm f}$ ($0 < x_{\rm f} < 1$), we obtain a set of four quantities $\{p(x_{\rm f}), q(x_{\rm f}), l(x_{\rm f}), t(x_{\rm f})\}$ as functions of C_1 and C_2. Integrating these equations from the center (($x = 0$, boundary conditions (20)) with the same C_1 and C_2 and trial values of $p(0) = p_{\rm c}$ and $t(0) = t_{\rm c}$ outward to the same point $x_{\rm f}$, we obtain a similar set of quantities that now depend on C_1, C_2, $p_{\rm c}$, and $t_{\rm c}$. Since $p(x)$, $q(x)$, $l(x)$, and $t(x)$ must be continuous at any point

in the star, the two sets must coincide at $x = x_f$. We can organize an iterative process for the four unknown parameters C_1, C_2, p_c, and t_c to find values of them that would satisfy the required continuity of the four variables p, q, l, and t at $x = x_f$. Of course, the result does not depend on x_f, but the domain of convergence of the iterations depends on this quantity.

As soon as C_1 and C_2 have been found, we can easily derive the luminosity L_0 and the radius R from Eqs. (17) and (18), which represent the mass–radius and mass–luminosity relations, respectively.

THE STRUCTURE OF VERY MASSIVE STARS

In this section, the above similarity theory is used to describe the structure of very massive stars. In this case, the opacity is dominated by the Thomson scattering, and we can set $\alpha = 0$, $\nu = 0$, and $\kappa = \kappa_0 = 0.2(1 + X)$ cm^2/g (X is the hydrogen mass fraction).

The mass–luminosity relation (Eq. (18)) can be rewritten as

$$L_0 = \frac{64\pi^2 ac}{3\kappa_0}\left(\frac{Gm_u}{k}\right)^4 \mu^4 M^3 C_2(\mu^2 M, \delta, \eta). \quad (22)$$

All of the models with physically reasonable exponents of the temperature in the energy generation law (i.e., $\eta \geq 4$) considered below have convective cores and radiative outer envelopes. Using Eqs. (12) and (15) and boundary conditions (21), we can make sure that C_2 is simply related to $\beta_s(\mu^2 M, \delta, \eta)$, the value of the parameter β at the stellar surface:

$$1 - \beta_s = 4BC_2. \quad (23)$$

The mass–luminosity relation can then be rewritten as

$$L_0 = L_{Ed}(1 - \beta_s), \quad (24)$$

where the Eddington critical luminosity L_{Ed} is given by

$$L_{Ed} \equiv \frac{4\pi cGM}{\kappa_0} = \frac{6.483 \times 10^4}{1 + X}\frac{M}{M_\odot}L_\odot. \quad (25)$$

Using the Runge–Kutta method with automatic control of the accuracy of calculations to solve Eqs. (12)–(16) from the stellar surface down to x_f and from the center up to x_f, we calculated a large number of models over a wide range of the parameter $\mu^2 M$ ($0 \leq \mu^2 M \leq 4000 M_\odot$). Typically, $x_f \approx 0.1$–0.3 are the best values that ensure the convergence of the iterations. However, the domain of convergence turned out to be rather narrow, at least for the Newton–Raphson iteration scheme used in our calculations. Therefore, when calculating such a sequence of models, we have to vary $\mu^2 M$ by no more than a few percent to ensure the convergence of the iterations.

Table 1 presents the most important properties of several selected models. For better perception, we use the same notations for dimensionless quantities as those for dimensional ones. The first row of Table 1 gives $\mu^2 M$, where M is measured in $M_\odot$. The next three rows contain the values of C_1 for three energy generation modes: the CNO cycle, the 3α reaction, and the pp chain. The fifth row gives the values of C_2. The next six rows contain the dimensionless values of the central density ρ_c, pressure P_c, temperature T_c, radiative temperature gradient ∇_{rc}, gravitational potential φ_c (in units of GM/R), and ratio β_c of the gas pressure to the total pressure. The next two rows list β_{conv} and β_s at the convective core boundary and the stellar surface, respectively. Finally, the last six rows present the dimensionless radius, mass, luminosity, density, pressure, and temperature at the convective core boundary.

The calculated sequences of one-parameter ($\mu^2 M$-dependent) dimensionless stellar models have such large convective cores that the bulk of the thermonuclear energy is released within the core. Therefore, $L(r)$ is almost equal to the luminosity L_0 outside the convective core. For such a strong temperature dependence of the energy generation as that in the CNO cycle ($\eta = 16$) and the 3α reaction ($\eta = 30$), we can set $l(x) = 1$ in Eq. (15) with a high accuracy. Consequently, Eq. (14) for $l(x)$ proves to be disentangled from remaining equations (12), (13), (15), and (16). The overall stellar structure can then be calculated by solving a truncated eigenvalue problem where the energy generation law is disregarded and the constant C_2 is considered as the only eigenvalue parameter. In particular, the C_2 eigenvalue found and, hence, L_0 (Eq. (22)), as well as the dimensionless stellar structure, cease to depend on the exponents δ and η. This truncated model is actually a model with a point energy source, Thomson opacity, and the radiation pressure that was first calculated by Henrich (1943) within a limited range of stellar masses ($0 \leq \mu^2 M \leq 119$). Our numerical results are in excellent agreement with his calculations, which are accurate to $\sim 1\%$. Once C_2 and the dimensionless functions $\sigma(x)$ and $t(x)$ have been calculated by solving the truncated problem, we can find the constant C_1 and the spatial distribution of the dimensionless luminosity $l(x)$:

Table 1. Structural properties of selected models

$\mu^2 M$	0	10	30	100	300	1000	4000
$\log_{10} C_1$ [a]	1.6347	3.4318	5.8102	8.8083	11.768	15.281	19.622
$\log_{10} C_1$ [b]	2.1212	5.8346	10.615	16.602	22.509	29.523	38.199
$\log_{10} C_1$ [c]	−0.4145	0.0220	0.5988	1.2961	1.9789	2.8051	3.8511
$\log_{10} C_2$	−3.2981	−3.5181	−3.9842	−4.7533	−5.5839	−6.5626	−7.7304
ρ_c	59.34	60.20	65.05	79.97	98.95	119.4	137.5
P_c	45.55	43.95	46.43	58.41	75.32	94.78	112.8
T_c	0.7677	0.5816	0.4097	0.2630	0.1699	0.1014	0.0539
∇_{rc} [a]	2.459	3.249	4.525	5.806	6.588	6.992	7.266
φ_c	−3.428	−3.389	−3.421	−3.602	−3.821	−4.034	−4.204
β_c	1	0.7967	0.5739	0.3601	0.2232	0.1278	0.0657
β_{conv}	1	0.8671	0.6689	0.4294	0.2600	0.1429	0.0704
β_s	1	0.9053	0.7085	0.4489	0.2673	0.1451	0.0709
r_{conv}	0.2832	0.3903	0.5107	0.6264	0.7077	0.7771	0.8380
m_{conv}	0.3120	0.5691	0.8063	0.9412	0.9823	0.99531	0.99895
L_{conv} [b]	0.7810	0.9799	0.9991	1.0000	1.0000	1.0000	1.0000
ρ_{conv}	31.39	16.02	5.985	1.661	0.5056	0.1429	0.0379
P_{conv}	15.76	5.830	1.446	0.2540	0.05324	0.01049	0.00184
T_{conv}	0.5022	0.3157	0.1616	0.06564	0.02736	0.01036	0.00342

[a] CNO cycle ($\delta = 1, \eta = 16$).
[b] 3α reaction ($\delta = 2, \eta = 30$).
[c] pp-chain ($\delta = 1, \eta = 4$).

$$C_1 = \left[\int_0^1 x^2\, \sigma^{1+\delta} t^\eta \mathrm{d}x\right]^{-1}, \qquad (26)$$

$$l(x) = C_1 \int_0^x x^2 \sigma^{1+\delta} t^\eta \mathrm{d}x.$$

If the energy generation is described by an arbitrary (but still strongly temperature-dependent) function, $\varepsilon = \varepsilon(\rho, T)$, then the stellar radius R must be considered as an eigenvalue instead of C_1; this can be found from the equation

$$L_0 = M \int_0^1 x^2 \sigma(x) \varepsilon \left[\frac{M}{4\pi R^3}\sigma(x), \mu\frac{m_u}{k}\frac{GM}{R}t(x)\right] \mathrm{d}x, \qquad (27)$$

where the luminosity L_0 for a given M and stellar chemical composition is determined by the value of C_2 obtained from the solution of the truncated problem.

The structural properties of the models presented in Table 1 correspond to three energy generation modes: the CNO cycle, the 3α reaction, and the pp chain. For the CNO cycle and the 3α reaction, the dimensionless luminosity L_{conv} at the convective core boundary is 1.0000 for all values of $\mu^2 M$. The same is true for the pp chain when $\mu^2 M \gtrsim 30$. All of the other parameters in Tables 1 and 2, except C_1 and ∇_{rc}, prove to be the same for the CNO cycle and the 3α reaction and, at $\mu^2 M \gtrsim 30$, for the pp chain as well. In practice, we may consider only the range $0 \leq \mu^2 M \lesssim 10$ as corresponding to the slight differences for the pp chain shown in Figs. 4–6. This also applies to the mass–luminosity relation. More specifically, for $\eta = 4$: $\log_{10} C_2 = -3.516$ and -3.285 at $\mu^2 M = 10$ and 0, respectively.

The constant C_1 and the central value of the logarithmic radiative temperature gradient $\nabla_{rc} = C_1 C_2 p_c \sigma_c^\delta t_c^{\eta-4}$ depend strongly on the energy generation law. The values of ∇_{rc} for the CNO cycle in Table 1 show that even for low masses ($\mu^2 M \lesssim 1$), when the radiation pressure can be ignored, there

Table 2. Integrated properties of selected models

$\mu^2 M$	0	10	30	100	300	1000	4000	Edd
E_g [a]	−1.227	−1.206	−1.205	−1.254	−1.318	−1.383	−1.435	−1.5
E_T	0.6137	0.6969	0.8350	1.016	1.166	1.292	1.387	1.450
I [b]	0.1561	0.1616	0.1630	0.1536	0.1412	0.1302	0.1222	0.113
I_{conv}	0.0139	0.0439	0.0908	0.1247	0.1309	0.1271	0.1214	–
t_s	2.438	2.517	2.580	2.624	2.649	2.665	2.677	2.723
t_{conv}	0.269	0.433	0.619	0.843	1.044	1.251	1.472	–
τ_c	1.441	1.426	1.475	1.664	1.896	2.132	2.329	2.587
$\langle\gamma\rangle$	5/3	1.532	1.453	1.401	1.3733	1.3555	1.3445	1.345
ω	2.804	2.109	1.629	1.288	1.058	0.840	0.627	0.667

[a] E_g reaches a maximum of −1.2005 at $\mu^2 M = 18.5$.
[b] I reaches a maximum of 0.1636 at $\mu^2 M = 21$.

is an appreciable excess of ∇_{rc} over ∇_A ($0.25 \leq \nabla_A \leq 0.4$) large enough to drive adiabatic convection along the stellar core. For the 3α reaction, this excess is even larger: $\nabla_{rc} = 11.05$ for $\mu^2 M = 0$. However, for the pp chain, we have $\nabla_{rc} = 0.594$, a value that exceeds the adiabatic gradient only by a factor of 1.5 ($\nabla_{Ac} = 0.4$ for $\mu^2 M \approx 0$). This is because the exponent $\eta = 4$ is not very far from the critical value of $\eta \approx 2$–3 necessary to ensure the existence of a convective core (Cowling 1934; Naur and Osterbrock 1953).

In our case of constant opacity, the convective core disappears at $\mu^2 M = 0$ if $\eta = 1.644$ ($\delta = 1$). However, it still exists for all nonzero $\mu^2 M$. For lower η, the convective core appears only at fairly large $\mu^2 M$. For instance, for $\eta = 1$ ($\delta = 1$), the convective core exists only at $\mu^2 M > 2.54$ ($M > 6.6 M_\odot$ for solar chemical composition).

Figures 1 and 2 show the radial density and tem-

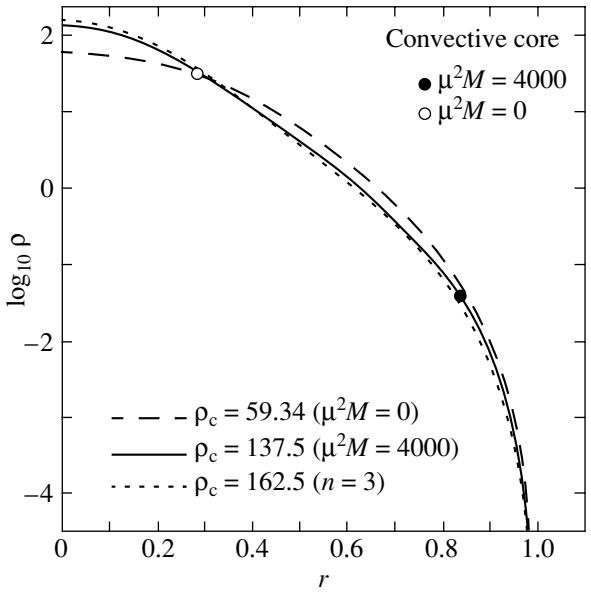


Fig. 1. Density vs. radius in the units defined by Eqs. (10) for $\mu^2 M = 0$ (dashed line) and 4000 (solid line) and for the $n = 3$ polytropic model (dotted line). The boundaries of the convective core are indicated for $\mu^2 M = 0$ and 4000.

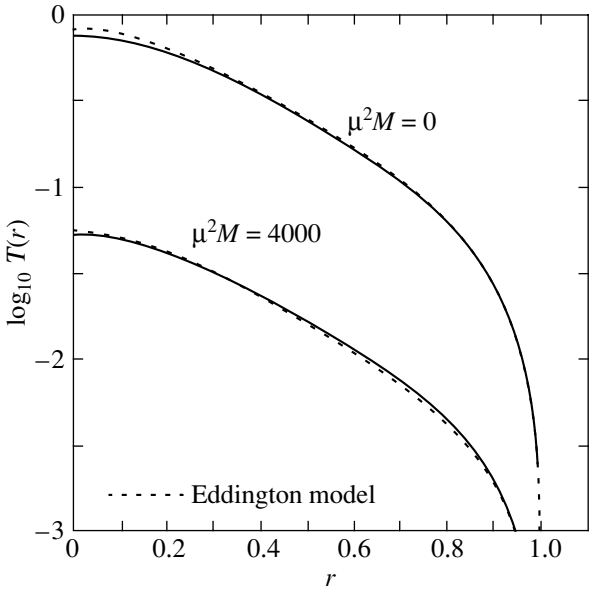


Fig. 2. Temperature versus radius in the units defined by Eqs. (10) for $\mu^2 M = 0$ and 4000. The Eddington model is indicated by the dotted lines.

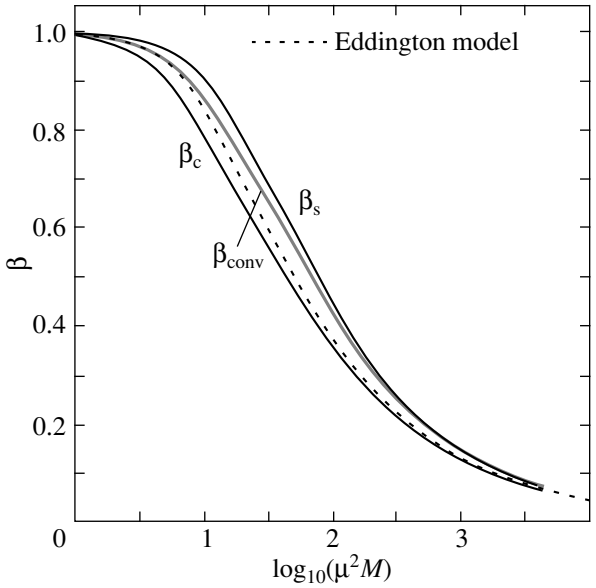


Fig. 3. Ratio of the gas pressure to the total pressure, β, at the center, β_c, the convective core boundary, β_{conv}, and the surface, β_s, versus $\mu^2 M$ (M is in $M_\odot$). The dotted line corresponds to Eddington standard model Eq. (29).

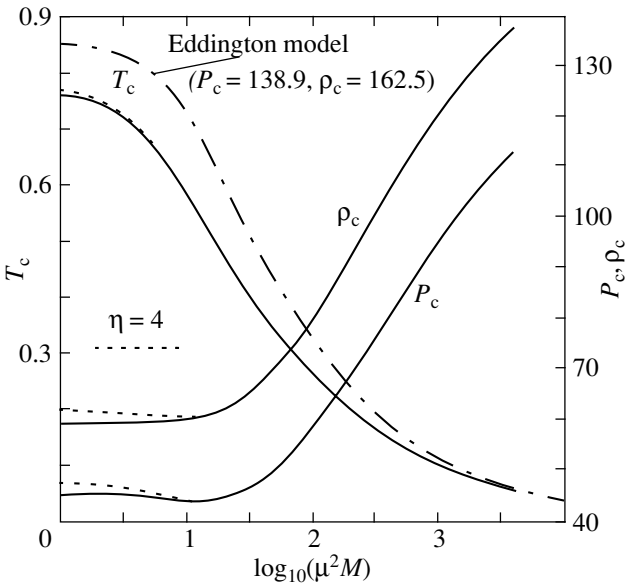


Fig. 4. Dimensionless central pressure P_c, density ρ_c (right vertical axis), and temperature T_c (left vertical axis) vs. $\mu^2 M$ (M is in $M_\odot$).

perature profiles for two extreme values of $\mu^2 M$ for the CNO cycle and the 3α reaction. With a high accuracy, these two energy generation modes lead to the same model for each given $\mu^2 M$ (see the discussion above).

It is instructive to compare our results with the Eddington standard model (Eddington 1926), which corresponds to a constant energy generation rate at constant opacity: $\varepsilon = \varepsilon_0$ ($\eta = 0, \delta = 0$). In this case, it immediately follows from Eq. (26) that $C_1 = 1$. Furthermore, it is easy to make sure that β is constant in this model. As a result, the pressure turns out to be related to the density along the radius by a power law:

$$P = K\rho^{4/3}, \tag{28}$$

$$K \equiv \frac{k}{m_u \mu}\left(\frac{3k}{a m_u \mu}\frac{1-\beta}{\beta^4}\right)^{1/3} = \text{const}.$$

Thus, the standard model is just a polytropic gas sphere with index $n = 3$. The value of β proves to be uniquely related to the mass M by the equation (Chandrasekhar 1939)

$$\frac{1-\beta}{\beta^4} = 0.01607 a G^3 \left(\frac{m_u}{k}\right)^4 (\mu^2 M)^2 \tag{29}$$

$$= 2.994 \times 10^{-3}\left(\mu^2 \frac{M}{M_\odot}\right)^2.$$

The constant C_2 as a function of $\mu^2 M$ and the mass–luminosity relation for the standard model are defined by Eqs. (23) and (24) with β_s substituted for β from Eq. (29). The Eddington standard model has no convective core at all—it is convectively stable for any $\mu^2 M$.

According to Fig. 3, β increases from the stellar center to the surface: $\beta_c < \beta_{conv} < \beta_s$, with β in the Eddington model always remaining between β_c and β_{conv}.

Figure 4 shows the central pressure P_c, density ρ_c, and temperature T_c as functions of $\mu^2 M$. At $\mu^2 M \lesssim 10$, these quantities for the pp chain (dotted lines) differ only slightly from those for the CN cycle and the 3α-reaction (solid lines). The dash–dotted line represents T_c for the Eddington model for which P_c and ρ_c do not depend on $\mu^2 M$: $P_c = 138.9$, $\rho_c = 162.5$ (the $n = 3$ polytrope!).

Figure 5 shows how the dimensionless radius of the convective core r_{conv}, the value of β at the convective core boundary β_{conv}, and the mass above the convective core ($1 - m_{conv}$) depend on $\mu^2 M$. At $\mu^2 M \gtrsim 40$, the convective core contains more than 85% of the stellar mass. We can fit m_{conv} (for all three energy generation modes!) by the following asymptotic relation (the dash–dotted line in Fig. 5):

$$m_{conv} = 1 - 9.075(\mu^2 M)^{-1.095}, \quad \mu^2 M \gtrsim 40. \tag{30}$$

Since the constant C_1 depends strongly on the energy generation mode (Table 1), it is useful to introduce a modified constant C_{1m}:

$$C_{1m}(\mu^2 M) \equiv (\mu^2 M)^{\frac{\delta+\eta}{3\delta+\eta}}[(1-\beta_s)C_1]^{-\frac{1}{3\delta+\eta}}, \tag{31}$$
(M in $M_\odot$).

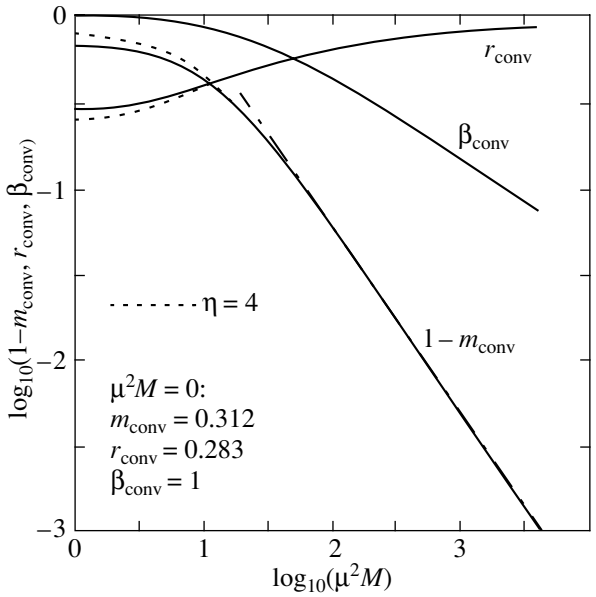


Fig. 5. Parameters of the convective core boundary vs. $\mu^2 M$ for the CNO cycle and the 3α reaction (solid lines) and for the pp chain (dotted lines). The values for the CNO cycle and the 3α reaction at $\mu^2 M = 0$ are shown explicitly.

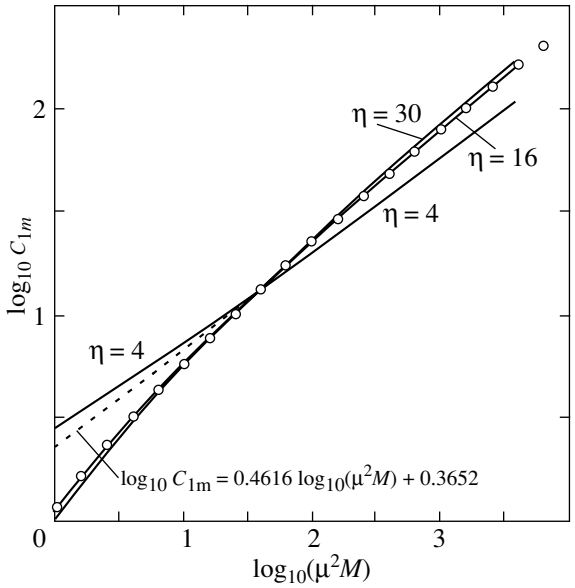


Fig. 6. Modified constant C_{1m} vs. $\mu^2 M$ for the CNO cycle, the 3α reaction, and the pp chain. The open circles and the dashed line indicate polynomial fits.

Using Eq. (24) for L_0 we then obtain the following expression for the stellar radius as a function of M and chemical composition:

$$\log R = \log C_{1m} + \log D + \log S, \quad (R \text{ in } R_\odot), \tag{32}$$

where

$$\log D = \frac{1}{3\delta+\eta}\log[0.2(1+X)\xi] - \frac{2\delta+\eta}{3\delta+\eta}\log\mu,$$

$$\log S = \frac{\eta}{3\delta+\eta}\log\left(\frac{Gm_u}{k}\right) - \frac{\delta+1}{3\delta+\eta}\log(4\pi)$$

$$+ \frac{1}{3\delta+\eta}\log\left(\frac{\varepsilon_{00}}{cG}\right) + \frac{\delta+\eta}{3\delta+\eta}\log M_\odot - \log R_\odot,$$

$$\xi \equiv X^2(\text{pp chain}), \quad XX_{CNO}(\text{CNO cycle}),$$

$$Y^3(3\alpha \text{ reaction}).$$

Here, the dependences of ε_0 on the mass fractions of hydrogen X, the CNO isotopes X_{CNO}, and helium Y are shown explicitly ($\varepsilon_0 = \varepsilon_{00}\xi$). For large values of η, C_{1m} is independent of the energy generation mode (cf. the CNO ($\eta = 16$) and 3α ($\eta = 30$) curves in Fig. 6) and can be fitted by the following polynomial (indicated by the open circles in Fig. 6):

$$y = 0.0720052 + 0.782547x - 0.120854x^2 \tag{33}$$
$$+ 0.0295923x^3 - 0.0030574x^4,$$

where $y \equiv \log_{10} C_{1m}$ and $x \equiv \log_{10}(\mu^2 M)$. A fit for the pp chain ($\eta = 4$) is indicated in Fig. 6 by the dashed line. It gives a good accuracy for $\mu^2 M \gtrsim 10$.

Figure 7 shows the luminosity L_0 in units of L_{Ed} as a function of $\mu^2 M$ (the solid curve described by Eq. (24)). The curve virtually holds for all three energy generation modes—the small difference for the pp chain at $\mu^2 M < 10$ is indistinguishable on the scale of the figure. For each given $\mu^2 M$, the Eddington model is always overluminous (by a factor of about 2 at $\mu^2 M < 10$).

For practical use, it makes sense to rewrite the mass–luminosity relation in terms of the variable λ:

$$\lambda \equiv \mu^2 \frac{M}{M_\odot}\frac{L_0}{L_{Ed}} = \frac{\mu^2 \kappa_0 L_0}{4\pi c G M_\odot} \tag{34}$$
$$= 1.5426 \times 10^{-5}\mu^2(1+X)\frac{L_0}{L_\odot}.$$

Connecting the asymptotics for small and large $\mu^2 M$ by a cubic spline (the open circles in Fig. 8) that ensures the continuity of the function and its first derivatives, we obtain the following analytical fit:

$$\log(\mu^2 M) = 0.9347 + \frac{1}{3}\log\lambda \quad (\log\lambda \leqslant -1.7), \tag{35}$$

$$\log(\mu^2 M) = 1.015209 + 0.449843\log\lambda$$
$$+ 0.0534969\log^2\lambda + 0.00754094\log^3\lambda$$
$$(-1.7 < \log\lambda < 1.7),$$

$$\mu^2 M = 10.125\left(1 + 0.0987654\lambda + \sqrt{1 + 0.1975\lambda}\right)$$
$$(\log\lambda \geqslant 1.7).$$

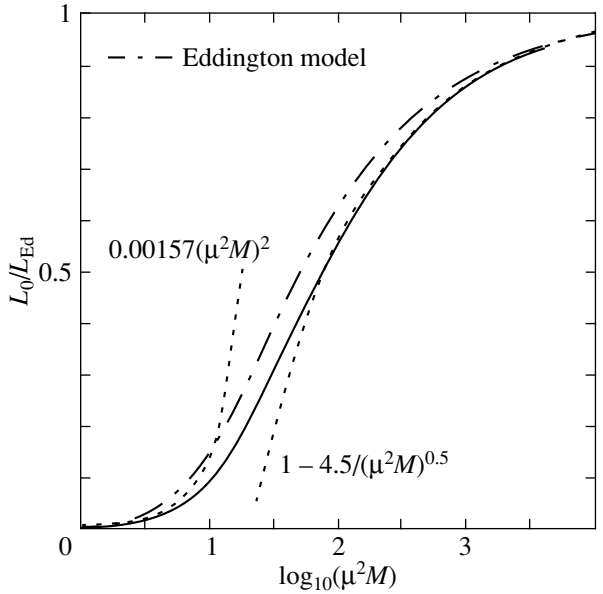


Fig. 7. Normalized mass–luminosity relation Eq. (24). The dotted curves indicate the asymptotics for small and large μ^2M. The Eddington model is represented by the dash–dotted line.

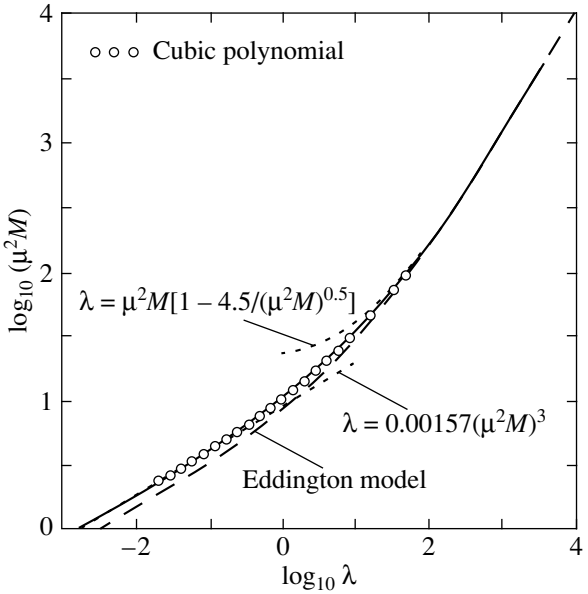


Fig. 8. Mass–luminosity relation (M is in $M_\odot$, see the text).

Equation (35) allows us to estimate M when the luminosity L_0 and the chemical composition components (μ and X) are known. In contrast, to estimate L_0 for given M and chemical composition, we can either solve Eq. (35) for λ or use the following practically accurate fit:

$$\lambda = 0.00157(\mu^2M)^3 \quad (\mu^2M \leqslant 2.4), \tag{36}$$

$$\log\lambda = -2.907029 + 3.552793\log(\mu^2M) - 0.7717945\log^2(\mu^2M) + 0.078623\log^3(\mu^2M) \quad (2.4 < \mu^2M < 100),$$

$$\lambda = \mu^2M\left(1 - \frac{4.5}{\sqrt{\mu^2M}}\right) \quad (\mu^2M \geqslant 100).$$

Concluding this section, we present a number of integrated properties of the stellar models in Table 2: the gravitational, E_g, and thermal, E_T, energies (in units of GM^2/R); the central moments of inertia of the whole star, $I = \int_0^M r^2dm$, and the convective core, I_{conv} (in units of MR^2); the time it takes for sound to propagate from the stellar center to the surface, $t_s = \int_0^R \frac{dr}{\sqrt{\gamma P/\rho}}$, and to the convective core boundary, t_{conv} (in units of $\sqrt{R^3/(GM)}$); the integrated absorption density of the star, $\tau_c = \int_0^R \rho r$ (in units of M/R^2); the mean adiabatic index, $\langle\gamma\rangle = \frac{\int_0^M \gamma(P/\rho)dm}{\int_0^M (P/\rho)dm}$; and an estimate of the fundamental angular frequency $\omega = \sqrt{(3\langle\gamma\rangle - 4)|E_g|/I}$ of radial pulsations (in units of $\sqrt{GM/R^3}$).

The last column in Table 2 presents the properties of the Eddington standard model at $\mu^2M = 4000$. Note that the dimensionless E_g, I, and τ_c for the standard model are determined by the $n = 3$ polytrope structure and do not depend on μ^2M.

Figure 9 shows E_g, E_T, I, and gravitational binding energy $E_b = -E_{tot} = -(E_g + E_T)$ as functions of μ^2M.

COMPARISON WITH DETAILED MODELS

To demonstrate the potentialities of the similarity theory, let us compare our results with the detailed models of massive main-sequence stars calculated by Schaller *et al.* (1992) and the models of helium and carbon–oxygen stars (Wolf–Rayet stars) studied by Langer (1989) and Deinzer and Salpeter (1964). Figures 10–13 display the results of such a comparison. The solid curves in Fig. 10 were obtained using Eq. (36) for the two compositions considered by Schaller *et al.* (1992): $X = 0.680$, $Y = 0.300$, $Z = 0.020$ (upper curve) and $X = 0.756$, $Y = 0.243$, $Z = 0.001$ (lower curve). We see excellent agreement for low metallicities ($Z = 0.001$) and satisfactory agreement for $Z = 0.02$. In the latter case, our models are slightly overluminous (by 25% at $M = 15M_\odot$). This natural result can be attributed to, on average,

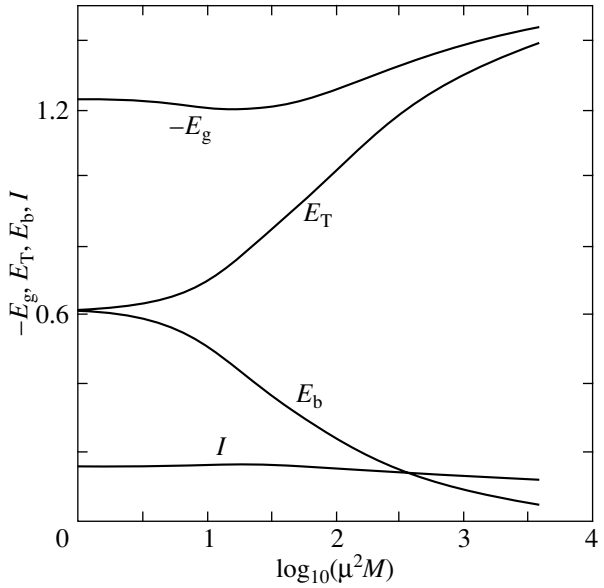


Fig. 9. Dimensionless gravitational, E_g, and thermal, E_T, energies, central moment of inertia, I, and gravitational binding energy, E_b, vs. $\mu^2 M$.

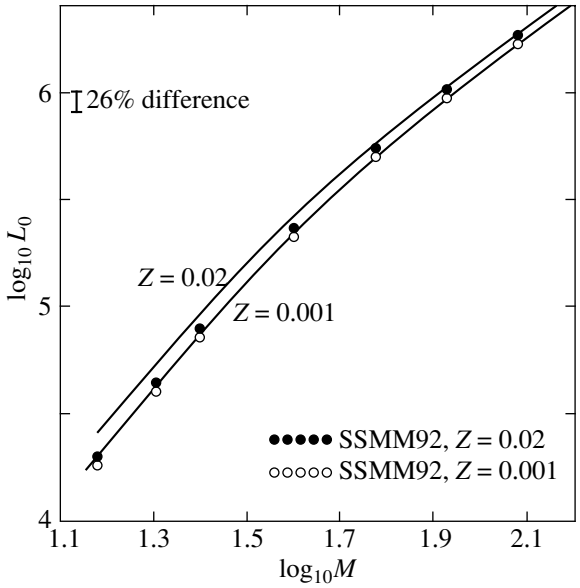


Fig. 10. Mass–luminosity relation (solid lines) in comparison with the detailed models by Schaller *et al.* (1992) (filled and open circles). M and L_0 are in solar units.

an ∼25% contribution from sources other than the electron scattering to the opacity.

The mass–radius relation for the same models is shown in Fig. 11. The detailed models have systematically larger radii (typically by ∼5% for $Z = 0.02$) than our models (solid lines). This can be explained by an appreciable increase in opacity in the stellar envelope due to absorption in atomic spectral lines (Imshennik and Nadyozhin 1967). The solid lines were calculated using Eqs. (32) and (33) by assuming that the CNO cycle is the main energy source. The energy generation law was taken from Caughlan and Fowler (1988) and fitted by

$$\varepsilon_{\rm CNO} = 8.43 \times 10^{-3} X_{\rm CNO} X \rho T_7^{16} \ {\rm erg\ g^{-1}\ s^{-1}} \quad (37)$$
$$(2.2 \leqslant T_7 \leqslant 3.6),$$

where $T_7 \equiv T/10^7$ K and ρ is in g cm^{-3}. Within this range for T_7, the accuracy of fit (37) is better than 10%.

A similar fit for the 3α reaction is given by the following power law:

$$\varepsilon_{3\alpha} = 4.95 \times 10^{-38} Y^3 \rho^2 T_7^{30} \ {\rm erg\ g^{-1}\ s^{-1}} \quad (38)$$
$$(T_7 > 8).$$

The accuracy of the fit is better than 10% at $T_7 > 8$.

The structural properties of our models are also in good agreement with the detailed models, as illustrated by Fig. 12 for the central value of β and the mass fraction of the convective core $m_{\rm conv}$.

Figure 13 shows the mass–luminosity relation for purely helium stars ($Y = 1$) calculated by Deinzer and Salpeter (1964) and Langer (1989) and for carbon–oxygen Wolf–Rayet stars with the mass fractions of carbon, oxygen, and helium equal to $X_{\rm C} = 0.113$, $X_{\rm O} = 0.867$, and $Y = 0.02$, respectively (Langer 1989), in comparison with the mass–luminosity relation for our models defined by Eq. (35) (solid lines). We should set $X = 0$ for the hydrogen mass fraction in Eq. (34) and use the following

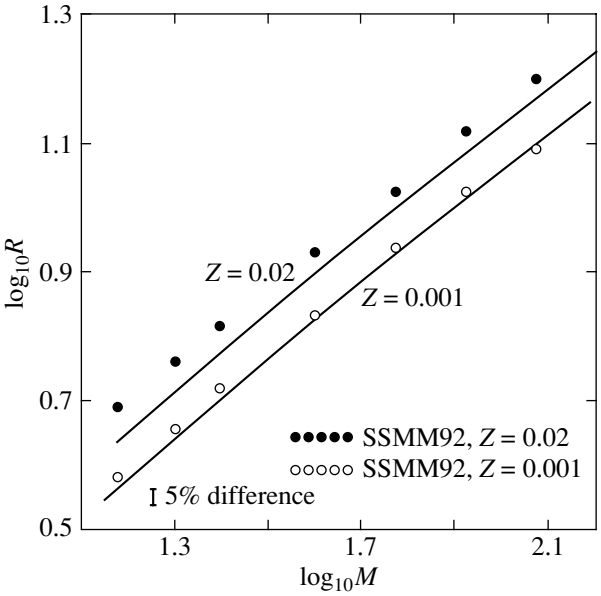


Fig. 11. Mass–radius relation (solid lines) in comparison with the detailed models by Schaller *et al.* (1992) (filled and open circles). M and R are in solar units.

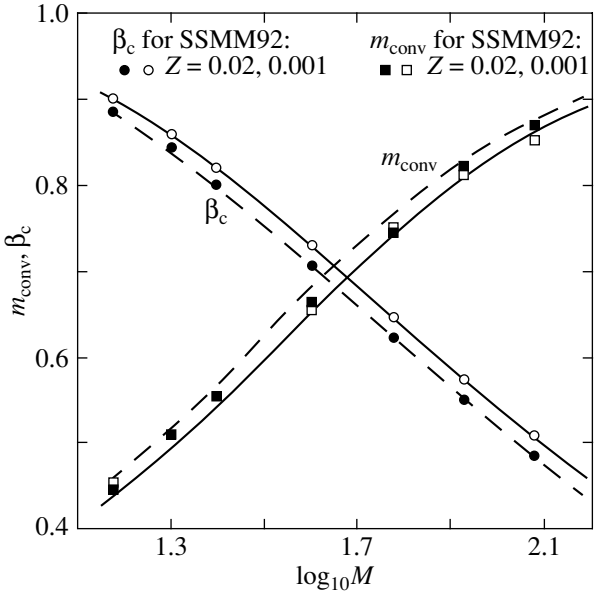


Fig. 12. Dimensionless mass of the convective core $m_{\rm conv}$ and central value of β (solid and dashed lines) in comparison with those from Schaller *et al.* (1992) (filled and open circles and squares). The solid and dashed lines are for $Z = 0.001$ and 0.02, respectively (M is in $M_\odot$).

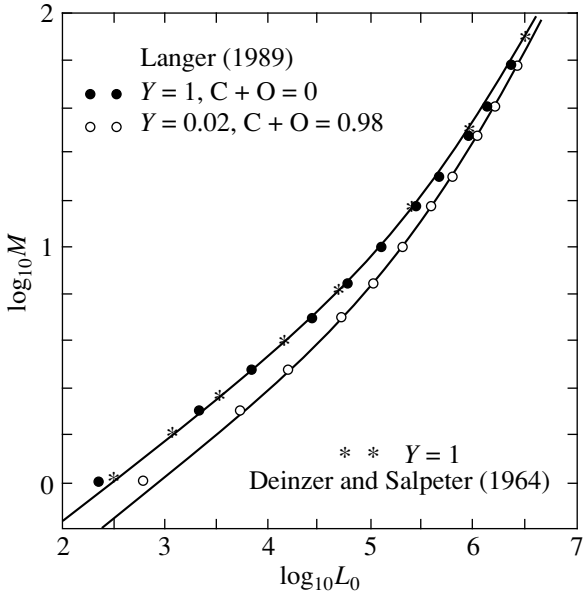


Fig. 13. Mass–luminosity relation for helium and carbon–oxygen Wolf–Rayet stars (solid lines) in comparison with the detailed models by Langer (1989) (filled and open circles) and Deinzer and Salpeter (1964) (asterisks). M and L_0 are in solar units.

expression for the mean molecular mass μ:

$$\mu = \frac{48}{36Y + 28X_{\rm C} + 27X_{\rm O}} \qquad (39)$$
$$(Y + X_{\rm C} + X_{\rm O} = 1).$$

Our results are in excellent agreement with the detailed models even for such a small mass as $1M_\odot$.

DISCUSSION AND CONCLUSIONS

It is interesting to apply the mass–luminosity relation given by Eqs. (35) and (36) to very massive stars observed in our Galaxy, the Magellanic Clouds and a number of nearby resolved galaxies. There is a large body of observational data for several dozen such stars (see, e.g., Figer *et al.* 1998; Puls *et al.* 1996; Humphreys and Davidson 1994; Sandage and Tammann 1974). A detailed analysis of the efficiency of the similarity theory in determining the properties of such stars from observations deserves a special paper. Here, as an example, we consider the Pistol star studied in detail by Figer *et al.* (1998). Assuming that this star initially had solar chemical composition ($X = 0.707$, $Y = 0.274$, $Z = 0.019$; Anders and Grevesse 1989) and the luminosity $L_0 = 10^{6.7\pm0.5}L_\odot$ (Najarro and Figer 1999), we can estimate its initial mass from Eq. (35) to be $M = 116$, 246, and $595M_\odot$ for $L_0 = 10^{6.2}$, $10^{6.7}$, and $10^{7.2}L_\odot$, respectively. The corresponding initial radii derived from Eqs. (32) and (33) are $R = 14$, 22, and $36R_\odot$ for the CNO-cycle energy generation rate given by Eq. (37). The spread in masses $(116–595)M_\odot$ corresponds to the spread (44–230) in parameter $\mu^2 M$, since $\mu = 0.618$ for solar chemical composition. Such stars have luminosities of $(0.4–0.7)L_{\rm Ed}$ (Fig. 7) and very large convective cores: $m_{\rm conv} = 0.86–0.976$ (Eq. (30) or Fig. 5). Estimating the initial mass of the Pistol star, we assumed that its luminosity did not change appreciably during its evolution. The star appears to be in a state close to hydrogen depletion in the convective core; i.e., it is about to leave the main sequence. According to Schaller *et al.* (1992), the increment in luminosity ΔL_0 within the main-sequence strip decreases with increasing mass M. Since $\Delta L_0 \approx 25\%$ for $M = 120M_\odot$ (solar metallicity), the Pistol star could initially be $\sim 10\%$ less massive than the estimates given above.

The properties of the stellar structure obtained here are also in satisfactory agreement with the detailed models of very massive, $(100–250)M_\odot$, initially zero-metallicity Population III stars calculated by the UCSC astrophysical group (Woosley 2005), especially when these stars eventually settle on the main sequence in the state of thermal equilibrium.

The similarity theory for massive stars formulated here yields the following important results.

(1) The simple approximate formula for the mass–luminosity relation given by Eqs. (30), (35), and (36)

is valid for various chemical compositions and is virtually independent of the energy generation mode (the pp-chain, the CNO cycle, or the 3α reaction).

(2) The overall structure of very massive stars described in terms of dimensionless variables (Eq. (11)) depends only on the parameter $\mu^2 M$, where μ is the mean molecular mass. This structure is virtually independent of the energy generation mode, be it the CNO cycle, the 3α reaction, or the pp chain.

(3) Although the stellar structure approaches that of the Eddington standard model with increasing $\mu^2 M$, the convergence proves to be rather slow. There are still noticeable discrepancies in some of the stellar parameters even for $\mu^2 M = 4000$. For instance, the central dimensionless pressure and density are, respectively, about 20 and 15% lower than those for the Eddington model (Fig. 4).

ACKNOWLEDGMENTS

We take the opportunity to thank the Max-Planck-Institut für Astrophysik for financial support and hospitality. The work was supported in part by the Russian Foundation for Basic Research, project no. 04-02-16793-a.

REFERENCES

1. E. Anders and N. Grevesse, Geochim. Cosmochim. Acta **53**, 197 (1989).
2. L. Biermann, Z. Astrophys. **3**, 116 (1931).
3. G. R. Caughlan and W. A. Fowler, At. Data Nucl. Data Tables **40**, 283 (1988).
4. S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (Univ. of Chicago Press, Chicago, 1939; Inostrannaya Literatura, Moscow, 1950).
5. H.-Y. Chiu, *Stellar Physics* (Blaisdell, Massachusetts, Toronto, London, 1968), Vol. 1.
6. T. G. Cowling, Mon. Not. R. Astron. Soc. **94**, 768 (1934).
7. J. P. Cox and R. T. Guili, *Principles of Stellar Structure* (Gordon & Breach, New York, 1968), Vol. 2.
8. W. Deinzer and E. E. Salpeter, Astrophys. J. **140**, 499 (1964).
9. E. A. Dibai ans S. A. Kaplan, *Dimensions and Similarity of Astrophysical Quantities* (Nauka, Moscow, 1976) [in Russian].
10. A. S. Eddington, *The Internal Constitution of the Stars* (Cambridge Univ. Press, Cambridge, 1926).
11. D. F. Figer, F. Najarro, M. Morris, *et al.*, Astrophys. J. **506**, 384 (1998).
12. R. M. Humphreys and K. Davidson, Publ. Astron. Soc. Pac. **106**, 1025 (1994).
13. L. R. Henrich, Astrophys. J. **98**, 192 (1943).
14. V. S. Imshennik and D. K. Nadezhin, Astron. Zh. **44**, 377 (1967) [Sov. Astron. **11**, 297 (1967)].
15. V. S. Imshennik and D. K. Nadezhin, Astron. Zh. **45**, 81 (1968) [Sov. Astron. **12**, 63 (1968)].
16. R. Kippenhahn and A. Weigert, *Stellar Structure and Evolution* (Springer-Verlag, Berlin, Heidelberg, 1990).
17. N. Langer, Astron. Astrophys. **210**, 93 (1989).
18. F. Najarro and D. F. Figer, Astrophys. Space Sci. **263**, 251 (1999).
19. P. Naur and D. E. Osterbrock, Astrophys. J. **117**, 306 (1953).
20. J. Puls, R.-P. Kudritzki, A. Herrero, *et al.*, Astron. Astrophys. **305**, 171 (1996).
21. A. Sandage and G. A. Tammann, Astrophys. J. **191**, 603 (1974).
22. L. I. Sedov, *Similarity and Dimensional Methods in Mechanics* (Nauka, Moscow, 1977; Academic, New York, 1959).
23. B. Strömgren, Handb. Astrophys. **7**, 121 (1936).
24. G. Schaller, D. Schaerer, G. Meynet, and A. Maeder, Astron. Astrophys., Suppl. **96**, 269 (1992).
25. M. Schwarzschild, *Structure and Evolution of Stars* (Inostrannaya Literatura, Moscow, 1961) [in Russian].
26. S. E. Woosley, private communication (2005).

Translated by D. Nadyozhin

Astronomy Letters, Vol. 31, No. 10, 2005, pp. 706–712. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 10, 2005, pp. 793–800.

On the Determination of Meridional Flow on the Sun by the Method of Tracers

S. V. Olemskoy* **and L. L. Kitchatinov**

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

Received March 30, 2005

Abstract—The standard methods for determining the meridional flow on the Sun from the motions of tracers are shown to give an error related to the latitudinal nonuniformity of the tracer distribution. We suggest a simple method for eliminating this error. Using this method to determine the meridional circulation from the motions of sunspots brings the result into agreement with helioseismological data on the meridional flow. The discussed effects can be important for observations of meridional flows on stars.

Key words: *Sun, magnetic fields, global flows.*

INTRODUCTION

In recent years, interest in global meridional flows on the Sun and solar-type stars has increased. This is mainly because it has been realized that meridional circulation plays a crucial role in the dynamics of large-scale magnetic fields. The observed equatorward migration of the sunspot zone may be attributable to the transport of the toroidal magnetic field by a meridional flow in the deep interiors of the Sun (Choudhuri *et al.* 1995; Dikpati and Gilman 2001; Bonanno *et al.* 2002). At the same time, the poleward migration of the poloidal magnetic field may be caused by a meridional flow on the solar surface (Ivanov and Obridko 2002; Obridko and Shelting 2003; Durrant *et al.* 2004). Meridional circulation also plays an important role in forming the inhomogeneous rotation of the Sun (Kitchatinov 2004) and stars (Kitchatinov and Rüdiger 2004). Therefore, observational data on the solar meridional flow are of great importance.

However, data obtained by different methods disagree. Doppler measurements show a poleward flow on the solar surface (Komm *et al.* 1993). Helioseismology confirms that this flow exists up to depths of $\sim$|!12 000 km, but there is also a relatively slow converging flow toward the latitude region with the highest frequency of occurrence of sunspots (Zhao and Kosovichev 2004). At the same time, the motions of sunspots reveal an exactly opposite picture: a meridional flow from the latitudes of maximum sunspot activity (Tuominen 1955, 1966; Tuominen *et al.* 1983; Lustig and Wöhl 1991; Brajsa and Wöhl 2001).

The goal of this paper is to show that the standard methods for determining the meridional flow from tracers, in particular, from the motions of sunspots are most likely subject to a methodological error. As we show in the next section, this error arises from the latitudinal nonuniformity of the tracer distribution. We suggest a simple method for eliminating it. Using this method can bring the meridional circulation determined from tracers into agreement with helioseismological data. This assertion is demonstrated for the meridional flow determined from Greenwich sunspot data.

DETERMINING THE MERIDIONAL FLOW FROM THE MOTIONS OF TRACERS

The content of this section applies to traces of an arbitrary nature, but, for definiteness, we will consider sunspots.

The displacements of sunspots in a finite time are used to determine the global flows. The corresponding mean velocity, i.e., the ratio of the displacement to the time, is averaged over an ensemble of tracers. The averaging is necessitated by the existence of a random component in the motions of sunspots (Vitinskiĭ *et al.* 1986). The random motions, along with the latitudinal nonuniformity of the sunspot distribution, lead to the error in determining the meridional flow mentioned in the Introduction.

Clearly, when the displacements are estimated, the final positions are determined from older sunspots than the initial positions (the displacement time can range from several days to several tens of days for recurrent sunspots). Since the latitudinal sunspot

*E-mail: olemskoy@list.ru

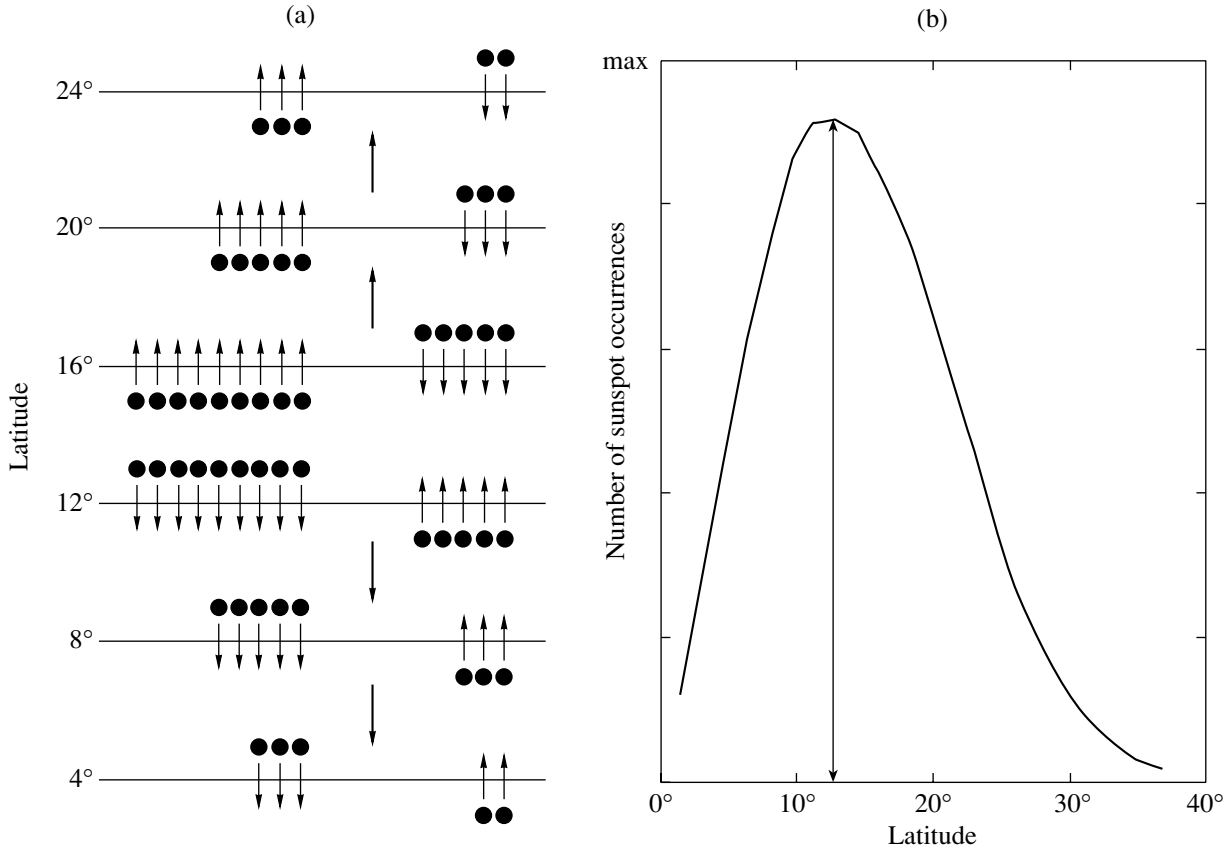


Fig. 1. (a) Influence of boundary effects on the determination of the net meridional motions. The dots denote the sunspots at the boundaries of the latitude zones, and the arrows indicate the directions of their motions. The heavy arrows indicate the predominant motion as a result of the boundary effects. (b) The characteristic nonuniform heliolatitudinal distribution of sunspots.

distribution generally has a maximum at a certain latitude λ^*, the frequency of occurrence of sunspots decreases from λ^* both poleward and equatorward. This distribution would be diffusively smoothed due to the random motions of sunspots; i.e., the distribution has a less distinct maximum and broader wings for older sunspots. If we disregarded this factor, but assumed that all of the changes in the latitudinal sunspot distribution are attributable solely to a regular meridional flow, then a false meridional flow from the latitude λ^* would be detected. Such a result was obtained by Richardson and Schwarzschild (1953) and Tuominen (1955, 1961); the latitude from which the flow originates follows λ^* as λ^* changes in the solar cycle (Tuominen *et al.* 1983; Lustig and Wöhl 1991; Brajsa and Wöhl 2001).

The false meridional flow velocity is estimated to be

$$V \sim \eta_{\mathrm{T}}/L \sim 1\text{–}10\ \mathrm{m\ s^{-1}}, \qquad (1)$$

where $\eta_{\mathrm{T}} \sim 10^9\ \mathrm{m^2\ s^{-1}}$ is the turbulent diffusion coefficient, $L = R_\odot \delta\lambda$ is the scale length of the latitudinal nonuniformity in the sunspot distribution, and $\delta\lambda$ is its angular scale in radians. The meridional flow velocity determined by the method of tracers from the motions of sunspots agrees with estimate (1).

Let us now consider in more detail how the false velocity appears in the meridional circulation determined by the method of tracers and how this can be avoided. As was noted above, the velocity for an individual sunspot is defined as

$$V = R_\odot \frac{\lambda(t_1) - \lambda(t_2)}{t_2 - t_1}, \qquad (2)$$

where $\lambda(t_2)$ is the latitude of the displacement under consideration at the final time t_2 and $\lambda(t_1)$ is the corresponding latitude at the initial time t_1. Velocity (2) is then averaged over an ensemble of tracers. To be more precise, the complete latitude range is broken down into a number of finite intervals and the averaging is performed over the sunspots that belong to individual latitude intervals. In this way, the meridional flow velocity is found as a function of the latitude.

However, the following question arises: to which latitude should velocity (2) be attributed? Clearly, the answer to this question can be of importance only if

Distribution of solar cycles over four phases

Cycle number	Cycle phases											
	1			2			3			4		
11										1875	1876	1877
12	1878	1879	1880	1881	1882	**1883**	1884		1885	1886	1887	1888
13	1889	1890	1891	1892	**1893**	1894	1895	1896	1897	1898	1899	1900
14	1901	1902	1903	1904	**1905**	1906	1907	1908	1909	1910	1911	1912
15	1913	1914	1915	1916		**1917**	1918		1919	1920	1921	1922
16	1923	1924	1925	1926		1927	**1928**		1929	1930	1931	1932
17	1933	1934	1935	1936		**1937**	1938	1939	1940	1941	1942	1943
18	1944	1945	1946									
11–18	2027			4751			4230			1895		
Odd	834			2758			2383			1117		
Even	1193			1993			1847			778		

Note. The years of solar maximum are highlighted in boldface. The first year in each cycle is the year of solar minimum. The lower part of the table gives the number of sunspot occurrences.

$\lambda(t_1)$ and $\lambda(t_2)$ belong to *different* latitude intervals, i.e., when a tracer crosses the boundary between the two chosen latitude intervals during its displacement in the time $t_2 - t_1$. The standard procedure implies that velocity (2) should be attributed to the mean value,

$$\overline{\lambda} = (\lambda(t_1) + \lambda(t_2))/2. \quad (3)$$

As we see from Fig. 1, the false circulation emerges precisely in this case. Clearly, the number of tracers entering one of the chosen latitude intervals through the boundary toward which the tracer density gradient is directed will be, on average, larger than the number of tracers entering through the other boundary. Therefore, there will be a false flow in the direction opposite to the tracer density gradient. The coherence time of the random motions of sunspots is of the order of the solar rotation time (Vitinskiĭ *et al.* 1986). This time, along with the random motions themselves, is most likely determined by the large-scale solar convection. Therefore, the motions remain almost coherent over the time $t_2 - t_1$, and the discussed boundary effects will manifest themselves in full measure. Note that the length of the chosen latitude intervals does not affect the false flow velocity (see estimate (1)). As the length of the intervals increases, the fraction of the tracers crossing its boundaries decreases, but at the same time, the nonuniformity in the tracer distribution on the scale of these intervals and the difference in the numbers of tracers entering through different boundaries increase.

However, there is a simple way to avoid the boundary effects and the related false circulation: velocity (2) should be attributed to the initial latitude λ_1. The false velocity (1) will then not emerge due to the diffusion through random motions, and the entry of tracers from the adjacent latitude intervals will be unimportant.

THE METHOD

In this paper, we used Greenwich data on the heliographic coordinates of nonrecurrent (short-lived) sunspots for the period 1875–1946. Following Tuominen *et al.* (1983), we distributed the years over four phases of the 11-year solar cycle, as shown in the table. The lower part of this table gives the number of sunspot occurrences in each phase for all cycles (11–18) and individually for the even and odd solar cycles.

To eliminate errors related to the decrease in the visibility function toward the solar limb, only the sunspots near the central meridian were considered in a number of papers. In this case, the limitation criteria are different: from $\pm 0.17\ R_\odot$ (Tuominen 1982, 1983) to $\pm 0.85\ R_\odot$ (Ward 1965, 1973; Hansmeier and Lustig 1986; Kambry *et al.* 1991). Here, we adopted the latter criterion and used the sunspot groups no farther than $\pm 76.5°$ from the central meridian. The reason is that we are dealing with nonrecurrent sunspots whose heliocoordinates are measured with a higher accuracy due to their smaller sizes. We also excluded the sunspots above the 40° parallel in both the northern and southern hemispheres (46

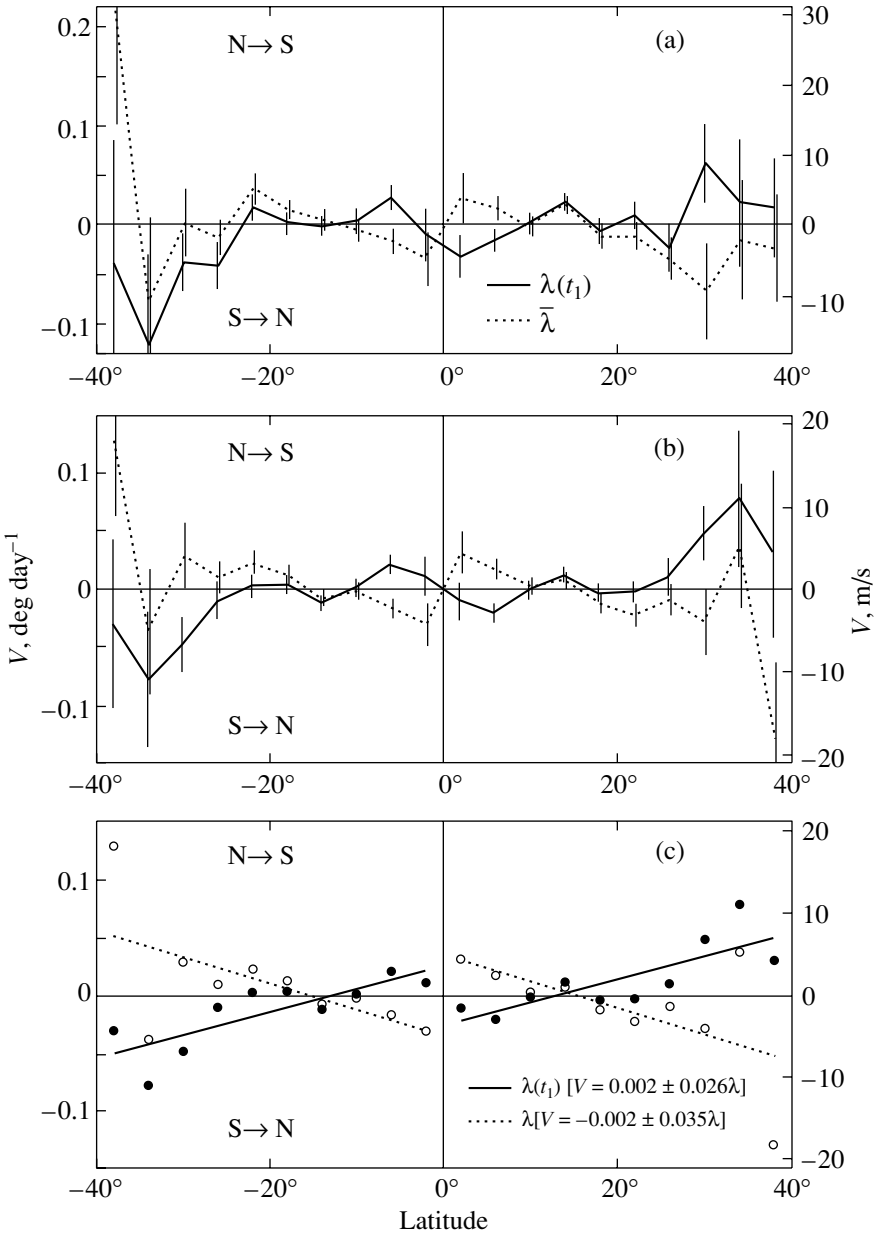


Fig. 2. Meridional motions of sunspots for solar cycles 11–18 (a) without and (b) with averaging over the hemispheres; (c) a linear fit to the data, the fitting equations for each type of lines are given in brackets. The solid and dotted lines represent the results of our calculations in which the sunspot displacements were attributed to their initial latitudes (the nonuniformity effect is eliminated) and to the mean sunspot latitude, respectively. The vertical bars indicate the standard error; N → S and S → N show that the motion is directed toward the south and north poles in the ranges of positive and negative values, respectively.

occurrences). Thus, the total number of sunspots after the preliminary data processing was 12 903.

The meridional displacement velocity of sunspot groups was determined using Eq. (2). Positive and negative velocities mean the motions toward the south and north poles, respectively. According to the standard technique, the velocity obtained is attributed to the mean latitude (3), which we did in the former case. In the latter case, to eliminate the sunspot distribution nonuniformity effect described above, we attribute the velocity obtained not to $\overline{\lambda}$, but to $\lambda(t_1)$, the latitude at the time of sunspot occurrence.

Subsequently, we divided the latitude zone of sunspot occurrence (0°–40°) into ten 4° intervals and found the arithmetic mean velocity for each interval. To estimate the confidence intervals of the mean values, we calculated their standard deviations. The calculations were performed separately for the northern and southern hemispheres as well as for the even and odd solar cycles.

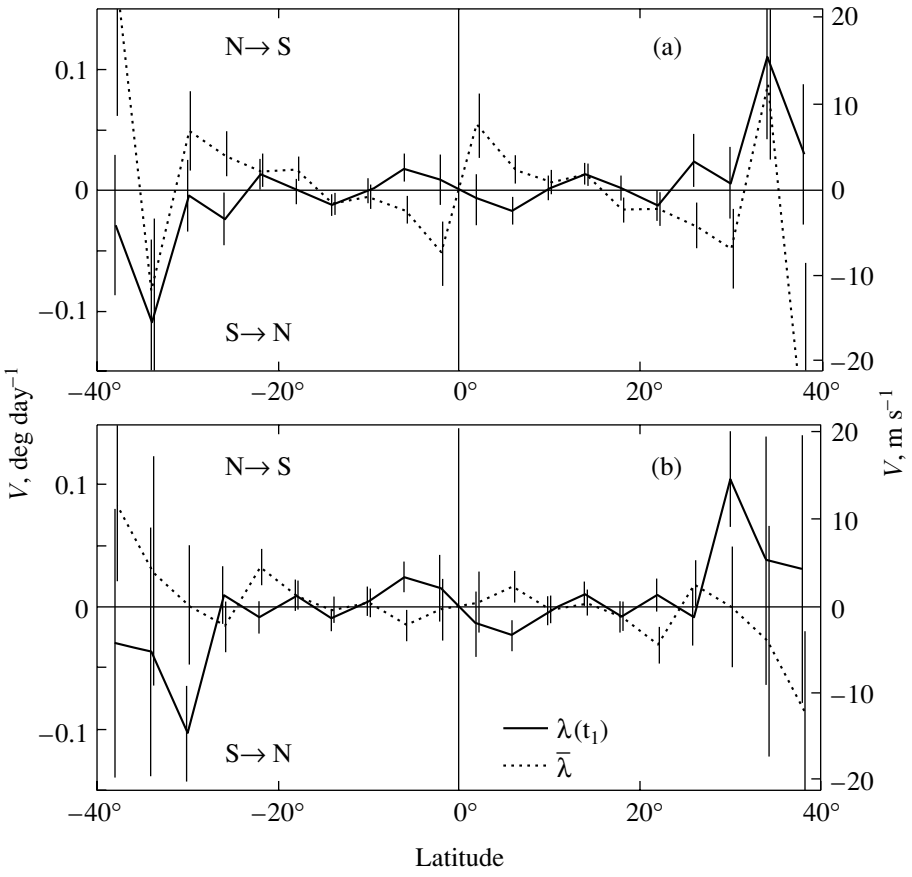


Fig. 3. Meridional flow found by the motions of sunspots averaged over the hemispheres for the (a) odd and (b) even solar cycles. The notation is the same as that in Fig. 2.

RESULTS AND DISCUSSION

In Fig. 2, the mean meridional motion of nonrecurrent sunspot groups is plotted against latitude without separation into even and odd solar cycles. The solid and dotted lines indicate the mean meridional flow velocity for the case where the meridional sunspot displacements were attributed to the initial, $\lambda(t_1)$, and mean, $\overline{\lambda}$, latitudes, respectively. We see that the sunspot motions are almost antisymmetric about the equator. Figure 2b and the figures that follow show the results averaged over the hemispheres in such a way that

$$|V(-\lambda)| = |V(\lambda)|.$$

The velocities are highest in the latitude range $35°-40°$, but these results are less reliable due to the small number of events, $\sim$10 (the error exceeds ± 0.05 deg day^{-1}), while the number of events for low latitudes is more than a thousand.

The meridional motion calculated using the standard technique is in good agreement with the results obtained by Tuominen (1955, 1961, 1982, 1983) for both recurrent and short-lived sunspot groups and with more recent results (Lustig *et al.* 1987, 1991; Balthasar and Fangmeier 1988; Kambry *et al.* 1991; Brajsa and Wöhl 2001). In this case, the flow found is the equatorward and poleward meridional flow from the latitudes $\pm 15°-20°$.

When the sunspot displacements are attributed to their initial latitudes, which is believed to eliminate the nonuniformity effect of their latitudinal distribution, the reverse is true. This result is found both for the complete sunspot statistics (Fig. 2) and for the even and odd solar cycles individually (Fig. 3). In this case, the mean meridional flow is directed toward the mid-latitudes of the sunspot zone. For greater clarity, Fig. 2c shows a linear fit to the data obtained. These data agree with the helioseismological data on the meridional flow in the upper (to a depth of $\sim$12 000 km) convection zone of the Sun (Zhao and Kosovichev 2004). Helioseismology reveals a poleward meridional flow with a velocity of $\sim$20 m s^{-1} (0.14 deg day^{-1}) and a flow toward the latitudes of the highest magnetic activity with a velocity of 2–8 m s^{-1} superimposed on it. Only the latter converging flow may be preserved at the depth beneath the solar surface where the sunspots are "anchored."

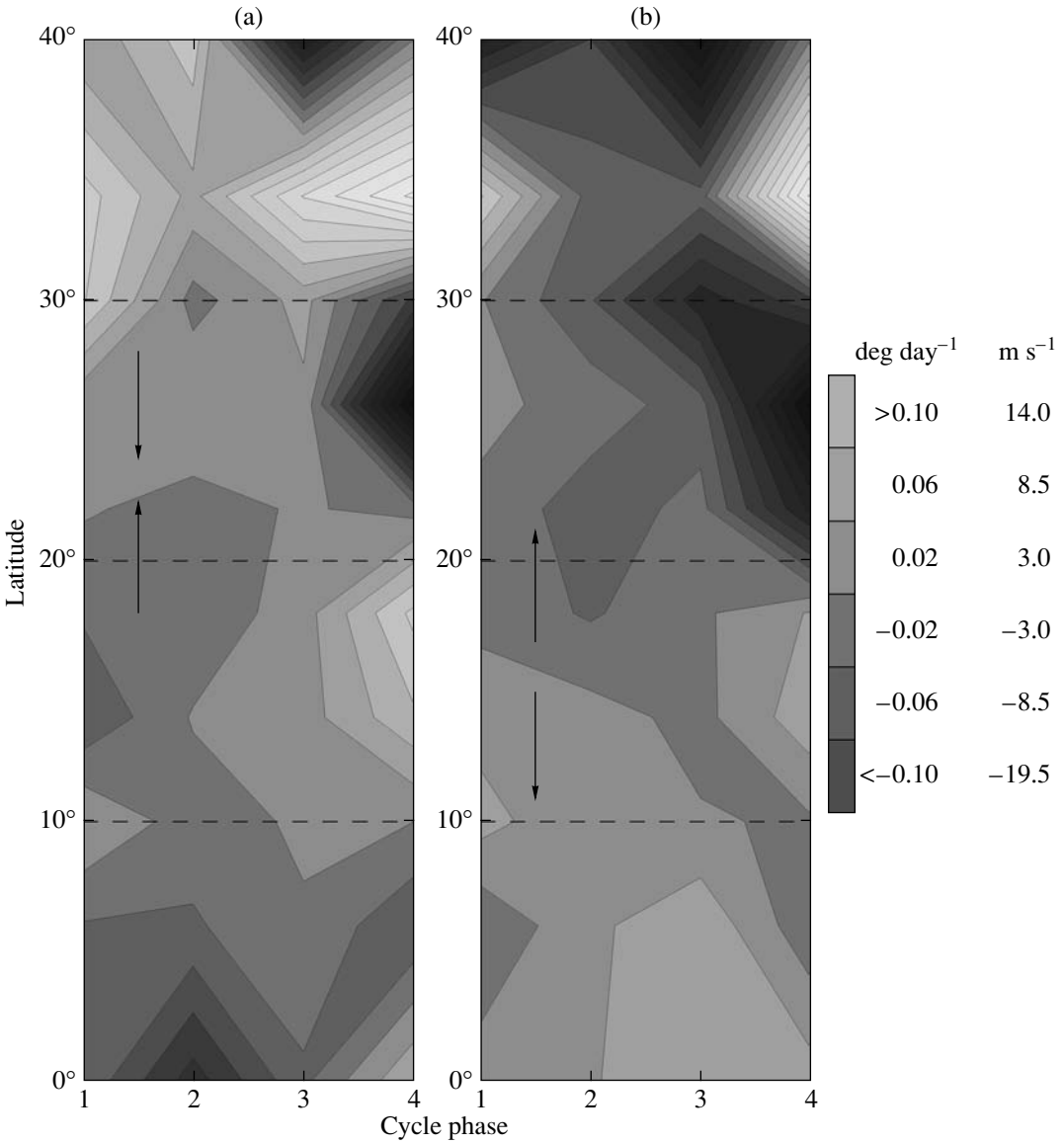


Fig. 4. Meridional flow detected by the motions of sunspots as a function of the phase of the 11-year solar cycle: (a) the results obtained when the sunspot displacements are attributed to their initial latitudes (the nonuniformity effect is eliminated) and (b) the results for the displacements attributed to the mean latitude (3). The arrows indicate the direction of the net meridional motion.

Studies of the latitudinal drift of large-scale magnetic fields (Ivanov and Obridko 2002; Ivanov *et al.* 2002; Obridko and Shelting 2003) reveal a strong latitude dependence of the drift velocity. The field migration from the equator to latitudes of 15°–20° is rather fast, but the drift velocity decreases sharply in the latitude range 20°–40°. At even higher latitudes, the drift velocity increases again. Assuming that both the large-scale and local concentrated magnetic fields are formed through a common generation process (Makarov and Tlatov 2000; Makarov *et al.* 2001), the meridional drift of the sunspot groups in which the errors due to the latitudinal nonuniformity of the sunspot formation are eliminated can be said to agree to some extent with the latitudinal drift of the large-scale magnetic field. The reversal of the direction of motion (Figs. 2b and 2c) in the latitude range 15°–40° may be reflected in a sharp decrease in the meridional drift velocity of the large-scale magnetic field in this latitude range.

At the same time, using large-scale magnetic field structures as tracers to determine the meridional circulation without eliminating the errors due to the latitudinal nonuniformity of these tracers again leads to a picture with poleward and equatorward meridional flow from a certain mean latitude (Latushko 1996).

The effects of the latitudinal nonuniformity in the distribution of tracers are probably the same irrespective of the nature of the tracers. It is hoped that the corresponding errors for tracers of any type can be eliminated by the simple method suggested in this paper. To all appearances, these factors should be taken into account in the planned measurements of the meridional circulation on stars by the Doppler imaging method (Mackay *et al*. 2004), which will eventually use the method of tracers. Note that the meridional flows on stars are much faster than those on the Sun (Mackay *et al*. 2004) and can play a crucial role in the evolution of magnetic fields and the formation of global flows (Kitchatinov and Rüdiger 2004).

An examination of the results obtained as a function of the phase of the 11-year cycle (Fig. 4) shows that the differences in the meridional circulation noted above are traceable in each phase of the cycle. The same was also observed when examining the even and odd solar cycles individually, but we failed to find any pattern related to the 22-year cycle. Helioseismological studies also revealed a displacement of the latitude to which matter flows equatorward following the migration of sunspot activity (Zhao and Kosovichev 2004).

REFERENCES

1. H. Balthasar and E. Fangmeier, Astron. Astrophys. **203**, 381 (1988).
2. A. Bonanno, D. Elstner, G. Rüdiger, and G. Belvedere, Astron. Astrophys. **390**, 673 (2002).
3. R. Brajsa and H. Wöhl, Harv. Obs. Bull. **24**, 125 (2001).
4. A. R. Choudhuri, M. Shussler, and M. Dikpati, Astron. Astrophys. **303**, L29 (1995).
5. M. Dikpati and P. A. Gilman, Astrophys. J. **559**, 428 (2001).
6. C. J. Durrant, J. P. R. Turner, and P. R. Wilson, Sol. Phys. **222**, 345 (2004).
7. A. Hansmeier and L. Lustig, Astron. Astrophys. **154**, 227 (1986).
8. E. V. Ivanov and V. N. Obridko, Sol. Phys. **206**, 1 (2002).
9. E. V. Ivanov, V. N. Obridko, and B. D. Shelting, in *Solar Variability: From Core to Outer Frontiers. 10th European Solar Physics Meeting*, Ed. by A. Wilson (ESA Publ. Divis., Noordwijk, 2002), Vol. 2, p. 851.
10. M. A. Kambry, J. Nishikawa, T. Sakurai, *et al*., Sol. Phys. **132**, 41 (1991).
11. L. L. Kichatinov, Astron. Zh. **81**, 176 (2004) [Astron. Rep. **48**, 153 (2004)].
12. L. L. Kitchatinov and G. Rüdiger, Astron. Nachr. **325**, 496 (2004).
13. R. W. Komm, R. F. Howard, and J. W. Harvey, Sol. Phys. **147**, 207 (1993).
14. S. Latushko, Sol. Phys. **163**, 241 (1996).
15. L. Lustig and H. Wöhl, Astron. Astrophys. **249**, 528 (1991).
16. L. Lustig and A. Hanslmeier, Astron. Astrophys. **172**, 332 (1987).
17. V. I. Makarov and A. G. Tlatov, Astron. Zh. **77**, 858 (2000) [Astron. Rep. **44**, 759 (2000)].
18. V. I. Makarov, A. G. Tlatov, and K. P. Sivaraman, Sol. Phys. **202**, 11 (2001).
19. D. H. Mackay, M. Jardine, A. C. Cameron, *et al*., Mon. Not. R. Astron. Soc. **355**, 1066 (2004).
20. V. N. Obridko and B. D. Shel'ting, Astron. Zh. **80**, 364 (2003) [Astron. Rep. **47**, 333 (2003)].
21. R. S. Richardson and M. Schwarzschild, Acad. Lincei Conv. **11**, 228 (1953).
22. J. Tuominen, Z. Astrophys. **37**, 145 (1955).
23. J. Tuominen, Z. Astrophys. **51**, 91 (1961).
24. J. Tuominen, Astrophys. J. **143**, 266 (1966).
25. J. Tuominen and J. Kyrölainen, Sol. Phys. **79**, 161 (1982).
26. J. Tuominen, I. Tuominen, and J. Kyrölainen, Mon. Not. R. Astron. Soc. **205**, 691 (1983).
27. Yu. I. Vitinskiĭ, M. Kopetskiĭ, and G. V. Kuklin, *Statistics of Sunspot Activity* (Nauka, Moscow, 1986) [in Russian].
28. F. Ward, Astrophys. J. **141**, 534 (1965).
29. F. Ward, Sol. Phys. **30**, 527 (1973).
30. J. Zhao and A. Kosovichev, Astrophys. J. **603**, 776 (2004).

Translated by V. Astakhov