Astronomy Letters, Vol. 28, No. 9, 2002, pp. 579–588. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 643–652.

A Kinematic Study of M51-Type Galaxies

S. A. Klimanov[1]*, V. P. Reshetnikov[1], and A. N. Burenkov[2]

[1] *Astronomical Institute, St. Petersburg State University, Bibliotechnaya pl. 2, Petrodvorets, 198904 Russia*

[2] *Special Astrophysical Observatory, Russian Academy of Sciences, Nizhni Arkhyz, Stavropol krai, 357147 Russia*

Received March 27, 2002

Abstract—Spectroscopic observations of twelve M51-type binary galaxies with the 6-m Special Astrophysical Observatory telescope are presented. We constructed the rotation curves for the primary galaxies of each binary system and determined the line-of-sight velocities of their companions from the Hα and [N II] 6583 Å emission lines. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *galaxies, galaxy groups and clusters, intergalactic gas*

INTRODUCTION

Binary systems similar to the M51 galaxy constitute rather rare and very interesting types of objects. Vorontsov-Vel'yaminov (1957) and Arp (1966) were the first to point out M51 type systems as a special class of binary galaxies. These systems consist of a spiral primary and a relatively small companion that is seen projected onto the tip of one of the spiral arms of the primary.

Klimanov and Reshetnikov (2001) presented a new sample of M51-type galaxies. Objects of this type are very poorly studied and for many of them neither kinematic data for the primary galaxy nor even the line-of-sight velocities of their companions are available. The principal aim of this work was to obtain observational data about M51-type galaxies. We will use these data to study the origin and evolution of binary galaxies with very different component masses.

OBSERVATIONS

Our program list included 12 binary systems drawn from the sample described by Klimanov and Reshetnikov (2001) (see the table for the list of objects).

Observations were made in February, March, and November 2000 with a high focal ratio spectrograph operating in the primary focus of the 6-m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences (Afanas'ev *et al.* 1995). The observations used a 1024×1024 Photometrics CCD with a pixel size of 24 μm attached to the Schmidt–Cassegrain camera of the spectrograph ($F = 150$ mm). The spectra were obtained using a spectrograph with a $2'' \times 120''$ slit and a 1302 lines/mm grating operating in the wavelength interval 6000–7200 Å. The scale along the slit was $0''.40$ per pixel and the reciprocal dispersion was equal to 1.21 Å pixel. The seeing during observations varied from $1''$ to $2''$. For each object we obtained several spectra with a characteristic exposure time of 10–20 min. ArNeHe comparison spectra were recorded between the spectra of program objects. The position angle of the spectrograph slit is given in the second column of the table.

We reduced the spectra using an ESO-MIDAS software package (LONG context). We constructed the rotation curves of galaxies from Hα (6562.8 Å) and [NII] (6583.4 Å) emission lines whose contours we fitted to Gaussian profiles. We did not measure the positions of other lines, because of their weakness. We estimated the accuracy of our reduction from the measurements of night-sky lines. The standard deviation of the wavelengths of the latter from the mean position never exceeded 0.33 Å which corresponds to $\sigma \leq 15$ km s^{-1} at the Hα wavelength. We estimated the internal errors of measurements from the dispersion of line-of-sight velocities measured at the center of the galaxy and in the neighboring pixels. Such a procedure is justified by the small separation between neighboring pixels ($0''.40$), which is much smaller than the seeing. One must bear in mind, however, that the gradient of the rotation curve may increase the internal error estimate thus obtained and, consequently, our estimates characterize only the upper boundary of the real error. Internal errors given

*E-mail: serg@gong.astro.spbu.ru

1063-7737/02/2809-0579$22.00

Parameters of galaxies

Name	*P.A*	$V_{\rm din}$, km s^{-1}	$V_{\rm phot}$, km s^{-1}	ΔV, km s^{-1}	R_V/R_{25}
MCG -01-01-70 a	175°	5991 ± 11	5971 ± 13	-76 ± 16	0.62
MCG -01-01-70 b	175	6067 ± 11	6071 ± 3	–	0.3
NGC 151 a	68	3732 ± 21	3745 ± 28	-277 ± 21	0.64
NGC 151 b	68	4009 ± 4	4008 ± 4	–	0.8
NGC 797 a	66	5664 ± 18	5662 ± 18	92 ± 39	0.71
NGC 797 b	66	–	5572 ± 35	–	0.4
NGC 2535	59	4079 ± 7	4082 ± 5	1 ± 21	0.47
NGC 2536	59	4078 ± 20	4071 ± 20	–	0.6
NGC 2535	163	4090 ± 8	4084 ± 10	15 ± 22	0.32
NGC 2536	163	4075 ± 20	4077 ± 20	–	0.4
UGC 6865 a	37	5930 ± 29	5882 ± 27	162 ± 29	0.94
UGC 6865 b	37	5768 ± 3	5768 ± 2	–	0.6
NGC 4088 a	55	775 ± 12	769 ± 9	159 ± 12	0.56
NGC 4088 b	55	616 ± 3	621 ± 3	–	0.9
NGC 4137 a	90	11171 ± 3	11172 ± 4	99 ± 8	1.06
NGC 4137 b	90	–	11072 ± 7	–	1.0
NGC 5278	68	7629 ± 17	7527 ± 12	58 ± 17	0.55
NGC 5279	68	7571 ± 4	7552 ± 4	–	0.6
UGC 10396 a	19	8680 ± 16	8652 ± 15	-154 ± 21	0.68
UGC 10396 b	19	8834 ± 14	8866 ± 3	–	0.7
UGC 11680 a	71	7806 ± 4	7807 ± 4	-89 ± 7	0.75
UGC 11680 b	71	7895 ± 6	7899 ± 5	–	0.5
NGC 7753	44	5142 ± 30	5112 ± 32	202 ± 34	0.52
NGC 7752	113	4940 ± 16	4940 ± 16	–	0.6
NGC 7757 a	110	2944 ± 12	2974 ± 5	-109 ± 30	0.76
NGC 7757 b	110	–	3053 ± 27	–	0.9

in the table do not exceed 35 km s^{-1} (their mean is 13 ± 9 km s^{-1}).

We used the rotation curves obtained as described above from Hα and [NII] lines to determine the positions of the centers of both components and their radial velocities. We computed the positions and velocities as the weighted means of the corresponding values inferred from Hα and [NII] lines with weights equal to the line intensities. To reveal an eventual relation between the dynamical and photometric parameters of the systems, we determined two centers—dynamical and photometric—for each galaxy. We found the position and velocity of the former from the condition of maximum symmetry of the rotation curve. The photometric center coincided with the maximum of the continuum spectrum. Note that the dynamical centers were determined only formally for a number of companion galaxies due to their small angular sizes and, consequently, small extent of their rotation curves.

RESULTS AND DISCUSSION

The figure shows the images of 12 M51-type systems taken from the DSS[1] with slit positions superimposed, and the rotation curves of these galaxies. North is at the top and West, on the right. The distances and velocities are given relative to the dynamical center of the primary galaxy.

[1]The Digitized Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166.

The main results are summarized in the columns of the table:

(1) Number of the galaxy according to NGC, UGC, or MCG catalogs. In the cases where the companion has no designation in the above catalogs the components of the system are indicated by letters a and b corresponding to the primary and its companion, respectively;

(2) Position angle of the slit during observations;

(3) Line-of-sight velocity of the dynamical center of the galaxy corrected for the Earth's orbital motion;

(4) Line-of-sight velocity of the photometric center of the galaxy corrected for the Earth's orbital motion;

(5) Difference of the radial velocities of the dynamical centers of the primary and its companion. In the cases where the small size of the companion made the determination of its dynamical center uncertain we used the velocity of its photometric center;

(6) Extent of the measured portion of the rotation curve expressed as a fraction of the radius of the galaxy (R_{25}) either reduced to the standard isophote of 25^m/sq. qrcsec (according to the RC3 catalog) or inferred from our measurements of DSS galaxy images.

The mean extent of our rotation curves is equal to $(0.66 \pm 0.20)R_{25}$ and $(0.64 \pm 0.22)R_{25}$ for the primary galaxies and their companions, respectively. Thus our rotation curves reach the region of maximum rotation velocity for the exponential disk of a typical bright spiral galaxy ($\sim 2h$, where h is the exponential disk scale length, or $\sim 2/3R_{25}$), and they can therefore be used for analyzing the global kinematics of these objects.

In this paper we briefly describe the main specific features of the spectra of the observed galaxies and their rotation curves, and compare our results with those published earlier. We adopted most of our comparison data from the NED database[2] where one can find the bibliographic references. Below we give references only to recent publications and to the detailed studies of objects. Hereafter all radial velocities are heliocentric line-of-sight velocities.

MCG-01-01-70. During observations the spectrograph slit passed almost along the minor axis of the galaxy. The Northern and Southern branches of the rotation curve exhibit local maxima at a galactocentric distance of 5″ where radial velocity of rotation reaches 50–60 km s^{-1}. So far, no rotation curves of this galaxy has been published. The radial velocity of the center of the primary—5985 ± 38 km s^{-1}—found by Huchra *et al.* (1993) agrees well with our result (see the table).

NGC 151. The slit passed through the major axis of the primary. The rotation curve is characterized by a nonmonotonic variation of velocity, which is most conspicuous on the companion side. The Hα line is almost invisible in the spectrum of the companion—its central intensity is lower than that of the [NII] line by a factor of 30. The Hα rotation curve of the primary was obtained by Mathewson and Ford (1996) (at a slit position angle of 75°). It reaches a plateau and exhibits smaller variations of rotation velocity. The velocity of the center of the primary computed by Mathewson and Ford—3738 ± 10 km s^{-1}—agrees well with our result.

NGC 2535–2536. We observed this system at two slit position angles and obtained separate spectra of the primary and companion (see the figure). When aligned along the position angle $P.A. = 59°$, the slit crosses the arms of the primary galaxy and the spectrum exhibits a characteristic clumpy pattern with several condensations in Hα and [NII] lines, which are likely to be associated with ~100 km s^{-1}-amplitude waves on the rotation curve. The rotation curve of the companion also exhibits nonmonotonic variations of V_r with a slightly smaller amplitude. The spectrum taken at the slit position $P.A. = 163°$ shows much fewer bright condesations. The velocity variations in this case have much smaller amplitude. At the same time, this rotation curve is not quite symmetric: its southeastern part at a distance of 8″–10″ from the center of the primary reaches a plateau (the maximum velocity V_r reaches ~100 km s^{-1}), whereas the northwestern part at about the same distance exhibits a sharp maximum beyond which the velocity also reaches a plateau. In addition, this part cannot be traced as far as the southeastern part (out to the galactocentric distances of 20″ and 30″, respectively). The rotation curve of the satellite is more symmetric than that of the main galaxy.

Amram *et al.* (1989) report a rotation curve for this system based on an analysis of a detailed velocity field of ionized gas obtained with a Fabry-Perot interferometer. The above authors also point out that the rotation curve of the primary reaches a plateau at a velocity of about 110 km s^{-1}, which remains approximately constant out to a galactocentric distance of 30″ and 20″ for the southeastern and northwestern sides of the central disk, respectively. Beyond this distance the rotation curve of Amram *et al.*, which is traced to greater galactocentric distances than our totation curve, exhibits a fast decrease of rotation velocity. The direction of rotation of NGC 2535 as determined from our data (southeastern part of the galaxy

[2]The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

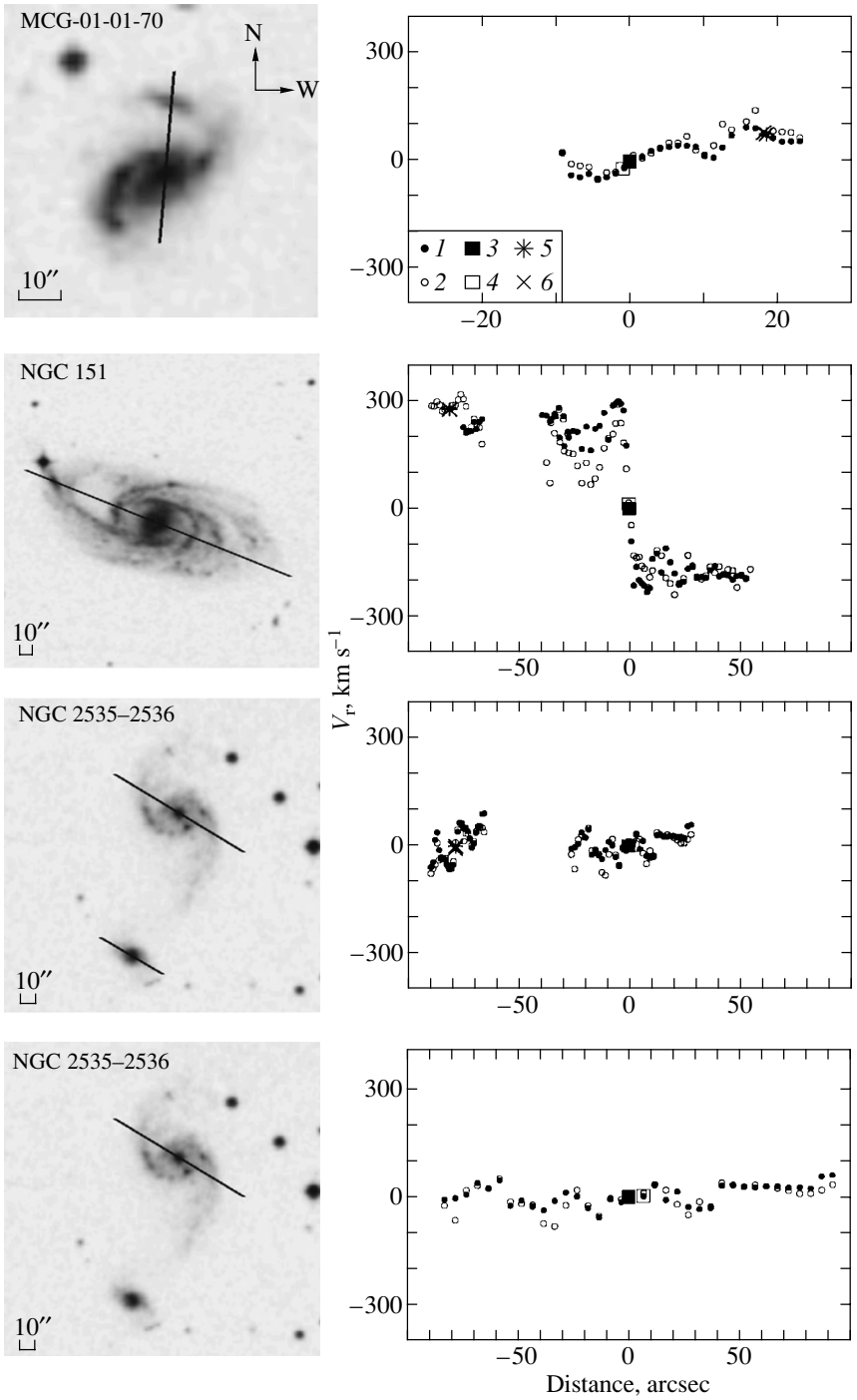


Figure. Images and rotation curves of M51-type galaxies. The lines indicate the positions of the slit during observations. Designations on the rotation curves: (*1*) velocity of rotation determined from Hα line; (*2*) velocity of rotation determined from [NII] line; (*3*) dynamical center of the primary galaxy; (*4*) photometric center of the primary galaxy; (*5*) dynamical center of the companion; and (*6*) photometric center of the companion. The distances and velocities are with respect to the dynamical center of the galaxy.

recedes and the northwestern part approaches) coincides with that indicated by Amram *et al.* (1989) (see Fig. 3 in the above paper). (Note, however, that Amram *et al.* give different rotation directions in their Figs. 3 and 4.)

As for the rotation curve of the satellite reported by Amram *et al.*, it exhibits the same main features as can be seen on our rotation curve: velocity decreases beyond 5″ from the center, and the curve flattens out. The radial velocities of the centers of the primary and

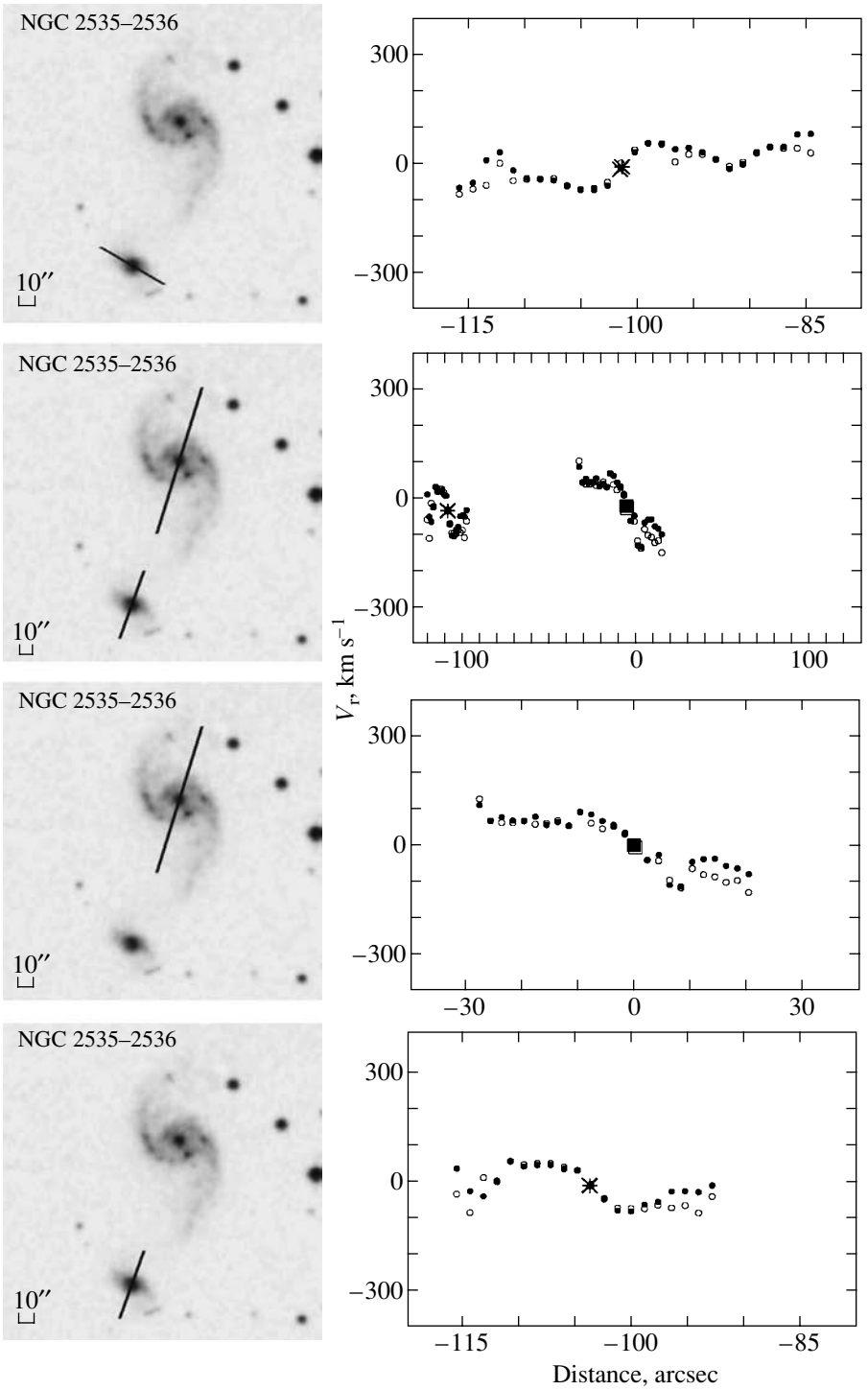


Figure. (Contd.)

the satellite determined by Amram *et al.* agree well with our results: 4095 ± 10 and 4085 ± 10 km s^{-1}, respectively.

NGC 797. The spectrum of this system exhibits a powerful continuum in the nuclear regions of both components. A strong absorption feature develops closer to the center, which virtually engulfs the Hα and [NII] lines, thereby hindering significantly the measurements. The emission spectrum shows a discontinuous pattern outside the nuclei and line intensities are extremely low in these regions. Both rotation curves exhibit strong local variations of rotation velocity. Van Moorsel (1983) obtained a 21-cm HI

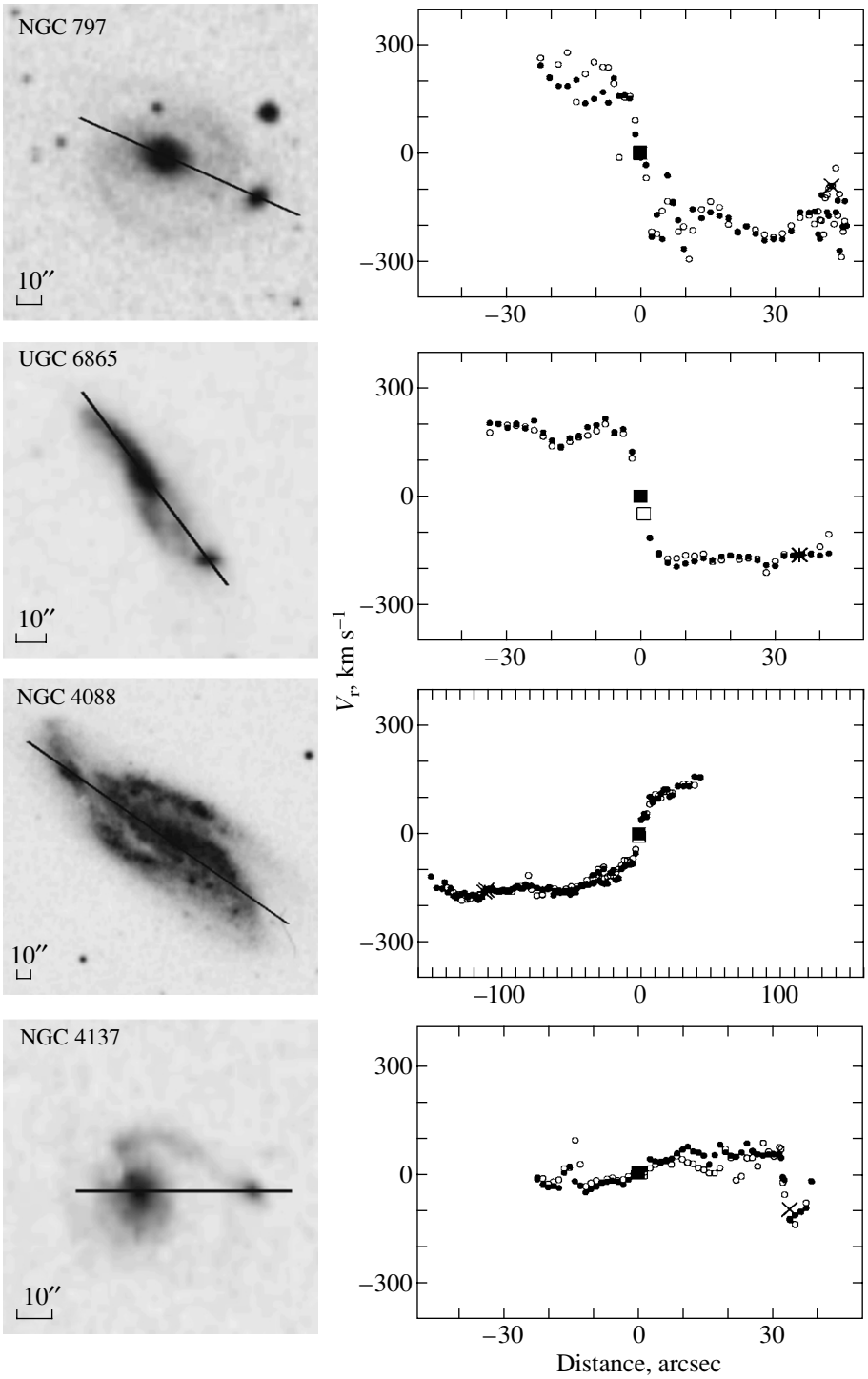


Figure. (Contd.)

line rotation curve for NGC 797, which extends out to a distance of 1′ from the center of the primary. The maximum rotation velocity corrected for the tilt of the galaxy inferred from the same rotation curve is equal to 270 km s^{-1} (at a distance of 40″ from the nucleus). The Updated Zwicky Catalog (UZC) (Falco *et al.* 2000) reports for the optical center of the primary a radial velocity of 5654 ± 4 km s^{-1}, which is close to our result.

UGC 6865. This system is strongly tilted with respect to the line of sight. The rotation curve flattens out beyond 8″–10″ from the center of the primary galaxy. However, the northern branch exhibits a local

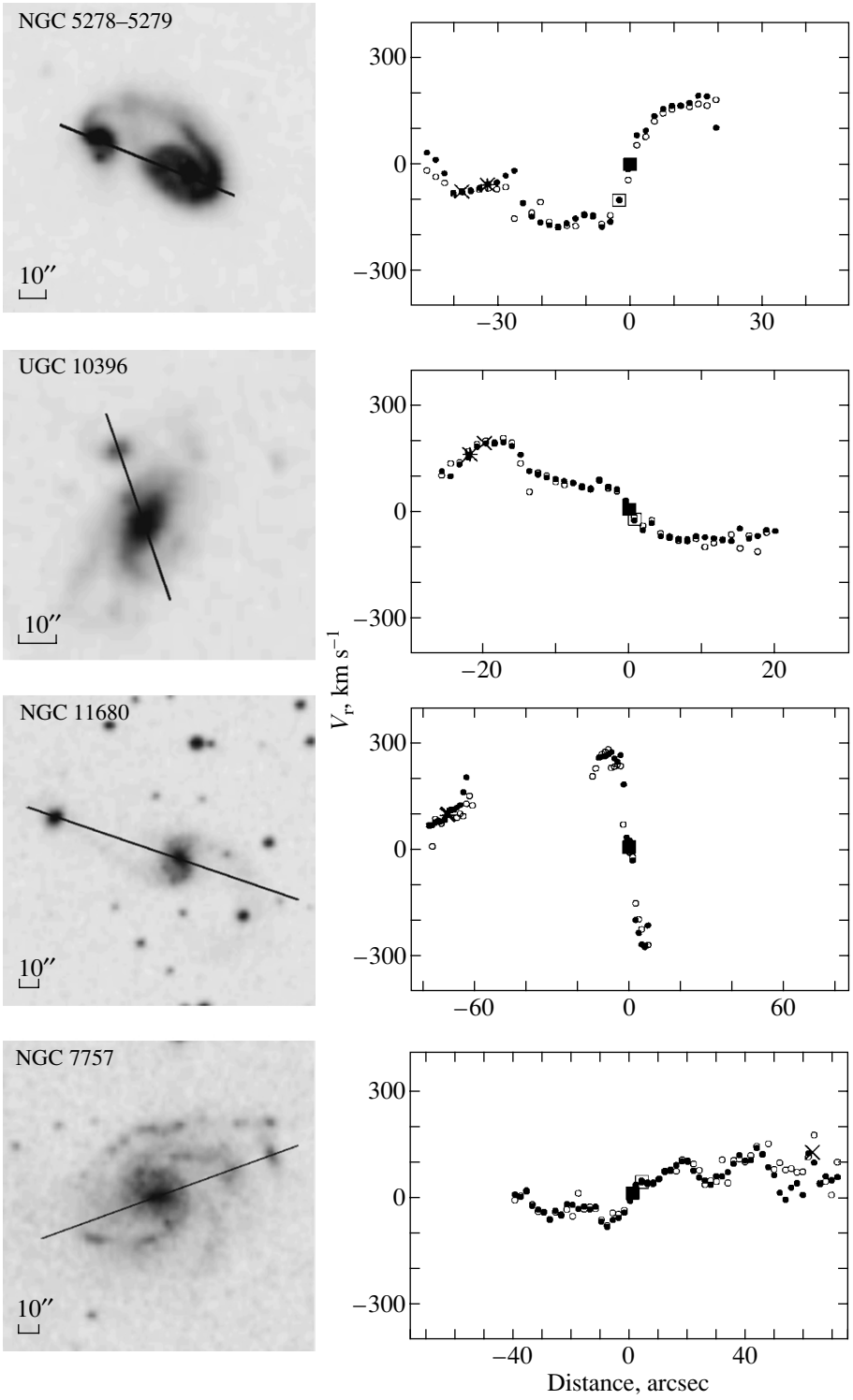


Figure. (Contd.)

velocity decrease with an amplitude of 80 km s^{-1}. This may be due to inhomogeneous internal absorption. Because of the latter the rotation curve appears asymmetric and our systemic radial velocities determined for the photometric and dynamic centers of the main galaxy differ by 50 km s^{-1}. The central radial velocities reported by different authors scatter substantially. For example, the LEDA database gives for the primary galaxy a radial velocity of 5845 ± 83 km s^{-1}, whereas Bushouse (1986) reports velocities of 5900

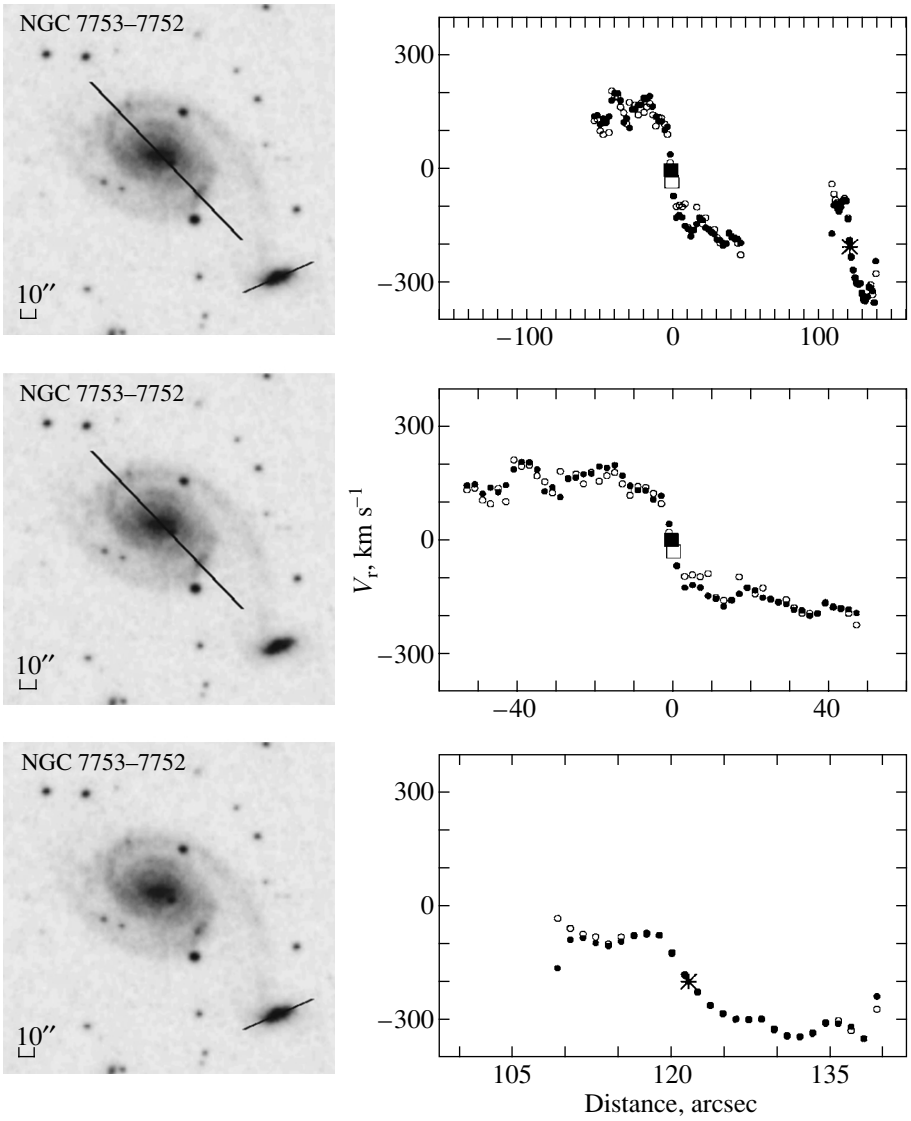


Figure. (Contd.)

and 5650 km s^{-1} for the primary and the satellite, respectively. No rotation-curve data could be found in the literature.

NGC 4088. We observed this galaxy with a slit passing through the major axis. Both emission lines broaden strongly in the nucleus of the galaxy. The large angular size of the galaxy prevented us from tracing the rotation curve for the southwestern part at more than 50″ from the center. The northeastern part of the curve flattens out beyond 40″ at a level of ∼150 km s^{-1}. Carozzi-Meysonnier (1978) obtained several spectra of this system at different position angles including $P.A. = 58°$. The northeastern branch of his rotation curve flattens out at ≈100 km s^{-1} at a galactocentric distance of ∼10″, as determined by the author, whereas the velocity of the southwestern part continues to increase monotonically. The position of the center of the galaxy determined by Carozzi-Meysonnier appears to be shifted relative to our position by ∼25″. The radial velocity of this center determined by the above author for the position angle considered is equal to 706 km s^{-1}. In addition, Carozzi-Meysonnier also points out the asymmetry of his rotation curve: the rotation velocity of the northeastern part of the curve is, on the whole, smaller by 100 km s^{-1} than that of the southwestern part. Our rotation curve appears much more symmetric. Carozzi-Meysonnier explains this asymmetry either by ejections from the active nucleus or by the influence of the neighboring galaxy NGC 4085. The velocities of the center of the primary galaxy reported by other authors scatter substantially. The UZG catalog (Falco *et al.* 2000) gives an optical velocity of 759 ± 4 km s^{-1}, which is close to our result.

NGC 4137. The satellite in this system has a very peculiar spectrum. It consists of two components separated by a small gap. The difference of the radial velocities of these components reaches 200 km s^{-1}. The rotation curve of the satellite consists of two branches, which correspond to these components. The rotation curve has a very large central gradient dV_r/dr. The rotation curve of the primary galaxy has velocity maxima of about $V_r = 70$ km s^{-1} and, on the whole, a symmetric shape. The UGC (Falco *et al.* 2000) gives a velocity of $11\,218 \pm 58$ km s^{-1} for the center of the primary galaxy, which is close enough to our result. No published rotation curves could be found in the literature.

NGC 5278–5279. This Markarian system (Mrk 271) exhibits a spectrum with very bright and broad emission lines in the nucleus. The spectral lines of both components are inclined in opposite directions, indicating that they rotate in opposite directions. The rotation curve of the primary galaxy flattens out, albeit with certain velocity variations. The main specific feature of this system are strongly noncoincident positions and radial velocities of the dynamic and photoimetric centers of both components (3″ and 100 km s^{-1} for NGC 5278 and 6″ and 20 km s^{-1} for NGC 5279, respectively). A detailed photometric analysis of the system performed by Mazzarella and Boroson (1993) revealed a number of regions or spots of enhanced brightness and the brightest of them is located in the satellite. The strong difference between the dynamic and photometric centers mentioned above may be explained by a similar inhomogeneity of the photometric properties of the two components. The latter may be due, in turn, to the interaction between the galaxies. Numerous determinarions of the radial velocities of the component centers yielded results that show a rather substantial scatter. The LEDA database gives the velocities of 7558 ± 79 and 7580 ± 46 km s^{-1} for the primary galaxy and the satellite, respectively. No published rotation curve could be found in the literature.

UGC 10396. The rotation curve of the primary galaxy flattens out beyond 5″ from the center, reaching a velocity of ~80 km s^{-1}. The system's components rotate in opposite directions. The dynamical and photometric centers of both components somewhat differ in position and radial velocity. The only determination of the central radial velocity found in the literature appears to come from Arkhipova and Esipov (1979). It is listed, e.g., in the UZC catalog (Falco *et al.* 2000) and is equal to 6185 ± 150 km s^{-1}. The causes of such a strong discrepancy between this value and our determination (~2500 km s^{-1}) are unclear.

UGC 11680. The satellite in this system is the Markarian galaxy Mrk 897. The spectra of both components in the vicinity of the nuclei exhibit strongly inclined broad emission lines superimposed on a strong continuum. The rotation curve of the primary galaxy can be confidently traced only at a small distance from the center—in fact only within the rigid-rotation portion in the broad emission-line region. The maximum observed velocity is equal to $V_r = 300$ km s^{-1}. The rotation curve of the system obtained by Keel (1996a) appears to show the same features as our rotation curve. The component radial velocities determined by Keel agree well with our results: 7791 ± 11 and 7894 ± 5 km s^{-1} for the primary and the satellite, respectively.

NGC 7757. The primary galaxy in this system has an asymmetric rotation curve: the observed rotation velocity of the eastern branch reaches a maximum of $V_r = 100$ km s^{-1} at a galactocentric distance of about 8″, and that of the western branch, at a distance of 18″. The rotation curve shows a specific behavior at its center: here velocity does not increase monotonically but the gradient dV/dr decreases locally to zero. No published rotation curves could be found for this galaxy. The UZC catalog (Falco *et al.* 2000) gives an optical radial velocity of 2960 ± 21 km s^{-1}, which agrees well with our results.

NGC 7753–7752. The primary galaxy and the satellite were observed separately and with different positions of the slit, which passed through the major axes of the components. The spectrum in the spiral-arm region appears discontinuous and clumpy and consists of several condensations. The rotation curve of the primary substantial velocity variations with an amplitude of ~100 km s^{-1}, which can be explained by inhomogeneous internal absorption when the slit crosses the spiral arms, or by the tidal interaction of the components. The same interaction can explain the different velocities of the photometric and dynamic centers (Marcelin *et al.* 1987). An analysis of the detailed velocity field allowed Marcelin et al. to obtain the rotation curves of both components. The rotation curve of the primary extends out to a distance of 100″ from the center (compared to 50″ for our rotation curve). The rotation velocity decreases beyond 50″. At smaller galactocentric distances the rotation curve of Marcelin *et al.* is, on the whole, similar to our rotation curve, and exhibits the same specific features. Thus Marcelin et al. point out velocity maxima exceeding 200 km s^{-1} at a galactocentric distance of 40″, and this feature agrees well with our results. In addition, Marcelin *et al.* traced their rotation curve somewhat farther from the center than we did in the case of our satellite's rotation curve, and it exhibits the same

characteristic features. The component radial velocities obtained by the above authors agree rather well with our data: 5160 ± 10 and 4940 ± 10 km s^{-1} for the primary and the satellite, respectively. Note that other radial-velocity determinations in this system are characterized by a large scatter of the results obtained.

To reveal the possible systematic errors in the velocities of galaxies, we compared our results with the those published in the NED database. The mean difference in the sense of the radial velocity of the dynamic center minus the NED velocity is -5 km s^{-1} with a dispersion of 55 km s^{-1} (averaged over 15 objects including some satellites). The same comparison for the photometric centers yielded -18 ± 44 km s^{-1} (15 objects). These results imply that our data are free of important systematic errors.

The mean difference of our velocities of dynamic and photometric centers averaged over 23 objects (including satellites) is equal to 16 km s^{-1} with a dispersion of 23 km s^{-1}. The corresponding difference for 36 binary galaxies from the catalog of Karachentsev (1972) with radial-velocity estimates given by Keel (1996b) is equal to 9 ± 37 km s^{-1}.

ACKNOWLEDGMENTS

We are grateful to V.V. Vlasyuk and A.G. Pramskoi from the Special Astrophysical Observatory (Russian Academy of Sciences) for help with the observations. We used the LEDA database (http://leda.univ-lyon1.fr).

REFERENCES

1. V. L. Afanas'ev, A. N. Burenkov, V. V. Vlasyuk, and S. V. Drabek, Otchet SAO RAN **234** (1995).
2. P. Amram, M. Marcelin, J. Boulesteix, and E. Le Coarer, Astron. Astrophys., Suppl. Ser. **81**, 59 (1989).
3. V. P. Arkhipova and V. F. Esipov, Pis'ma Astron. Zh. **5**, 265 (1979) [Sov. Astron. Lett. **5**, 140 (1979)].
4. H. Arp, Astrophys. J., Suppl. Ser. **14**, 1 (1966).
5. H. Bushouse, Astron. J. **91**, 255 (1986).
6. N. Carozzi-Meyssonnier, Astron. Astrophys. **63**, 415 (1978).
7. E. Falco, M. Kurtz, M. Gellar, *et al.*, Publ. Astron. Soc. Pac. **111**, 438 (1999).
8. J. Huchra, D. Latham, L. da Costa, *et al.*, Astron. J. **105**, 1637 (1993).
9. I. D. Karachentsev, Soobshch. SAO **7**, 3 (1972).
10. W. Keel, Astron. J. **111**, 696 (1996a).
11. W. Keel, Astrophys. J., Suppl. Ser. **106**, 27 (1996b).
12. S. A. Klimanov and V. P. Reshetnikov, Astron. Astrophys. **378**, 428 (2001).
13. M. Marcelin, E. Lecoarer, J. Boulesteix, *et al.*, Astron. Astrophys. **179**, 101 (1987).
14. D. Mathewson and V. Ford, Astrophys. J., Suppl. Ser. **107**, 97 (1996).
15. J. Mazzarella and T. Boroson, Astrophys. J., Suppl. Ser. **85**, 27 (1993).
16. G. van Moorsel, Astron. Astrophys., Suppl. Ser. **54**, 1 (1983).
17. B. A. Vorontsov-Vel'yaminov, Astron. Zh. **34**, 8 (1957).

Translated by A. Dambis

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 589–593. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 653–658.

Two Period–Radius Relations for Classical Cepheids: Determining the Pulsation Mode and the Distance Scale

M. E. Sachkov*

Institute of Astronomy, Russian Academy of Sciences, ul. Pyatnitskaya 48, Moscow, 119017 Russia

Received March 20, 2002

Abstract—Based on our radial-velocity measurements and on published photometric observations, we calculated the radii of 64 classical Cepheids that were previously assumed to be fundamental-mode pulsators. Our detailed analysis of the period–radius diagram shows that the sample of Cepheids with pulsation periods shorter than 9 days probably contains a significant fraction (up to 30%) of stars pulsating in the first overtone. This fact leads to incorrect luminosity estimates for Cepheids and may be partly responsible for the discrepancy between the short and long distance scales. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *variable stars, Cepheids, pulsation mode, distance scale*

INTRODUCTION

An investigation of Cepheids, radially pulsating variable stars, remains one of the most important problems in astrophysics and galactic astronomy. These stars, along with open clusters, are known to be the most important distance indicators in the Universe. The distance scale for classical Cepheids is based on the existence of a clear relation between the pulsation period and the mean absolute magnitude (or the luminosity averaged over the pulsation period) for these variable stars. The period–luminosity relation is calibrated by using Cepheids that are members of open clusters, whose distances are determined from photometric data. Thus, the Cepheid distance scale is eventually based on the distances to young open clusters. Berdnikov *et al.* (1996) estimated the distance modulus to the Large Magellanic Cloud on this scale to be $(m - M)_0 = 18\overset{m}{.}25 \pm 0\overset{m}{.}12$. This estimate matches the so-called short distance scale.

A major result of the HIPPARCOS project was the catalog of distances to 118 000 stars. Having analyzed the HIPPARCOS trigonometric parallaxes for classical Cepheids (the samples include from 20 to 200 stars with different errors in the parallaxes), Feast and Catchpole (1998) concluded that the Cepheid distance scale should be lengthened. The distance modulus to the Large Magellanic Cloud in this paper is $18.70 \pm 0\overset{m}{.}10$. However, the distance scale is still a problem: the Cepheid parallaxes are so small that the result depends on subtleties of the statistical approach to the initial data.

The method of statistical parallaxes gives another chance of solving the problem of the distance scale for classical Cepheids without invoking data on their trigonometric parallaxes. Its basic idea is to match the tangential and radial velocities for a sample of objects. Rastorguev *et al.* (1999) applied this method to two samples of Cepheids, with pulsation periods of more than 9 days and of less than 9 days, and obtained different estimates for the distance scale. The first sample of Cepheids agrees well with the short distance scale, while the distance scale for the Cepheids with pulsation periods less than 9 days must be lengthened. The authors explained this difference by the probable error in identifying the pulsation mode: some Cepheids pulsating in the first overtone were, probably, mistaken for Cepheids pulsating in the fundamental mode. Recall that the observed period of brightness (or radial-velocity) variations for a Cepheid may refer both to the fundamental mode and to one of the overtones and it cannot be identified without an additional investigation. An error in identifying the pulsation mode leads to an incorrect estimate of the Cepheid luminosity and, hence, to errors in the distance scale.

The radial pulsation mode characterizes the number of nodal concentric surfaces inside the star that are not involved in pulsations and that are located at the boundary between two regions moving in opposite radial directions. The stars pulsating in the fundamental mode have no such surfaces; the first-overtone pulsators have one such surface. Thus, being among the fundamental physical parameters, the pulsation mode and radius of the star are important not only for

*E-mail: msachkov@inasan.rssi.ru

stellar-pulsation studies; they are also related to the problem of distance scale.

TWO PERIOD–LUMINOSITY RELATIONS FOR CLASSICAL CEPHEIDS

By analogy with RR Lyrae stars, the existence of two period–luminosity relations for classical Cepheids (for two groups of stars pulsating in the fundamental mode and in the first overtone) was assumed by Christy (1966) and Stobie (1969). Böhm-Vitense (1994) even concluded that most of the variables among the Cepheids with pulsation periods less than 8 days were first-overtone pulsators.

The separation of Cepheids into two groups by their pulsation modes is most pronounced in the period–extinction-independent magnitude diagram (1333 objects) for stars in the Large Magellanic Cloud (Udalski *et al.* 1999). An additional advantage for studying Cepheids in this galaxy is that they are all located virtually at the same distance from us, which significantly simplifies the analysis of their peculiarities. The difference between the two period–magnitude relations corresponds to the ratio of the pulsation periods of the first overtone, P_1, and the fundamental mode, P_0: $P_1/P_0 \approx 0.71$ (Alcock *et al.* 1995).

The dependences of Fourier parameters (e.g., R_{21} and φ_{21}) for the light and radial-velocity curves on the logarithm of period (Alcock *et al.* 1999; Ogloza and Moskalik 2000) are used to separate the objects into fundamental-mode and first-overtone pulsators (which is particularly important for Galactic Cepheids located at different distances). Here, we also turned to this method. However, such studies are difficult to carry out for some ranges of periods ($0.6 < \log P < 0.8$, $0.2 < \log P < 0.4$). In our view, analysis of the period–radius relation to determine the pulsation mode seems promising in this region; it was successfully applied to a sample of low-amplitude Cepheids (Sachkov 1997). The radii are in absolute units, solar radii, and do not depend on information about the Cepheid distances and interstellar extinction. Note also that this method is the only way of determining the pulsation mode for Galactic Cepheids that does not depend on the analysis of the Fourier coefficients for the light curves.

THE METHOD FOR CALCULATING THE RADII

The radii of radially pulsating stars can be estimated by the Baade–Wesselink method from photometric and spectroscopic observations. Among the numerous modifications of this method, the modification of Balona (1977) is particularly remarkable, because it uses direct observational data (brightness, color, and radial-velocity measurements) and assumes only the existence of a linear (in a more accurate approximation, quadratic) relation between the logarithm of effective temperature and the normal color and the bolometric correction in a limited temperature range. In addition, Balona's modification does not require any knowledge of *a priori* information on the interstellar extinction. Our improved implementation of Balona's method using the nonlinear maximum-likelihood method and analysis of the errors in the method were described previously (Sachkov *et al.* 1998). For the binary Cepheids in our sample (RX Cam, DL Cas, SU Gyg, VZ Cyg, MW Cyg, S Sge, V350 Sgr), we separated their radial-velocity variations into pulsational and orbital ones (Gorynya *et al.* 1996) and used the pulsational velocities for our calculations.

OBSERVATIONAL DATA

To derive the system of radii for classical Cepheids, and to eliminate the effects of systematic observational errors, and, in particular, to analyze the period–radius relation, it is important that the observational data used be homogeneous. In addition, it is important that the light, color, and radial-velocity curves contain a sufficient number of measurements to reveal all features (e.g., humps). In particular, to achieve the highest accuracy in calculating the radii, the light and color curves should contain at least 50 measurements (with an accuracy of no less than $0^m.02$) and the radial-velocity curve should contain at least 20 measurements (with an accuracy of ~ 0.5 km s^{-1}). We used the spectroscopic observations obtained during 1987–2001 with a correlation spectrometer by a group of observers at the Institute of Astronomy of the Russian Academy of Sciences (Moscow) and at the Sternberg Astronomical Institute (SAI, Moscow). Most of the observations were published in Gorynya *et al.* (1998). The photoelectric BV measurements were taken from the SAI database (Berdnikov 1995). The bulk of the observational data meets the above requirements.

RESULTS AND DISCUSSION

To plot the period–radius diagram (see the figure), we selected 64 classical Cepheid with the largest amount of available observational data. In the GCVS (Kholopov 1985), these are all designated as fundamental-mode pulsators (DCEP).

Based on the period–radius diagram, we separated all stars into two groups with different period–radius relations. The significance of this separation

Table 1. Cepheids pulsating in the first overtone

Star	$\log P_1$	$R/R_\odot$	σ_R	$\Delta\phi_1$	$\sigma(\Delta\phi_1)$	Amplitude, V	$M-m$
IR Cep	0.3251	31	4	−0.25	0.05	$0^m.37$	$0^p.42$
SU Cyg	0.5850	40	3	−0.40	0.10	0.76	0.28
V379 Cas	0.6340	50	8	−0.51	0.10	0.33	0.46
VZ Cyg	0.6870	53	3	−0.56	0.19	0.67	0.28
CF Cas	0.6879	51	4	−0.53	0.18	0.56	0.28
V Lac	0.6975	52	5	−0.59	0.10	0.92	0.24
V386 Cyg	0.7207	50	4	−0.61	0.17	0.69	0.29
CR Ser	0.7243	51	4	−0.60	0.15	0.76	0.27
SW Cas	0.7357	55	2	−0.62	0.18	0.67	0.28
RZ Gem	0.7426	55	4	−0.46	0.35	0.94	0.20
V733 Aql	0.7908	59	6	−0.68	0.08	0.46	0.37
RS Cas	0.7990	65	6	−0.43	0.35	0.78	0.27
BK Aur	0.9032	72	7	−0.47	0.35	0.66	0.32

is confirmed by the χ^2 test with a probability higher than 96%.

This separation can be easily explained by assuming that the stars pulsate in different modes: group 1 in the fundamental mode and group 2 in the first overtone. Taking into account the ratio of the first-overtone period P_1 to the fundamental-mode period P_0 [which is close to $P_1/P_0 \approx 0.71$ for Galactic Cepheids (Alcock *et al.* 1995)], we can write the following relations:

$$\log R = 1.09 + 0.73\log P_0 \quad \text{for group 1,}$$
$$\pm.01 \quad \pm.02$$

the lower dotted line in the figure;

$$\log R = 1.20 + 0.74\log P_1 \quad \text{for group 2,}$$
$$\pm.06 \quad \pm.06$$

the upper dotted line in the figure;

or

$$\log R = 1.09 + 0.74\log P_0 \quad \text{for group 2.}$$
$$\pm.06 \pm .06$$

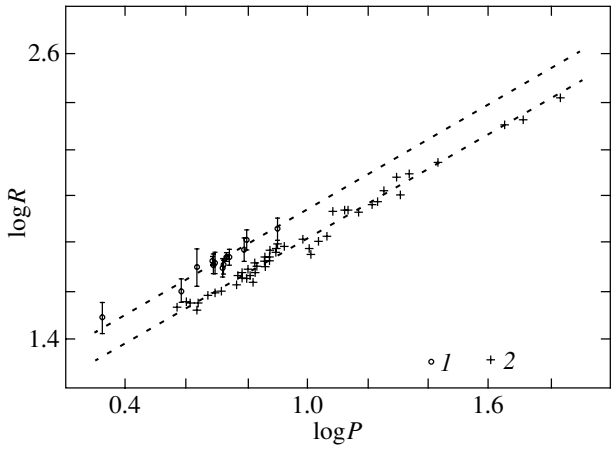
Since the first and third formulas matched when the periods recalculated to the fundamental pulsation mode were used for the group-2 objects, our assumption that the group-2 objects are first-overtone pulsators receives confirmation. Our results are in good agreement with the period–radius relation derived by Laney and Stobie (1995): $\log R = 1.07 + 0.75\log P_0$.

Table 1 gives data on the Cepheids that, according to our data, pulsate in the first overtone; Table 2 gives data on the Cepheids pulsating in the fundamental mode. The radii are in solar radii and the periods are in days. The tables also give parameters of the light curves we used: the amplitudes in V and the durations of the brightness rise from minimum to maximum, $M-m$ (in fractions of the period). Based on these parameters, we separated out the subclass of the so-called low-amplitude Cepheids (DCEPS) in the GCVS with amplitudes less than $0^m.5V$ and with almost symmetric light curves ($M-m \approx 0.4$–0.5). It was previously suggested that these stars could be first-overtone pulsators (Kholopov 1985). However, variables pulsating in the fundamental mode are also encountered among them (Sachkov 1997). Most of the Cepheids in Table 1 cannot be classified as DCEPS.

As was mentioned above, we analyzed the Fourier parameters to check our results. More specifically, we fitted the observed light and radial-velocity curves by Fourier series and then determined the first-order

Table 2. Cepheids pulsating in the fundamental mode

Star	$\log P_0$	$R/R_\odot$	σ_R	$\Delta\phi_1$	$\sigma(\Delta\phi_1)$	Amplitude, V	$M-m$
RT Aur	0.5714	34	3	−0.25	0.07	$0^m.78$	$0^P.27$
ST Tau	0.6058	36	3	−0.26	0.08	0.77	0.26
SY Cas	0.6097	35	2	−0.26	0.05	0.79	0.27
Y Lac	0.6358	33	2	−0.18	0.06	0.70	0.28
V402 Cyg	0.6399	36	2	−0.29	0.06	0.57	0.30
RY CMa	0.6700	38	3	-0.40	0.11	0.73	0.25
V1154 Cyg	0.6924	39	2	−0.23	0.06	0.38	0.36
V350 Sgr	0.7122	40	4	−0.22	0.08	0.71	0.28
FM Cas	0.7641	42	5	−0.22	0.09	0.58	0.30
MW Cyg	0.7748	46	4	−0.31	0.09	0.71	0.28
KL Aql	0.7859	47	7	−0.30	0.06	0.74	0.29
FM Aql	0.7863	47	2	−0.31	0.04	0.73	0.28
V538 Cyg	0.7866	45	5	−0.23	0.07	0.56	0.33
X Vul	0.8006	45	3	−0.24	0.05	0.77	0.27
RR Lac	0.8072	49	5	−0.28	0.05	0.78	0.30
AW Per	0.8104	46	5	−0.23	0.06	0.28	0.27
BB Sgr	0.8219	43	6	−0.28	0.04	0.61	0.31
V495 Cyg	0.8276	47	5	-0.29	0.15	0.45	0.31
U Sgr	0.8289	52	6	−0.22	0.05	0.74	0.30
V496 Aql	0.8330	50	6	−0.25	0.09	0.37	0.46
V600 Aql	0.8596	55	2	−0.21	0.06	0.65	0.30
V459 Cyg	0.8604	50	5	−0.30	0.15	0.69	0.30
V336 Aql	0.8636	53	4	−0.27	0.05	0.72	0.30
V1344 Aql	0.8737	59	2	−0.29	0.04	0.27	0.40
BB Her	0.8755	53	6	−0.31	0.05	0.65	0.33
RS Ori	0.8789	55	6	−0.33	0.08	0.81	0.29
W Gem	0.8983	60	5	−0.23	0.05	0.80	0.29
RX Cam	0.8983	58	5	−0.27	0.04	0.73	0.30
U Vul	0.9025	60	4	−0.32	0.05	0.71	0.32
DL Cas	0.9031	62	4	−0.32	0.04	0.57	0.33
S Sge	0.9234	61	3	−0.31	0.06	0.71	0.31
DD Cas	0.9918	65	7	−0.34	0.08	0.61	0.45
BZ Cyg	1.0061	60	10	−0.25	0.07	0.50	0.50
ζ Gem	1.0064	60	7	−0.29	0.14	0.49	0.50
AN Aur	1.0124	57	6	−0.33	0.09	0.69	0.50
Z Lac	1.0368	64	6	−0.24	0.07	0.96	0.42
RX Aur	1.0653	67	10	−0.26	0.08	0.67	0.44
RY Cas	1.0841	85	8	−0.40	0.10	0.96	0.40
V916 Aql	1.1284	87	4	−0.29	0.11	0.90	0.42
TT Aql	1.1384	87	2	−0.26	0.04	1.10	0.35
RW Cas	1.1701	85	4	−0.33	0.09	1.15	0.34
X Cyg	1.2144	91	5	−0.22	0.05	1.00	0.34
CD Cyg	1.2323	95	7	−0.30	0.14	1.18	0.30
CP Cep	1.2518	105	10	−0.25	0.12	0.78	0.37
RU Sct	1.2944	120	10	−0.42	0.11	1.08	0.33
VX Cyg	1.3039	101	10	−0.24	0.07	0.95	0.34
WZ Sgr	1.3394	125	5	−0.26	0.15	1.09	0.33
T Mon	1.4317	140	10	−0.24	0.13	0.99	0.30
SV Vul	1.6534	200	30	−0.36	0.06	1.03	0.23
GY Sge	1.7126	210	20	−0.25	0.05	0.65	0.28
S Vul	1.8325	260	40	−0.20	0.10	0.56	0.35



The period–radius diagram: (*1*) probable first-overtone pulsators (shown with their individual errors); (*2*) fundamental-mode Cepheids (their individual errors do not exceed the errors for the probable first-overtone pulsators). The radii are in solar units; the periods are in days.

phase shift $\Delta\phi_1 = \phi_1^{Vr} - \phi_1^{\rm mag}$ (Ogloza and Moskalik 2000). This parameter does not depend on the pulsation period for Cepheids pulsating in the fundamental mode, being $\sim$0.28 (Sachkov 2000). For Cepheids pulsating in the first overtone, the first-order phase shift varies with period from -0.24 to -0.71. The computed first-order phase shifts confirm our assumption that the Cepheids from Table 1 probably pulsate in the first overtone. For three stars, RZ Gem, RS Cas, and BK Aur, $\Delta\phi_1$ was computed with large errors, because the observational data were insufficient. The parameter $\Delta\phi_1$ and its error $\sigma(\Delta\phi_1)$ are also presented in Tables 1 and 2.

CONCLUSIONS

Thus, based on our analysis of the period–radius diagram, we can classify 13 of the 64 stars studied as first-overtone pulsators. The pulsation periods for these Cepheids are shorter than 9 days, in agreement with the pulsation theory that prohibits the existence of overtone pulsators with periods longer than 10 days. The fraction of the probable first-overtone pulsators among the Cepheids with pulsation periods shorter than 9 days in our sample is $\approx$0.3, in good agreement with the assumptions of Rastorguev *et al.* (1999). It should be noted, however, that the fraction of the first-overtone pulsators among all Galactic Cepheids may be different (our sample is incomplete). Nevertheless, the existence of Cepheids pulsating in the first overtone that were previously mistaken for fundamental-mode pulsators may be considered to be proven. A wrong identification of the pulsation mode can significantly contribute to the overestimation of the Cepheid luminosities, which accounts for the conclusion of Feast and Catchpole (1998) that the distance modulus to the Large Magellanic Cloud must be increased to $18^m.70$.

ACKNOWLEDGMENTS

I wish to thank A.S. Rastorguev and A.K. Dambis for valuable remarks. I am also grateful to N.N. Samus' and N.A. Gorynya for cooperation during the observations. This study was supported in part by grants from the American Astronomical Society and the Federal Science and Technology Program "Astronomy."

REFERENCES

1. C. Alcock, R. A. Allsman, T. S. Axelrod, *et al.*, Astron. J. **109**, 1653 (1995).
2. C. Alcock, R. A. Allsman, D. Alves, *et al.*, Astrophys. J. **511**, 185 (1999).
3. L. A. Balona, Mon. Not. R. Astron. Soc. **178**, 231 (1977).
4. L. N. Berdnikov, Astron. Soc. Pac. Conf. Ser. **83**, 349 (1995).
5. L. N. Berdnikov, O. V. Vozyakova, and A. K. Dambis, Pis'ma Astron. Zh. **22**, 936 (1996) [Astron. Lett. **22**, 838 (1996)].
6. E. Böhm-Vitense, Astron. J. **107**, 673 (1994).
7. R. F. Christy, Astrophys. J. **145**, 340 (1966).
8. M. W. Feast and R. M. Catchpole, Mon. Not. R. Astron. Soc. **286**, 1 (1998).
9. *General Catalog of Variable Stars,* Ed. by P. N. Kholopov (Nauka, Moscow, 1985, 4th ed.).
10. N. A. Gorynya, A. S. Rastorguev, and N. N. Samus', Pis'ma Astron. Zh. **22**, 38 (1996) [Astron. Lett. **22**, 33 (1996)].
11. N. A. Gorynya, N. N. Samus', M. E. Sachkov, *et al.*, Pis'ma Astron. Zh. **24**, 939 (1998) [Astron. Lett. **24**, 815 (1998)].
12. C. D. Laney and R. S. Stobie, Mon. Not. R. Astron. Soc. **274**, 117 (1995).
13. W. Ogłoza and P. Moskalik, Astron. Soc. Pac. Conf. Ser. **203**, 235 (2000).
14. A. S. Rastorguev, E. V. Glushkova, A. K. Dambis, and M. V. Zabolotskikh, Pis'ma Astron. Zh. **25**, 689 (1999) [Astron. Lett. **25**, 595 (1999)].
15. M. E. Sachkov, Inf. Bull. Var. Stars, No. 4522, 1 (1997).
16. M. E. Sachkov, A. S. Rastorguev, N. N. Samus', and N. A. Gorynya, Pis'ma Astron. Zh. **24**, 443 (1998) [Astron. Lett. **24**, 377 (1998)].
17. M. E. Sachkov, Astron. Soc. Pac. Conf. Ser. **203**, 240 (2000).
18. R. S. Stobie, Mon. Not. R. Astron. Soc. **144**, 485 (1969).
19. A. Udalski, I. Soszyński, M. Szymański, *et al.*, Acta Astron. **49**, 223 (1999).

Translated by N. Samus'

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 594–600. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 659–666.

The Generation of Ionization-Shock Front Oscillations by a Variable Radiation Flux

K. V. Krasnobaev*

Moscow State University, Vorob'evy gory, Moscow, 119899 Russia

Received March 20, 2002

Abstract—The effect of a time-varying radiation flux incident on an ionization front on the generation of ionization-shock front oscillations in the interstellar medium is analyzed analytically and numerically. We take into account both variations in the flux of ionizing radiation directly from the source that produces the ionization front and the absorption of energetic photons by the post-front plasma. Based on our calculations, we show that the time dependence of the radiation flux can be an additional factor (apart from small inhomogeneities in the interstellar medium) that contributes to the amplification of oscillations and to the kinetic energy input to the observed turbulent motions in H II regions. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *interstellar medium, gaseous nebulae, shock waves, ionization fronts, instabilities, H II regions*

INTRODUCTION

The studies of the stability of radiative shock waves in the interstellar medium performed to date show that the perturbation amplitude and spectrum significantly depend on the properties of the medium through which the shock front (S-front) propagates and on the formation conditions for the S-front itself. In particular, the following factors are of importance: variations of the gas cooling rate with temperature (Strickland and Blondin 1995), the existence of a magnetic field (Kimoto and Chernoff 1997) various types of discontinuities (Krasnobaev 2001a, 2001b) in the interstellar medium, and the fact that the flow is not one-dimensional (Walder and Follini 2000).

As applied to the dynamics of an ionization–shock front (IS-front), we have found previously (Krasnobaev 2001b) that even small density variations in the preshock gas can produce perturbations in the parameters of the medium comparable to those observed in H II regions. In this case, however, we assumed that the plasma behind the ionization front (I-front) was transparent to radiation and that the flux density of the energetic photons incident on the I-front, Φ, was constant. However, this condition can be violated if the intensity of the radiation from the source that produces the I-front varies with time or if the optical depth of the region occupied by plasma is large enough for Lyman continuum photons. For example, it is clear that Φ can depend on time near young hot stars where the IS-front is formed at the hydrodynamic expansion stage of an H II region (Spitzer 1978).

To determine how large the variations in Φ must be for the IS-front oscillations emerging under their effect to be able to contribute appreciably to the observed irregular structure of H II regions, we analytically and numerically analyze the propagation of the IS-front when Φ is a periodic function of time. In this case, much attention is given to the generation conditions for the fundamental oscillation mode, because, as numerous calculations show (see the above papers on the stability of radiative shock waves), the oscillations in gas parameters are largest when this type of oscillation develops.

SMALL IONIZATION–SHOCK FRONT OSCILLATIONS GENERATED BY A VARIABLE RADIATION FLUX

The IS-front is a complex of discontinuities that consists of an I-front preceded by a shock wave. This system of fronts is characteristic, in particular, of H II regions at the hydrodynamic expansion stage (Spitzer 1978). In this case, the adiabatic discontinuity is immediately followed by a zone of radiative gas cooling with a characteristic scale L_c, in which the temperature T and velocity u of the particles decrease and the density of the medium ρ increases.

The I-front propagates relative to the shock-compressed gas at velocity U, which is related to

*E-mail: kraskons@uuws05.math.msu.su

1063-7737/02/2809-0594$22.00

ρ and Φ by the standard relation $\rho U = m_{\rm H}\Phi$ ($m_{\rm H}$ is the mass of the hydrogen atom). In general, Φ differs from the flux density Φ_∞ of the photons emitted by the source, because the emission is absorbed by plasma. Since the absorption is difficult to accurately take into account, the studies of the general properties of motions in H II regions are commonly restricted to two limiting cases: optically transparent and optically thick H II regions. In the former case, we may clearly set $\Phi = \Phi_\infty$. In the latter case, taking into account the condition of ionization balance, we obtain for the one-dimensional plane flow geometry considered below (Spitzer 1978)

$$\Phi = \Phi_\infty - \int_{x_{\rm I}}^{l} \alpha(T)(\rho/m_{\rm H})^2 dx, \qquad (1)$$

where x is the variable of integration over space; $x_{\rm I}$ and l are the coordinates of the I-front and the source, respectively, with $x_{\rm I} \le l$; and $\alpha(T)$ is the coefficient of photorecombination to all of the hydrogen atomic levels except the ground level.

Physically, Eq. (1) is based on the fact that for the diffuse (i.e., produced by the gas itself in an H II region) radiation in an optically opaque medium, the photorecombinations to the ground level are balanced by the photoionizations from this level. In that case, however, the number of photorecombinations to all hydrogen levels except the ground level under ionization-balance conditions must be equal to the number of photoionizations under external radiation. As a result, the high-energy photons from the external source and the lower-energy photons that freely escape from the H II region will be absorbed. Numerical calculations with allowance for the angular and spectral composition of the radiation (Rubin 1968; Krasnobaev 1970) show that formula (1) slightly overestimates the contribution of the diffuse radiation in the inner (optically thin) parts of the H II region, being accurate enough at the boundary of the expanding ionized gas.

Let us use Eq. (1) to determine the conditions under which the variations of Φ_∞ with time t have the largest effect on the motion of the S- and I-fronts. To this end, we turn to a two-front model that allows a number of qualitative features in the behavior of the IS-front to be investigated (Krasnobaev 2001b). According to this model, the IS-front is replaced with a system of discontinuities: one is an isothermal shock wave (i.e., the thickness L_c of the relaxation zone is disregarded compared to the distance L between the adiabatic discontinuity and the I-front) and the other is a weak D-type I-front.

Below, we choose the following motion as the main one. Assume, for simplicity, that the medium ahead of the IS-front is homogeneous and that the velocities of the S- and I-fronts are equal and time-independent. In that case, the gas density, velocity, and the isothermal speed of sound c in a coordinate system where the fronts are at rest are constant and equal to ρ_n, u_n, and c_n ($n = 0, 1, 2$); the values of n refer to the gas parameters in the H I region, in the layer between the I- and S-fronts, and in the H II region, respectively. We choose the x axis to be perpendicular to the shock plane and parallel to the inflow velocity. Denote the coordinates of the S- and I-fronts by $x_{\rm S} = 0$ and $x_{\rm I} = L$. For the main flow, it follows from the boundary conditions at the I-front and from equality (1) that $\Phi = \Phi_0 = \rho_0 u_0/m_{\rm H}$, $\Phi_\infty = {\rm const}$. In addition, since the shock wave is assumed to be isothermal, $c_1 = c_0$.

Now, let the photon flux density Φ_∞ undergo small periodic (in time) perturbations, so that $\Phi_\infty = \Phi_{\infty 0} + A\Phi_0 \exp(i\omega t)$, where A is a constant and ω is the frequency. Linearizing the gas-dynamical equations for an isothermal flow and the conditions at discontinuities relative to the chosen main flow, we derive a relationship between A and the perturbation amplitudes of the S- and I-front coordinates.

However, an important feature that arises when Eq. (1) is linearized should be noted. The point is that under our assumption of a uniform flow behind the I-front, it would be natural to assume that $(l - x_{\rm I}) \to \infty$. At the same time, for large $(l - x_{\rm I})$, there is generally no limit of the integral that emerges when Eq. (1) is linearized.

Indeed, denote the small deviations of Φ, ρ, and $x_{\rm I}$ from their values in the main flow by Φ', ρ', and $x'_{\rm I}$, respectively. We then have from Eq. (1)

$$\Phi' = A\Phi_0 \exp(i\omega t) - \frac{2\alpha\rho_2}{m_{\rm H}^2}\int_{x{\rm I}}^{l} \rho' dx + \alpha(\rho_2/m_{\rm H})^2 x'_{\rm I}, \qquad (2)$$

$$\rho' \propto \exp[i(\omega t - k_2 x)], \qquad k_2 = \omega/c_2.$$

Formula (2) takes into account the fact that the linearized gas-dynamical equations have solutions of the running-wave type with frequency ω.

We see from relation (2) that for real ω and for $(l - x_{\rm I})k_2 \gg 1$, the integral on the right-hand side of (2) is a rapidly varying function and its limit for $l \to \infty$ does not exist. Nevertheless, even an arbitrarily small damping of ρ' for $x \to \infty$ allows the asymptotic value of the integral to be calculated [see, e.g., the monograph of Nayfeh (1981)]. Physically, however, the dissipation of wave motions in H II regions may result from radiative cooling (Krasnobaev 2000), so

using the asymptotic expression for the integral in formula (2) seems to be justified.

Denoting the displacement amplitudes of the S- and I-fronts, respectively, by L_0D and L_0F (here, the choice of the spatial scale L_0 is arbitrary) and substituting (2) in the boundary conditions at the I- and S-fronts, we obtain the sought-for relationship between A, D, and F in the form

$$D = -\frac{ic_0\gamma_S}{\omega L_0}A; \qquad F = -\frac{ic_0\gamma_I}{\omega L_0}A;$$

$$\gamma_S = \frac{2M_0^2}{M_0^2-1}\frac{q_0\exp(iq(M_0-1)/(2M_0))}{1-\exp(iq)}(a+b\gamma_I); \quad (3)$$

$$\gamma_I = \frac{M_2-(1+M_2-2iM_2\beta)a}{(1+M_2-2iM_2\beta)b-(c_0/c_2)[1-i\beta M_2/(1+M_2)]}; \quad (4)$$

$$q_0 = \frac{M_0^2+1}{M_0^2-1}\frac{(1+M_2)^2}{1+M_2^2}; \qquad q = \frac{2\omega L}{c_0(1-M_1^2)}; \qquad M_n = \frac{u_n}{c_n};$$

$$a = \frac{1-\exp(iq)}{[(M_0-1)+(M_0+1)\exp(iq)]q_0-2i\beta[1-\exp(iq)]}; \quad b = \left(M_0-\frac{i\beta}{1+M_2}\frac{c_0}{c_2}\right)a;$$

$$\beta = \frac{1+M_2}{2\pi M_2}\frac{\tau}{\tau_r}; \qquad \tau = \frac{2\pi}{\omega}; \qquad \tau_r = \frac{m_H}{\alpha\rho_2}.$$

Clearly, the variations in the flux of ionizing radiation are largest at L at which the amplification coefficients $|\gamma_S|$ and $|\gamma_I|$ are largest. To determine these values, we note that according to formulas (3) and (4) for perturbations of fixed frequency ω, $|\gamma_S|$ and $|\gamma_I|$ are periodic functions of q. Therefore, it will suffice to consider the behavior of the amplification coefficients in the segment $[-\pi, \pi]$.

When $\beta = 0$, expressions (3) and (4) can be analyzed analytically. As a result, the maxima and minima of $|\gamma_S|$ and $|\gamma_I|$ are reached, respectively, at $q = 2m\pi$ and $q = (2m+1)\pi$, where $m = 0, \pm1, \ldots$. The quantities $|\gamma_S|$ and $|\gamma_I|$ corresponding to these values of q can then be easily determined from (3) and (4).

If, alternatively, $\beta \neq 0$, then it becomes appreciably more difficult to analyze the q dependence of γ_S and γ_I. Let us first estimate the values of β, which is proportional to the ratio of the perturbation period τ to the photorecombination time τ_r, that are of considerable interest.

It would be natural to choose an upper limit for τ to be close to the period of the fundamental oscillation mode of the IS-front. In that case, for typical conditions in the interstellar medium (assuming, for example, that $u_0 = 17$ km s^{-1}, $\lambda = \rho_r/\rho_2 = 80$, and the adiabatic Mach number $M_\infty = 10$), according to our previous calculations (Krasnobaev 2001a), we may set $\tau \sim 1.5\times10^{13}m_H/\rho_0$, where τ is in seconds.

We take $\alpha = 5\times10^{13}$ cm^3 s^{-1} as the characteristic photorecombination coefficient. This value of α corresponds to a temperature of about 4.5×10^3 K in the H II region. At higher temperatures, α decreases, causing β to decrease.

The derived estimates for τ and α allow us to restrict our analysis to the range $0 \leq \beta \leq 5$.

Figure 1 shows the numerically calculated dependence of $|\gamma_S|$ and $|\gamma_I|$ on q (the numbers near the curves correspond to different β). Naturally, the amplification coefficients are largest for $\beta = 0$. In this case, the maxima of $|\gamma_S|$ and $|\gamma_I|$ decrease by no more than 20% compared to those obtained by analyzing the interaction of the IS-front with isothermal waves in an H I region (Krasnobaev 2001b).

The maxima of $|\gamma_S|$ and $|\gamma_I|$ decrease with increasing β. Nevertheless, even at $\beta = 5$, the relative perturbations of the front positions (or velocities) are more than twice the relative variations in the flux density of the external radiation. Therefore, just as is the case for the interaction of the IS-front with an inhomogeneous interstellar medium (Krasnobaev 2001b), one might expect a time-varying flux of energetic photons to stimulate (at least for moderately large β) nonlinear oscillations of the IS-front. To quantitatively analyze the effect of Φ variability on the amplification of oscillations, we turn to numerical simulations.

NONLINEAR OSCILLATIONS OF THE IONIZATION–SHOCK FRONT IN A VARIABLE RADIATION FIELD

To take into account the variability of the ionizing radiation flux incident on the I-front, let us consider a plane one-dimensional unsteady flow of gas with volume heat losses through radiative cooling. The cooling rate $\Lambda(\rho, T)$ and the unperturbed motion of the S- and I-fronts are assumed to be the same as those in our previous calculations of the interaction of the IS-front with interstellar inhomogeneities (Krasnobaev 2001b). In general, the IS-front is assumed to be nonstationary, so the photon flux density referred to ρ_0u_0/m_H is $\Phi_0^* = \Phi_0m_H/\rho_0u_0 = 1 -$

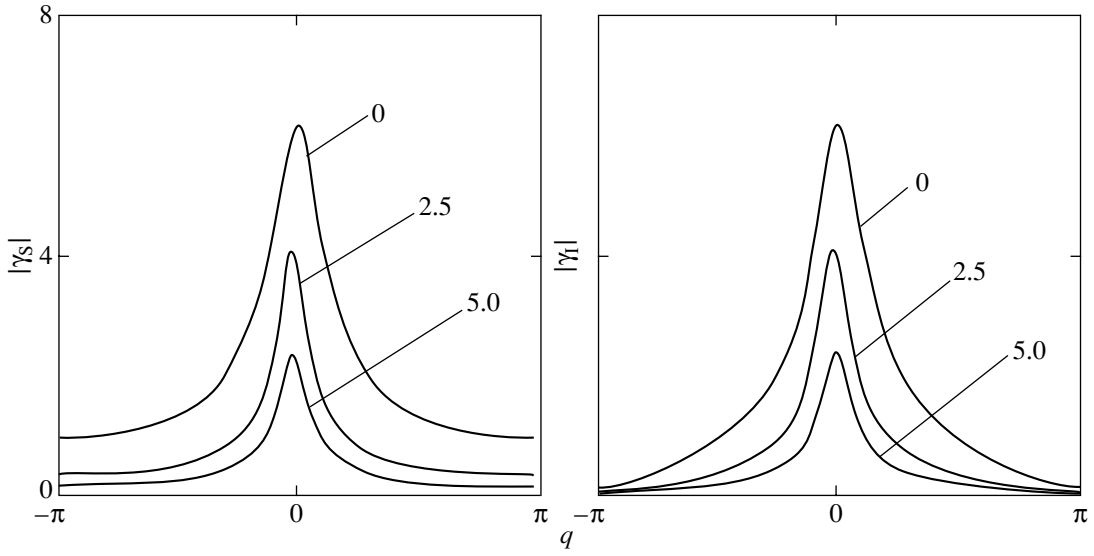


Fig. 1. Amplification coefficients versus q for various β.

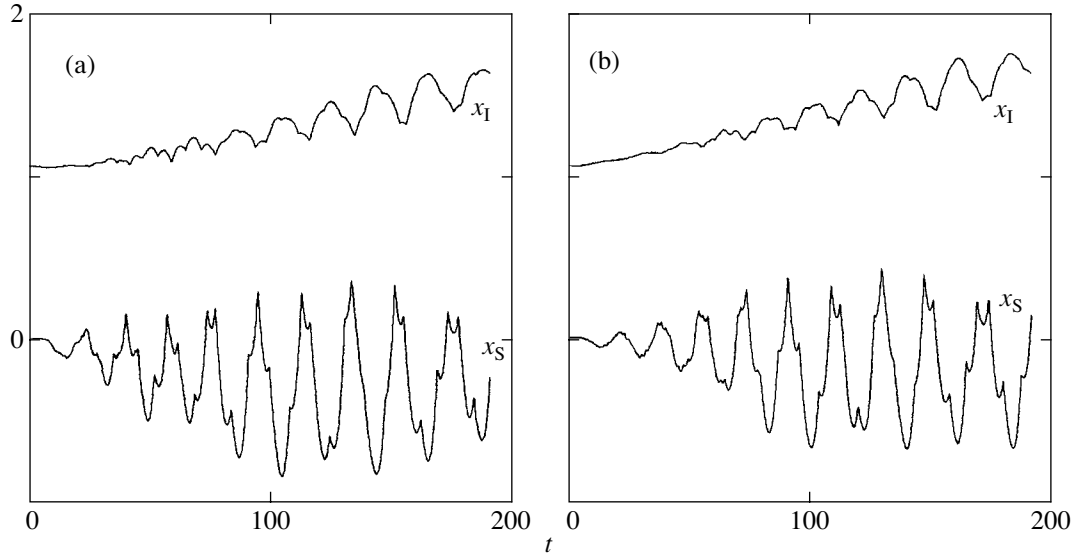


Fig. 2. Trajectories of the shock and the I-front in a variable external radiation field without (a) and with (b) photorecombinations.

$v_I M_0^2/u_0$, where v_I is the velocity difference between the S- and I-fronts.

However, the x_I and ρ perturbations now cause variations in Φ and the integral relation (1) must be used in the boundary conditions at the I-front. In our calculations, the distance l at which, according to (1), Φ_∞ should be specified, was a parameter and its value in each calculation was fixed for the unperturbed I-front. Since the effect of the photorecombination-produced absorption of Φ is clearly at a maximum for values of l close to half the wavelength of the perturbations generated in the H II region, we present below the results of our calculations precisely for these values of l.

We emphasize that our calculations were performed using nonlinear gas-dynamical equations by taking into account the actual dependence of the cooling rate on the flow temperature and nonuniformity in the region between the fronts.

Figure 2 gives an idea of the effect of photorecombination on the dynamics of the IS-front. In this figure and below, the coordinates of the S-front, $x_S(t)$, and I-front, $x_I(t)$, are referred to L_c and the time t is measured in units of L_c/u_0. The following quantity

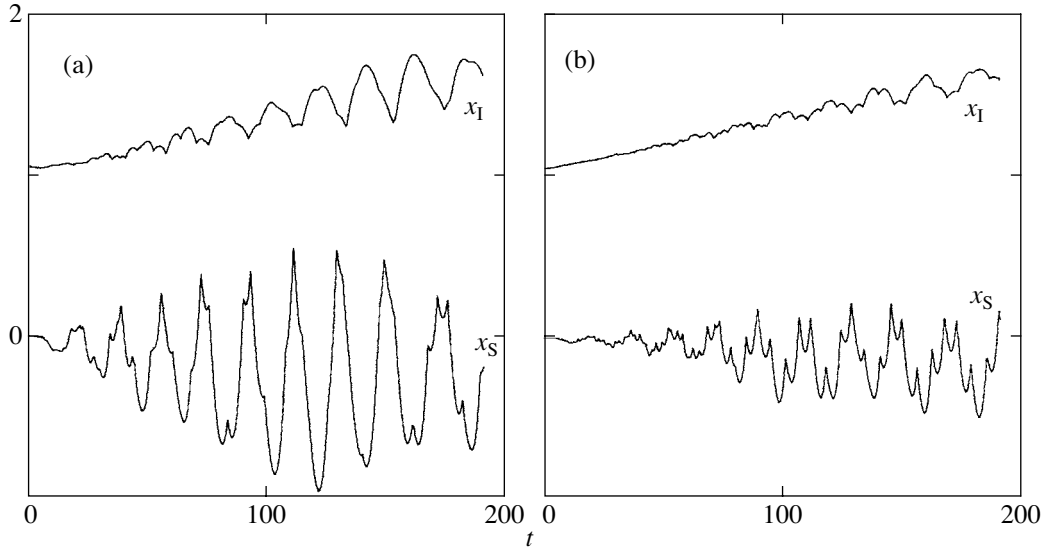


Fig. 3. Influence of the ratio of the neutral-gas and plasma densities, λ, on the I- and S-front oscillations: (a) $\lambda = M_0^2$ and (b) $\lambda = 80$.

was chosen as L_c:

$$L_c = \frac{k_B T_S}{m_H} \frac{\rho_S u_S}{\Lambda(\rho_S, T_S)},$$

where k_B is the Boltzmann constant; and ρ_S, u_S, and T_S are the gas parameters immediately behind the adiabatic discontinuity.

Figure 2a shows the front coordinates when the photon absorption is ignored (i.e., $\Phi = \Phi_\infty$) and the variability period of the radiation flux from the external source is $\tau_w = 18L_c/u_0$ and $A = 0.1$ ($\lambda = M_0^2 = \gamma M_\infty^2$; below, for definiteness, we take $\gamma = 5/3$, $M_\infty = 10$, $\Phi_0^* = 1/2$, and $u_0 = 17$ km s^{-1}).

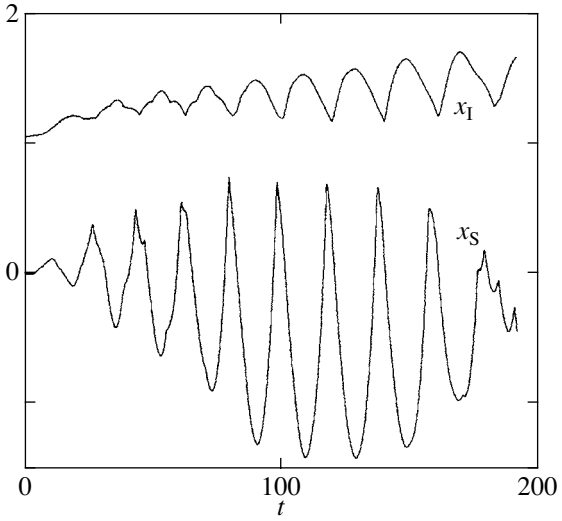


Fig. 4. Front oscillations in an inhomogeneous medium with an opaque plasma behind the I-front.

We can see that qualitatively and quantitatively, the Φ variability affects the development of nonlinear oscillations, much as is the case in an inhomogeneous medium (Krasnobaev 2001b). However, since the I-front is now the source of perturbations in the gas, its generated pressure waves accelerate rather than decelerate the shock.

Allowance for the absorption, naturally, results in a decrease of the oscillation amplitude, as evidenced by the $x_S(t)$ and $x_I(t)$ plots in Fig. 2b for the same values of τ_w, A, and λ as above ($\beta = 1.15$ corresponds to them).

The oscillation amplitude is significantly affected by λ. It decreases with decreasing λ (i.e., with increasing charged-particle density behind the I-front). This is illustrated in Fig. 3, where $A = 0.2$ and τ_u is the same as that in Fig. 2. Comparison of the curves in Figs. 2 and 3 with the same λ shows that an increase in A contributes to the growth of the front oscillation amplitude. However, for $\lambda = 80$ ($\rho_2 \approx 2\rho_0$, $\beta = 5.18$), the $x_S(t)$ and $x_I(t)$ perturbations are small. Thus, photorecombinations can play an important role in attenuating the oscillations if $\rho_2 > \rho_0$. Note only that the opposite inequality, $\rho_2 \le \rho_0$, generally holds for expanding H II regions. Therefore, the condition $\rho_2 > \rho_0$ can be satisfied only for flows where there is a mechanism that maintains a relatively high ionized-gas pressure (e.g., when an H II region is formed in an envelope with a density gradient).

It is also of interest to investigate the stabilizing effect of photorecombinations on the interaction of the IS-front with small interstellar inhomogeneities. It is qualitatively clear that reducing the I-front displacement, photorecominations thereby increase the

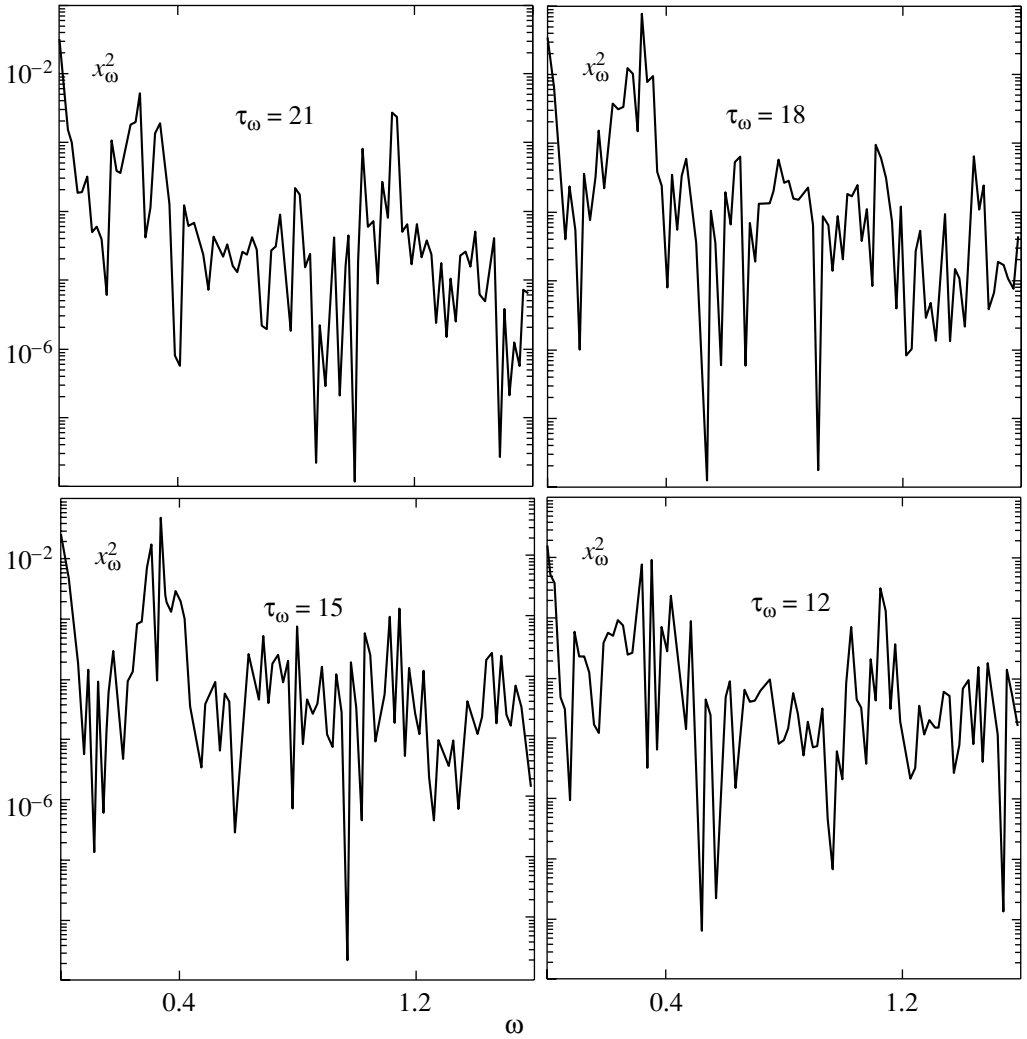


Fig. 5. Spectra of the shock oscillations generated by an external radiation source.

amplitude of the pressure waves reflected from the I-front.

This assumption is completely confirmed by calculations. In particular, Fig. 4 shows $x_{\rm S}(t)$ and $x_{\rm I}(t)$ for $A = 0$ (there are no variations of Φ_∞ with time), $\lambda = 80$, and $\tau_e = 17L_c/u_0$ (here, τ_e is the period of the isothermal waves propagating in an unperturbed interstellar medium; the amplitude of the relative density variations in the wave is taken to be 10^{-1}). Now, the characteristic features are the decrease in the mean velocity of the I-front (compared to its value for $\Phi = {\rm const}$) and the development of shock-front oscillations with an amplitude that even slightly exceeds the value reached in an optically transparent medium (Krasnobaev 2001b).

In conclusion, let us consider the τ_w dependence of the shock oscillation spectrum. As our calculations show, the principal features of the $x_{\rm S}(t)$ spectrum do not depend too strongly on the specific oscillation generation mechanism. Therefore, we take into account only the variations in external radiation flux for $\Phi = \Phi_\infty$ and restrict our analysis to the parameters $\lambda = \gamma M^2_\infty$ and $A = 0.1$ (see Fig. 2).

The $x_{\rm S}(t)$ spectra are shown in Fig. 5 for various τ_w (x_ω and ω are, respectively, the Fourier component and the frequency referred to L_c and u_0/L_c). They reflect the emergence of a mean displacement in the S-front noted above; the frequencies in the vicinity of $\omega_d \approx 0.35u_0/L_c$ close to the fundamental oscillation frequency of an isolated shock remain dominant as τ_w varies over the range $(12{-}21)L_c/u_0$. As τ_w increases (in the above range), the amplitudes of the harmonics with frequencies ω_d gradually increase and then rapidly decrease. Another characteristic feature is the Doppler splitting of the Fourier components, which is absent for a stationary IS-front.

Our calculations lead us to the following conclusions.

In the approximation of an optically transparent plasma layer behind the I-front, nonlinear oscillations of the IS-front are generated by a time-varying radiation flux almost as effectively as in an inhomogeneous interstellar medium.

The photon absorption through photorecombinations in an H II region significantly attenuates the oscillations only if the particle density in the ionized gas exceeds the particle density in the unperturbed medium.

When the IS-front interacts with low-amplitude waves propagating in the interstellar medium, photorecombinations have no stabilizing effect on the development of shock oscillations. On the contrary, the oscillation amplitude even (although slightly) increases.

In general, our calculations show that variability of the ionizing radiation flux incident on the I-front can be an additional factor (apart from small inhomogeneities in the interstellar medium) that contributes to the penetration of finite-amplitude pressure waves into an H II region. The corresponding neutral- and ionized-gas velocity perturbations are in agreement with the velocities of random motions observed in H II regions.

Thus, instability of the IS-front can serve as a mechanism of energy input to turbulent motions. It follows from our calculations that the neutral- and ionized-gas velocity perturbations can reach 1–3 km s^{-1} if the shock velocity is 10–17 km s^{-1} (the gas temperature behind the adiabatic discontinuity and the temperature in the H II region vary over the ranges $(2.33–6.74) \times 10^3$ K and $(0.44–1.31) \times 10^4$ K, respectively). Note that a velocity dispersion of this order of magnitude was obtained from observations by O'Dell and Wen Zheng (1992) near the ionization front in the Orion Nebula.

Nevertheless, the question as to the origin of the observed velocity irregularities is still an open question. The reason is that, apart from knowing the structure function of the velocities, we need information on the shock parameters and on the gas cooling ahead of the I-front. It is also difficult to compare the role of the above enhancement of inhomogeneities with the other types of instability investigated previously (thermal instability, Rayleigh–Taylor and Kelvin–Helmholz instabilities, Richtmeier–Meshkov instability). This is because the analysis was restricted either to the results of a linear theory or to qualitative conclusions [see, e.g., O'Dell and Wen Zheng (1992) and Keto and Ho (1989)], while we obtained the characteristics of the fluctuations in parameters of the medium for a nonlinear saturation of small initial perturbations.

REFERENCES

1. E. R. Keto and P. T. P. Ho, Astrophys. J. **347**, 349 (1989).
2. P. A. Kimoto and D. F. Chernoff, Astrophys. J. **487**, 728 (1997).
3. K. V. Krasnobaev, Astron. Zh. **47**, 1047 (1970) [Sov. Astron. **14**, 840 (1971)].
4. K. V. Krasnobaev, Astrophys. Space Sci. **274**, 307 (2000).
5. K. V. Krasnobaev, Pis'ma Astron. Zh. **27**, 112 (2001a) [Astron. Lett. **27**, 91 (2001a)].
6. K. V. Krasnobaev, Pis'ma Astron. Zh. **27**, 667 (2001b) [Astron. Lett. **27**, 568 (2001b)].
7. A. H. Nayfeh, *Introduction to Perturbation Techniques* (Wiley, New York, 1981; Mir, Moscow, 1984).
8. C. R. O'Dell and Wen Zheng, Astrophys. J. **387**, 229 (1992).
9. R. H. Rubin, Astrophys. J. **153**, 761 (1968).
10. L. Spitzer, Jr., *Physical Processes in Interstellar Medium* (Wiley, New York, 1978; Mir, Moscow, 1981).
11. R. Strickland and J. M. Blondin, Astrophys. J. **449**, 727 (1995).
12. R. Walder and D. Follini, Astrophys. Space Sci. **274**, 343 (2000).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 601–616. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 667–684.

Disk Wind in Young Binaries with Low-Mass Secondary Components: Optical Observational Manifestations

V. P. Grinin[1,2*] and L. V. Tambovtseva[1]

[1] *Pulkovo Astronomical Observatory, Russian Academy of Sciences, Pulkovskoe sh. 65, St. Petersburg, 196140 Russia*

[2] *Crimean Astrophysical Observatory, p/o Nauchnyi, Crimea, 334413 Ukraine*

Received March 20, 2002

Abstract—We consider a model of a young binary with a low-mass secondary component. Mass accretion from the remnants of the protostellar cloud onto the binary components is assumed to take place in accordance with current models; i.e., it proceeds mainly onto the low-mass component. The accretion is accompanied by mass outflow (disk wind), whose low-velocity component can be partially captured by the primary component. As a result, an asymmetric common envelope is formed. Its densest part is involved in the orbital motion of the secondary and can periodically shield the primary component of the binary from the observer. Assuming a standard dust-to-gas ratio for the disk wind (1 : 100), we calculated the possible photometric effects from such eclipses and showed that they could be observed even at moderate accretion rates onto the low-mass binary component, $\sim 10^{-8}$–$10^{-9} M_\odot$ per year. In this case, the parameters of the minima depend on the model of the disk wind, on the ratio of its characteristic velocity to the orbital velocity of the secondary, and on its orbital inclination to the line of sight. These results can form the basis for interpreting a wide range of phenomena observed in young stars, such as the activity cycles in UX Ori stars, the unusually broad minima in some young eclipsing systems, etc., and for searching for substellar objects and massive protoplanets. In addition, the peripheral parts of the gas and dust disk around a young binary can fall within the shadow zone produced by the opaque part of the common envelope. In such cases, a shadow from the common envelope must be observed on the disk; this shadow must move over the disk following the orbital motion of the low-mass component. Detection and investigation of such structures in the images of protoplanetary disks may become a method of searching for protoplanets and studying binaries at early stages of their evolution. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *young binary stars, accretion, disk wind, cyclic phenomena.*

INTRODUCTION

Recent studies show that binarity of young stars is a common phenomenon. It has a significant effect on the evolution of their surrounding matter, with this effect depending on the separation between the binary components [see Beckwith *et al.* (1990); Mathieu *et al.* (2000); and references therein]. The physical basis for the interaction of a binary with the remnants of the protostellar cloud seems clear enough: the periodic gravitational perturbations generated by the revolution of the binary components around a common center of mass. Such binaries have long been studied theoretically and these studies originate from an investigation of the protoplanetary disks perturbed by giant planets (Goldreich and Tremain 1982; Lin and Papaloizou 1993; Artimowicz and Lubow 1994, 1996; Bate and Bonnell 1997; Kley 1999; see also Lubow and Artimowicz 2000). Calculations show that a matter-free cavity is produced by tidal perturbations in the inner region of the gas and dust disk around a young binary [below called a CB (circumbinary) disk]. The characteristic size of the cavity depends on the component mass ratio $q = m_2/m_1$ and orbital eccentricity; it is equal to triple the binary semimajor axis in order of magnitude. Mass flows from the CB disk periodically penetrate into this cavity under the effect of viscosity and gravitational perturbations; they maintain the accretion activity of the primary component (Artimowicz and Lubow 1996).

An accretion disk is also formed around the secondary component. In this case, as was shown by Bate and Bonnell (1997), the ratio of the accretion rates onto the secondary and primary components, $Q = \dot{m}_2/\dot{m}_1$, depends on the component mass ratio q and on the specific angular momentum j_d of the accreted matter: the larger j_d is, the larger Q is. If j_d is larger than the orbital specific angular momentum and if $q \ll 1$, then Q can be much larger than unity.

*E-mail: VG1723@spb.edu

1063-7737/02/2809-0601$22.00

Since a young binary primarily accretes matter with a low specific angular momentum during its evolution, the parameter j_d must increase with time.

A situation that is unusual at first glance arises: the low-mass component of the binary can become a major "consumer" of matter from the remnants of the protostellar cloud as it evolves. As was shown by one of us (Grinin 2000a; paper I), this may lead to a large-scale redistribution of the matter surrounding the binary. The point is that by accreting matter from the CB disk, the low-mass component loses part of it in the form of a disk wind to become a source of matter that can rise high above the binary plane. As a result, an extended asymmetric gas–dust envelope is formed in the binary. Calculations show (Grinin 2002; paper II) that the dust contained in the envelope can absorb an appreciable fraction of the radiation from the primary component and can be a dominant source of near-infrared (1–10 μm) binary radiation. In this paper, we consider the photometric effects that can be observed at optical wavelengths due to the orbital motion of the common envelope.

DISK WIND AS A SOURCE OF MATTER IN YOUNG BINARIES

Because of the conservation of angular momentum, disk accretion is always accompanied by mass outflow from the accretion disk, with which an excess angular momentum of the accreted matter is carried away (Shakura and Sunyaev 1973; Lynden-Bell and Pringle 1974). The mass outflow originates from the disk surface and produces the so-called disk wind (Blandford and Payne 1982). This phenomenon was studied most extensively in T Tauri stars from forbidden-line profiles (Kwan and Tademaru 1988). According to Hirth *et al.* (1997), the gas velocity in the wind periphery ($r > 1$ AU) in stars of this type can be only a few tens of kilometers per second. Therefore, if such a star is a member of a binary with a more massive primary component, then the matter of its wind disk can be partially or completely captured by the primary even if the separation between the binary components is of the order of several AU. As a result, an asymmetric common envelope emerges.

Note that apart from the low-velocity wind component discussed above, high-velocity mass flows (bipolar outflows and jets) are observed in young stars. These are formed in the parts of the accretion disk closest to the star [see Eislöffel *et al.* (2000) and references therein] and are also seen in the profiles of forbidden lines (Hartigan *et al.* 1995; Hirth *et al.* 1997). However, calculations indicate (Goodson *et al.* 1999) that the contribution of this wind component to the total mass loss is relatively small, ~20%. Therefore, the high-velocity component is disregarded in the simplified model of a disk wind considered below.

The formation of a common envelope in the way described above depends on the accretion rate onto the low-mass component ($\dot{M}_a = \dot{m}_2$) and on disk-wind parameters. One of these parameters, $f = \dot{M}_w/\dot{M}_a$, is the ratio of the mass-loss rate in the disk wind, $\dot{M}_w$, to the accretion rate onto the low-mass component. It specifies the fraction of the matter accreted from the CB-disk onto the secondary component that is returned to the binary in the form of a disk wind. Below, as in paper II, we take $f = 0.2$, typical of magnetocentrifugal models of accretion disks (Shu *et al.* 1993).

The second parameter V_w/V_k is the ratio of the terminal wind velocity V_w to the orbital velocity of the secondary component V_k. It specifies the fraction of the disk-wind matter p captured by the primary component of the binary.

THE MODEL

Given the aforesaid, let us consider the following model.

(1) A young binary is surrounded by a CB-disk whose matter is accreted onto its components. Since, according to current estimates (see, e.g., Natta *et al.* 2000), the circumstellar-disk masses generally do not exceed $0.1M_\odot$, the self-gravitation of such disks may be ignored.

(2) The component mass ratio is $q \ll 1$. Note that, according to Duquennoy and Mayor (1991), the distribution function of binary component mass ratios has a maximum at $q \approx 0.2$; it is reached mostly due to long-period systems with periods $P > 1000$ days (Mazeh *et al.* 1992). Therefore, the above constraint on q does not narrow the validity range of our model too much for long-period binaries.

(3) Mass accretion from the remnants of the protostellar cloud is assumed to proceed mainly onto the low-mass component. Its orbit is assumed to be circular. In this case, the accretion rate $\dot{M}_a$, a parameter of the problem, does not depend on the orbital phase. For the same reason, the rate of outflow from the accretion disk of the low-mass component is also the same at all orbital phases.

(4) The disk wind is assumed to be azimuthally homogeneous in the coordinate system of the secondary component. Thus, we ignore the effect of the tidal force from the primary component. Note that, as our estimates show (see paper II), this effect is actually small in sufficiently wide pairs. Although a wind velocity close to the terminal velocity is reached at large distances from the secondary (comparable to the Hill radius in order of magnitude), to simplify our

calculations, we assume the mass outflow to originate from a point source that coincides in position with the secondary.

When we pass to the coordinate system of the primary, the wind particle velocity vector $\mathbf{V}_w$ is added to the orbital velocity vector of the secondary $\mathbf{V}_k$. As a result, the distribution of initial particle velocities

$$\mathbf{V}_0 = \mathbf{V}_w + \mathbf{V}_k \quad (1)$$

becomes anisotropic.

We consider two wind models. One model assumes that mass outflow from the accretion disk takes place at a fixed angle α to the disk symmetry axis (this angle is called below the wind opening angle). The other model assumes that the matter density in the wind and its velocity V_w do not depend on latitude within the wind opening angle (from $\alpha_{\min}$ to $\alpha_{\max}$). The condition for wind mirror symmetry about the accretion-disk plane is also assumed to be satisfied. The wind velocity V_w, in units of the orbital velocity of the secondary, is a parameter of the problem. We take $\alpha_{\min}$ to be 45°. Note that such an opening angle of the circumstellar disk has recently been obtained by Men'shchikov *et al.* (1999) for HL Tau, one of the best-studied T Tauri stars. In the wind model with a fixed outflow angle, we consider two cases: $\alpha = 60°$ and 45°. The latter value is close to the angle that corresponds to the peak in the latitude distribution of the mass outflow rate in the magnetocentrifugal wind models of Goodson *et al.* (1999).

Thus, our model is the limiting case of a young binary in which a common envelope is produced mainly by the secondary component.

The Method of Calculation

In this case, the problem reduces to calculating the trajectories of the particles ejected by the disk wind from the low-mass component during its orbital motion in the gravitational field of the primary component. As in paper II, for its solution, we use the ballistic approximation in which the motion of each particle is assumed to be independent of other particles. The application of this method in our case is justified by the fact that the mechanism of mass outflow from an accretion disk involving MHD turbulence facilitates the formation of a clumpy wind structure composed of weakly interacting fragments of various scales. The capture of the disk wind by the primary leads to the fact that some of the wind fragments move along the trajectories that initially rise above the binary equatorial plane and subsequently sink and cross it. Because of the assumed wind mirror symmetry relative to the equatorial plane, the same fragments cross it from below. During collisions with one another, the vertical fragment velocities cancel each other out. As a result, a geometrically thin gas and dust disk is formed in the central part of the binary,[1] whose matter is accreted onto the binary components. Modeling this disk is a complex gas-dynamical problem and is beyond the scope of this paper.

To calculate the particle number density in the common envelope, we broke down the orbit of the secondary component into discrete segments at 5° steps. The orbital displacement of the component by one step was accompanied by an isotropic ejection of test particles either at a fixed angle to the accretion-disk plane or within the assumed opening angle of the disk wind: $\alpha_{\max} > \alpha > \alpha_{\min}$. Depending on the wind model, the number of particles ejected at each step varied between 72 and 144. The initial particle velocity vector at each orbital step was determined from relation (1). The motion of each particle was calculated until it escaped from the binary (the limiting distance from the primary is taken to be 20 orbital radii of the secondary) or crossed the binary equatorial plane. Our calculations show that to obtain a steady-state mass distribution in the common envelope, it will suffice to follow the motion of the ejected particles during several orbital periods of the binary.

THE SPATIAL STRUCTURE OF THE COMMON ENVELOPE

Figure 1 shows the models of common envelopes calculated by the method described above. The dimensionless wind velocity V_w was assumed to be 0.25, 0.5, 1, and 1.5. The orbital radius of the secondary was taken to be unity. We consider three cases: in one case, the mass outflow from the secondary is isotropic within the range of angles $75° > \alpha > 45°$; in the other two cases, the outflow takes place at a fixed angle, $\alpha = 45°$ and 60°, to the disk symmetry axis. Since the disk wind is assumed to be mirror-symmetric about the accretion-disk plane, Fig. 1a shows only the upper half of the common envelope. In order not to complicate the figures, we do not show the disk wind produced by the accretion disk of the primary. Because of the assumed large disparity between the accretion rates onto the secondary and primary components, its effect on the common-envelope structure and brightness modulation of the binary cannot be large.[2]

[1] Actually, because of the velocity and density fluctuations in the wind, the vertical velocities completely cancel out only on the average. As a result, the disk must have a finite thickness.

[2] Because of the larger mass of the primary, its contribution may be more significant in the formation of high-velocity outflows. During collisions with them, part of the common envelope must be swept up from the binary in polar directions and this factor must be taken into account in a more careful modeling.

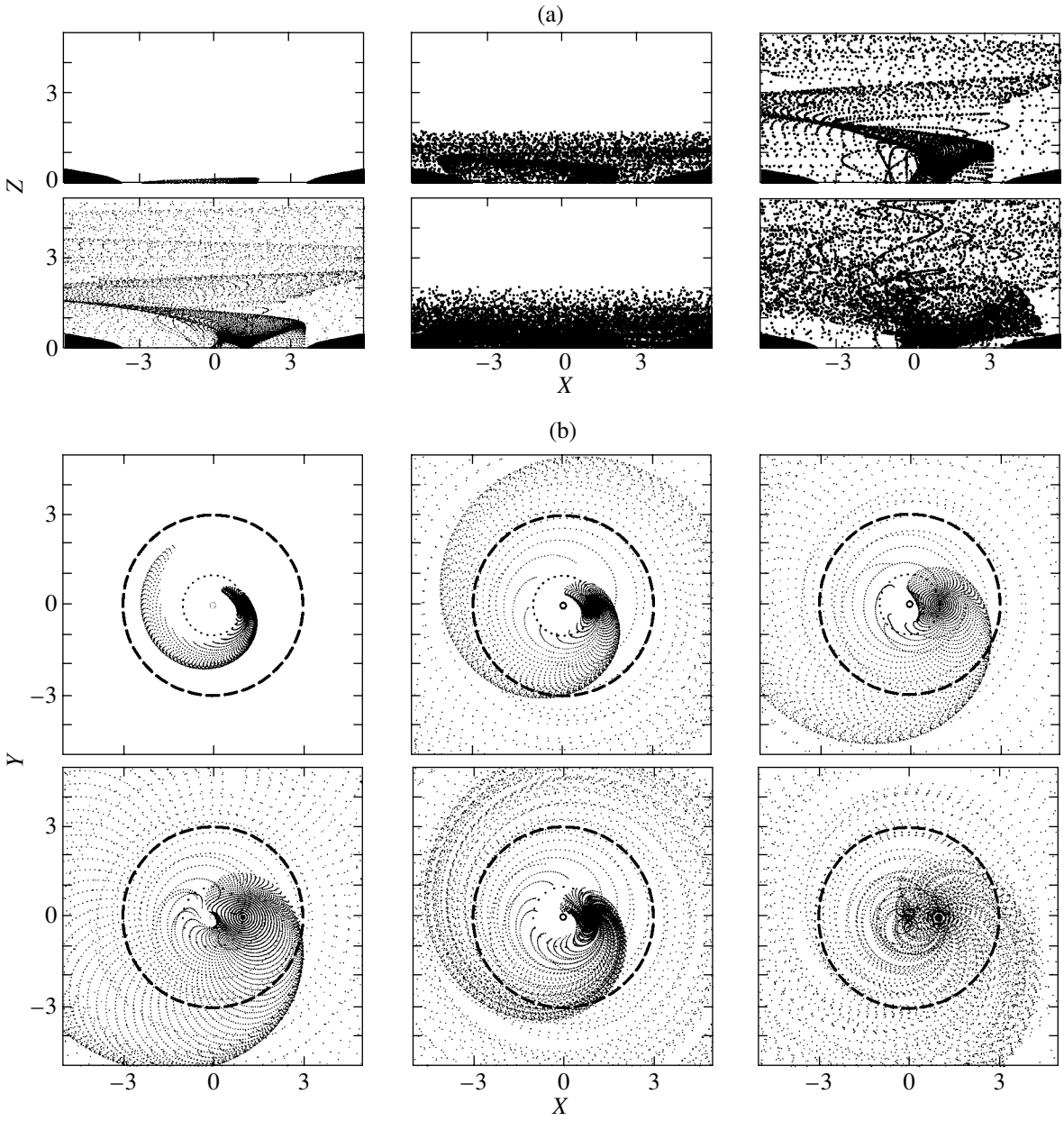


Fig. 1. The particle distributions in the models of common envelopes around young binaries in projection onto the XZ plane (a) and onto the binary equatorial XY plane (b). The coordinates of the primary and secondary components in the XY plane are (0,0) and (1,0), respectively. The thin and thick dotted lines indicate, respectively, the orbit of the secondary and a schematic inner boundary of the CB disk (its section is schematically shown in Fig. 1a). The model parameters, from left to right, are as follows: upper panel: $V_w = 0.25$, 0.5, 1, $\alpha = 45°$; lower panel: $V_w = 1$, $\alpha = 60°$; $V_w = 0.5$, $\alpha = 45°–75°$; $V_w = 1.5$, $\alpha = 45°–75°$.

We see from Fig. 1 that the orbital motion of the secondary component in the binary produces an asymmetric common envelope whose densest part has a shape resembling the tail of a comet as it passes near the Sun. At low wind velocities, this envelope fills a small volume near the low-mass component. At velocities comparable to the orbital velocity, it becomes more extended, the wind particles rise high above the binary plane and form a decaying spiral pattern on its periphery typical of some models for close binaries (Mastrodemos and Morris 1999). At velocities higher than the escape velocity, the disk-wind matter leaves the binary. No common envelope is formed in this case.

The Inner Region

As was pointed out in the Introduction, under tidal perturbations, the central region of a young binary is

rapidly freed from matter and the accretion activity of the components is maintained by two mass flows that periodically penetrate into this region from the CB-disk (Artimowicz and Lubow 1996). As was estimated by these authors, the central cavity is filled with matter about 10% full through this replenishment; i.e., it remains almost empty as before. In our model, the capture of the disk wind from the secondary by the primary produces an additional mass inflow into this region. Because of the mirror symmetry of the disk wind about the binary equatorial plane, the captured wind fragments fall to this plane from above and from below and the vertical velocities cancel each other out during their collisions. At the same time, the azimuthal velocity is preserved. As a result, a layer of matter is formed in the equatorial plane; subsequently, it is accreted onto the binary components. Analysis of the particle distribution in this region indicates that it depends on the disk wind velocity V_w and the angle α at which the particles are ejected from the accretion disk of the secondary. At high velocities, almost all particles leave the binary and the filling of the central cavity with matter is negligible in this case. A similar result, but for a different reason, is obtained at low wind velocities: in this case, the matter captured by the primary is concentrated mainly near the orbit of the secondary (Fig. 1). Thus, the filling of the neighborhood of the primary with matter is at a maximum in the models with intermediate velocities: $V_w \approx 0.5$–1.5. The situations where some of the wind fragments have both positive and negative angular momenta are theoretically possible. In that case, a subsystem of gas with a low angular momentum can emerge in the immediate vicinity of the primary.

As an illustration of this peculiarity, Fig. 2 shows the azimuthal velocity of the particles at the times of their fall to the equatorial plane as a function of the distance from the point of fall to the primary. The calculations were performed for three cases:

$$V_w = 1, \quad \alpha = 45^\circ; \qquad V_w = 1.5, \quad \alpha = 45^\circ;$$
$$V_w = 1.5, \quad \alpha = 60^\circ.$$

In the first and second cases, all and almost all particles have direct rotation, respectively; in the third case, an appreciable fraction of the particles falling to the equatorial plane at distances $r = 0.4$–0.5 from the center have reverse rotation. Our calculations indicate that approximately the same number of particles fall within this region with the same, in absolute value, but positive azimuthal velocity. They add up to form a ring around the primary composed of matter with a nearly zero angular momentum. Matter from this region will be accreted onto the primary in a regime close to free fall, moving in highly eccentric orbits. As we will see below, this peculiarity is of great interest in the case of UX Ori stars.

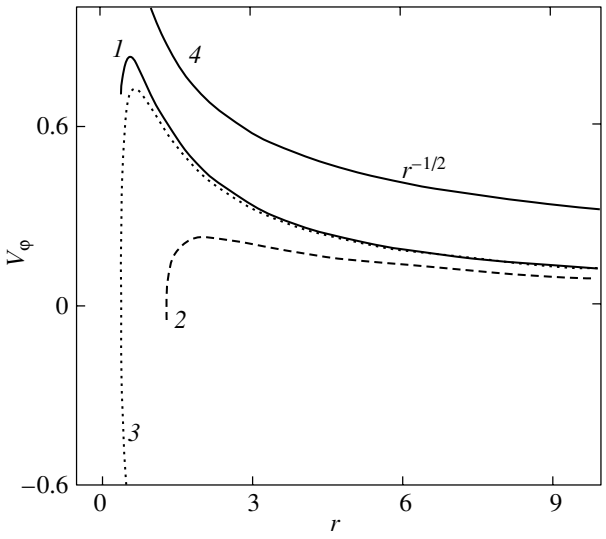


Fig. 2. Azimuthal velocity of the wind particles captured by the primary at the time of their fall to the binary equatorial plane versus distance to the primary in the models: (*1*) $V_w = 1$, $\alpha = 45^\circ$; (*2*) $V_w = 1.5$, $\alpha = 45^\circ$; (*3*) $V_w = 1.5$, $\alpha = 60^\circ$; (*4*) circular Keplerian velocity.

We also see from Fig. 2 that in all three models, the azimuthal velocity of the matter at the time of its fall to the equatorial plane is lower than the circular Keplerian velocity at the same distance from the center. This implies that the capture of the wind-disk matter by the primary must cause the accretion rate onto the primary to increase.

OPTICAL PROPERTIES OF COMMON ENVELOPES

As a result of the orbital motion of the secondary component and its associated densest part of the common envelope, a modulation of the particle column density between the observer and the primary component of the binary must be observed. Let us estimate the possible optical effects produced by dust grains in the common envelope. Denote the column density of test particles in the common envelope in the direction that makes the angle θ with the binary equatorial plane by $N(\theta, \phi)$. Its behavior with orbital phase ϕ determines the structure of the photometric minima that must be observed when there is a sufficient amount of dust on the line of sight.

The Phase Dependence $N(\theta, \phi)$

In Fig. 3, the column density N of test particles is plotted against azimuth ϕ for three orbital inclinations to the line of sight: $\theta = 5^\circ$, 10°, and 15°. The column section is a square whose side is taken to be 0.2 of the orbital radius of the secondary. We consider nine

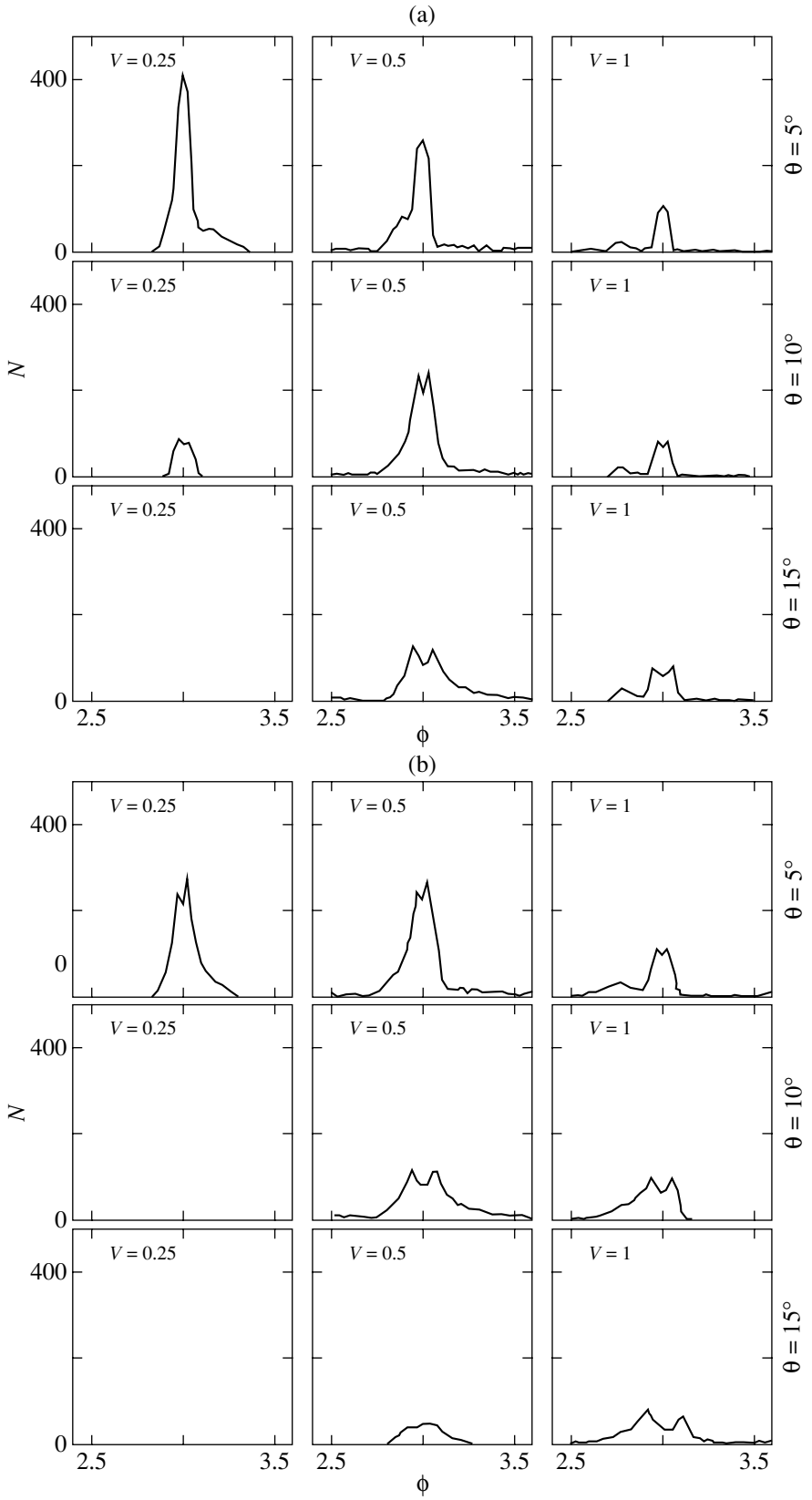


Fig. 3. (a) The phase dependence of the column density of test particles, $N(\theta, \phi)$, in the models with $\alpha = 45°$. The angle θ between the line of sight and the binary equatorial plane and the dimensionless wind velocity V_w are indicated in the plots (the subscript w was omitted for brevity); (b) same as case (a) for the wind model with the opening angle $\alpha = 60°$; (c) same as case (a) for the wind model with the opening angle $\alpha = 45°–75°$.

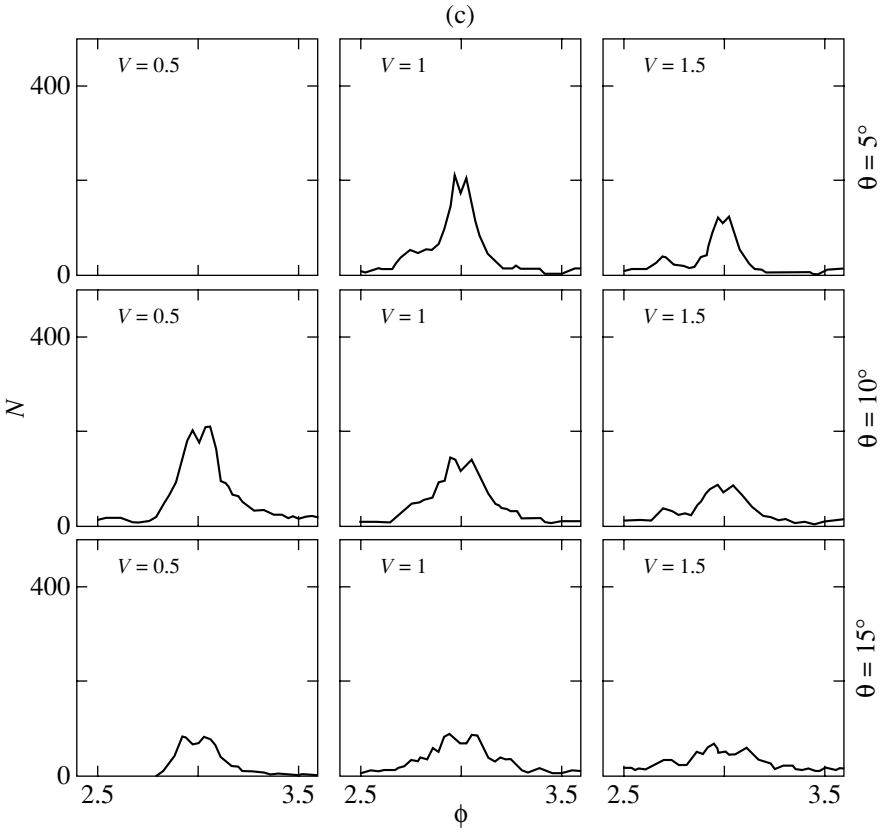


Fig. 3. (Contd.)

models that differ by the dimensionless wind velocity V_w and opening angle. The behavior of N with ϕ depends mainly on θ and the wind velocity V_w. At low V_w, the common envelope is compact, particularly in the vertical direction (Fig. 1a). Therefore, an appreciable particle column density is observed only at small angles θ and at orbital phases close to $\phi = 1,\ 2,\ 3, \ldots, K \ldots$, when the secondary passes at a minimum distance from the line of sight. In this case, the dependence of N on ϕ is highly asymmetric and exhibits a steep rise and a slower fall.

As the wind velocity V_w increases, the envelope becomes more extended both in the XY plane and in the vertical direction. As a result, an appreciable column density can also be observed at an appreciable inclination of the binary plane to the line of sight. The pattern of asymmetry in $N(\theta, \phi)$ also changes: the increase in N during ingress occurs more slowly than its decrease during egress. In addition, in the models with $V_w = 1$ and 1.5 at small θ, the dependence of N on ϕ has two maxima. They are not related to the splitting of $N(\theta, \phi)$ near the phase $\phi = K$ ($K = 1,\ 2,\ 3, \ldots$) caused by the passage of the cone of the disk wind from the secondary along the line of sight but are attributable to a peculiar mass distribution in the inner region of the common envelope.

To elucidate this peculiarity, Fig. 4 shows the particle distribution in the model with $V_w = 1$ and $\alpha = 45°$ near the equatorial plane ($|z| < 0.3$) at the time $t = K$ when the secondary component, having made K turns, is at the point with the coordinates $x = 1$ and $y = 0$. The system of almost concentric circumferences around the secondary is formed by the disk-wind particles that were ejected within several steps before the time $t = K$ and that have not risen above the equatorial plane higher than $|z| = 0.3$. These particles give the primary, slightly split minimum (Fig. 5). Some of them are captured by the primary and they must fall to the binary equatorial plane during their motion. We see from Fig. 4 that these particles form a complex configuration of points with a crowding in the upper part of the figure. This crowding of particles gives a minimum leading in time (in Fig. 5, it is observed near phase $\phi \approx K - 0.25$, where K is an integer). Analysis indicates that these particles were ejected by the disk wind in the direction opposite to the orbital motion of the secondary. That the particles moving in their orbits at a velocity lower than the orbital velocity give a minimum leading in time can be explained by the fact that they were ejected during the *preceding* orbital passage of the secondary.

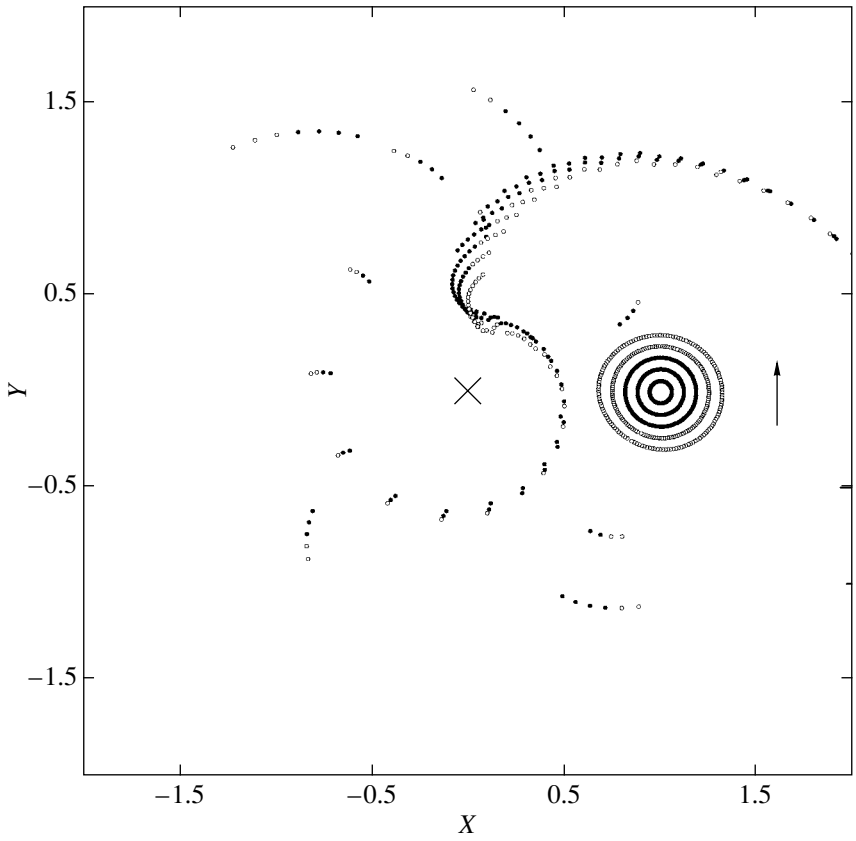


Fig. 4. The distribution of the particles in the common envelope in projection onto the binary equatorial XY plane that are in the $|z| < 0.3$ layer at the time corresponding to the phase $\phi = K$. The model parameters are $V_w = 1$, $\alpha = 45°$; the secondary is at the point with the coordinates $y = 0$ and $x = 1$ (see the text for details). The arrow indicates the direction of orbital motion of the secondary component.

Our calculations show that the presence of a double maximum in the phase dependence of the particle column density and, accordingly, a double photometric minimum (Figs. 5a, 5b) depends on the wind model, including the particle distribution in ejection angle α. For example, in the model with $V_w = 1$ and $\alpha = 45°$–$75°$, this feature is smoothed out (Fig. 3c) and a gradual increase in $N(\theta, \phi)$ that ends with a steeper decrease is observed instead of the double maximum.

Thus, we see that the shapes of the minima produced by eclipses of the primary component by an asymmetric common envelope can vary widely. They depend both on the model of the disk wind from the secondary (the wind opening angle and dimensionless velocity) and on the inclination of the binary plane to the line of sight.

The Amplitudes of Photometric Minima

To estimate the amplitude of optical variability in a binary, we assume that the matter of its CB-disk is a mixture of gas and dust with the standard dust-to-gas ratio for the interstellar medium (1 : 100). It would also be natural to take the same dust-to-gas ratio for the accretion disk of the secondary component (except for its inner hot region). Safier (1993) showed that as a result of collisions with gas atoms, the dust grains in the disk must be entrained by the atoms and they must be present approximately in the same proportion in the disk wind. Hence, for our estimates, we take the same dust-to-gas ratio (1 : 100) in the wind and in the common envelope produced by it.

Taking this into account, we calculated the dust distribution in a binary for the following parameters.

(1) The accretion rate onto the low-mass component is $\dot{M}_a = 10^{-8}$, 3×10^{-9}, and $10^{-9} M_\odot$ per year. According to Calvet *et al.* (2000), $\dot{M}_a \sim 10^{-8} M_\odot$ per year correspond to the mean accretion rate onto a classical T Tauri star. The values of $\dot{M}_a \sim 10^{-8}$–$10^{-9} M_\odot$ per year are obtained in current models for the formation of giant planets in the phase of intense accretion (Kley 1999).

(2) We assumed the standard ratio of graphite and silicate particles for the interstellar medium: $N_{\rm sil}/N_{\rm gra} = 1.12$ (Draine and Lee 1984) and consider

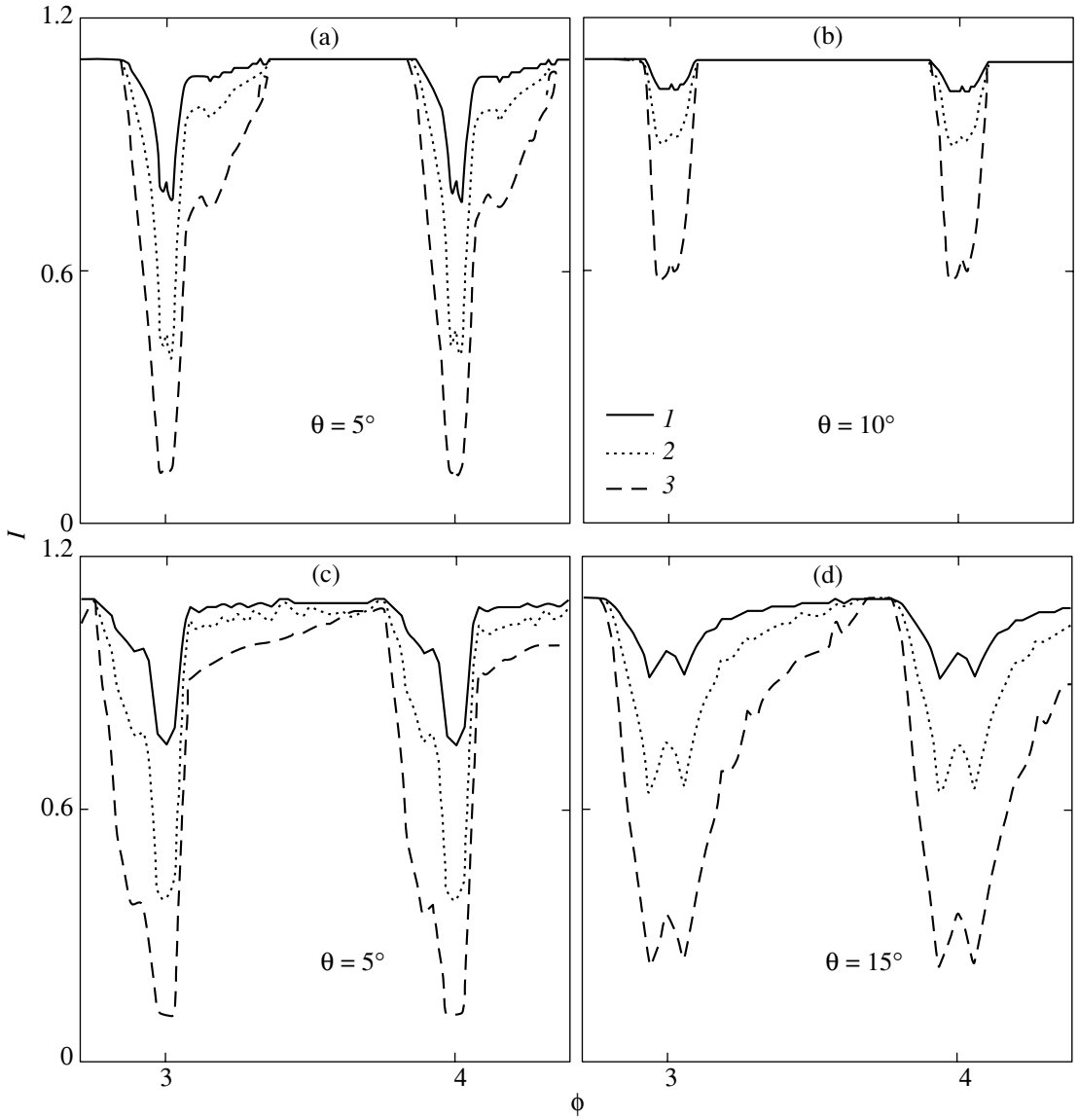


Fig. 5. Theoretical light curves for the models of common envelopes. The opening angle of the disk wind in all cases is $\alpha = 45°$, the accretion rate onto the low-mass component is $\dot{M}_a = 10^{-9} M_\odot$/yr (*1*), $3 \times 10^{-9} M_\odot$/yr (*2*), and $10^{-8} M_\odot$/yr (*3*) for $V_w = 0.25$ (a, b) and $V_w = 0.5$ (c, d). The angle θ between the line of sight and the binary plane is indicated in the plots.

spherical grains of radius $a = 0.1$ μm and density 3 g cm^{-1}. Such particles were chosen, because, as estimated by various authors (Voshchinnikov and Grinin 1991; Rostopchina *et al.* 1997; Natta and Whitney 2000), the reddening of young stars due to variable circumstellar extinction and the wavelength dependence of their intrinsic polarization due to light scattering by circumstellar dust are produced by grains of approximately the same composition as in the interstellar medium that differ only by slightly larger sizes.

We passed from the column density of test particles $N(\theta, \phi)$ calculated in the preceding section to the column density of actual grains $N_\mathrm{d}(\theta, \phi)$ with the above parameters by rescaling. In this procedure, we took into account the ratio of the total number of actual particles ejected by the wind during one orbital period to the corresponding number of test particles. We also took into account the difference in the column sections: the column density of test particles was calculated for a square section, 0.2×0.2 of the orbital radius in size, while the grain column density in the formula for the optical depth of a common envelope

$$\tau(\theta, \phi) = N_\mathrm{d}(\theta, \phi) \pi a^2 Q_\mathrm{ext} \qquad (2)$$

was calculated for a section with an area of 1 cm^2. In this case, the orbital radius of the secondary was taken to be 10 AU. The calculations were performed for a wavelength of 0.55 μm, which corresponds to the

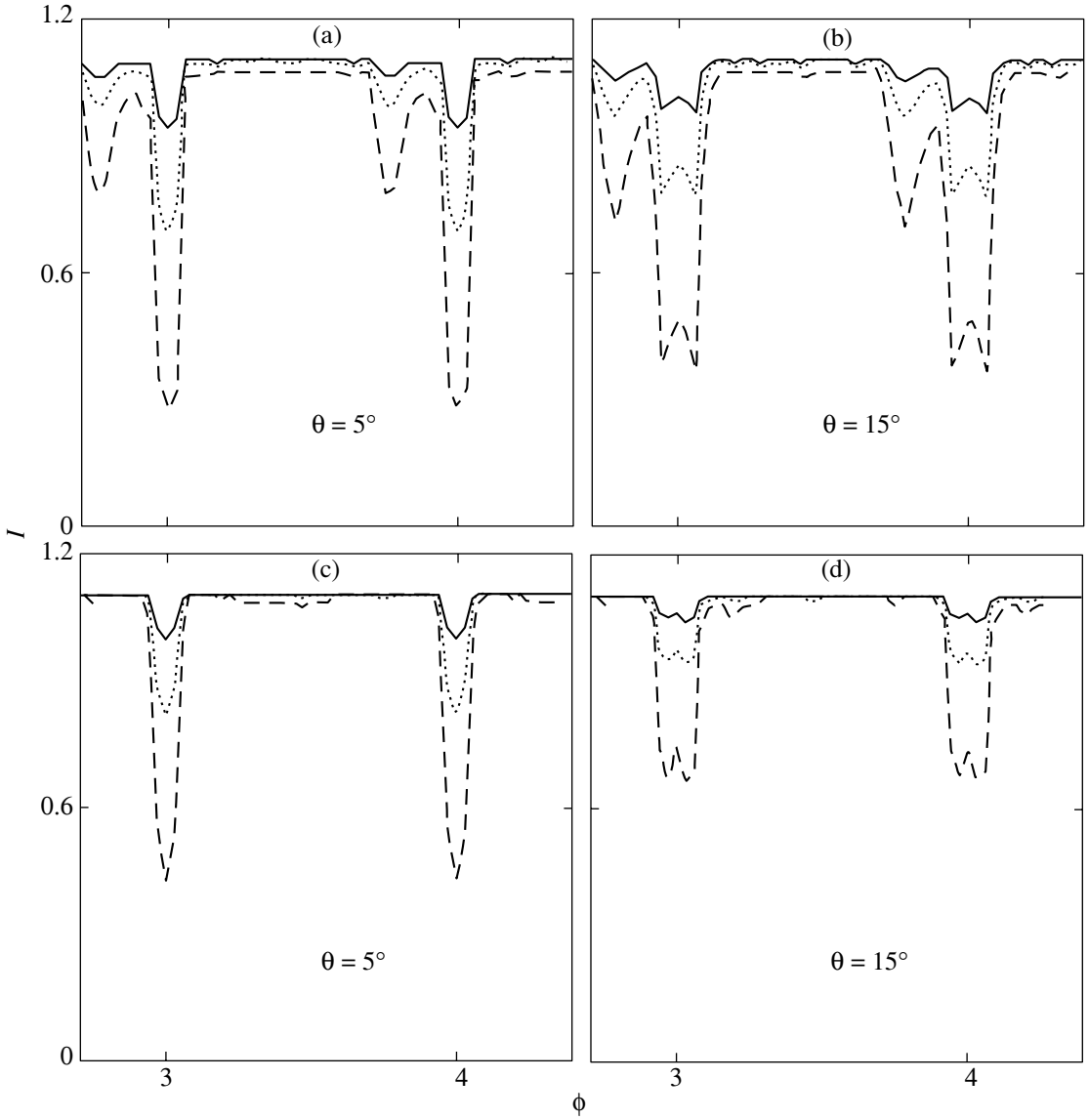


Fig. 6. Same as Fig. 5 for $V_w = 1$ (a, b) and $V_w = 2$ (c, d).

maximum transmission of the V photometric band. The quantity $Q_{\rm ext}$ was calculated by using the Mie theory. The required optical parameters of graphite and silicate (we used the so-called astrosilicate) were taken from Draine (1985).

Figures 5 and 6 show the models of photometric minima calculated using the formula

$$I(\theta, \phi) = I_* e^{-\tau(\theta,\phi)} + I_{\rm sc}. \qquad (3)$$

Here, I_* is the intensity of the stellar radiation (we took $I_* = 1$) and $I_{\rm sc}$ is the intensity of the radiation scattered by circumstellar dust.

It should be noted that the latter quantity includes the scattered radiation from the asymmetric common envelope and, therefore, depends on θ and orbital phase ϕ. However, since the intensity of the scattered radiation from young stars is generally low compared to the intensity of their direct radiation, we disregard the dependence of $I_{\rm sc}$ on phase ϕ and assume that $I_{\rm sc} = {\rm const} = 0.1 I_*$. Note that in eclipsing binaries with unseen secondary components, scattered light limits the amplitude of the minima during total eclipses, just as is the case for UX Ori stars. The occasional minima in the latter are attributable to the screening of the stars by circumstellar dust clouds (Grinin 1988). In this case, $I_{\rm sc}$ in Eq. (3) is equal in order of magnitude to the intensity of the scattered radiation from the protoplanetary disks of UX Ori stars.

We see from Figs. 5 and 6 that the photometric effect produced by the orbital motion of the densest part of the common envelope can be significant even

at moderate accretion rates for young stars, $\dot{M}_a = 10^{-8} M_\odot$ per year. In this case, the parameters of the minima (their shape, depth, and duration) depend on the dimensionless wind velocity, accretion rate onto the secondary, and inclination θ of the binary equatorial plane to the line of sight. The larger $\dot{M}_a$ is, the larger is the angle θ at which a noticeable photometric effect is possible. The lower the wind velocity V_w, the smaller the inclination angle θ at which eclipses can be observed. The longest duration of the minima comparable to the orbital period is obtained in the models with $V_w \approx 1$.

At an initial particle velocity higher than the escape velocity at the distance of the orbit of the secondary, all of the disk-wind matter leaves the binary. In such binaries, no common envelope is formed and fadings of the primary can be observed only when the cone of the disk wind from the secondary crosses the line of sight. The minima are almost symmetric (Figs. 6c, 6d) and their depth decreases with increasing wind velocity. It is easy to show that the eclipse duration, in fractions of the orbital period, in such models is

$$\Delta t = \frac{1}{\pi} \tan\theta \tan\alpha, \qquad (4)$$

where θ is the angle between the line of sight and the binary equatorial plane and α is the opening angle of the disk-wind cone. Hence, $\Delta t \approx 5 \times 10^{-3} \theta \cdot \tan\alpha$ at small θ (θ is in degrees).

DUSCUSSION

Thus, our calculations show that the brightness of the primary component in a binary can be subjected to a periodic modulation due to extinction variations in the common envelope produced by the disk wind from the secondary. Hence, several important results follow. First, the detection probability of eclipsing binaries among young stars can be nonzero even when the binary plane is appreciably inclined to the line of sight and no eclipsing effect in its classical form is possible. This circumstance is particularly important when a young binary is surrounded by an extended CB-disk, because the components of such binaries are unseen during equator-on observations due to the strong absorption of radiation in the disk.

Another important result is that the eclipse duration for young binaries can be much longer than that for ordinary eclipsing binaries.

Examples of Binaries with Long-Duration Eclipses

GW Ori. The classical T Tauri star GW Ori is an example of such a young pair. Mathieu *et al.* (1991) discovered the secondary component in this star by variations in the radial velocity V_r with a 242-day period and semiaplitude of $\approx$5 km s^{-1}. The pattern of V_r variations with phase shows that the orbital eccentricity is nearly zero.

The spectroscopic period was confirmed by the photometric observations of Shevchenko *et al.* (1998) GW Ori was found to be an eclipsing binary with long-duration (about 0.1 of the period) eclipses. In attempting to explain this observational fact, these authors assumed that the eclipses were produced by the dust that filled the Roche lobe around the secondary component. However, this assumption led to an overly large (for T Tauri stars) mass of the primary component ($3.3 M_\odot$). As a result, the authors came to serious conflicts with the current theory for the formation and evolution of young stars.

If the eclipses in GW Ori are assumed to be produced by the disk wind from the secondary and by the common envelope formed by it, then the above contradiction is removed. Comparison of the observed light curve for the star (Fig. 1 in Shevchenko *et al.* 1998) with the theoretical light curves shown in Figs. 5c and 6d indicates that the duration and shape of the minima are in satisfactory agreement with the disk-wind model with $V_w = 2$ or $V_w = 0.25$. The latter value is appropriate only at small inclinations of the binary plane to the line of sight, which appears to be inacceptable for GW Ori.

Clearly, investigation of such binaries is of great interest not only in determining the parameters of the stars that are binary members but also in studying the structure of the disk winds from young stars. Thus, for example, if formula (4) is used, then at an eclipse duration for GW Ori of 0.1 of the orbital period and at the opening angle of the wind from the secondary $\alpha = 45°$, it follows from this formula that the angle between the binary plane and the line of sight is $\theta = 17°$. This value is almost equal to one of the two inclination angles of the GW Ori binary plane discussed by Mathieu *et al.* (1991) (15°)[3] and only 10° smaller than the second value (27°) that Mathieu *et al.* (1991) believe to be more likely. If the latter is true, then in this case, according to (4), the opening angle for the cone of the wind from the secondary component of GW Ori must be $\alpha \approx 30°$.

For GW Ori, the change in the amplitude of the photometric minima from eclipse to eclipse pointed out by Shevchenko *et al.* (1998) is of great interest. The calculations show that in our model, apart from the main cycle attributable to the orbital motion of the

[3]This value is equal to the inclination i of the GW Ori rotation axis obtained by Bouvier and Bertout (1989) from the observed brightness modulation in the star due to its surface irregularity.

secondary and its associated densest part of the common envelope, there are longer cycles attributable to tangential motions of the peripheral parts of the common envelope. However, the extinction variations on the line of sight produced by them are small compared to the amplitude of the main cycle. Therefore, different factors, for example, a change in the accretion rate onto the secondary component, are most likely responsible for the observed change in the amplitude of the minima for GW Ori. The latter may be due to the presence of a third body suspected by Mathieu *et al.* (1991) from radial-velocity variations in the binary center of mass.

The Eclipsing Binary KH 15D. Another example of a young binary with long-duration eclipses is the WTTS (weak T Tauri star) star KH 15D. The eclipses of this star occur with a 48-day period and have a large amplitude ($\approx 3^m$ in the I band) and a two-component structure (Kearns and Herbst 1998). The total duration of the minima is about a third of the orbital period. This makes it difficult to explain such eclipses by a circumstellar dust disk or ring around the secondary, because the radius of this ring or disk must be comparable to the orbital radius of the secondary. Structures of such sizes are unstable against tidal perturbations and must rapidly break up.

As in the case of GW Ori, this problem does not arise if the eclipses in KH 15D are assumed to be attributable to the disk wind from the secondary. In this case, the neutral pattern of fadings in this star suggests that the dust structure that periodically appears on the line of sight consists of large particles (Hamilton *et al.* 2001). Such particles have a highly forward-elongated scattering diagram (van de Hulst 1957) and this may be the cause of the brightenings in KH 15D above the normal level observed by Kearns and Herbst (1998) in the breaks between successive fadings of the object.

If we again use formula (4), then for an eclipse duration of KH 15D equal to about a third of the orbital period and for $\alpha = 45°$, we find that the plane of this binary has a large (about 40°) inclination to the line of sight.

UX Ori Stars

UX Ori stars are among the photometrically most active young stars. Their brightness varies irregularly with an amplitude $\Delta V \approx 2$–3. Several independent proofs that (apart from their youth) the optimal orientation of the circumstellar disks around these stars at a small angle to the line of sight is mainly responsible for their high photometric activity and that variable circumstellar extinction is mainly responsible for their light variations have been obtained to date [see the review article by Grinin (2000b) and references therein). Therefore, by studying the photometric and polarimetric activity of these stars, we can obtain information on the fine structure of their surrounding protoplanetary disks that is difficult or impossible to obtain by other methods. In particular, analysis of such observations indicates that the circumstellar gas–dust clouds that occasionally shield the stars from the observer move in highly eccentric orbits and can appear in the immediate vicinity of the stars, at distances of the order of or smaller than 1 AU (Grinin 1994). However, it was unclear how and where the clouds with such orbits are formed.

In the preceding section, we showed that the capture of the disk wind from the low-mass component could give rise to a ring of matter with a low specific angular momentum near the primary component. From here, the matter will fall to the primary component of the binary while moving in highly eccentric, almost radial orbits. Thus, in our model, there is the possibility for the formation of circumstellar clouds with a low specific angular momentum, which deserves further study.

Another important conclusion drawn from long-term series of photometric and polarimetric observations of UX Ori stars is that the mass distribution in their circumstellar disks significantly deviates from axial symmetry (Grinin *et al.* 1998; Rostopchina *et al.* 1999, 2000; Shakhovskoi *et al.* 2002). As a result, apart from irregular variability, large-scale cycles with time scales from one to ten or more years are seen in the light and linear polarization curves. The above authors suggested that these cycles result from the modulation of optical properties of the circumstellar disks produced by the orbital motion of the low-mass components. A similar view on the nature of cyclic activity in stars of this type has also been recently put forward by Bertout (2000).[4] Our calculations show that this assumption is quite feasible and points out the specific mechanism capable of producing a large-scale brightness modulation in a young star.

As an illustration, Fig. 7 shows the light curve of BF Ori, a typical star of this subclass. Comparison with theoretical curves (Fig. 5b) indicates that the model of a binary with a low-mass secondary component and $V_w \approx 0.5$ gives approximately the same (in shape) minima characterized by a fast decline in brightness and a slower rise. Their total duration is comparable to the orbital period.

[4]Note an alternative approach to interpreting the cyclic activity of UX Ori stars based on the hypothesis that the stars are eclipsed by protocomets [see Herbst and Shevchenko (1999), Shevchenko and Ezhkova (2001) and references therein].

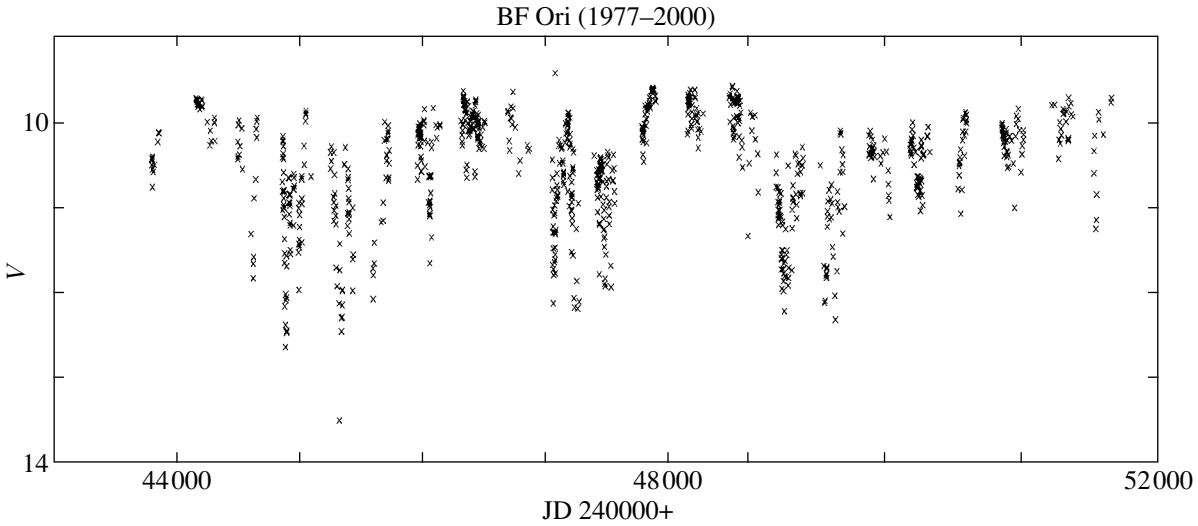


Fig. 7. Cyclic activity of UX Ori stars [the light curve of BF Ori is from Grinin *et al.* (2000)].

The typical duration of the photometric cycles for UX Ori stars is several years, which corresponds to the orbital radii of the low-mass components in the range ~3–5 AU for stellar masses of about 2–3 $M_\odot$. The longest cycle is observed in CQ Tau, 22–25 years (Shakhovskoi *et al.* 2002) and corresponds to an orbital radius of about 10 AU. The orbits of binaries with such parameters generally have an appreciable eccentricity (Mayor *et al.* 2001). Therefore, for the activity cycles of UX Ori stars to be analyzed in more detail, the models considered above must be generalized to elliptical orbits. Apart from the orbital eccentricity, yet another parameter appears in these models, the angle between the line of sight and the orbital semimajor axis, which introduces an additional variety in the theoretical light curves.

Concluding the discussion of the role of binarity in organizing the activity cycles of young stars, it should be noted that the mechanism of photometric variability considered here (attributable to a periodic extinction modulation on the line of sight) is not the only possible one. In the model of Artimowicz and Lubow (1996), the accretion rate onto the binary components can also be subjected to a periodic modulation. In our model, such a modulation can also take place in binaries with elliptical orbits because of periodic variations in the probability p of the disk wind from the secondary being captured by the primary component of the binary (paper I). Therefore, cyclic photometric variability attributable to a modulation of the accretion rate can be observed in young cool stars whose optical emission, along with their photospheric emission, also includes the emission from hot accreted gas (classical late-type T Tauri stars). Examples of such binaries were considered by Mathieu *et al.* (2000) and Lamzin *et al.* (2001). Thus, in pure form, the model of cyclic variability considered here is applicable to binaries with hotter stars as the primary components (e.g., Herbig Ae/Be stars), whose optical variability is attributable to circumstellar extinction variations and the role of other mechanisms is minor. Most of the UX Ori stars belong to such stars.

Giant Protoplanets at the Stage of Intense Accretion

Giant protoplanets at the stage of intense accretion can also have gas and dust disks, as suggested by the developed system of satellites around Jupiter and Saturn's rings. Current calculations show (Kley 1999) that the accretion rate onto a Jupiter-like protoplanet can reach values, $\sim\dot{M}_a \approx 10^{-8}$–$10^{-9} M_\odot$ per year, comparable to the accretion rate onto some T Tauri (Calvet *et al.* 2000) and Herbig Ae/Be (Tambovtseva *et al.* 2001) stars. As a result, an accretion disk is formed around the protoplanet; this disk can also be a source of the disk wind (Quillen and Trilling 1998). If the parameter f, the ratio of the outflow and accretion rates, is assumed to be in this case of the same order of magnitude as that in our model, then the results of our numerical simulations can be applied to such objects.[5] It follows from these

[5]It should be noted that because of the great disparity between the masses of the protoplanet and the central star, the accretion rate onto the star can be much higher than the accretion rate onto the protoplanet. Therefore, the effect of stellar wind on the dynamics of circumstellar matter must be more significant in this case.

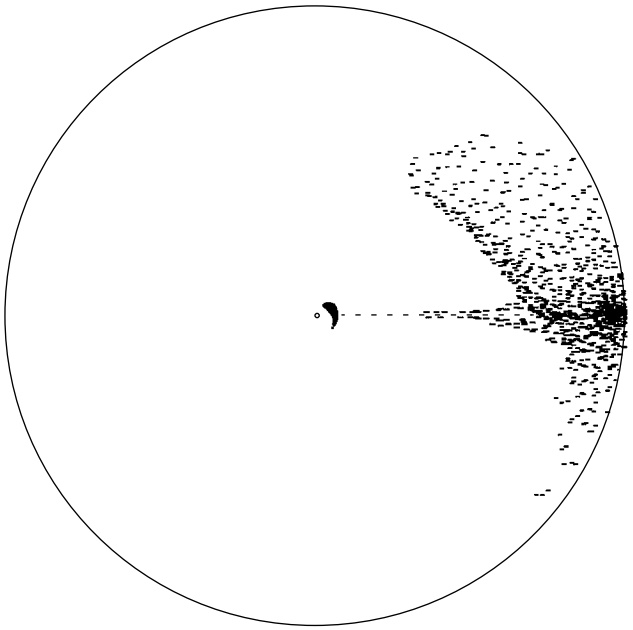


Fig. 8. An illustration to the effect of a moving shadow on the periphery of the CB-disk formed by the common envelope in the model with $\alpha = 45^\circ$ and $V_w = 0.25$.

results that a protoplanet at the stage of intense accretion can be detected optically through observations of periodic eclipses of the central star by the disk wind from the protoplanet. In this case, because of the optimal orientation of their circumstellar disks relative to the line of sight, UX Ori stars are the most suitable objects for such searches. It may well be that the cycles of photometric variability in these stars discussed in the preceding section are attributable to mass accretion by giant protoplanets or brown dwarfs, because no spectroscopic evidence of ordinary binarity has yet been found in most of these stars.

Shadow Effects

The described periodic extinction variations on the line of sight caused by the orbital motion of the common envelope produce shadow zones moving in space in which, in addition to the observer, other objects can be located. One such object can be the CB-disk itself or, to be more precise, its peripheral zone. Far from the secondary, the structure of the CB-disk must generally be the same as that around a single young star. According to Kenyon and Hartmann (1987), the ratio of the geometric disk thickness h to the current radius r is a slowly increasing function of r. When such a disk is observed pole-on, a shadow from the disk wind of the secondary and the opaque part of the common envelope can be seen in its image (Fig. 8); this shadow must move over the CB-disk following the orbital motion of the secondary. This prediction can be verified in the immediate future, because such powerful interferometer telescopes as VLTI designed for observations with high spatial resolutions are in the stage of completion. Detection of moving shadows on CB-disks and their monitoring may become a method of studying binaries at the earliest stages of their evolution. It may well be that the asymmetric images of the CB-disks around some pole-on young binaries (see, e.g., McCabe and Ghez 2001) are attributable to this effect. Therefore, it is of interest to obtain new images of such objects in an effort to detect possible changes in the pattern of disk asymmetry with time.

CONCLUSIONS

We obtained our results using a simple model of a young binary in which the formation of a common envelope was considered in the ballistic approximation, with the disk wind from the primary component being disregarded. For this reason, our calculations are approximate. It follows from the calculations that the optical emission from the primary component of the binary could be subjected to a periodic modulation attributable to the absorption of emission in the densest part of the gas–dust envelope associated with the low-mass component. The source of matter in this envelope is the low-velocity component of the disk wind formed via mass accretion from the remnants of the protostellar cloud. This conclusion complements the results of paper II, in which we showed that the thermal radiation from such envelopes could play an important role in producing infrared excesses of young stars. It thus follows that a photometric monitoring can be used to search for and study the low-mass components of young binaries, including substellar objects and giant protoplanets. The latter are difficult to detect in the vicinity of young stars by other methods. The most suitable objects for such searches are photometrically active, young UX Ori stars whose circumstellar disks are inclined at a small angle to the line of sight. It may well be that the activity cycles of these stars discussed above owe their origin precisely to such objects.

In addition, eclipses of the primary component in a young binary by the opaque part of its common envelope makes the detection of young eclipsing systems possible in those cases where the orbital plane of the secondary makes a large angle with the line of sight and the eclipsing effect in its classical form (when one star shields the other) cannot be observed. We assume that the young binary GW Ori and, possibly, KH 15D, whose common feature is a long duration of their eclipses, belong to such systems.

Apart from the photometric cycles, the orbital motion of the densest part of the common envelope may be responsible for the appearance of a moving shadow in the peripheral parts of the CB-disks. Young, photometrically quiet stars with large infrared excesses are best suited to searches for such structures, because the circumstellar disks of such objects are highly inclined to the line of sight (Grinin and Rostopchina 1996) and, therefore, are convenient for studying their surface brightness. Moving shadows can also be observed in the images of reflection nebulae, which are commonly associated with young objects. Therefore, interestingly, the variability of such nebulae has already been pointed out in the past [see Section 3.2 in the book by Vorontsov-Vel'yaminov (1948) and references therein]. We even know cases where the bright reflection nebula of R Mon disappeared in the photographs almost completely. It may well be that the fadings of such a large object as the reflection nebula resulted from the shielding of the star illuminating this nebula by the opaque part of its common envelope. Therefore, a photometric monitoring of the images of such nebulae in an effort to detect periodic variations in their surface brightness is of great interest.

Apart from the above, purely photometric effects, the dust component of the common envelope can also contribute significantly to the scattered light and intrinsic polarization of young stars at optical wavelengths (which is currently attributed to the scattered radiation from circumstellar disks or disk-like dust envelopes). Studying the polarization properties of asymmetric common envelopes is of considerable interest in detecting a high linear polarization at the deep minima of some UX Ori stars (Rostopchina *et al.* 2001), which cannot be explained in terms of axisymmetric models for circumstellar envelopes. We plan to analyze this and several other problems in following papers.

ACKNOWLEDGMENTS

We wish to thank Antonella Natta and all the participants of the Star Formation Workshop at the Arcetri Observatory (Florence) for a discussion of the results. We are also grateful to S.A. Lamzin and A.B. Makalkin for a number of valuable remarks. This study was supported by a grant from the Presidium of the Russian Academy of Sciences "Nonstationary Phenomena in Astrophysics."

REFERENCES

1. P. Artimowicz and S. H. Lubow, Astrophys. J. **421**, 651 (1994).
2. P. Artimowicz and S. H. Lubow, Astrophys. J. Lett. **467**, L77 (1996).
3. M. R. Bate and I. A. Bonnell, Mon. Not. R. Astron. Soc. **285**, 33 (1997).
4. S. V. W. Beckwith, A. I. Sargent, R. S. Chini, and R. Güsten, Astron. J. **99**, 924 (1990).
5. C. Bertout, Astron. Astrophys. **363**, 984 (2000).
6. R. D. Blandford and D. G. Payne, Mon. Not. R. Astron. Soc. **199**, 883 (1982).
7. J. Bouvier and C. Bertout, Astron. Astrophys. **211**, 99 (1989).
8. N. Calvet, L. Hartmann, and S. E. Strom, in *Protostars and Planets IV,* Ed. by V. Mannings, A. P. Boss, and S. S. Russell (Univ. of Arizona Press, Tucson, 2000), p. 377.
9. B. T. Draine, Astrophys. J., Suppl. Ser. **57**, 587 (1985).
10. B. T. Draine and H. M. Lee, Astrophys. J. **285**, 89 (1984).
11. A. Duquennoy and M. Mayor, Astron. Astrophys. **248**, 485 (1991).
12. J. Eislöffel, R. Mundt, T. P. Ray, and L. F. Rodriguez, in *Protostars and Planets IV,* Ed. by V. Mannings, A. P. Boss, and S. S. Russell (Univ. of Arizona Press, Tucson, 2000), p. 841.
13. P. Goldreich and S. Tremaine, Annu. Rev. Astron. Astrophys. **20**, 249 (1982).
14. A. P. Goodson, K. H. Böhm, and R. Winglee, Astrophys. J. **524**, 142 (1999).
15. V. P. Grinin, Pis'ma Astron. Zh. **14**, 65 (1988) [Sov. Astron. Lett. **14**, 27 (1988)].
16. V. P. Grinin, Astron. Soc. Pac. Conf. Ser. **62**, 63 (1994).
17. V. P. Grinin, Astrofizika **43**, 603 (2000a).
18. V. P. Grinin, Astron. Soc. Pac. Conf. Ser. **219**, 216 (2000b).
19. V. P. Grinin, Astron. Zh. (2002) (in press) [Astron. Rep. (2002)].
20. V. P. Grinin and A. N. Rostopchina, Astron. Zh. **73**, 194 (1996) [Astron. Rep. **40**, 171 (1996)].
21. V. P. Grinin, A. N. Rostopchina, and D. N. Shakhovskoĭ, Pis'ma Astron. Zh. **24**, 925 (1998) [Astron. Lett. **24**, 802 (1998)].
22. V. P. Grinin, A. N. Rostopchina, and D. N. Shakhovskoy, in *Birth and Evolution of Binary Stars: Poster Proceedings of the IAU Symposium No. 200, Potsdam, 2000,* Ed. by H. Zinnecker and R. D. Mathieu, p. 115.
23. C. M. Hamilton, W. Herbst, C. Shih, and A. J. Ferro, Astrophys. J. Lett. **554**, L201 (2001).
24. P. Hartigan, S. E. Edwards, and L. Ghandour, Astrophys. J. **452**, 736 (1995).
25. W. Herbst and V. S. Shevchenko, Astron. J. **118**, 1043 (1999).
26. G. A. Hirth, R. Mundt, and J. Solf, Astron. Astrophys., Suppl. Ser. **126**, 437 (1997).
27. K. M. Kearns and W. Herbst, Astron. J. **116**, 261 (1998).
28. S. J. Kenyon and L. Hartmann, Astrophys. J. **323**, 714 (1987).
29. W. Kley, Mon. Not. R. Astron. Soc. **303**, 696 (1999).
30. J. Kwan and E. Tademaru, Astrophys. J. Lett. **332**, L41 (1988).

31. S. A. Lamzin, S. Yu. Melnikov, K. M. Grankin, and O. V. Ezhkova, Astron. Astrophys. **372**, 922 (2001).
32. D. N. C. Lin and J. C. B. Papaloizou, in *Protostars and Planets III,* Ed. by E. H. Levy and J. I. Lunine (Univ. of Arizona Press, Tucson, 1993), p. 749.
33. S. H. Lubow and P. Artimowicz, in *Protostars and Planets IV,* Ed. by V. Mannings, A. P. Boss, and S. S. Russell (Univ. of Arizona Press, Tucson, 2000), p. 731.
34. D. Lynden-Bell and F. E. Pringle, Mon. Not. R. Astron. Soc. **168**, 603 (1974).
35. N. Mastrodemos and M. Morris, Astrophys. J. **497**, 303 (1998).
36. R. D. Mathieu, F. C. Adams, and D. W. Latham, Astron. J. **101**, 2184 (1991).
37. R. D. Mathieu, A. M. Ghez, E. K. N. Jensen, and M. Simon, in *Protostars and Planets IV,* Ed. by V. Mannings, A. Boss, and S. S. Russell (Univ. of Arizona Press, Tucson, 2000), p. 559.
38. M. Mayor, S. Udry, J.-L. Halbwachs, and F. Arenou, *The Formation of Binary Stars: Proceedings of the IAU Symposium No. 200, Potsdam, 2000,* Ed. by H. Zinnecker and R. D. Mathieu, p. 45.
39. T. Mazeh, D. Goldberg, A. Dequennoy, and M. Mayor, Astrophys. J. **401**, 265 (1992).
40. C.-E. McCabe and A. M. Ghez, in *The Formation of Binary Stars: Proceedings of the IAU Symposium No. 200, Potsdam, 2001,* Ed. by H. Zinnecker and R. D. Mathieu, p. 245.
41. A. B. Men'shchikov, T. Henning, and O. Fisher, Astrophys. J. **519**, 257 (1999).
42. A. Natta and B. Whitney, Astron. Astrophys. **364**, 633 (2000).
43. A. Natta, V. P. Grinin, and V. Mannings, in *Protostars and Planets IV,* Ed. by V. Mannings, A. Boss, and S. S. Russell (Univ. of Arizona Press, Tucson, 2000), p. 559.
44. A. C. Quillen and D. E. Trilling, Astrophys. J. **508**, 707 (1998).
45. A. N. Rostopchina, V. P. Grinin, A. Okazaki, *et al.*, Astron. Astrophys. **327**, 145 (1997).
46. A. N. Rostopchina, V. P. Grinin, and D. N. Shakhovskoĭ, Pis'ma Astron. Zh. **25**, 291 (1999) [Astron. Lett. **25**, 243 (1999)].
47. A. N. Rostopchina, V. P. Grinin, and D. N. Shakhovskoĭ, Astron. Zh. **77**, 420 (2000) [Astron. Rep. **44**, 365 (2000)].
48. A. N. Rostopchina, V. P. Grinin, and D. N. Shakhovskoĭ, Astron. Zh. **78**, 60 (2001) [Astron. Rep. **45**, 51 (2001)].
49. P. N. Safier, Astrophys. J. **408**, 115 (1993).
50. N. I. Shakura and R. A. Sunyaev, Astron. Astrophys. **24**, 337 (1973).
51. D. N. Shakhovskoĭ, V. P. Grinin, and A. N. Rostopchina, in press (2002).
52. V. S. Shevchenko, K. N. Grankin, S. Yu. Mel'nikov, and S. A. Lamzin, Pis'ma Astron. Zh. **24**, 614 (1998) [Astron. Lett. **24**, 528 (1998)].
53. V. S. Shevchenko and O. V. Ezhkova, Pis'ma Astron. Zh. **27**, 47 (2001) [Astron. Lett. **27**, 39 (2001)].
54. F. Shu, J. Najita, D. Galli, E. Ostriker, and S. Lizano, in *Protostars and Planets III,* Ed. by E. H. Levy and J. I. Lunine (Univ. of Arizona Press, Tucson, 1993), p. 3.
55. J. M. Stone, C. F. Gammie, S. A. Balbus, and J. F. Hawley, in *Protostars and Planets IV,* Ed. by V. Mannings, A. P. Boss, and S. S. Russel (Univ. of Arizona Press, Tucson, 2000), p. 589.
56. L. V. Tambovtseva, V. P. Grinin, B. Rodgers, and O. V. Kozlova, Astron. Zh. **78**, 514 (2001) [Astron. Rep. **45**, 442 (2001)].
57. H. C. van de Hulst, *Light Scattering by Small Particles* (Wiley, New York, 1957; Inostrannaya Literatura, Moscow, 1961).
58. B. A. Vorontsov-Vel'yaminov, *Gaseous Nebulae and Novae* (Akad. Nauk SSSR, Moscow, 1948), p. 38.
59. N. V. Voshchinnikov and V. P. Grinin, Astrofizika **34**, 181 (1991).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 617–619. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 685–687.

A Model for the Dust Envelope of V1853 Cygni

M. B. Bogdanov*

Saratov State University, ul. Astrakhanskaya 83, Saratov, 410026 Russia

Received February 20, 2002

Abstract—We computed a model for the dust envelope of the protoplanetary nebula V1853 Cyg by using data on its fluxes from the ultraviolet to the far infrared. The spherically symmetric envelope was assumed to be composed of silicate grains with the standard MRN size distribution; their number per unit volume is inversely proportional to the distance squared. The optical depth of the envelope, whose inner boundary lies at a distance of 7.6×10^{16} cm from the central star, is 0.18 at a wavelength of 0.55 μm. The grain temperature at the inner boundary of the envelope is 110 K. The distance to V1853 Cyg is estimated to be 4.1 kpc. The current mass-loss rate of the object was found by solving a self-consistent problem of radiative transfer and dust motion in the envelope to be $2.2 \times 10^{-5}\ M_\odot\ \mathrm{yr}^{-1}$. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *stars—structure and evolution; protoplanetary nebulae, dust envelopes, mass loss*

INTRODUCTION

The interesting object V1853 Cyg (= LSII+34°26 = IRAS 20462+3416) belongs to the stars at the evolutionary stage that follows the asymptotic giant branch (post-AGB) and precedes the formation of a planetary nebula. Such objects were called protoplanetary nebulae (Volk and Kwok 1989). The duration of this stage is so short that any evolutionary change in the effective temperature of the star can become noticeable in the lifetime of one generation of observers. Investigation of such objects will make it possible to elucidate the mechanism for the formation of a strong stellar wind, which causes the stellar envelope to be ejected. This is of great importance in studying the final evolutionary stages of moderate-mass stars and the formation of the interstellar medium.

Parthasarathy (1993) was the first to justify the fact that V1853 Cyg belongs to protoplanetary nebulae. Using IRAS observations, he concluded that this is a post-AGB B supergiant surrounded by a cold dust envelope. Based on UBV photometry and on analysis of the spectra taken with the 1.25-m and 6-m telescopes, Arkhipova *et al.* (2001) studied the photometric and spectroscopic features of the object in detail. Also given in their paper are references to previous studies.

Here, our goal is to compute a model for the dust envelope of V1853 Cyg and to estimate its mass-loss rate.

*E-mail: BogdanovMB@info.sgu.ru

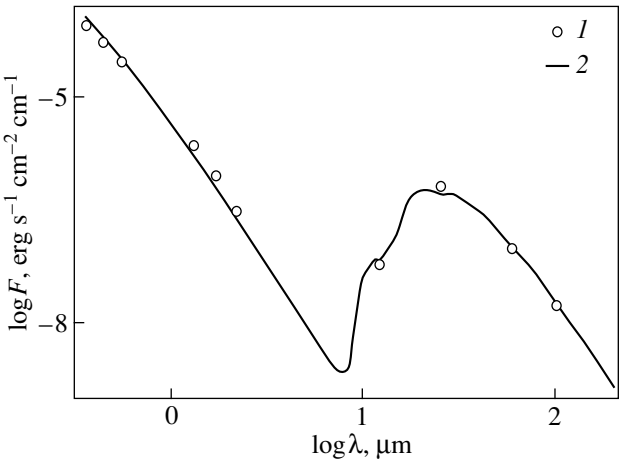
OBSERVATIONAL DATA

For comparison with the model, we used the mean UBV magnitudes of the star obtained by Arkhipova *et al.* (2001) in 1999–2000. These are in good agreement with the mean magnitudes observed in 1973 (Drilling 1975), suggesting that the object evolves slowly. The amplitude of the rapid chaotic brightness variability, which is probably caused by stellar-wind fluctuations, is modest, ranging from $0\overset{m}{.}3$ to $0\overset{m}{.}4$ in different bands. In addition, we used JHK photometry for the star (Garcia-Lario *et al.* 1997) and IRAS data on its fluxes in the 12, 25, 60, and 100 μm bands.

The region around V1853 Cyg exhibits significant interstellar extinction. To make a correction for its effect, we used the color excess $E(B-V) = 0\overset{m}{.}24$ derived by Parthasarathy (1993) from the intensity of the λ 2200 Å band. To estimate the total extinction in V, $A_V = RE(B-V)$, we took $R = 3.3$ (Allen 1973). After the correction for interstellar extinction, we determined the fluxes $F(\lambda)$ (in erg s^{-1} cm^{-2} cm^{-1}) from the magnitudes by using extraatmospheric fluxes from Straizys (1977). The observed fluxes are represented by circles in the figure.

For objects like V1853 Cyg, the radiation during free–free transitions in the ionized matter of the forming nebula can also contribute to the far-IR flux. However, our estimate (Bogdanov 2000) for V886 Her, a similar protoplanetary nebula, shows that this contribution is minor. Since the degree of excitation and the electron density for V1853 Cyg are much lower than those for V886 Her (Arkhipova *et al.* 2001), the effect of this radiation may be ignored.

1063-7737/02/2809-0617$22.00



(*1*) Observed fluxes from the protoplanetary nebula V1853 Cyg versus wavelength; (*2*) the spectral energy distribution for the computed model star with a dust envelope.

According to Arkhipova *et al.* (2001), the effective temperature of the central star is $T_{\rm eff} = 20\,000$ K and the slow evolution suggests that its mass is comparatively low, $M \approx 0.6\ M_\odot$. Using the evolutionary tracks for planetary-nebula nuclei computed by Blocker (1995), we may take the stellar luminosity in this case to be $L = 6300\ L_\odot$.

COMPUTING A MODEL FOR THE DUST ENVELOPE

The standard assumptions were used to compute envelope models for V1853 Cyg. It was assumed that the spherically symmetric dust envelope had a sharp inner boundary at distance r_1 from the center and that the grain number density decreased in inverse proportion to the distance squared up to the envelope outer boundary at $r_2 = 1000 r_1$. The optical constants of the grain material were chosen to be identical to those for warm silicates (Ossenkopf *et al.* 1992) and the grain size distribution $n(a)$ was described by the MRN model (Mathis *et al.* 1977): $n(a) \propto a^{-q}$ for spherical-grain radii $a_{\rm min} \le a \le a_{\rm max}$ with the parameters $q = 3.5$, $a_{\rm min} = 0.005\ \mu$m, and $a_{\rm max} = 0.25\ \mu$m. The spectral energy distribution for the central source was assumed to be the Planck one with the temperature equal to the effective temperature of the star, $T_{\rm eff} = 20\,000$ K; its luminosity is $L = 6300\ L_\odot$.

We solved the problem of radiative transfer in the dust envelope by using the DUSTY code (version 2.0) for grids of 30 points in radius and 99 wavelengths within the range 0.01 μm to 3.6 cm. The algorithm that forms the basis for this code was described by Ivezic and Elitzur (1997, 1999). The input model parameters were the dust temperature at the inner boundary, T_1, and the envelope optical depth at a wavelength of 0.55 μm, τ_V. Having computed the spectral energy distribution for the model, we chose the distance d at which the sum of the squares of the deviations of the observed and model fluxes was at a minimum. The following optimal model parameters were obtained: $T_1 = 110$ K, $r_1 = 7.6 \times 10^{16}$ cm, $\tau_V = 0.18$, and $d = 4.1$ kpc. The derived temperature and distance are in good agreement with the estimates of Parthasarathy (1993): $T_{\rm d} \approx 100$ K and $d = 3.0$–4.6 kpc.

As we see from the figure, the model satisfactorily describes the observations, with the deviations of the observed fluxes being significantly smaller than those for our previously computed model of the dust envelope around the protoplanetary nebula V886 Her (Bogdanov 2000). This is apparently because the model assumptions and primarily spherical symmetry of the envelope are more appropriate for the object in question.

The derived r_1 and d give an angular radius of the envelope inner boundary equal to $1\farcs2$. This value agrees with the Hubble Space Telescope observations by Ueta *et al.* (2000), who found a reflecting nebula about $3''$ in size around V1853 Cyg. Clearly, intrinsic radiation from the dust in this nebula is observed in the mid- and far infrared.

To estimate the parameters of the stellar wind from V1853 Cyg produced by radiation pressure on the dust and the subsequent momentum transfer to the gaseous medium, we used the gas-dynamical mode of the DUSTY code with the derived envelope optical depth $\tau_V = 0.18$. In this case, a self-consistent procedure of solving the problem of radiative transfer and dust motion in the stellar envelope is implemented (Ivezic and Elitzur 1995). The grain density $\rho_{\rm d}$ and the gas-to-dust mass ratio $r_{\rm gd}$ in the envelope were assumed to be 3 g cm^{-3} and 200, respectively. Based on these data, we computed the total mass-loss rate, $\dot{M} = 2.2 \times 10^{-5}\ M_\odot$ yr^{-1}, with the DUSTY code. If necessary, $\dot{M}$ can be easily recalculated for different values of $\rho_{\rm d}$ and $r_{\rm gd}$, because $\dot{M} \propto (\rho_{\rm d} r_{\rm gd})^{1/2}$ (Ivezic and Elitzur 1995). Note that this result was obtained for current parameters of the star and the dust envelope. In addition, we disregarded the radiation pressure on the neutral and ionized components of the nebula. Therefore, the derived $\dot{M}$ is a lower limit for the mass-loss rate.

CONCLUSIONS

Despite the comparatively small amount of observational data, close agreement between the observed fluxes and the model spectral energy distribution gives hope that the derived parameters of the

dust envelope around V1853 Cyg are fairly realistic. This envelope appears to have been ejected by the star when it was leaving the asymptotic giant branch. If the stellar mass is assumed to be $M \approx 0.6 M_\odot$, then, according to the evolutionary calculations of Blocker (1995), $\Delta t \sim 1300$ yr must have elapsed since that time. The velocity $V = r_1/\Delta t$ was roughly estimated to be 19 km s^{-1}, a value that is fairly close to the observed expansion velocities of planetary nebulae.

It is of interest to compare the envelope parameters for V1853 Cyg with the values for V886 Her, a similar object. In this case, $T_1 = 410$ K, $r_1 = 4.3 \times 10^{15}$ cm, and $\tau_V = 0.48$ (Bogdanov 2000). The dust envelope is much closer to the star and is denser and hotter, suggesting that the envelope was ejected recently. The mass of the central star of V886 Her is $M \approx 0.7\ M_\odot$; it is characterized by a rapid evolutionary change in its effective temperature (Arkhipova *et al.* 1999). Using data from Blocker (1995), we can find that the time is $\Delta t \approx 100$ yr and that the envelope expansion velocity is $V = 12$ km s^{-1}. The mass-loss rate for V886 Her is lower than that for V1853 Cyg, being $\dot{M} = 4.5 \times 10^{-6}\ M_\odot$ yr^{-1} (Bogdanov 2000).

ACKNOWLEDGMENTS

I wish to thank the group of astrophysicists at the Kentucky University (Z. Ivezic, M. Nenkova, and M. Elitzur) for the opportunity to use the DUSTY code designed to solve the radiative transfer equation and the SIMBAD database staff for access to the observational data for V1853 Cyg. This study was supported by grants from the Federal Science and Technology Program "Astronomy," the Education Ministry of Russia, and the Program "Universities of Russia."

REFERENCES

1. C. W. Allen, *Astrophysical Quantities* (Athlone Press, London, 1973; Mir, Moscow, 1977).
2. V. P. Arkhipova, N. P. Ikonnikova, R. I. Noskova, *et al.*, Pis'ma Astron. Zh. **25**, 30 (1999) [Astron. Lett. **25**, 25 (1999)].
3. V. P. Arkhipova, N. P. Ikonnikova, R. I. Noskova, *et al.*, Pis'ma Astron. Zh. **27**, 841 (2001) [Astron. Lett. **27**, 719 (2001)].
4. T. Blocker, Astron. Astrophys. **299**, 755 (1995).
5. M. B. Bogdanov, Astron. Zh. **77**, 772 (2000) [Astron. Rep. **44**, 685 (2000)].
6. J. S. Drilling, Astron. J. **80**, 128 (1975).
7. P. Garcia-Lario, A. Manchado, W. Pych, and S. R. Pottasch, Astron. Astrophys., Suppl. Ser. **126**, 479 (1997).
8. Z. Ivezic and M. Elitzur, Astrophys. J. **445**, 415 (1995).
9. Z. Ivezic and M. Elitzur, Mon. Not. R. Astron. Soc. **287**, 799 (1997).
10. Z. Ivezic and M. Elitzur, Mon. Not. R. Astron. Soc. **303**, 864 (1999).
11. J. S. Mathis, W. Rumpl, and K. H. Nordsieck, Astrophys. J. **217**, 425 (1977).
12. V. Ossenkopf, T. Henning, and J. S. Mathis, Astron. Astrophys. **261**, 567 (1992).
13. M. Parthasarathy, Astrophys. J. **414**, L109 (1993).
14. V. Straizys, *Multicolor Stellar Photometry* (Mokslas, Vilnius, 1977; Pachart Publ. House, Tucson, 1992).
15. T. Ueta, M. Meixner, and M. Bobrovsky, Astrophys. J. **528**, 861 (2000).
16. K. M. Volk and S. Kwok, Astrophys. J. **342**, 345 (1989).

Translated by N. Samus'

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 620–625. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 688–694.

A Minimum in the Light Curve of the Classical Symbiotic Star YY Her in 2001

E. A. Kolotilov, A. A. Tatarnikova*, S. Yu. Shugarov, and B. F. Yudin

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

Received March 21, 2002

Abstract—We present new photometric $UBVRI$ observations of the classical symbiotic star YY Her in 2001. A deep secondary minimum with an amplitude of $0^m.5$ in V was found in its light curve during this period. The color characteristics of the primary and secondary minima are compared. We show that the star reddens during its primary minimum and blues during its secondary minimum in the $U-B$ and $B-V$ color indices. In the red spectral band, the color of YY Her does not change as it passes through its secondary minimum. We estimate photometric parameters of the radiation sources that disappear from the field of view when the star is at minimum light. Based on observational data, we investigate various models for the formation of the secondary minimum. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *stars—variable and peculiar; symbiotic stars.*

INTRODUCTION

YY Her belongs to the classical symbiotic stars whose prototype is Z And. This subclass of symbiotic stars is characterized by a distinctive type of novalike outbursts (Kenyon 1986). The last bright outburst of YY Her occurred in 1993, while in the late 1990s the star returned to quiescence [see Tatarnikova *et al.* (2001) and references therein].

Analysis of the visual light curves revealed a periodicity with $P = 590$ days and amplitude $\leq 0^m.3$ (Munari *et al.* 1997a). An example of classical symbiotic stars for which the radial-velocity curves along with the light curves were constructed shows that their periodic light variations result from the orbital motion of these binary systems; the brightness is at a minimum when the red giant is at inferior conjunction, i.e., in front of the hot component. This led Munari *et al.* (1997a) to conclude that the orbital period of YY Her is $P_{\rm orb} \approx 590$ days.

Our photometric UBV observations started in 1995 (Tatarnikova *et al.* 2001) show that the amplitude of this minimum, which is called below the primary minimum, increases with decreasing wavelength. Its origin is attributed to an eclipse of the hot component and its circumstellar envelope by the red giant, although the heating of the red-giant hemisphere facing the hot component by the latter can apparently also contribute to the fading in quiescence (Tatarnikova *et al.* 2001).

The outburst activity of classical symbiotic stars suggests that their cool components fill their Roche lobes (Yudin 1987). A convincing confirmation of this suggestion could be the detection of a periodicity with $P = P_{\rm orb}/2$ attributable to an ellipsoidal shape of the lobe-filling red giant in the light curve of the star.

Such analysis of the visual light curve for YY Her (see Munari *et al.* 1997a) revealed a harmonic with a period of $\sim$283 days and amplitude of $\sim 0^m.1$ (Tatarnikova *et al.* 2000).

In 1998, a shallow ($\Delta V \approx 0^m.2$) secondary minimum was found in the light curves of YY Her in several optical spectral bands; it was located between two adjacent primary minima (Hric *et al.* 2001; Tatarnikova *et al.* 2001). Its origin is attributed either to an eclipse of part of the red-giant surface by the circumstellar envelope of the hot component (Hric *et al.* 2001) or to an ellipsoidal shape of the cool component (Tatarnikova *et al.* 2001).

Here, we present the photometric observations of YY Her in 2001, when according to the ephemerides of this binary system, another secondary minimum must have been observed. It was actually recorded, with its amplitude being unprecedentedly large. We analyze the photometric parameters of the secondary minima and compare them with similar parameters of the primary minima. We also analyze various hypotheses of the nature of the secondary minimum.

OBSERVATIONS

The photoelectric $UBVR'I'$ observations are carried out with two photometers attached to the 0.6-m

*E-mail: aat@sai.msu.ru

and 1.25-m telescopes at the Crimean Station of the Sternberg Astronomical Institute. The detectors are a photomultiplier (UBV) and a CCD array ($BVR'I'$). The UBV bands correspond to the broadband photometric system of Johnson (1966). At the same time, the $R'I'$ bands do not correspond to any standard photometric system. The central wavelengths and widths are, respectively, $\sim$0.76 and $\sim$0.24 μm for R' and $\sim$0.9 and $\sim$0.3 μm for I'. In R' and I', the instrumental system was tied in to Johnson's system (R, I) using the observations of photometric standards in M 67. However, for stars with an emission spectrum, this procedure is partly conditional; i.e., it would be incorrect to identify our spectral bands with Johnson's bands.

The standard stars were HD 168957 ($U = 6^m.35$, $B = 6^m.91$, $V = 7^m.01$) and the star denoted on the finding chart in Munari *et al.* (1997a) by the letter G ($V = 13^m.08$, $R' = 12^m.44$, $I' = 11^m.94$). The measurement errors do not exceed $0^m.03$. The latest observations are given in Tables 1 and 2.

Figure 1 shows the $UBVR'I'$ light curves of YY Her constructed by using previously published data (Tatarnikova *et al.* 2001). Also shown in this figure are the data from Hric *et al.* (2001), where the observations are given in an updated Johnson–Kron–Cousins system ($B_cV_cR_cI_c$). In order to make their light curves coincident with our light curves in Fig. 1, we increased the tabulated B_c and V_c magnitudes by $0^m.3$ and $0^m.2$ and decreased the R_c and I_c magnitudes by $0^m.2$ and $0^m.52$, respectively. In other words, our photometric systems are related as follows: $B = B_c + 0.3$, $V = V_c + 0.2$, $R' = R_c - 0.2$, and $I' = I_c - 0.52$.

To perform a comparative analysis of the photometric parameters for the secondary and primary minima, we begin the next section with analysis of our photometric observations of the primary minimum in 2000.

ANALYSIS OF THE OBSERVATIONS

The Primary Minimum

We see from Fig. 1 that the amplitude of brightness decline at the primary minimum decreases with increasing wavelength and, accordingly, the star becomes redder. For the minimum of 2000, the amplitude ratio is $\Delta U/\Delta V \approx 1.8$, $\Delta B/\Delta V \approx 1.2$, $\Delta R'/\Delta V \approx 0.6$, and $\Delta I'/\Delta V \approx 0.2$.

Figure 2 shows the color–magnitude diagrams [($U-B$, V), ($B-V$, V), ($V-R'$, V), ($V-I'$, V), ($R'-I'$,V)] constructed separately for the primary and secondary minima of the star. YY Her reddens in all color indices as it passes through its primary minimum.

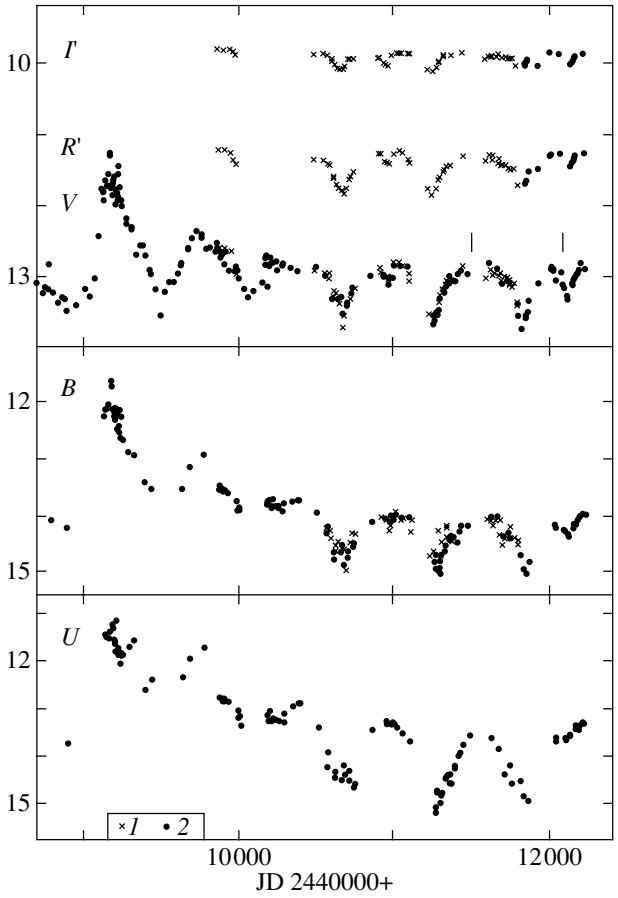


Fig. 1. The $UBVR'I'$ light curves of YY Her: (*1*) the data from Hric *et al.* (2001) reduced to our photometric system; (*2*) our data, including those published previously [see Tatarnikova *et al.* (2001) and references therein]. The vertical bars in the V light curve indicate the middle points of the secondary minima in 1998 and 2001.

The decline in brightness at minimum can be described as the appearance of a deficit in its radiation (by analogy with an infrared excess). We denote this deficit by PD and SD for the primary and secondary minima, respectively. During the minimum light of YY Her at its primary minimum in 2000, when the PD brightness was at a maximum, its magnitudes were $U_{pd} \approx 13^m.80$, $B_{pd} \approx 14^m.60$, $V_{pd} \approx 13^m.62$, $R'_{pd} \approx 12^m.49$, and $I'_{pd} \approx 11^m.94$. The PD color indices corrected for interstellar reddening with the color excess $E(B-V) = 0.2$ (Munari *et al.* 1997b) are $(U-B)_{0,pd} \approx -0.93$ and $(B-V)_{0,pd} \approx 0.78$.

Figure 3 shows the color variations of the PD with its visual brightness as the star passes through the last three primary minima (Fig. 1). We see that as the PD brightness rises, i.e., as the brightness of the star approaches its minimum, the PD becomes redder. This fact may serve as a confirmation of the following assumption made by Tatarnikova *et al.* (2001): at its

Table 1. Photoelectric UBV observations of YY Her

JD +2440000	U	B	V
12047	$13^m.73$	$14^m.24$	$12^m.95$
12048	13.66	14.21	12.97
12055	–	14.26	13.10
12106	13.66	14.30	13.16
12115	13.70	14.32	13.22
12135	13.38	14.28	13.32
12138	13.62	14.42	13.39
12168	13.38	14.27	13.18
12172	13.46	14.20	13.14
12191	13.43	14.17	13.06
12195	13.50	14.13	13.04
12202	13.38	14.13	13.01
12212	13.33	14.07	12.96
12223	13.35	14.03	12.86

Table 2. Photoelectric $BVR'I'$ observations of YY Her

JD +2440000	B	V	R'	I'
11865	–	$13^m.59$	$11^m.76$	$10^m.06$
11866	–	13.59	11.74	10.05
11867	–	13.64	11.76	10.09
11868	–	13.64	11.73	10.08
11887	$14^m.87$	13.40	11.58	10.00
11888	14.86	13.40	11.59	10.02
11952	–	13.14	11.54	10.10
12029	–	12.91	11.35	9.91
12033	–	12.94	11.33	9.90
12090	–	12.99	11.33	9.93
12162	14.27	13.14	11.51	10.08
12173	14.26	13.17	11.46	10.05
12176	–	13.08	11.43	10.03
12178	–	13.08	11.42	10.01
12188	–	13.02	11.39	9.99
12191	14.21	13.02	11.36	9.97
12247	14.04	12.95	11.33	9.93

primary minimum, the brightness of YY Her declines not only because the circumstellar envelope of the hot component is shielded by the red giant but also because the side of the red giant facing the hot component has a higher temperature and, accordingly, is optically brighter than its opposite side.

The Secondary Minimum

Let us first consider the parameters of the secondary minimum in UBV. We see from Fig. 1 that the secondary minimum is most pronounced in V, which is also reflected in the color–magnitude diagram (Fig. 2). YY Her in this spectral band becomes bluer as it passes through its secondary minimum and reddens as it passes through its primary minimum.

In V, the secondary minimum of 2001 appears symmetric about the vertical straight line that passes through the date of minimum light (Fig. 1). In other words, its descending and ascending branches are similar. This minimum is similar in shape to the two preceding minima and to the plot of the function $y = \text{const} \times |\sin x|$. Its depth is $\Delta V \approx 0.5$ and the maximum SD brightness, which corresponds to the minimum light of YY Her, is $V_{\rm sd} \approx 14^m.0$, while at the secondary minimum of 1998, $\Delta V \approx 0.25$ and $V_{\rm sd} \approx 14^m.6$.

In B, the light curve for the minimum of 2001 becomes asymmetric. After egress from its minimum, the star is brighter than during ingress. When calculating the maximum SD brightness, we assume that the star would have a magnitude intermediate between its magnitudes before and after the eclipse if it were outside the eclipse at this time. In this case, the SD magnitude is $B_{\rm sd} \approx 15^m.7$ and, accordingly, its color is $(B-V)_{0,\rm sd} \approx 1.5$. If the primary minimum is assumed to be a superposition of the secondary minimum and some addition, then the colors of the latter are $(U-B)_0 \approx -1.2$ and $(B-V)_0 \approx 0$. These are close to the corresponding colors of the dense gaseous envelope without any intense forbidden lines in its spectrum. An appreciable part of the envelope is eclipsed by the red giant at primary minimum, as suggested by its significant amplitude in U.

The $(B-V)_{0,\rm sd}$ color index for the secondary minimum of 1998 is $\sim$2.0, as estimated from our data, and $\sim$1.9, as estimated from the data of Hric *et al.* (2001). If we reduce the latter estimate to our photometric system, as was described above, then it will match our estimate. Of course, this match should not be overstated. If an optically thick (in this wavelength range) radiation source disappears from the observer's field of view during the secondary minimum, then according to the $(B-V)_{0,\rm sd}$ color, it must be cool enough. Its color temperature is $\sim$3300 K for the minimum of 2001.

Figure 3 shows the color variations of the SD with

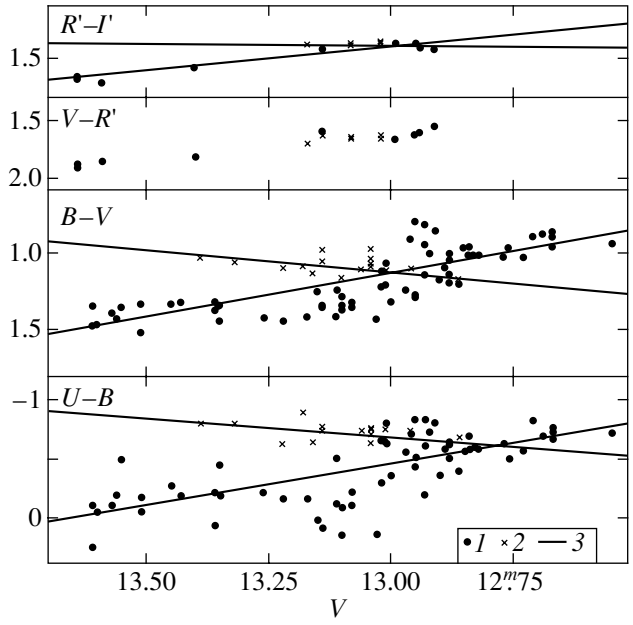


Fig. 2. $U-B$, $B-V$, $V-R'$, and $R'-I'$ colors versus V magnitude: (*1*) and (*2*) the data for the primary and secondary minima of the star, respectively; (*3*) the corresponding linear fits, which clearly show the color variations.

its visual brightness as the star passes through the minimum of 2001. We see from this figure that as the SD brightness rises, i.e., as the brightness of the star approaches its minimum, the SD becomes redder as does the PD.

The U light variations in YY Her during its secondary minimum are difficult to interpret in terms of this phenomenon. If we had at our disposal only observations in this spectral band, in which the gaseous envelope mainly radiates, then we would most likely conclude that the observed light variations in the star resulted from variations (for some reasons) in the envelope brightness.

The Secondary Minimum in $V_cR_cI_c$ and $VR'I'$

The depth of the secondary minimum in 1998 is $\Delta V_c \approx 0.26$, $\Delta R_c \approx 0.16$, and $\Delta I_c \approx 0.16$. Unfortunately, we have no $R'I'$ observations of YY Her during the maximum decline in its brightness in 2001. These were started during egress of the star from its minimum, when the fading was $\Delta V \approx 0.21$, $\Delta R' \approx 0.16$, $\Delta I' \approx 0.16$ and, accordingly, when the SD magnitudes were $V_{sd} \approx 14\overset{m}{.}7$, $R'_{sd} \approx 13\overset{m}{.}4$, $I'_{sd} \approx 12\overset{m}{.}0$.

Thus, as the star passes through its secondary minimum, its V_c-R_c and $V-R'$ color indices increase; i.e., the star reddens. At the same time, its color does not change in the red spectral band, as suggested by the constant R_c-I_c and $R'-I'$ color

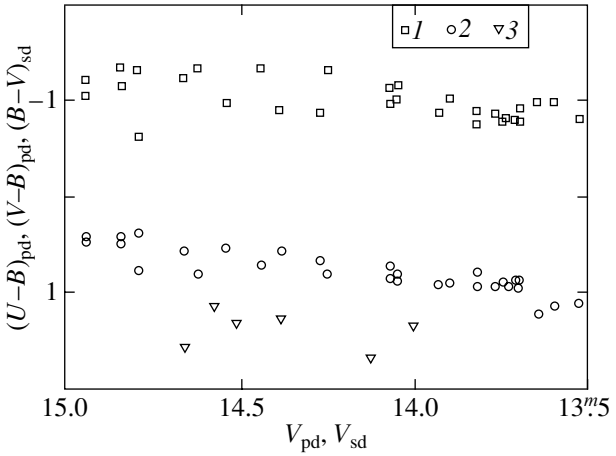


Fig. 3. The $(U-B)_{pd}-V_{pd}$ (*1*), $(B-V)_{pd}-V_{pd}$ (*2*), $(B-V)_{sd}-V_{sd}$ (*3*) diagrams.

indices (Fig. 2). The latter implies that the SD color in the red spectral band is the same as the color of the red giant, which almost completely determines the stellar flux in this spectral band when YY Her is in quiescence. The conclusion that the SD energy distribution is at a maximum in I' can be directly reached by calculating the corresponding fluxes from its magnitudes given above and by constructing the energy distribution in $(\lambda, F(\lambda))$ coordinates.

As was pointed out above, the R' and I' observations of YY Her in 2001 were started during egress of the star from its minimum and the fading was $\Delta V \approx 0.2$, while $\Delta V \approx 0.5$ at the minimum. Assuming that the $\Delta I'/\Delta V$ ratio in this segment of the light curve was the same as it was at the time when $\Delta V \approx 0.21$ ($\Delta I'/\Delta V \approx 0.75$), we find that the magnitude of the SD it had at the minimum visual light of YY Her was $I' \approx 11\overset{m}{.}4$. Since its energy distribution is similar to the energy distribution of the red giant, the decrease in bolometric flux during the secondary eclipse in 2001 may be said to have accounted for $\sim$20% of the bolometric flux from the cool component.

The Nature of the Secondary Minimum

Hric *et al.* (2001) assumed that the secondary minimum was produced by an eclipse of part of the red-giant surface by the circumstellar envelope of the hot component. However, this envelope is optically thin in continuum, as follows, in particular, from the presence of a Balmer jump in its emission. Besides, $\sim$20% of the red-giant surface should be screened. Such a large and hot screen would completely dominate in the optical emission from YY Her.

However, the screen can be cool and optically thick if it is associated with an accretion disk that has been exhausted during the accretion outburst

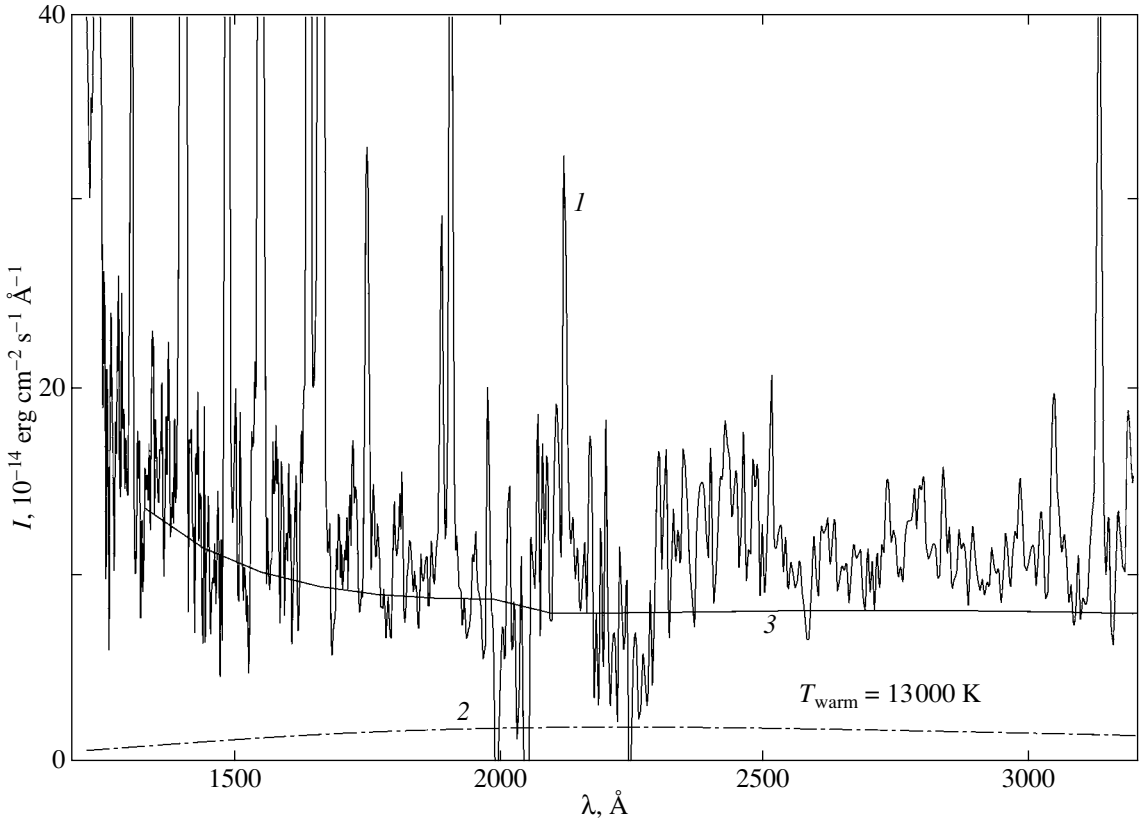


Fig. 4. (*1*) The May 25, 1980 spectrum of YY Her ($\varphi = 0.27$) corrected for interstellar reddening with $E(B-V) = 0^m.2$. (*2*) The spectral energy distribution for the warm component. (*3*) The computed continuum energy distribution for the star in a model that includes the contributions from the hot component, the gaseous nebula, and the warm component. The hot-component parameters were taken from Munari *et al.* (1997b).

of the star caused by its instability and that again began to build up when the star passed to quiescence (Duschl 1986a, 1986b). Note that the binary system YY Her is seen at a large angle (Tatarnikova *et al.* 2001), so the disk is seen almost edge-on (but the hot component is open for the observer). Accordingly, for this reason, it will appear as a cool radiation source.

Until the disk, while building up, reaches the critical surface density at which it eruptively passes to an active state, the accretion rate is low (the disk is passive) and the accretion luminosity of the hot component can be appreciably lower than its intrinsic luminosity if it is structurally the hot nucleus of a young planetary nebula. In this case, the minimum (quiescent) brightness of the hot component is determined by its intrinsic luminosity. Accordingly, after its outburst, the star will each time return to the same fixed appearance (Tatarnikova *et al.* 2000) typical of the symbiotic stars that are not undergoing novalike outbursts.

The star V443 Her may serve as an analog of YY Her in quiescence. These stars have the same orbital period, the same spectral type of cool component, and the same ratio of bolometric luminosities of the hot component (for YY Her in quiescence) and the red giant. The only difference is that V443 Her exhibits no novalike activity.

Tatarnikova *et al.* (2001) assumed that the secondary minimum resulted from an ellipsoidal shape of the cool component of YY Her. In that case, the deepening of the secondary minimum with the brightness of the star outside its minima preserved should be attributed to variations in the shape of the cool component and in the temperature of individual parts of its surface. Recall that the secondary minimum deepened by a factor of $\sim$2 over two orbital periods. This implies that the restructuring, say, on the thermal scale of the red giant must be eruptive in nature and may serve as a precursor of a new outburst of YY Her. In this hypothesis, the similarity between the shapes of the primary and secondary eclipses is something of a mystery.

Of particular interest is the fact that the amplitude of the secondary minimum in V is appreciably larger than in R and I. In I, as was mentioned above, the cool component gives an overwhelming contribution to the total flux. In V, apart from the contribution of the red giant, there is an addition due to the radiation

from the gaseous nebula. In that case, the amplitude of the secondary minimum in V must be smaller than in I (if the temperature distribution over the cool-component disk is disregarded). The cool screen may eclipse the hotter part of the red giant (the reflection effect); hence the large amplitude in the visual band. In this case, however, the SD must become bluer as the center of the secondary minimum is approached.

Munari *et al.* (1997b) introduced the concept of a warm component, as applied not only to the active state of YY Her but also to its quiescence. Thus, the IUE spectrum taken on May 25, 1980, at phase $\varphi \approx 0.27$ (see Fig. 4) cannot be represented by the standard three-component model (the hot component, the gaseous nebula, and the cool component). An additional radiation source whose continuum energy distribution is modeled by a Planck function with temperature 10000–13000 K is required to be introduced. For a temperature of 13000 K, the dereddened magnitudes of the warm component are $U_{\rm warm} = 13^m.78$, $B_{\rm warm} = 14^m.72$, $V_{\rm warm} = 14^m.73$. However, three months later, this radiation source ceased to show up in the UV continuum energy distribution. The spectrum taken on August 12, 1980, ($\varphi \approx 0.41$) exhibits no deviations from the standard model (see Munari *et al.* 1997b).

Subtracting the radiation source with the above magnitudes from the UBV estimates for the maximum of 2001 yields a secondary minimum with a depth $\Delta U \approx 0^m.25$, $\Delta B \approx 0^m.25$, $\Delta V \approx 0^m.11$. However, the U light curve can often reflect the light variations in the gaseous envelope attributable to intrinsic variability of the hot subdwarf. Munari *et al.* (1997b) showed that such variations take place for the quiescent state of YY Her. In addition, the above amplitudes of the minimum strongly depend on the assumed temperature of the warm component. The latter can be determined more or less accurately only when simultaneous UV and optical spectroscopic observations are available. Unless the presence or absence of a near-UV excess is attributed to the appearance or disappearance of the warm component, it should be assumed that the variability of this excess results from different areas of the warm-component projection onto the plane of the sky. Note that in this case, the warm component cannot be an accretion disk. This radiation source must be a nonspherical component located between the hot and cool components [of the hot-line type introduced by Bisikalo *et al.* (1998) for cataclysmic variables].

Optical spectrophotometry reveals a modest contribution from the warm component at the maxima of 1995 and 1996. In both cases, the excess produced by the warm component appeared after egress from the primary minimum ($\varphi \approx 0.25$) and disappeared near the secondary minimum ($\varphi \approx 0.5$). There is no doubt that the presence or absence of such a warm component will in no way affect the magnitudes in R and I, where the photometric variability is attributable to the ellipsoidal shape of the red giant. In the visual band, however, both these effects can play a role.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16235).

REFERENCES

1. D. V. Bisikalo, A. A. Boyarchuk, V. M. Chechetkin, *et al.*, Mon. Not. R. Astron. Soc. **300**, 39 (1998).
2. W. J. Duschl, Astron. Astrophys. **163**, 56 (1986a).
3. W. J. Duschl, Astron. Astrophys. **163**, 61 (1986b).
4. L. Hric, K. Petrik, and P. Niarchos, Inf. Bull. Var. Stars, No. 5046 (2001).
5. H. L. Jonhson, Annu. Rev. Astron. Astrophys. **4**, 193 (1996).
6. S. J. Kenyon, *The Symbiotic Stars* (Cambridge Univ. Press, Cambridge, 1986), p. 283.
7. U. Munari, M. Rejkuba, M. Hazen, *et al.*, Astron. Astrophys. **323**, 113 (1997a).
8. U. Munari, E. A. Kolotilov, A. A. Popova, and B. F. Yudin, Astron. Zh. **74**, 898 (1997b) [Astron. Rep. **41**, 802 (1997b)].
9. A. A. Tatarnikova, V. F. Esipov, E. A. Kolotilov, *et al.*, Pis'ma Astron. Zh. **27**, 825 (2001) [Astron. Lett. **27**, 703 (2001)].
10. A. A. Tatarnikova, M. Rejkuba, L. M. Buson, *et al.*, Astron. Zh. **77**, 220 (2000) [Astron. Rep. **44**, 190 (2000)].
11. B. F. Yudin, Astrophys. Space Sci. **135**, 143 (1987).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 626–631. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 695–700.

The Direction of Propagation of the Solar Dynamo Wave

L. L. Kitchatinov*

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

Received March 19, 2002

Abstract—We propose a solution to one of the oldest problems in the solar-dynamo theory: explaining the equatorward drift of magnetic activity in the solar cycle. The well-known suggestion that the dynamo waves propagate along the surfaces of constant angular velocity is shown to be restricted to an isotropic medium. Allowance for the rotation-induced anisotropy in turbulent diffusion leads to an equatorward deviation of the wave phase velocity from the isorotational surface. Estimates for the dynamo waves are illustrated with two-dimensional numerical models in a spherical geometry. The model with anisotropic diffusion also shows an equatorward drift of the toroidal magnetic field when the rotation is radially uniform.

Key words: *Sun, magnetic fields, rotation, turbulence, MHD*

INTRODUCTION

One of the difficulties of the solar-dynamo theory has long been a mismatch between the propagation direction of dynamo waves expected from the internal rotation of the Sun and the observed equatorward drift of activity (Parker 1987). The dynamo waves migrate along the so-called isorotational surfaces (the surfaces of constant angular velocity; Yoshimura 1975). Therefore, before the emergence of helioseismology, the fact that the angular velocity in the solar interior increases with depth was considered to be almost obvious. Only in this case was it possible to achieve agreement with observations. However, helioseismology shows that the angular velocity depends on the depth only slightly (Schou *et al.* 1998).

A thin layer at the base of the convection zone where the radial rotation nonuniformity is significant constitutes an exception. However, the assumption of the dynamo action in this layer leads to even greater difficulties. First, the toroidal field at the base of the convection zone must reach $\sim 10^5$ G to account for the active regions observed on the solar surface (Schüssler *et al.* 1994). For the differential rotation to generate such a strong toroidal field during the solar cycle, the poloidal field must have a strength of more than 100 G, i.e., two orders of magnitude larger than the observed strength. Second, as can be easily estimated, the Lorentz force from such strong fields will destroy the differential rotation in less than one year. Third, the angular velocity at the base of the convection zone at low latitudes decreases with depth. In this case, the dynamo waves drift equatorward only when the helicity of convective motions is positive (a negative alpha-effect). The dynamo equations then give a phase relation between the poloidal and toroidal field opposite to the observed one (Stix 1976). An alternative to the dynamo action throughout the convection zone can hardly be found. Note that large-scale fields in a turbulent medium are subjected to magnetic buoyancy only slightly (Kichatinov and Pipin 1993a, 1993b). However, the problem of the latitudinal drift of magnetic activity mentioned above remains.

Here, our goal is to draw attention to the fact that the validity of the generally accepted suggestion that the dynamo waves propagate along isorotational surfaces is restricted to the simplest case of an isotropic medium. The solar rotation breaks the isotropy. Under the effect of Coriolis forces, the characteristic scales of the convective motions along and across the rotation axis differ significantly (Gillman and Miller 1986). As a result, a difference in the effective magnetic diffusion coefficients arises for these two directions (Kitchatinov *et al.* 1994). Note that the anisotropy emerges for the same reason (rotation) as the dynamo effect. This naturally leads to the problem of dynamo waves in an anisotropic medium.

As we show in the next section, the anisotropy in turbulent diffusion affects the threshold dynamo number and, more importantly, gives rise to a phase velocity component normal to the isorotational surface. For the solar conditions, this gives an additional equatorward drift of the dynamo waves.

Below, we illustrate these results by numerical solutions of the two-dimensional dynamo problem. Our

*E-mail: kit@iszf.irk.ru

calculations including anisotropy show an equatorward drift of the toroidal field even when the angular velocity depends only on latitude but not on radius.

DYNAMO WAVES IN AN ANISOTROPIC MEDIUM

Dynamo waves, probably, give the simplest example of a hydromagnetic dynamo. The validity of the wave approximation is restricted by the condition of the field spatial scale being small compared to the system size. This scale ratio is not typical of actual objects. Nevertheless, the results of global (that do not assume the field scale to be small but that take into account the object geometry and boundary conditions), generally numerical models are excellently explained in terms of dynamo waves. In the wave approximation, great progress can be made by using analytic methods and the physics of the corresponding dynamo mechanisms can be elucidated (Kuzanyan and Sokoloff 1995, 1997; Galitskiĭ and Sokoloff 1998). Therefore, we also begin our discussion of the role of anisotropy with an analysis of dynamo waves.

We proceed from the induction equation for the mean field (see, e.g., Krause and Rädler 1980),

$$\partial\mathbf{B}/\partial t = \mathrm{curl}(\mathbf{V}\times\mathbf{B} + \boldsymbol{\alpha}\mathbf{B} - \eta\mathrm{curl}\mathbf{B} - \eta_\parallel \mathbf{e}\times((\mathbf{e}\cdot\nabla)\mathbf{B})). \quad (1)$$

In this equation, η means the isotropic turbulent-diffusion component of the magnetic field, $\eta_\parallel$ includes anisotropy and means the additional diffusion along a unit vector $\mathbf{e}$; i.e., the diffusion coefficient is $\eta+\eta_\parallel$ for the direction of $\mathbf{e}$ and η for all of the directions normal to $\mathbf{e}$. All coefficients in Eq. (1) are assumed to be constant, irrespective of the coordinates, except for the velocity $\mathbf{V}$, which represents a flow with a uniform shear:

$$\mathbf{V} = \mathbf{j}(\mathbf{G}\cdot\mathbf{r}), \quad (2)$$

where $\mathbf{j}$ is a unit vector along the flow, $\mathbf{G}$ is the coordinate-independent shear vector normal to $\mathbf{j}$, and $\mathbf{r}$ is the radius vector. The alpha effect of cyclonic convection (Krause and Rädler 1980) is represented in Eq. (1) by the tensor $\boldsymbol{\alpha}$, which can also contain anisotropy. However, below, we restrict our analysis to the so-called $\alpha\Omega$-dynamo (Krause and Rädler 1980). Since only one of the components of the tensor $\boldsymbol{\alpha}$ is significant for it, the anisotropy of the alpha effect is of no importance.

Since the direction of $\mathbf{j}$ corresponds to the longitude on the Sun and since the preferential direction of anisotropy $\mathbf{e}$ coincides with the rotation axis, we assume $\mathbf{e}$ to be normal to $\mathbf{j}$. The magnetic field is assumed to be uniform along $\mathbf{j}$ and to be a superposition of the toroidal (B) and poloidal components. The latter is determined by the toroidal potential A:

$$\mathbf{B} = \mathbf{j}B + \mathrm{curl}(\mathbf{j}A). \quad (3)$$

The linear dynamo problem reduces to the eigenvalue problem for plane waves:

$$B, A \sim \exp(\sigma t + \mathrm{i}\mathbf{k}\cdot\mathbf{r}), \quad \mathbf{k}\cdot\mathbf{j} = 0. \quad (4)$$

Substituting (2)–(4) in Eq. (1) yields

$$\sigma B = -\mathrm{i}((\mathbf{k}\times\mathbf{G})\cdot\mathbf{j})A - (\eta k^2 + \eta_\parallel(\mathbf{k}\cdot\mathbf{e})^2)B \quad (5)$$
$$\sigma A = \alpha B - (\eta k^2 + \eta_\parallel(\mathbf{k}\cdot\mathbf{e})^2)A.$$

Here, α is the azimuthal component of the tensor $\boldsymbol{\alpha}$: $\alpha = \boldsymbol{\alpha}_{mn}j_m j_n$. Recall that we consider the $\alpha\Omega$-dynamo; i.e., the toroidal field is generated solely by a shear flow, not by the alpha effect.

From the condition of consistency of Eqs. (5), we determine the eigenvalues that can correspond to undamped, $\mathrm{Re}(\sigma)\geq 0$, waves,

$$\sigma = -\eta k^2 - \eta_\parallel(\mathbf{k}\cdot\mathbf{e})^2 + \begin{cases} (1+\mathrm{i})\sqrt{\frac{1}{2}\alpha((\mathbf{G}\times\mathbf{k})\cdot\mathbf{j})} & \text{for} \quad \alpha((\mathbf{G}\times\mathbf{k})\cdot\mathbf{j}) > 0 \\ (1-\mathrm{i})\sqrt{-\frac{1}{2}\alpha((\mathbf{G}\times\mathbf{k})\cdot\mathbf{j})} & \text{for} \quad \alpha((\mathbf{G}\times\mathbf{k})\cdot\mathbf{j}) < 0. \end{cases} \quad (6)$$

Separating out the real and imaginary parts in the eigenvalue, $\sigma = \gamma - \mathrm{i}\omega$, we find the wave frequency,

$$\omega = \mp\sqrt{\frac{1}{2}|\alpha((\mathbf{G}\times\mathbf{k})\cdot\mathbf{j})|}, \quad (7)$$

where the upper and lower signs correspond to the upper and lower rows in formula (6). The relation for the dimensionless increment $\hat{\gamma} = \gamma/(\eta k^2)$ is

$$\hat{\gamma} = -1 - a\cos^2(\theta-\phi) + \sqrt{|\mathcal{D}\sin\phi|}. \quad (8)$$

Here, $a = \eta_\parallel/\eta$ is the anisotropy parameter; θ and ϕ are the angles formed by the rotation axis ($\mathbf{e}$) and the wave vector $\mathbf{k}$, respectively, with the shear vector $\mathbf{G}$;

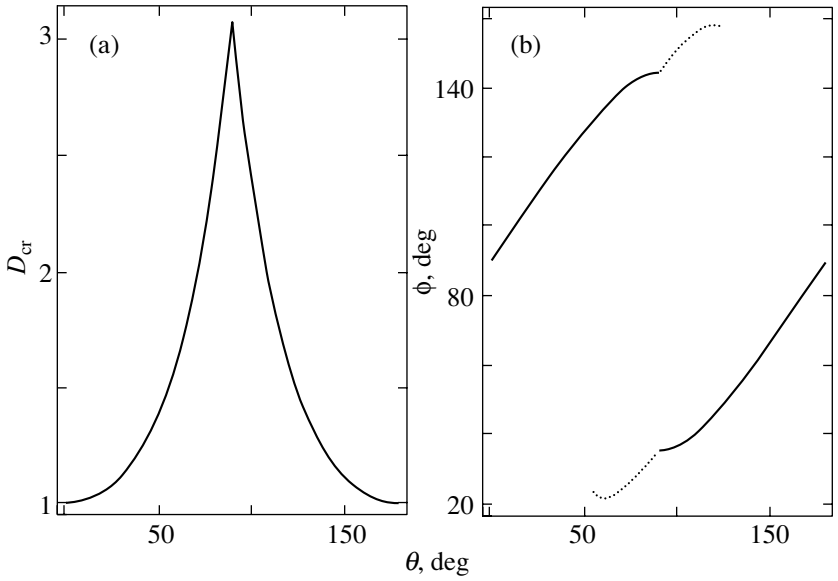


Fig. 1. (a) Critical dynamo number (9) *versus* angle θ between the anisotropy direction and the shear vector $\mathbf{G}$; (b) the same diagram for the angle ϕ between the dynamo-wave velocity (10) and the vector $\mathbf{G}$. The dotted lines correspond to the secondary (not largest) maximum of increment (8). The anisotropy parameter is $a = 1$. The angles are in degrees.

and

$$\mathcal{D} = \frac{\alpha G}{\eta^2 k^3} \tag{9}$$

is the dynamo number.

The observed latitudinal migration of solar activity is attributed to the propagation of dynamo waves. In this case, we must decide precisely which wave velocity, the phase or group velocity, is meant. The group velocity at which the wave energy is transferred is believed to have a deeper meaning. Indeed, for wave perturbations in the absence of dissipation, energy is conserved and the direction of the energy flux is of great importance. The situation with the dynamo wave is completely different. Such waves continuously exchange energy with the turbulent medium. The dynamo-wave energy is not conserved. By the latitudinal migration of solar activity, we mean time variations of the latitude with the highest formation rate of active regions. In other words, the motion of the dynamo-wave crest is observed. This is, of course, the phase velocity

$$\mathbf{v}_\mathrm{p} = \mathbf{k}\omega/k^2. \tag{10}$$

In the absence of anisotropy ($a = 0$), the field strengthens when the dynamo number exceeds the critical value $\mathcal{D}_\mathrm{cr} = 1$. In this case, the phase velocity for neutrally stable perturbations is directed along the normal to the shear vector $\mathbf{G}$, which corresponds to the well-known wave propagation along isorotational surfaces (Yoshimura 1975). The choice between the two possible propagation directions on this surface is determined by the sign of the dynamo number.

Let us now consider the effect of anisotropy. For large-scale solar convection, the Coriolis number is $\Omega^* = 2\tau\Omega \simeq 6$ (Durney and Latour 1978); τ is the rotation time of a convective cell. For such Ω^*, rotation significantly perturbs the convection and a quasi-linear theory (Kitchatinov *et al.* 1994) yields similar (in magnitude) coefficients η and $\eta_\parallel$; i.e., the anisotropy parameter is $a = \eta_\parallel/\eta \sim 1$. Without loss of generality, we may assume the dynamo number to be positive and the angles θ and ϕ to be within the range 0 to π. In Fig. 1, the critical dynamo number and the angle ϕ between the wave phase velocity (10) and shear vector $\mathbf{G}$ are plotted against the angle θ between the vector $\mathbf{G}$ and the anisotropy direction (rotation axis). As can be seen, anisotropy increases $\mathcal{D}_\mathrm{cr}$, which is attributable to the additional diffusion $\eta_\parallel$. The velocity $\mathbf{v}_\mathrm{p}$ acquires a component normal to the surfaces of constant flow velocity.

This change in the direction of wave propagation can be explained by the fact that it reduces the diffusion losses. For a fixed magnitude of the wave vector, the effect of the additional diffusion along $\mathbf{e}$ will decrease as the angle between this direction and the velocity $\mathbf{v}_\mathrm{p}$ approaches $\pi/2$. Only for parallel (or antiparallel) orientation of the anisotropy and shear vectors do the waves propagate as usual along the surfaces of constant flow velocity. For the other anisotropy orientations, the waves propagate at an angle to the isorotational surfaces while tending to minimize the diffusion losses ($\theta - \phi \to \pi/2$). As we see from Fig. 1, this angle is not small. The discontinuity in the θ dependence of ϕ near $\theta = \pi/2$ stems

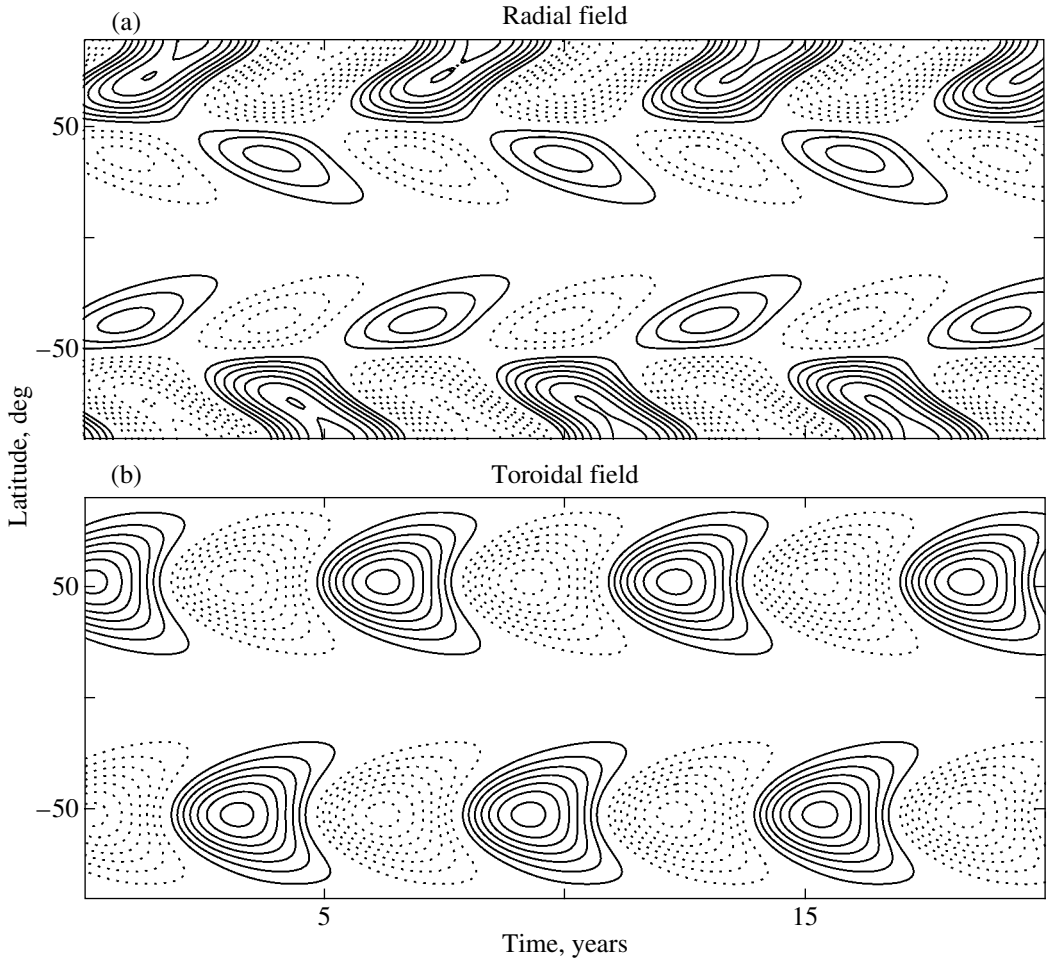


Fig. 2. Maunder diagrams for the model with isotropic diffusion ($a = 0$). Isolines of the radial (a) and azimuthal (b) field components on the surface are shown. The solid and dotted lines indicate the positive and negative levels, respectively. There is virtually no latitudinal migration of the toroidal field.

from the fact that there are two maxima here in the ϕ dependence of increment γ (8); the position of each of them continuously changes with θ but the larger maximum, which determines $\mathcal{D}_{\rm cr}$, passes from one branch to the other as θ passes through $\pi/2$. The value of ϕ corresponding to the smaller (secondary) maximum of increment (8) is indicated in Fig. 1 by the dotted line.

If the rotation rate in the solar convective envelope is assumed to depend on the latitude alone, then the vector $\mathbf{G}$ is directed toward the equator. The dynamo-wave velocity also deviates equatorward from the radial direction, which is required to reconcile theory with observations.

GLOBAL MODELS

Let us illustrate the results of the previous section with numerical solutions of the dynamo equations in a spherical geometry.

We take the convection zone, i.e., a spherical layer with the upper boundary $r_{\rm e} = R_\odot = 6.96 \times 10^{10}$ cm and the base $r_{\rm i} = 0.7R_\odot$ to be the region of dynamo action. The mean diffusion coefficient η_0 between the directions along and across the rotation axis is assumed to be constant, irrespective of the coordinates,

$$\eta_0 = \eta + \eta_\parallel/2 = 10^{12}\ \ {\rm cm}^2\ {\rm s}^{-1}. \qquad (11)$$

The coefficient $\alpha = \boldsymbol{\alpha}_{\phi\phi}$ of the alpha effect is assumed to depend on colatitude θ (below, θ is the colatitude) as $\cos\theta$ and to be independent of depth:

$$\alpha = \frac{\alpha_0 \cos\theta}{1 + (B/B_{\rm eq})^2}. \qquad (12)$$

Here, we took into account the simplest nonlinearity in the form of suppression of the alpha effect by the magnetic field; the magnitude of the equipartition field $B_{\rm eq}$ is taken to be 3000 G.

Let us consider axisymmetric (longitude-independent) solutions of the induction equation (1) with

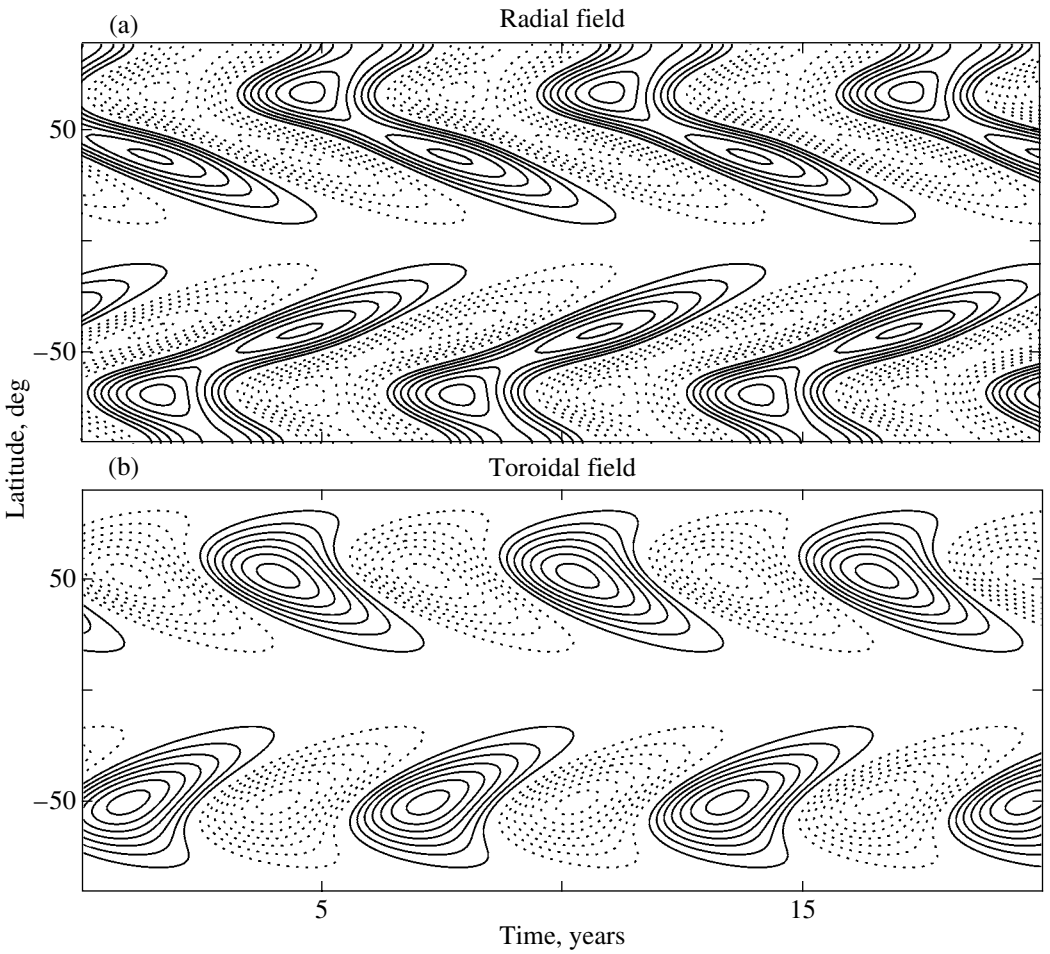


Fig. 3. Same as Fig. 2 for anisotropic diffusion ($a = 1$). Anisotropy results in an equatorward drift of the toroidal field.

boundary conditions. For the base of the convection zone, we use the condition for the interface with a superconducting medium:

$$\boldsymbol{\varepsilon} \times \mathbf{n} = 0 \quad \text{for} \quad r = r_{\rm i}, \tag{13}$$

where $\mathbf{n}$ is the normal vector to the surface and $\boldsymbol{\varepsilon}$ is the so-called mean EMF (see, e.g., Krause and Rädler 1980); i.e., the vector under the curl sign on the right-hand side of the induction equation (1). On the outer surface, we use the nonlinear condition (Kitchatinov *et al.* 2000) for the toroidal field that corresponds to the field escape through this surface because of instability against the formation of magnetic loops:

$$\varepsilon_\theta + \frac{\eta_0}{R_\odot}\frac{B}{B_0} = 0 \quad \text{for} \quad r = R_\odot. \tag{14}$$

The threshold value $B_0 = 200$ G was chosen to agree with the observed strengths (200–600 G) of the fields that escape through the solar surface when active regions are formed (Zwaan 1992; Lites *et al.* 1998).

In the absence of a proper boundary condition for the poloidal field, we use the standard (quasi-vacuum) condition that requires the vertical orientation of the field lines on the surface,

$$\partial rA/\partial r = 0 \quad \text{for} \quad r = R_\odot. \tag{15}$$

To consider the role of anisotropy in a pure form, we eliminate the effect of radial rotation nonuniformity on the latitudinal drift of the field by assuming the angular velocity to be uniform in depth and latitude-dependent, in agreement with the Doppler measurements of Howard and Harvey (1970):

$$\Omega = 2.78(1 - 0.13\cos^2\theta - 0.16\cos^4\theta)\mu \ \text{rad s}^{-1}. \tag{16}$$

Recall that the field generation takes place when the dynamo number exceeds some critical value. This critical value depends on the anisotropy parameter $a = \eta_\parallel/\eta$. Below, we compare the results for the isotropic ($a = 0$) and anisotropic ($a = 1$) cases. The critical values of $C_\alpha = \alpha_0 R_\odot/\eta_0$ for these cases are 2.69 and 2.51, respectively. The solar dynamo

is assumed to be slightly supercritical. The results presented below correspond to dynamo numbers exceeding the threshold values by 10%, i.e., $C_\alpha = 2.96$ and $C_\alpha = 2.76$ for the isotropic and anisotropic case, respectively.

Figure 2 shows the computed field isolines in latitude–time coordinates (Maunder diagram) in the absence of anisotropy ($a = 0$). In essence, these calculations reproduce the result of Köhler (1973) that there is no latitudinal drift of the field in an isotropic medium when the rotation is radially uniform. There is only a latitudinal diffusion "spread" of the field.

Note that the cycle period in the model under discussion is approximately a factor of 3 shorter than the observed period. The computed period can be increased (or decreased) by decreasing (increasing) the diffusion coefficient η_0. Such model adjustment is unlikely to be meaningful. The assumed $\eta_0 = 10^{12}$ cm^2 s^{-1} was already slightly underestimated. The computed cycle appears to be too short, because there is no proper boundary condition for the poloidal field (Kitchatinov *et al.* 2000).

Figure 3 shows the results of our calculations including anisotropy. As expected from our estimates for the dynamo waves, an equatorward drift of the field emerges. Note once again that it was obtained when the angular velocity was radially uniform. In all likelihood, allowance for the rotation-induced anisotropy in turbulent diffusion of the magnetic field solves one of the oldest problems in dynamo theory concerning the origin of the equatorward drift of magnetic activity in the solar cycle.

ACKNOWLEDGMENTS

I wish to thank V.V. Pipin for helpful discussions. This work was supported by the Russian Foundation for Basic Research (project no. 02-02-16044).

REFERENCES

1. B. R. Durney and J. Latour, Geophys. Astrophys. Fluid Dyn. **9**, 241 (1978).
2. V. M. Galitskiĭ and D. D. Sokolov, Astron. Zh. **75**, 144 (1998) [Astron. Rep. **42**, 127 (1998)].
3. P. A. Gillman and J. Miller, Astrophys. J., Suppl. Ser. **61**, 585 (1986).
4. R. Howard and J. Harvey, Sol. Phys. **12**, 23 (1970).
5. L. L. Kitchatinov and V. V. Pipin, Pis'ma Astron. Zh. **19**, 557 (1993a) [Astron. Lett. **19** (1993a)].
6. L. L. Kitchatinov and V. V. Pipin, Astron. Astrophys. **274**, 647 (1993b).
7. L. L. Kitchationov, V. V. Pipin, and G. Rüdiger, Astron. Nachr. **315**, 157 (1994).
8. L. L. Kitchatinov, M. V. Mazur, and M. Jardine, Astron. Astrophys. **359**, 531 (2000).
9. H. Köhler, Astron. Astrophys. **25**, 467 (1973).
10. F. Krause and K. Radler, *Mean Field Magnetohydrodynamics and Dynamo* (Akademie-Verlag, Berlin, 1980; Mir, Moscow, 1984).
11. K. M. Kuzanyan and D. D. Sokoloff, Geophys. Astrophys. Fluid Dyn. **81**, 113 (1995).
12. K. M. Kuzanyan and D. D. Sokoloff, Sol. Phys. **173**, 1 (1997).
13. B. W. Lites, A. Skumanich, and V. Martinez Pillet, Astron. Astrophys. **333**, 1053 (1998).
14. E. N. Parker, Sol. Phys. **110**, 11 (1987).
15. J. Schou, H. M. Antia, S. Basu, *et al.*, Astrophys. J. **505**, 390 (1998).
16. M. Schüssler, P. Caligari, A. Ferriz-Mas, and F. Moreno-Insertis, Astron. Astrophys. **281**, L69 (1994).
17. M. Stix, Astron. Astrophys. **47**, 243 (1976).
18. H. Yoshimura, Astrophys. J. **201**, 740 (1975).
19. C. Zwaan, in *Sunspots: Theory and Observations,* Ed. by J. H. Thomas and N. O. Weiss (Kluwer, Dordrecht, 1992), p. 75.

Translated by G. Rudnitskii

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 632–640. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 701–710.

Effects of Nonlinear Alfvén Wave Interaction on Particle Diffusive Shock Acceleration. The Earth's Bow Shock

E. G. Berezhko, S. I. Petukhov, and S. N. Taneev*

Institute of Cosmophysical Research and Aeronomy, Siberian Branch, Russian Academy of Sciences, pr. Lenina 31, Yakutsk, 677891 Russia

Received January 29, 2002

Abstract—We investigate the role of nonlinear Alfvén-wave interaction in the diffusive shock acceleration of solar-wind ions at the Earth's bow shock. Allowance for the nonlinear wave interaction through induced scattering and two-quanta absorption at plasma parameters $\beta \lesssim 0.1$ is shown to limit the Alfvén-wave amplitude δB to $\delta B \lesssim B$, whereas the quasi-linear approach predicts the generation of waves with amplitudes much larger than the diffusive shock magnetic field strength B. The nonlinear interaction results in spectral wave energy transfer to lower frequencies, which yields a significant increase in the particle acceleration rate. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *solar wind, Alfvén waves, Earth's bow shock, diffusive acceleration.*

INTRODUCTION

The development of a theory for diffusive shock acceleration [see, e.g., the review article by Berezhko and Krymsky (1988) and the monograph by Berezhko *et al.* (1988)], as applied to interplanetary phenomena, is needed to gain a complete understanding of how the energetic-ion spectra are formed at shock fronts. The numerous measurements performed at the Earth's bow-shock front suggest an intensive generation of energetic ions (the so-called diffusive component), which is almost always accompanied by a significant increase in Alfvén turbulence (see, e.g., Childers and Russell 1972; Trattner *et al.* 1994). Previously (Berezhko and Taneev 1991, 1992; Berezhko *et al.* 1997), we studied the acceleration of ions at the Earth's bow shock in the quasi-linear approximation. Our calculations showed that for some sets of solar wind parameters, the amplitude of the waves generated by accelerated particles, δB, can reach or even exceed the interplanetary magnetic field (IMF) strength B. In this case, the quasi-linear approach becomes inapplicable and the nonlinear wave interaction must be taken into account.

Under typical conditions of a quasi-parallel shock (we restrict our analysis to this simple case), the accelerated particles (ions) excite mostly Alfvén waves propagating along the diffusive shock magnetic field. For these conditions, the principal mechanisms of nonlinear Alfvén-wave interaction are induced scattering and two-quanta absorption (Fedorenko *et al.* 1990, 1995; Zirakashvili 2000).

Here, we study the diffusive shock acceleration of ions at the Earth's bow-shock front by taking into account the nonlinear Alfvén-wave interaction.

THE MODEL

The problem under consideration was formulated in the quasi-linear approximation previously (Berezhko and Taneev 1991, 1992; Berezhko *et al.* 1997; see also Lee 1982). Therefore, we present here only its main content with the inclusion of elements that take into account the nonlinear Alfvén-wave interaction.

Since the diffusive component is observed under typical conditions of a quasi-parallel shock (the angle ψ between **B** and **u** is less than 45°), we assume that the IMF **B** is directed radially, just as the solar-wind velocity **u**, along the x axis of the reference frame whose origin is associated with the shock front (to be more precise, with the forward part of the shock, which in our model is assumed to be plane and perpendicular to the solar-wind velocity). Given that the ions are highly magnetized ($\kappa_\parallel \gg \kappa_\perp$), the equation for their distribution function $f(x, v, t)$ in the upstream region ($x < 0$) can be represented as

$$\frac{\partial f}{\partial t} = \frac{\partial}{\partial x}\left(\kappa_\parallel \frac{\partial f}{\partial x}\right) - u\frac{\partial f}{\partial x} - \frac{f}{\tau_\perp}, \qquad (1)$$

where $\kappa_\parallel$ and $\kappa_\perp$ are the parallel and perpendicular (with respect to IMF) diffusion coefficients of the ions, respectively; and v is their velocity. Below, we disregard the Alfvén velocity c_a compared to u; this is

*E-mail: taneev@ikfia.ysn.ru

justified for the solar-wind conditions ($c_a \ll u$). The last term in Eq. (1) describes the particle escape from the acceleration region through diffusion across IMF field lines with the time scale

$$\tau_\perp = a^2/\kappa_\perp,$$

where a is the effective transverse size of the shock, i.e., of the acceleration region $y^2 + z^2 \le a^2$. As previously (Berezhko and Taneev 1991, 1992; Berezhko *et al.* 1997), the size of the acceleration region is taken to be $a = 4.2R_E$, where R_E is the Earth's radius.

At the shock front ($x = 0$), at which the solar-wind velocity decreases by a factor of σ (σ is the compression ratio at the shock front), Eq. (1) is supplemented with the boundary condition

$$\frac{u}{q} v \frac{\partial f}{\partial v} + \kappa_\parallel \frac{\partial f}{\partial x} = Q_0, \qquad (2)$$

where

$$Q_0 = u \frac{N_{\rm inj}}{4\pi v_{\rm inj}^2} \delta(v - v_{\rm inj}) H(t) \qquad (3)$$

is the source concentrated at the shock front that injects into the acceleration $uN_{\rm inj}$ particles with velocity $v_{\rm inj}$ per unit front surface area per unit time, $q = 3\sigma/(\sigma - 1)$. The Heaviside step function $H(t)$ in this expression indicates that the source switched on at time $t = 0$.

Presently, there is no comprehensive theory of the shock transition in a strong quasi-parallel collisionless shock wave that could predict the number of injected particles. Therefore, we assume that some fraction $\eta \ll 1$ of solar-wind particles are drawn into diffusive shock acceleration: $N_{\rm inj} = \eta N$, where N is the number density of the solar-wind ions.

The diffusion of particles is attributable to their resonant interaction with MHD waves, in our case, with Alfvén waves:

$$\kappa_\parallel = \frac{v^2 B^2}{32\pi^2 \omega_B E_w \left(k = \rho_B^{-1}\right)}, \qquad (4)$$
$$\kappa_\parallel \kappa_\perp = \rho_B^2 v^2/3.$$

Here, $\rho_B = v/\omega_B$ is the gyroradius, ω_B is the gyrofrequency, and $E_w(k) = d(\delta B^2/8\pi)/d\ln k$ is the energy density of the Alfvén waves. The particles interact only with waves whose wave number k is equal to the reciprocal of the particle gyroradius.

The spectrum of the Alfvén waves in the undisturbed solar wind,

$$E_{w0} = E_w(x = -\infty, k, t = 0) \qquad (5)$$

is a mixture of the $E_w^+(E_w^-)$ waves propagating away from (toward) the Sun relative to the coordinate system associated with the solar wind, $E_w = E_w^+ + E_w^-$. Note that specifying the boundary conditions corresponding to the unperturbed solar wind at $x = -\infty$ should not lead to misunderstanding. An infinitely large distance from the shock front actually implies a distance $|x| \gtrsim 100R_{\rm E} \sim 10^{11}$ cm to which no accelerated particles penetrate and, accordingly, no Alfvén waves are excited. At the same time, this distance is much smaller than the Earth's orbital radius, $\sim 10^{13}$ cm, which justifies the plane-wave geometry used here.

Induced scattering and two-quanta absorption are taken into account by adding the last two terms to the transfer equation for Alfvén turbulence, which describes the modification of the background Alfvén-wave spectrum by accelerated particles:

$$\frac{\partial E_w^\pm}{\partial t} + u \frac{\partial E_w^\pm}{\partial x} = \mp \Gamma E_w^\pm + \Gamma_{NL}^\pm E_w^\pm + Q^\pm. \qquad (6)$$

Here,

$$\Gamma = \frac{12\pi^3 c_a}{kc^2} \sum_s \frac{(Ze)^2}{Am} \qquad (7)$$
$$\times \int_{v_{\min}}^{\infty} dv' v' \left(1 - \frac{\omega_B^2}{k^2 v'^2}\right) \kappa_\parallel \frac{\partial f}{\partial x}$$

is the increment of wave excitation by the accelerated particles; $v_{\min} = \max(v_{\rm inj}, \omega_B/k)$; c is the speed of light; Z and A are the charge and mass numbers of the ion of sort s, respectively (to save space, the index of the ion sort in the corresponding quantities was omitted); e and m are the proton charge and mass; c_a is the Alfvén velocity;

$$\Gamma_{NL}^\pm = -\frac{kc_a^2}{E_B} \int_0^\infty S(k, k') E_w^\mp(k') dk' \qquad (8)$$

is the nonlinear increment whose kernel for a plasma parameter $\beta \lesssim 1$ is (Fedorenko *et al.* 1995)

$$S(k, k') = a_0 \left\{ \frac{k - k'}{k + k'} \exp\left[-\frac{1}{2\beta} \left(\frac{k - k'}{k + k'} \right)^2 \right] \right. \qquad (9)$$
$$\left. + \left| \frac{k + k'}{k - k'} \right| \exp\left[-\frac{1}{2\beta} \left(\frac{k + k'}{k - k'} \right)^2 \right] \right\},$$

where $a_0 = \pi^{3/2}/(32\beta\sqrt{2\beta})$; $E_B = B^2/8\pi$ is the IMF energy density; $\beta = (v_{\rm T}/c_a)^2$ is the plasma parameter, $v_{\rm T}$ is the thermal ion velocity in the undisturbed solar wind.

Allowance for the nonlinear interaction between Alfvén waves inevitably transforms their spectrum even in the absence of accelerated particles. As a result, for $\Gamma = 0$, the boundary spectrum $E_{w0}(k)$ specified at $x = -\infty$ and the spectrum at the shock

front $E_w(x=0,k)$ turn out to be different. This is inadmissible, because, as was noted above, although the entire upstream region $x<0$ under consideration is formally infinitely large, it corresponds to a relatively small solar-wind region. To avoid this discrepancy, we introduced the following source into the transfer equation for waves;

$$Q^{\pm}=kc_{\rm a}^2\frac{E_w^{\pm}}{E_B}\int\limits_0^{\infty}S(k,k')E_{w0}^{\mp}(k')dk'. \qquad (10)$$

This source, on the one hand, ensures the required equality $E_w(x=0,k)=E_{w0}(k)$ for $\Gamma=0$ and, on the other hand, has no significant effect on the wave evolution near the shock front, because $E_{w0}(k)$ is relatively small.

The formulated problem (1)–(10) can be solved numerically.

RESULTS OF THE CALCULATIONS

The calculations whose results are shown in Figs. 1–3 were performed for the typical solar-wind velocity $u=400$ km s^{-1}, Alfvén velocity $c_{\rm a}=60$ km s^{-1}, proton number density $N=4$ cm^{-3}, IMF strength $B=6\times10^{-5}$ G, and shock compression ratio $\sigma=3.5$. In the frequency range $\nu=10^{-3}$–10^{-1} Hz under consideration (here, $\nu=ku/2\pi$ is the frequency perceived by a stationary observer), the observed Alfvén-wave spectrum near the Earth's orbit is [see Russell (1972) and the review article by Tu and Marsh (1995)] $E_w(\nu)=E_w(k=2\pi\nu/u)/\nu\propto\nu^{-3/2}$. Therefore, given the observed amplitudes, the background Alfvén-wave spectrum was taken in the form

$$E_{w0}^{\pm}(k)=E_0^{\pm}\,(k/k_0)^{-1/2}, \qquad (11)$$
$$E_0=E_0^{+}+E_0^{-}=6.87\times10^{-14}\ {\rm erg\,cm^{-3}},$$

where $k_0=\omega_B/v_{\rm inj}$ is the wave number corresponding to the waves that resonantly interact with protons with velocity $v_{\rm inj}$; ω_B is the proton gyrofrequency. The calculation is consistent with the assumption that in the background spectrum, there are a factor of 2 more waves propagating away from the Sun than those with the opposite direction ($E_0^{+}=2E_0^{-}$). Analysis of experimental data indicates that this proportion is typical of the slow equatorial solar wind near the Earth's orbit (Horbury 1999). The fraction (injection rate) of the particles injected into the acceleration is taken to be $\eta=7.6\times10^{-3}$; according to measurements (Trattner *et al.* 1994), this is a typical value. In our calculations, apart from protons, we also take into account the acceleration of α-particles whose number density is assumed to be $0.05N=0.2$ cm^{-3}. The solar-wind protons and α-particles have the same velocity distribution. Therefore, the energy of the injected protons and α-particles was taken to be $\varepsilon_{\rm inj}=5$ and 20 keV, respectively.

Since the nonlinear Alfvén-wave interaction significantly depends on the plasma parameter β, the calculations given below were performed for $\beta=0.01, 0.1, 1$, which may occur in the solar wind. Different solar-wind temperatures correspond to different β.

Figures 1a and 1b show the calculated differential (with respect to the kinetic energy) intensity of the accelerated protons,

$$J(\varepsilon)=v^2f(v,t)/m$$

and Figs. 1c and 1 d show the Alfvén-wave spectrum,

$$E_w(\nu)=E_w\,(k=2\pi\nu/u)/\,\nu$$

at the shock front for six times: $t/t_0=5\times10^{-3}, 0.9, 2.8, 4.8, 8.0, 36.7$, where $\nu=ku/2\pi$ is the wave frequency perceived by a stationary observer. The characteristic acceleration time scale $t_0=\kappa_{\|0}(v_{\rm inj})/u^2=0.48$ h is determined by the proton diffusion coefficient

$$\kappa_{\|0}(v)=\frac{v^2B^2}{32\pi^2\omega_BE_{w0}\left(k=\rho_B^{-1}\right)},$$

that corresponds to the background turbulence. At $t=36.7t_0$, the solution differs from the steady-state solution by no more than one percent.

Figures 1e and 1f show the energy density of the Alfvén waves excited by the accelerated particles,

$$W=\frac{1}{E_B}\int\limits_0^{\infty}[E_w(\nu,t)-E_{w0}(\nu)]d\nu,$$
$$E_w(\nu,t)=E_w^{+}(\nu,t)+E_w^{-}(\nu,t)$$

and the pressure of the accelerated particles,

$$P_{\rm s}=\frac{4\pi m_{\rm s}}{3\rho u^2}\int\limits_{v_{\rm inj}}^{\infty}dvv^4f_{\rm s}(v,t)$$

at the shock front referred to the dynamic pressure ρu^2 as a function of time, where ρ is the solar-wind density and s is the sort of ions.

Figures 1a, 1c, 1e and 1b, 1d, 1f show the results of our calculations performed in the quasi-linear approximation and with allowance for the nonlinear Alfvén-wave interaction for $\beta=0.1$, respectively. Comparison of the two calculations indicates that the nonlinear interaction significantly changes the temporal behavior of the proton acceleration. As we see from comparison of Figs. 1c, 1e and Figs. 1d, 1f, allowance for the nonlinear interaction leads to a

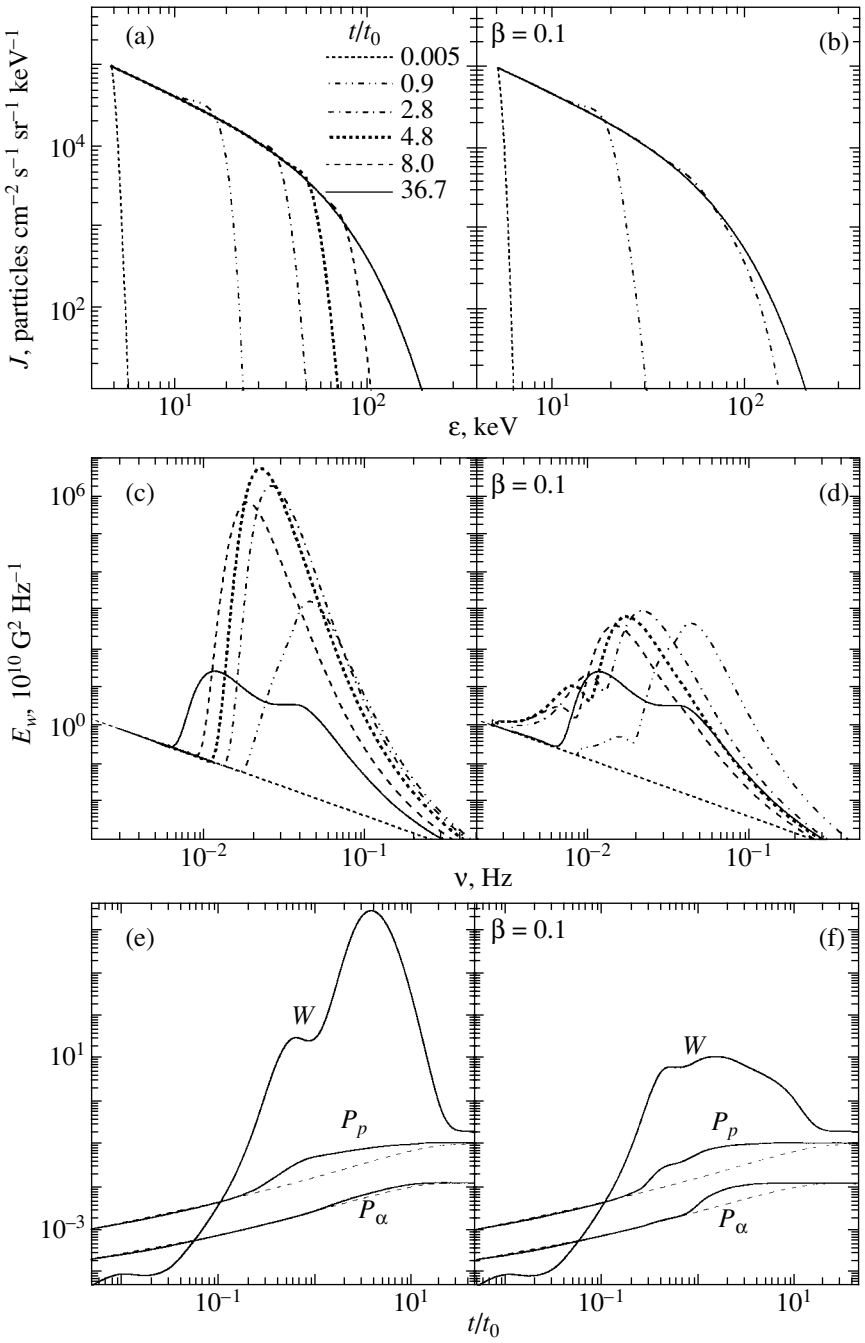


Fig. 1. (a, b) The differential intensity of the accelerated protons, (c, d) the Alfvén-wave spectrum and the wave energy density W, and (e, f) the pressures of the accelerated protons, P_p, and α-particles, P_α, at the Earth's bow-shock front for three calculations: quasi-linear (a, c, e), nonlinear (b, d, f) at $\beta = 0.1$, and linear (e, f) (dashes) for six times: $t/t_0 = 5 \times 10^{-3}, 0.9, 2.8, 4.8, 8.0, 36.7$. The calculations were performed with the following set of parameters: the solar-wind velocity $u = 400$ km s^{-1}, solar-wind proton density $N = 4$ cm^{-3}, IMF strength $B = 6 \times 10^{-5}$ G, compression ratio at the shock front $\sigma = 3.5$, size of the acceleration region $a = 4.2\ R_E$, particle injection rate $\eta = 7.6 \times 10^{-3}$, amplitude of the background Alfvén-turbulence spectrum $E_0 = 6.87 \cdot 10^{-14}$ erg cm^{-3}, $E^+_{w0} = 2E^-_{w0}$, and time scale $t_0 = 0.48$ h.

significant restriction of the Alfvén turbulence level at intermediate times: in comparison with the quasi-linear calculation (Fig. 1c), during the period $2.8 \lesssim t/t_0 \lesssim 8$, the wave energy density $E_w(\nu)$ near the spectral peak ($\nu \simeq 2 \times 10^{-2}$ Hz) and the total energy content W of the waves excited by the accelerated particles decrease by almost a factor of 10^4. In this case, the energy content W does not exceed 10, which

corresponds to a maximum wave amplitude $\delta B \simeq 3B$. During the initial stage of the acceleration process ($t/t_0 < 0.3$) and as a steady state is approached ($t/t_0 > 20$), the effect of nonlinear interaction is negligible, because the Alfvén-wave amplitude calculated in the quasi-linear approximation is low, $W < 1$. The latter, in turn, is attributable to the low energy content of the accelerated particles (see Fig. 1e). Recall that the extremely high generation rate of Alfvén waves at intermediate time (in our case, for $1 \lesssim t/t_0 \lesssim 10$) is a direct result of overacceleration, which is purely a nonstationary phenomenon (Berezhko and Taneev 1991, 1992; Berezhko *et al.* 1997).

The pattern of the steady-state Alfvén-wave spectrum can be easily understood by analyzing Eqs. (6) and (7) for the E_w^- waves excited by the accelerated particles. First, let us turn to the spectral range $\nu < \nu_{\rm inj}$. Given the relation $\nu = ku/2\pi$ and the resonance condition $k = \omega_B/v$, the wave spectrum $E_w(k)$ may be considered to be a function of the velocity, $E_w(v = \omega_B/k)$. In the steady-state case, ignoring the nonlinear interaction, Eq. (6) can be approximately integrated if we substitute $E_w(x, v)$ for the energy density $E_w(x, v')$ under the integral in expression (7). In this case, assuming, for simplicity, that $E_w \simeq E_w^-$ and taking into account only the proton contribution, the solution to (6) at the shock front ($x = 0$) is

$$E_w(v) = E_{w0}(v) + \frac{8\pi c_{\rm a} m v}{u} \times \int_v^\infty dv' v'^3 \left(1 - \frac{v^2}{v'^2}\right) f_0(v'), \qquad (12)$$

where $f_0(v) = f(x = 0, v)$ is the proton distribution function at the shock front. Thus, we see that when the number of accelerated particles with velocity v is large enough so that $mv^5 f c_{\rm a}/u \gg E_w$, the self-consistent Alfvén turbulence $E_w \gg E_{w0}$ dominates in the corresponding ranges of wave numbers $k = \omega_B/v$ and frequencies $\nu = ku/2\pi$.

In the energy range $\varepsilon_{\rm inj} = 5$ keV to $\varepsilon = \varepsilon_{\rm max} \simeq 40$ keV, the steady-state proton spectrum is close to a power law (Figs. 1a, 1b), $J \propto \varepsilon^{1-q/2}$, where the index $q = 3\sigma/(\sigma - 1) \simeq 4.2$ specifies the velocity dependence of the distribution function, $f \propto v^{-q}$. In our case, $q = 4.2$. According to (12), the wave spectrum $E_w \propto \nu^{6-q} = \nu^{-1.8}$ corresponds to this spectrum in the corresponding frequency band $\nu = \nu_{\rm max} - \nu_{\rm inj} = (1.3\text{–}4) \times 10^{-2}$ Hz. Comparison of this spectrum with the calculated spectrum reveals their difference: our calculation shows an excess near the frequency $\nu_{\rm max} = 1.3 \times 10^{-2}$ Hz and a reduction relative to the purely power law at slightly higher frequencies, $\nu \simeq 2.5 \times 10^{-2}$ Hz. Quantitatively, these features can be understood if it is considered that expression (7) for the increment can be written by discarding the constant factors as

$$\Gamma(x, v) \propto \frac{E_w(x, v)}{E_w(x, v_*)} \times \int_v^\infty dv' v'^3 \left(1 - \frac{v^2}{v'^2}\right) \frac{\partial f}{\partial x}, \qquad (13)$$

where $v_* > v$. In deriving expression (12), we assumed that $v_* = v$. The presence of an additional factor, $\delta = E_w(x, v)/E_w(x, v_*)$, must cause $E_w(v)$ to increase in the velocity range where expression (12) gives a maximum of $E_w(v)$, because $\delta > 1$ in this case. By contrast, $\delta < 1$ in the intermediate range $v_{\rm inj} < v < v_{\rm max}$, because $E_w(v) < E_w(v_*)$ and the factor δ leads to a reduction of the spectrum.

The proton energies $\varepsilon > \varepsilon_{\rm max}$ at which the proton spectrum exponentially falls off correspond to frequencies below $\nu_{\rm max} = 1.3 \times 10^{-2}$ Hz. Therefore, as ν decreases in the range $\nu < \nu_{\rm max}$, the self-consistent wave spectrum $E_w(\nu)$ rapidly approaches the background spectrum $E_{w0}(\nu)$ (Figs. 1c and 1d).

Note that relations (12) and (13) can also be used to qualitatively interpret the behavior of the wave spectra at intermediate times. In this case, one should take into account the fact that, because of the overacceleration effect, the high-energy tail of the accelerated-proton spectrum at intermediate stages falls off much more steeply than that of the steady-state spectrum (see Fig. 1a). As a result, the factor $\delta = E_w(v)/E_w(v_*)$ in expression (13) becomes large as v approaches the end of the steady-state part of the spectrum. The latter produces a significant excess in the amplitude of the Alfvén waves in the region of their spectral peak at intermediate stages above the steady-state spectrum (see Fig.1c).

The behavior of the self-consistent spectrum $E_w(\nu)$ at frequencies $\nu > \nu_{\rm inj}$ can be understood similarly.

The accelerated-particle pressure (Figs. 1e and 1f) monotonically increases with time without exceeding $0.1\rho u^2$, which justifies the neglect of shock modification due to the particle pressure backreaction.

Apart from the calculations under consideration, Figs. 1e and 1f show a linear calculation performed by assuming that no Alfvén waves are generated by the accelerated particles, i.e., for $\Gamma = 0$. In this case, as we see from Fig. 1f, an equilibrium spectrum of α-particles is established with a delay from that of protons. This is because the number of accelerated α-particles is too small to excite the Alfvén waves resonantly interacting with them to a nonlinear level.

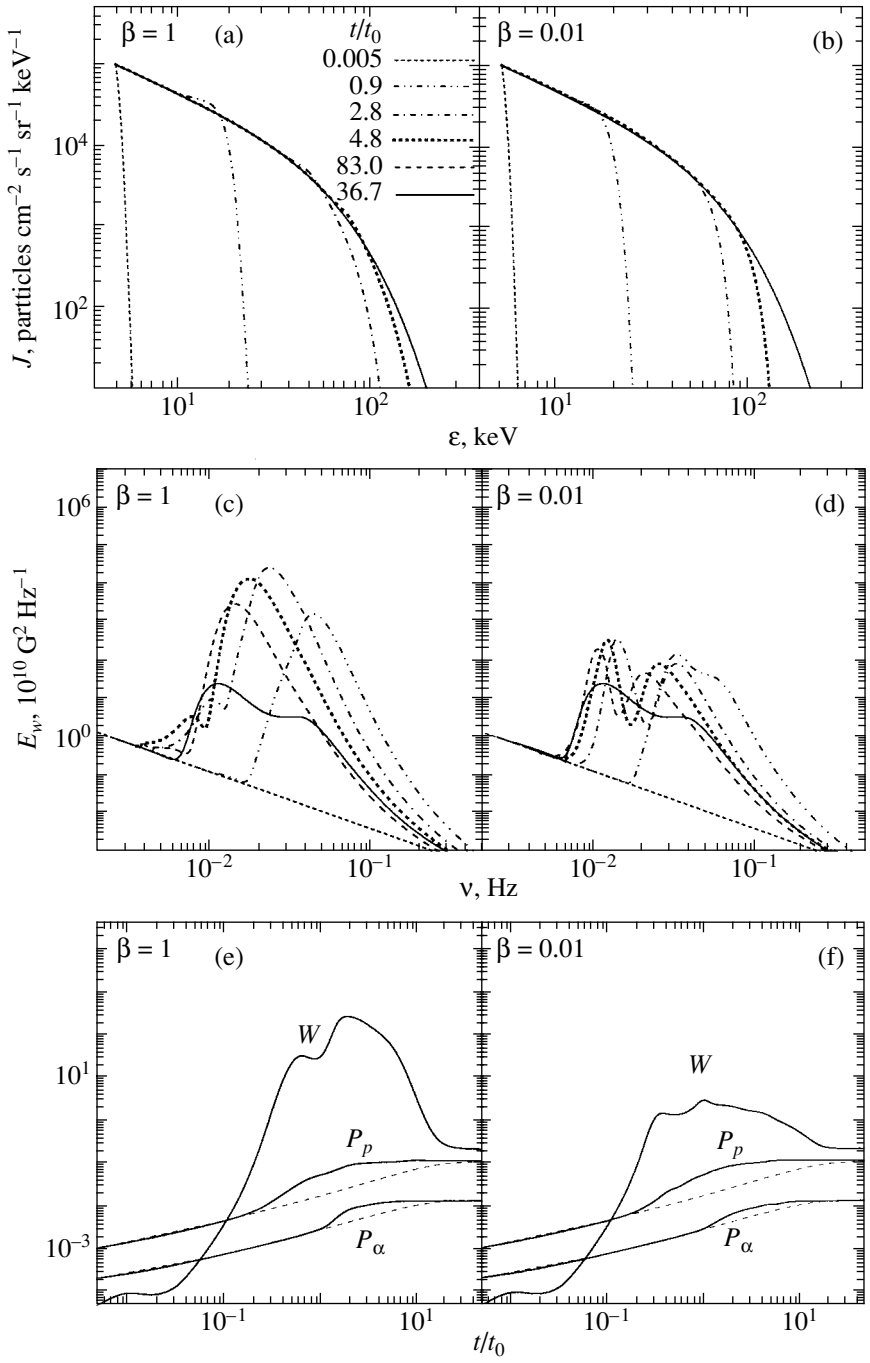


Fig. 2. Same as Figs. 1b, 1d, 1f for $\beta = 1$ (a, c, e) and $\beta = 0.01$ (b, d, f).

Therefore, the dynamics of the α-particles is determined by their interaction with the proton-excited waves; α-particles with the same velocity as protons have a factor of 2 larger gyroradius and interact with waves with a factor of 2 lower frequency. As we see from Figs. 1c and 1d, the excited Alfvén waves have such a dynamics that their frequency range continuously extends to lower frequencies. Since the steady-state spectrum of the accelerated particles in our case is a function of the kinetic energy per charge, $mv^2/(2Z)$ (Lee 1982), the α-particle spectrum is formed via a nonlinear factor more slowly than the proton spectrum.

The nonlinear interaction changes not only the amplitude of the waves but also their spectral distribution. During intensive wave excitation through the nonlinear interaction, an additional peak located at lower frequencies relative to the main peak is formed

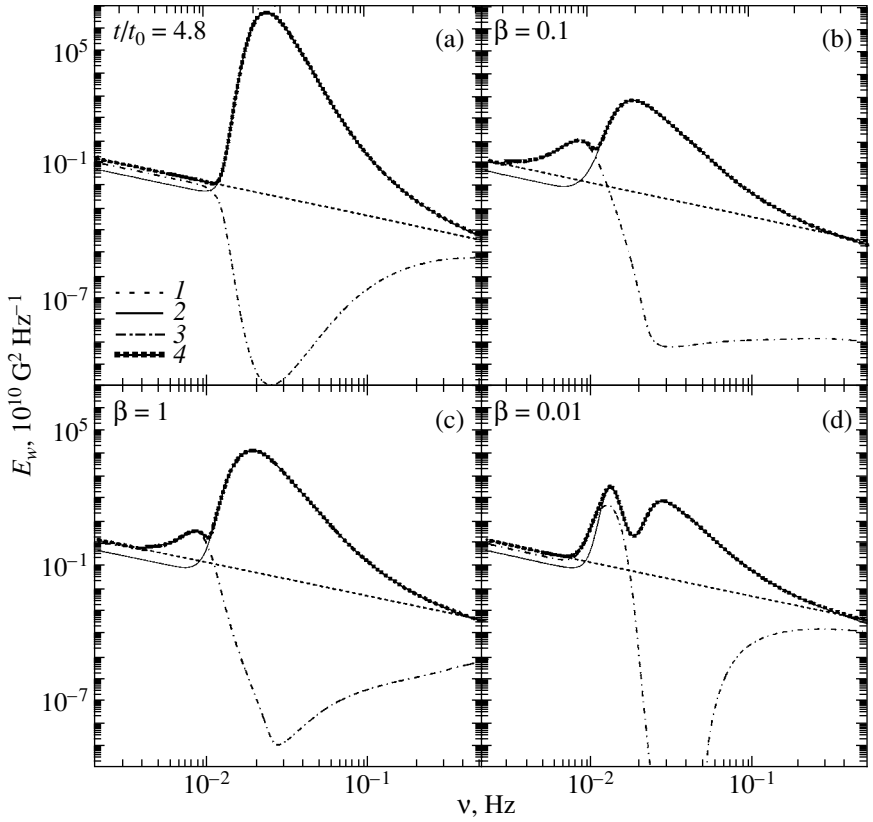


Fig. 3. The spectra of the Alfvén waves corresponding to (a) the quasi-linear and (b, c, d) nonlinear calculations at $\beta = 0.1, 1,$ and 0.01, respectively, for the time $t/t_0 = 4.8$ that corresponds to the maximum amplitude of the generated Alfvén waves in the quasi-linear calculation; (*1*) for background turbulence E_{w0}, (*2*) for the waves propagating away from the shock front E_w^-, (*3*) for the waves traveling toward the shock front E_w^+, and (*4*) the total wave spectrum $E_w = E_w^+ + E_w^-$.

in the wave spectrum (see Fig. 1d). The spectral wave energy transfer toward lower frequencies causes the particle acceleration rate to increase significantly: as we see from Figs. 1a and 1b, the accelerated-proton spectrum in the nonlinear case is formed several times faster than that in the quasi-linear case.

The spectral wave energy transfer is the most significant effect of nonlinear wave interaction relevant for the particle acceleration development. As was shown previously for interplanetary shocks (Berezhko *et al.* 1998), the high generation rate of Alfvén turbulence alone is not enough for a significant increase of the maximum energies in the accelerated-particle spectra compared to what is provided by background turbulence in the linear approximation. An increase in the particle acceleration rate is possible at a considerable generation rate of low-frequency waves that resonantly interact with particles of extremely high energies. This, in turn, is achieved in the presence of spectral transfer, because Alfvén waves are effectively generated by lower-energy particles at higher frequencies.

Comparing the two presented calculations (quasi-linear and nonlinear), it may be noted that at intermediate times, $1 \lesssim t/t_0 \lesssim 10$, the amplitude and total energy content of the Alfvén waves in the nonlinear case (Figs. 1d and 1f) are much lower than those in the quasi-linear case (Figs. 1c and 1e). The higher particle acceleration rate in the nonlinear case is attributable solely to spectral transfer.

It is important to note that the possibility of achieving a higher acceleration rate is of particular importance in elucidating the origin of the Galactic cosmic rays. The point is that the main source of cosmic rays with energy $\varepsilon \lesssim 10^{15}$ eV in the Galaxy is justifiably assumed to be diffusive shock acceleration at shock fronts from supernova explosions (e.g., Berezhko and Krymsky 1988; Berezhko *et al.* 1988). Cosmic rays of the above energies can be effectively accelerated only when Alfvén waves are intensively generated over a wide frequency range. As we showed here, this is possible only in the presence of effective spectral wave energy transfer toward lower frequencies.

The quasi-linear approach predicts a weak dependence of the acceleration on the relationship between

the Alfvén waves oppositely propagating along the diffusive shock magnetic field (Berezhko and Taneev 1992; Berezhko *et al.* 1992, 1997). However, when the excited waves reach large amplitudes, this relationship becomes important, because the oppositely propagating waves are involved in induced scattering, the principal mechanism of nonlinear interaction for $\beta \ll 1$ (Fedorenko *et al.* 1990, 1995). The waves produced by the nonlinear interaction propagating toward the shock front form an additional low-frequency peak in Fig. 1d. The waves excited by the accelerated particles and propagating away from the shock front form the main spectral peak in the quasi-linear and nonlinear cases. In contrast, the oppositely propagating waves in this frequency range are absorbed by the accelerated particles.

The dependence of nonlinear interaction on the system parameters is illustrated by the calculations performed for different values of the plasma parameters β. Figure 2 shows the results of our calculations for $\beta = 1$ (Figs. 2a, 2c, 2e) and $\beta = 0.01$ (Figs. 2b, 2d, 2f) for the same set of the remaining solar-wind parameters that was used in the calculations shown in Fig. 1. We see that the degree of restriction of the amplitude of the waves excited through the nonlinear interaction decreases with increasing β. Thus, for $\beta = 1$, the energy content of the Alfvén waves reaches $W = 300$ (Fig. 2e), which is a factor of 30 larger than that for $\beta = 0.1$ considered above. In this case, the factor $\beta^{-3/2}$ in the expression for the kernel $S(k, k')$ plays a major role. At the same time, spectral transfer provides almost the same high acceleration rate as that for $\beta = 0.1$ (see Figs. 1b and 2a).

In contrast, for $\beta \ll 0.1$, the energy content of the waves excited by particles is limited to $W = 3$ (Fig. 2f), which is almost independent of β. This fact can be easily explained: for $\beta \ll 1$, the nonlinear increment $\Gamma^{\pm}_{NL}$ reduces to a differential form, which does not depend on β (Fedorenko *et al.* 1995).

The effect of spectral transfer is significant over the entire range $\beta \lesssim 1$ and causes a considerable increase in the particle acceleration rate almost equally for all β. Some differences show up in the form of the secondary peak in the wave spectrum $E_w(\nu)$ resulting from the nonlinear interaction: for $\beta \ll 0.1$, it is narrower and sharper than for $\beta \gtrsim 0.1$ (see Figs. 1d and 2d).

The range of plasma parameters β in which intensive spectral wave energy transfer is expected appears to depend on the energy content that the particles can achieve during their acceleration. One might expect that in comparison with the case considered, spectral transfer takes place in a much wider range of β for more extended and stronger shock waves, where the achievable energy content of the accelerated particles is much higher.

Figure 3 shows the Alfvén-wave spectra chosen for the time $t/t_0 = 4.8$ when their generation rate is at a maximum for the quasi-linear calculation (see Fig. 1c). Figure 3a corresponds to the quasi-linear calculations, while Figs. 3b, 3c, and 3d correspond to the nonlinear calculations with $\beta = 0.1, 1, 0.01$, respectively.

Figure 3a shows an almost symmetric, relative to the background turbulence E_{w0} (dashed lines), behavior of the spectra for the Alfvén waves propagating toward the shock E^+_w (dash–dotted lines) and away from the shock E^-_w (solid lines). For the nonlinear calculations with $\beta \gtrsim 0.1$ (Figs. 3b and 3c), we clearly see local maxima in the low-frequency range of the E^+_w wave spectra, which are produced by the wave energy transfer to this part of the spectrum through their nonlinear interaction with E^-_w waves.

The E^+_w and E^-_w wave spectra shown in Fig. 3d for the calculation with $\beta = 0.01$ exhibit a more complex behavior in the region of the left local maximum of total turbulence E_w (the lines of heavy dots). Both the E^+_w and E^-_w spectra have local maxima in this frequency range that differ in amplitude and are almost coincident in frequency. This suggests energy transfer from one spectrum to the other in the narrow frequency range of their nonlinear interaction.

Note a peculiar feature in the temporal behavior of the Alfvén-wave energy density W: the presence of two local maxima for all calculations (see Figs. 1e, 1f, 2e, 2f). For example, in Fig. 1e for the quasi-linear calculation, these are the maxima with amplitudes ~ 30 and 3×10^3 for times $t_1 \sim 0.6t_0$ and $t_2 \sim 4t_0$, respectively.

During the initial period $t \leq t_1$, when the number of accelerated protons is still not large enough, they excite mostly waves with frequencies $\nu > \nu_{\rm inj}$, where $\nu_{\rm inj} \approx 3.7 \times 10^{-2}$ Hz is the frequency of the waves that resonantly interact with protons with injection energy $\varepsilon_{\rm inj}$. The time $t = t_1$ corresponds to the time when the leftward-displacing frequency of the maximum in the Alfvén-wave spectrum, $\nu_{\rm max}$, reaches $\nu_{\rm inj}$. Subsequently, the increase in wave energy content is mainly produced with a short delay by the overacceleration effect (Berezhko and Taneev 1991): at intermediate times, the number of accelerated particles in some energy range exceeds the then established steady-state level (see Figs. 1a, 1b and 2a, 2b). Although this excess is relatively small, it produces a considerable increase in the generation rate of Alfvén waves: at intermediate times, $t_1 < t < t_2$, the wave amplitude in the quasi-linear approximation is several orders of magnitude higher than their steady-state value (Fig. 1c).

The wave energy content reaches a maximum at $t = t_2$, when the formation of a power-law part of the proton spectrum ends (see Fig. 1a) and the subsequent development of the acceleration process forms an exponential part of the spectrum ($\varepsilon > 40$ keV in Fig. 1a), because the particle escape from the acceleration region described by the last term in Eq. (1) increases in importance with their energy.

Note also that numerical simulations of collisionless quasi-parallel shock waves clearly confirm the existence of overacceleration (see Fig. 4 in Scholer *et al.* 1999).

CONCLUSIONS

Our studies of the self-consistent diffusive shock acceleration of ions, as applied to typical conditions at the Earth's bow shock, have shown that allowance for the nonlinear A'fvén-wave interaction through induced scattering and two-quanta absorption at a solar-wind plasma parameter $\beta \lesssim 1$ significantly affects the acceleration dynamics.

The nonlinear interaction significantly restricts the amplitude of the waves excited by accelerated particles at $\delta B \lesssim B$ for $\beta \lesssim 0.1$. The restriction effect decreases with increasing β for $\beta > 0.1$, where the Alfvén-wave amplitude reaches $\delta B \sim B\beta^{3/2}$.

The nonlinear interaction produces spectral wave energy transfer toward lower frequencies.

Since the generation of low-frequency Alfvén waves causes an increase in the scattering efficiency of increasingly high-energy particles, the spectral transfer is accompanied by an increase in the particle acceleration rate.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 99-02-16325 and 00-15-96669).

REFERENCES

1. E. G. Berezhko and G. F. Krymskiĭ, Usp. Fiz. Nauk **154**, 49 (1988) [Sov. Phys. Usp. **31**, 27 (1988)].
2. E. G. Berezhko and S. N. Taneev, Kosm. Issled. **29**, 582 (1991).
3. E. G. Berezhko and S. N. Taneev, in *Proceedings of the 1st SOLTIP Symposium, Prague, 1992,* Vol. 1, p. 67.
4. E. G. Berezhko, V. K. Elshin, G. F. Krymskiĭ, and S. I. Petukhov, *Generation of Cosmic Rays by Shock Waves* (Nauka, Novosibirsk, 1988).
5. E. G. Berezhko, S. N. Taneev, and S. I. Petukhov, Izv. Akad. Nauk, Ser. Fiz. **61**, 1137 (1997).
6. E. G. Berezhko, S. I. Petukhov, and S. N. Taneev, Pis'ma Astron. Zh. **24**, 151 (1998) [Astron. Lett. **24**, 122 (1998)].
7. D. D. Childers and C. T. Russell, in *Solar Wind,* Ed. by C. P. Sonett *et al.* (NASA, Washington, 1972), NASA SP-308, p. 375.
8. V. N. Fedorenko, V. M. Ostryakov, A. N. Polyudov, and V. D. Shapiro, Fiz. Plazmy **16**, 443 (1990) [Sov. J. Plasma Phys. **16**, 253 (1990)].
9. V. N. Fedorenko, N. A. Gun'ko, and E. V. Frolushkina, Fiz. Plazmy **21**, 907 (1995) [Plasma Phys. Rep. **21**, 857 (1995)].
10. T. S. Horbury, in *Plasma Turbulence and Energetic Particles in Astrophysics,* Ed. by M. Ostrowski and R. Schlickeiser (Cracow, 1999), p. 115.
11. M. Lee, J. Geophys. Res. **87**, 5063 (1982).
12. C. T. Russell, in *Solar Wind,* Ed. by C. P. Sonett *et al.* (NASA, Washington, 1972), NASA SP-308, p. 365.
13. M. Scholer, H. Kucharek, and K.-H. Trattner, Ann. Geophys. **17**, 583 (1999).
14. K. J. Trattner, E. Möbius, M. Scholer, *et al.*, J. Geophys. Res. **99**, 13 389 (1994).
15. C.-Y. Tu and E. Marsh, Space Sci. Rev. **73**, 1 (1995).
16. V. I. Zirakashvili, Zh. Éksp. Teor. Fiz. **117**, 932 (2000) [JETP **90**, 810 (2000)].

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 9, 2002, pp. 641–650. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 9, 2002, pp. 711–720.

Peculiarities of Uranus's Satellite System

M. A. Vashkov'yak* **and N. M. Teslenko**

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Miusskaya pl. 4, Moscow, 125047 Russia

Received May 15, 2002

Abstract—The absence of Uranus's equatorial satellites in the region of approximately equal influence of its oblateness and solar perturbations is explained in terms of an improved physical model. This model is more complete than the previously studied case of an integrable averaged problem. The model improvement stems from the fact that the inclination of Uranus's equator to the ecliptic differs by 90° and that the orbital evolution of Uranus due to secular planetary perturbations is taken into account. The lifetime of Uranus's hypothetical satellites in orbits with semimajor axes 1.3–7 million km can be estimated by numerically integrating the evolution equations to be $\sim 10^4$ yr. This is the time scale on which the evolution of the orbits leads to their intersection with the orbits of inner satellites. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *Uranus's satellites, orbital evolution*

INTRODUCTION. STATEMENT OF THE PROBLEM

A distinctive feature of the satellite systems of the giant planets is the distribution of their satellites in orbital semimajor axis a, in particular, their breakdown into two sets, inner and outer satellites. Figure 1 shows this distribution for the satellite systems of Jupiter (NSJ—Number of Satellites of Jupiter), Saturn (NSS—Number of Satellites of Saturn), and Uranus (NSU—Number of Satellites of Uranus). In Jupiter's system, the outer satellites are clearly separated into two groups of satellites with prograde (inclinations $i < \pi/2$) and retrograde ($i > \pi/2$; Fig. 1a) orbits; in Saturn's system, there is a range of semimajor axes in which both prograde and retrograde satellite orbits exist (Fig. 1b); and in Uranus's system, all of the outer satellites discovered to date possess retrograde motions (Fig. 1c). In addition, in all three systems, there are ranges of semimajor axes free from satellite orbits. It seems of interest to establish the possible mechanism of avoidance of these circumplanetary regions by satellites.

Here, we consider only the satellite system of Uranus. It, along with the systems of the other giant planets, consists of a thin ring of inner regular satellites (their nearly circular orbits lie almost in the equatorial plane) and outer satellites that move in elliptical orbits of appreciable eccentricity with large inclinations to the equatorial plane. No celestial objects have yet been found in the intermediate region, $0.6 \leq a \leq 7$ million km. One of us (Vashkov'yak 2001) offered a celestial-mechanical explanation for the absence of Uranus's regular satellites in this region. Here, we analyze the evolution of satellite orbits under the combined effect of solar attraction and Uranus's oblateness in the following model statement: the equatorial plane of Uranus in which the satellites move makes the angle $I = 90°$ (which is close to the actual value of about 98°) with the plane of its heliocentric orbit; the circular orbit of Uranus lies in the plane of the ecliptic. Using one of the integrable cases of the well-known Hill averaged problem with oblateness of the central planet (Lidov 1963a, 1963b; Lidov and Yarskaya 1974), we were able to explain the absence of equatorial satellites for $a \geq 1.3$ million km.

Here, we analyze the problem for the actual equatorial inclination of Uranus by assuming that its heliocentric orbit (or the Uranocentric orbit of the Sun) evolves due to secular planetary perturbations. To determine the location of Uranus's equator, we use the right ascension α_0 and declination δ_0 of its north pole. These spherical coordinates are referred to the mean geoequator and to the equinox of the initial date. Thus, the rotation parameters of the planet are specified independently of its orbital parameters. Using the procedure of double averaging over the motion of the satellite and the perturbing body (Sun), we can show that the secular part of the perturbing function in the Hill approximation is

$$W = \frac{\mu}{a}[\beta(W_0 + W_1 \sin 2i_1 + W_2 \sin^2 i_1 + e_1 W_3) + \alpha W_4], \quad (1)$$

*E-mail: vashkov@spp.keldysh.ru

1063-7737/02/2809-0641$22.00

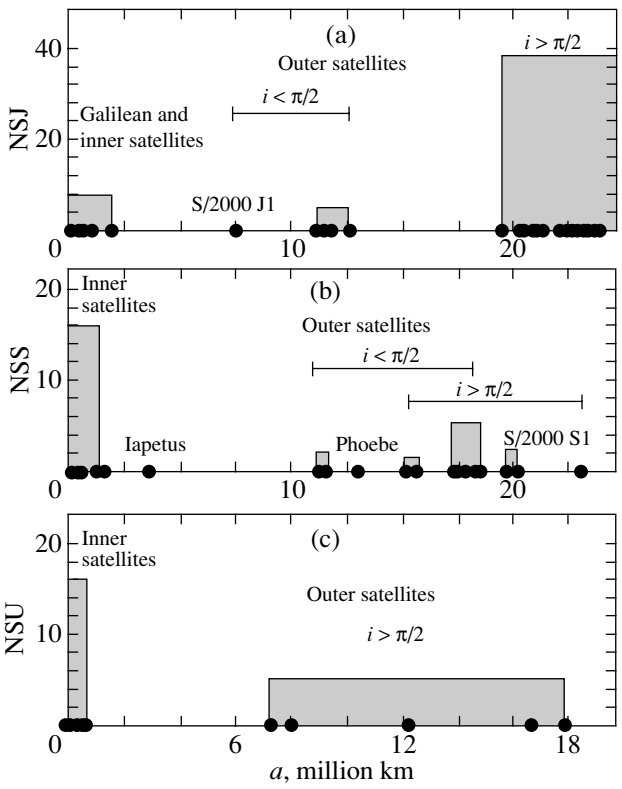


Fig. 1. The distribution of the satellites of three giant planets in orbital semimajor axis: (a) the system of 39 Jupiter's satellites, (b) the system of 30 Saturn's satellites, and (c) the system of 21 Uranus's satellites.

where

$$\alpha = -\frac{3c_{20}a_0^2}{8a^2}, \quad \beta = \frac{3\mu_1 a^3}{16\mu a_1^3(1-e_1^2)^{3/2}} \qquad (2)$$

are dimensionless parameters;

$$W_0 = 2e^2 + \sin^2 i(5e^2\cos 2\omega - 2 - 3e^2); \qquad (3)$$

$$W_1 = -[5e^2\sin 2\omega\sin(\Omega_1-\Omega) + (5e^2\cos 2\omega - 2 - 3e^2)\cos(\Omega_1-\Omega)\cos i]\sin i; \qquad (4)$$

$$W_2 = -3e^2 - 0.5\sin^2 i(5e^2\cos 2\omega - 2 - 3e^2)[3 + \cos(2\Omega_1 - 2\Omega)] + 5e^2[\cos 2\omega\cos(2\Omega_1 - 2\Omega) + \sin 2\omega\sin(2\Omega_1 - 2\Omega)\cos i]; \qquad (5)$$

$$W_3 = \frac{5ae}{16a_1(1-e_1^2)}\{(4+3e^2)[(5\sin^2 i - 4)\cos\omega \times \cos(\Omega - g_1) + (4 - 15\sin^2 i)\sin\omega\cos i \times \sin(\Omega - g_1)] + 35e^2\sin^2 i[-\cos 3\omega\cos(\Omega - g_1) + \sin 3\omega\cos i\sin(\Omega - g_1)]\}, \qquad (6)$$

$$g_1 = \Omega_1 + \omega_1;$$

$$W_4 = (1-e^2)^{-3/2}\{-2/3 + \zeta^2(1+\cos 2i) + \sin^2 i[(\xi^2+\eta^2) + (\eta^2-\xi^2)\cos 2\Omega] - 2\eta\zeta\sin 2i\cos\Omega - 4\xi\sin i\sin\Omega[\cos\varepsilon(\eta\sin i\cos\Omega - \zeta\cos i) + \sin\varepsilon(\zeta\sin i\cos\Omega + \eta\cos i)] - \sin 2\varepsilon[2\eta\zeta(\sin^2 i\sin^2\Omega + \cos 2i) + (\zeta^2-\eta^2)\sin 2i\cos\Omega] - 2\sin^2\varepsilon[(\zeta^2-\eta^2) \times (\sin^2 i\sin^2\Omega + \cos 2i) - 2\eta\zeta\sin 2i\cos\Omega]\}; \qquad (7)$$

$\xi = \cos\delta_0\cos\alpha_0$, $\eta = \cos\delta_0\sin\alpha_0$, and $\zeta = \sin\delta_0$ are the direction cosines of Uranus's rotation axis; μ and μ_1 are the products of the gravitational constant by the masses of Uranus and the Sun, respectively; c_{20} is the coefficient of the second zonal harmonic of Uranus's gravitational field; a_0 is its mean equatorial radius; ε is the inclination of the Earth's equator to the plane of the ecliptic; a, e, i, ω, Ω, a_1, e_1, i_1, ω_1, and Ω_1 are the Keplerian Uranocentric orbital elements for the satellite and the Sun, respectively. All angular elements are referred to the plane of the ecliptic and to the direction of the point of vernal equinox. We derived expression (1) under the assumptions of $a/a_1 \ll 1$ and $e_1 \ll 1$ and the function W_3 for $\sin i_1 \ll 1$.

The quantity g_1 in the function W_3, which is the pericenter longitude for the planetocentric orbit of the Sun, differs from the heliocentric perihelion longitude for the planet by 180°. The latter, together with the elements e_1, i_1, and Ω_1, is specified by the Lagrange–Brauer–Wurkom theory (Sharaf and Budnikova 1967).

For the evolution system with the perturbing function (1) to be properly used, we must estimate the possible variations in α_0 and δ_0 with time due to the evolution of the planet's axial rotation. The principal effects are the effect of the moment of gravitational forces and the tidal effect from celestial bodies. The former causes the planet's rotation axis to precess with a constant inclination, while the latter causes secular changes in the angular velocity of axial rotation and in the inclination of this axis.

To estimate the precession period for the rotation axis of the planet as a rigid body, we use the following relation from Beletskii (1975):

$$T^* = \frac{2\pi}{\omega_0}\left(\frac{3}{2}\frac{|A-C|}{C}\frac{\omega_0}{\omega^*}\cos\Delta\right)^{-1}, \qquad (8)$$

where ω_0 is the angular velocity of the planet's orbital motion; ω^* is the angular velocity of its rotation; Δ is the angle between the rotation axis and the normal to the plane of the ecliptic; A and C are the principal central moments of inertia for the planet. If the mass distribution within the planet is assumed to be uniform, then

$$\frac{|A-C|}{C} = \frac{5}{2}|c_{20}|.$$

Substituting the current parameters of Uranus in Eq. (8) yields $T^* \sim 3 \times 10^9$ yr.

The evolution of the proper rotation of planets under the effect of a tidal torque was analyzed in detail by Beletskiĭ (1978). The evolution time scale is estimated by τ^*, which is measured in years of the planet under consideration:

$$\tau^* = \frac{4}{45\pi}\frac{\mu}{\mu_1}\left(\frac{a_1}{a_0}\right)^3\frac{\omega^*}{\omega_0}Q, \qquad (9)$$

where Q is the dimensionless effective dissipative function (according to present views, $Q \sim 10$–10^2 for the terrestrial planets and $Q \sim 10^5$–10^6 for the giant planets). Substituting the current parameters of Uranus in Eq. (9) yields $\tau^* \sim 10^{21}$ yr (terrestrial).

In our problem, the evolution of Uranus's satellite orbits is considered on intervals no longer than 10^6 yr. As our series of calculations show, the lifetimes of hypothetical satellites are even shorter. Since the time scales T^* and τ^* are longer by many orders of magnitude, we assume α_0 and δ_0 to be constant and their numerical values to be known: $\alpha_0 = 76.^\circ051$, $\delta_0 = 14.^\circ855$ (Abalakin *et al.* 1976).

When the orbital evolution of a satellite and its lifetime are analyzed, the principal parameter is the pericenter distance, $q = a(1-e)$. In the course of evolution, e can reach values close to unity and, accordingly, q (at constant a in the averaged problem) can become smaller than Uranus's radius a_0 or the orbital radii of its most massive inner satellites $a^{(j)}$ ($j = 1, 2, \ldots, 5$). When these orbits are approximated by Gaussian mass rings, the effective cross section of Uranus increases significantly. In our evolution problem, by the lifetime of a satellite we mean the time scale on which q decreases from a to the orbital radius of at least the outermost inner satellite, Oberon ($a^{(5)} \approx 0.6$ million km). Under the orbit intersection conditions arising in this case, the probability of close encounters and collisions of hypothetical satellites with inner satellites increases significantly. As a result, the hypothetical satellites must either have moved to farther orbits or fallen to the inner satellites, significantly augmenting their masses.

Here, our goal is to confirm (by numerically analyzing the evolution of satellite orbits in a more complete physical model) the "death" mechanism for Uranus's hypothetical satellites previously described in the model problem in the intermediate range of semimajor axes, $1.3 \le a \le 7$ million km.

RESULTS OF ANALYTICAL AND NUMERICAL STUDIES

A Brief Description of the Model Problem

Before describing our numerical calculations, we recall the principal results of analysis of the integrable model problem ($I = i = 90^\circ$, $\sin\Omega = 0$, $e_1 = i_1 = 0$) (Lidov 1963a, 1963b; Lidov and Yarskaya 1974; Vashkov'yak 2001). The families of trajectories in the (ω, e) plane in this integrable case are symmetric about the $\omega = 0^\circ$, 90°, 180°, and 270° straight lines and have different structures, depending on the semimajor axis a of the satellite orbit (or on the ratio of the parameters α and β).

If $\alpha < 3\beta$ (or $a < 1.1$ million km for the Sun–Uranus system), then ω changes monotonically, e varies within a narrow range, and $q(t) > a^{(5)} > a^{(4)} \ldots > a^{(1)} > a_0$ for any $t > 0$ at e_0 close to zero.

If $\alpha > 3\beta$ (or $a > 1.1$ million km), then there are two singular points in the (ω, e) phase plane:

(1) $\omega = \omega_s = \arcsin\sqrt{\frac{2}{5}\left(1 + \frac{\alpha}{2\beta}\right)}$, $e = 0$ saddle-type point;

(2) $\omega = \pm 90^\circ$, $e = e^* = \sqrt{1 - \left(\frac{\alpha}{3\beta}\right)^{2/5}}$ center-type point.

In this case, the (ω, e) plane breaks down into two regions with circulation and libration variations of ω separated by the limiting solution, the separatrix. This special curve passes through the points with the (ω_s, 0) and ($\pm 90^\circ$, e_s) coordinates. The value of $e_s > e^*$, along with $q_s = a(1-e_s)$, is determined by the α/β ratio alone. Since $\alpha/\beta \sim a^{-5}$, $e^* \to 1$ as a increases and even at $a \approx 1.3$ million km for any initial e_0 and ω_0, q decreases to $a^{(5)}$ in a finite time; i.e., the evolution leads to an intersection of the hypothetical satellite orbit with the orbit of Oberon. For a from the range 1.3–7 million km, there are also intersections with the orbits of the remaining inner satellites; the minimum of $q(t) \le q^* = a(1-e^*)$ is reached at $\omega(t) = \pm 90^\circ$ and $q_{\min} < a^{(1)}$. The time scale of orbital evolution from an initially nearly circular orbit to a highly elliptical orbit intersecting the orbits of inner satellites is $\sim 10^4$ yr.

The Evolution of Satellite Orbits on Time Scales of $\sim 10^4$ *yr*

Based on the above results, one might expect that at I close to 90° and for time-varying orbital elements of Uranus, the qualitative behavior of the phase trajectories in projection onto the (ω, e) plane will be preserved (at least for the trajectories that do not pass near the separatrix) and the maximum values of e (the minimum values of q) will be reached at ω close to $\pm 90^\circ$.

Under this assumption, we performed the first series of calculations. For semimajor axes in the chosen

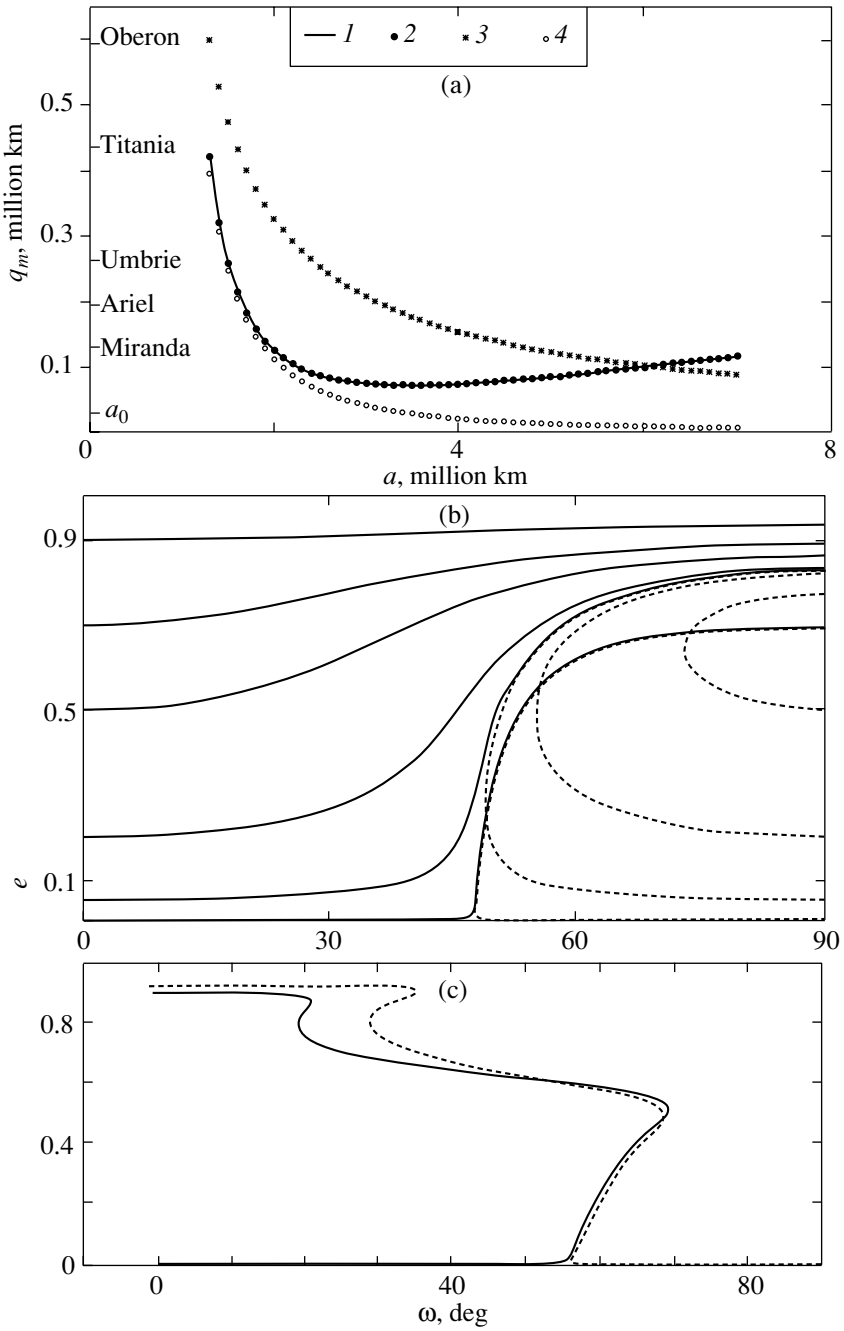


Fig. 2. Evolution parameters for satellite orbits on time scales of $\sim 10^4$ yr: (a) pericenter distance q_m for $\omega = 90^\circ$ versus semimajor axis a for (*1*) $\omega_0 = 0$, (*2*) for $\omega_0 = 90^\circ$, (*3*) $q^*(a)$, (*4*) $q_s(a)$; (b) projections of the family of phase trajectories onto the (ω, e) plane for $a = 1.5$ million km (solid lines $\omega_0 = 0^\circ$; dashed lines $\omega_0 = 90^\circ$); (c) projections of the phase trajectories close to the separatrix onto the (ω, e) plane for $a = 1.3$ million km (solid lines $\omega_0 = 0^\circ$; dashed lines $\omega_0 = 90^\circ$).

range $1.3 < a < 7$ million km, the evolution system in elements with the perturbing function W (1)–(7) was numerically integrated until τ at which $\omega(\tau) = 90^\circ$. The initial ω_0 was taken to be zero and 90° in the ω circulation and libration regions, respectively; the initial value of $e_0 = 0.03$. The angular ecliptic elements i_0 and Ω_0 were chosen from the condition for coincidence of the initial satellite orbital plane with Uranus's equatorial plane. This condition is given by

$$\cos i_0 = \cos\varepsilon \sin\delta_0 - \sin\varepsilon \cos\delta_0 \sin\alpha_0, \qquad (10)$$

$$\sin i_0 = \sqrt{1 - \cos^2 i_0},$$

$$\sin\Omega_0 = \cos\delta_0 \cos\alpha_0 / \sin i_0,$$

$$\cos\Omega_0 = (\cos\varepsilon \cos i_0 - \sin\delta_0) / \sin i_0 / \sin\varepsilon.$$

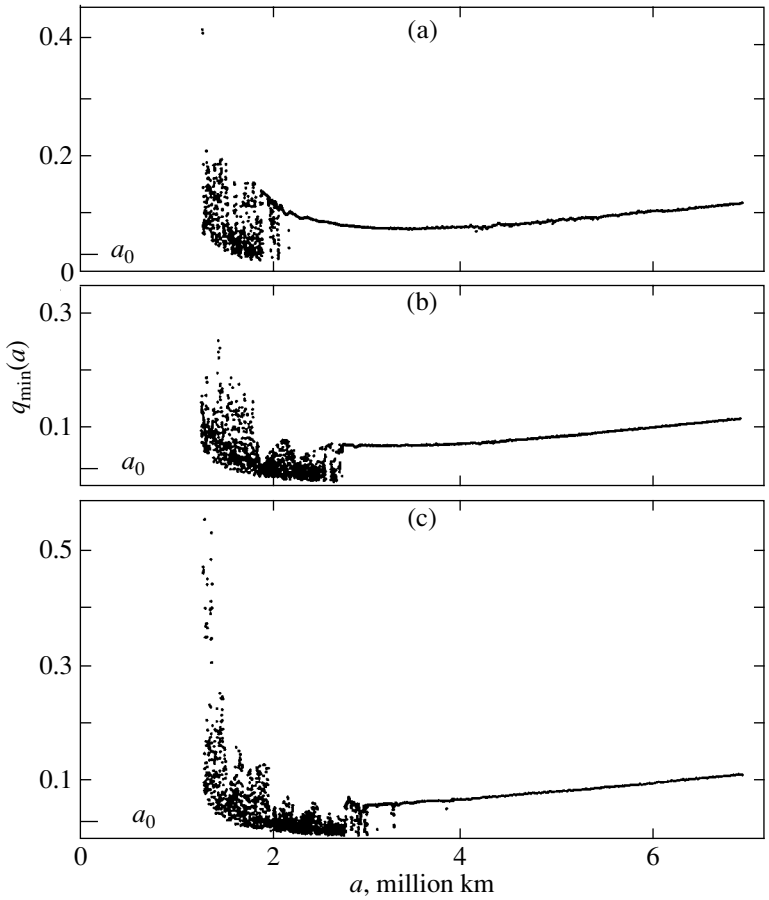


Fig. 3. The dependences $q_{\min}(a)$ on the time interval $T = 400$ thousand years for $\omega_0 = 0°$: (a) $e_0 = 0.03$, (b) $e_0 = 0.001$; (c) $e_0 = 0.0001$.

With the assumed values of α_0 and δ_0, we obtain $i_0 \approx 97\overset{\circ}{.}9$, $\Omega_0 \approx 166\overset{\circ}{.}4$.

Under the chosen initial conditions, $e_m = e(\tau)$ is a function of the semimajor axis a alone. Figure 2a shows almost identical dependences $q_m(a) = a(1 - e_m(a))$ for $\omega_0 = 0$ (solid curve) and for $\omega_0 = 90°$ (filled circles). The markers on the vertical axis indicate the values of $a^{(j)}$ ($j = 1, 2, \ldots, 5$). The dependences for the integrable problem are shown for comparison: $q_s(a)$ (open circles) and $q^*(a)$ (asterisks). In contrast to the model dependences that asymptotically approach zero, the two dependences $q_m(a)$ have a flat minimum at $a \approx 3$ million km and a sloping asymptote. The slope of this asymptote or the derivative $dq_m(a)/da$ can be estimated by considering the limiting case of the problem $\alpha = 0$, i.e., the Hill double-averaged problem integrable for any initial inclination i_0 (Lidov 1961). At low values of the constant of one of the integrals in this problem,

$$c_1 = (1 - e^2)\cos^2 i, \qquad (11)$$

$dq_m(a)/da \approx 5c_1/6$. For $e_0 \approx 0$, $i_0 \approx 98°$, it is ~ 0.016. The parameters e_m and e_s differ by a value of the same order of magnitude. The difference from the model problem also shows up in the behavior of the phase trajectories in projection onto the (ω, e) plane. By way of illustration, Fig. 2b shows a family of these trajectories for $a = 1.5$ million km. The solid and dashed lines correspond to $\omega_0 = 0°$ and $90°$, respectively. The apparent mutual intersections of the phase trajectories stem from the fact that the problem is nonintegrable. In addition, the trajectories are slightly asymmetric about the $\omega = 0°$, $90°$, $180°$, and $270°$ straight lines in the integrable model problem. Calculations show that for a given a, the nearly circular equatorial orbit of Uranus's hypothetical satellite rapidly evolves from $e_0 = 0.001$ to a highly eccentric elliptical orbit with an inclination to the ecliptic of about $155°$; q reaches a maximum $q_m \approx 0.463$ million km at $\tau \approx 33.8$ thousand years for $\omega_0 = 0$ and $q_m \approx 0.468$ million km at $\tau \approx 34.0$ thousand years for $\omega_0 = 90°$. In this case, the time it takes for q to reach a minimum for $e_0 = 0.01$–0.8 is even shorter.

Although the structure of the phase trajectories shown in Fig. 2b admits self-intersections, it nevertheless possesses a certain regularity in that the max-

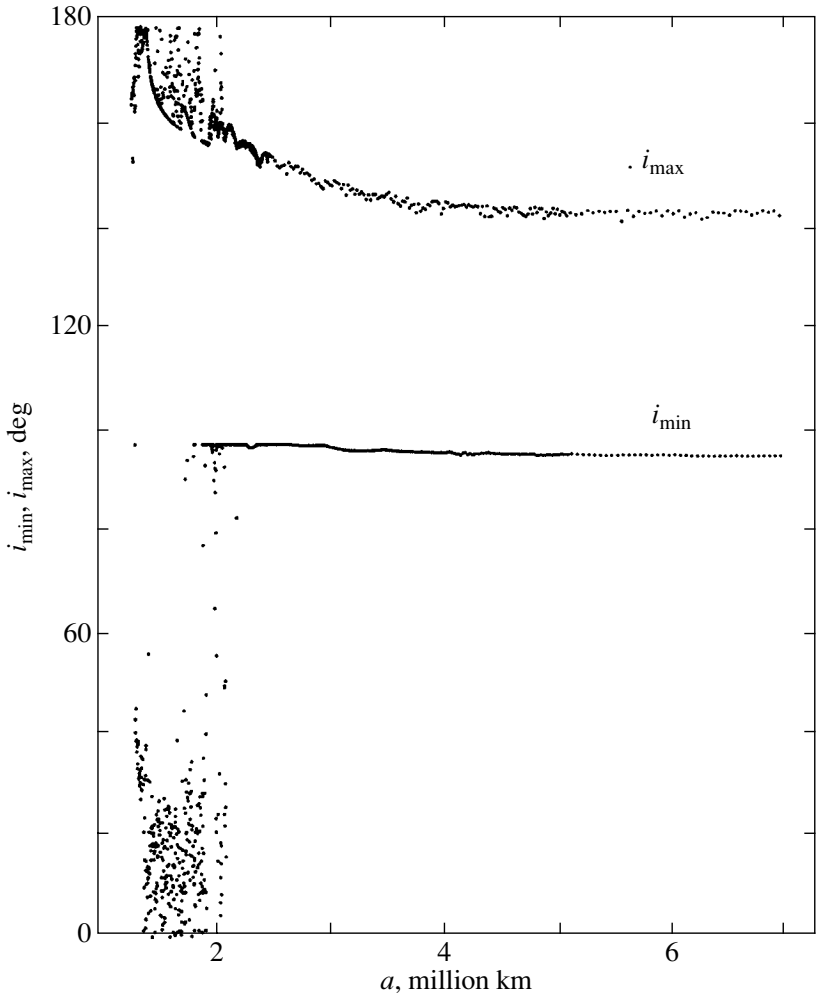


Fig. 4. Extreme i versus a on the time interval $T = 400$ thousand years for $e_0 = 0.03$, $\omega_0 = 0^\circ$, $i_0 \approx 98^\circ$.

ima of e for e_0 from the range 0.001–0.8 are reached at $\omega_0 \approx 90^\circ$. However, in the region of approximately equal influence of the perturbing factors under consideration ($\alpha \approx \beta$, $a \approx 1.3$ million km), the phase trajectories close to the separatrix ($e_0 < 0.03$) acquire chaotic properties. In this case, the first (in time) maximum of e can also be reached at $\omega \neq 90^\circ$. Our calculations for $a = 1.3$ million km and $e_0 = 0.001$ (Fig. 2c) show that the minima of q, nevertheless, do not exceed the values obtained for $e_0 = 0.03$.

The Evolution of Satellite Orbits on Time Scales of the Order of Uranus's Orbital Evolution Period

Since the characteristic period of the secular planetary perturbations in Uranus's orbit is $T \approx 400$ thousand years, the difference between the results of the first series of calculations described above and the results of the model problem was mainly affected by the nonorthogonality of Uranus's orbit to the ecliptic ($I \neq 90^\circ$). The effect of variations in the elements e_1, i_1, ω_1, and Ω_1 on the evolution of satellite orbits can be appreciable only on time scales of the order of T.

The second series of calculations with the same set of initial data as in the preceding subsection was performed precisely on this time interval, $0 \le t \le T$. Instead of e_m and q_m, we determined $e_{\max}(a) = \max\limits_{0 \le t \le T} e(a,t)$, $q_{\min}(a) = a[1 - e_{\max}(a)]$, $i_{\min}(a)$, $i_{\max}(a)$ for each a. The values of $q_{\min}(a)$ [million km] for $\omega_0 = 0^\circ$ are shown in Fig. 3: (a) $e_0 = 0.03$, (b) $e_0 = 0.001$, and (c) $e_0 = 0.0001$. To avoid unjustified complications, the points were plotted separately at constant steps in a without connecting them by a smooth curve. For the other initial values $\omega_0 = 90^\circ$, 180°, and 270°, the form of the dependences $q_{\min}(a)$ is virtually the same.

In all portions of Fig. 3, we clearly see transitions from a random variation in $q_{\min}$ to a more or less regu-

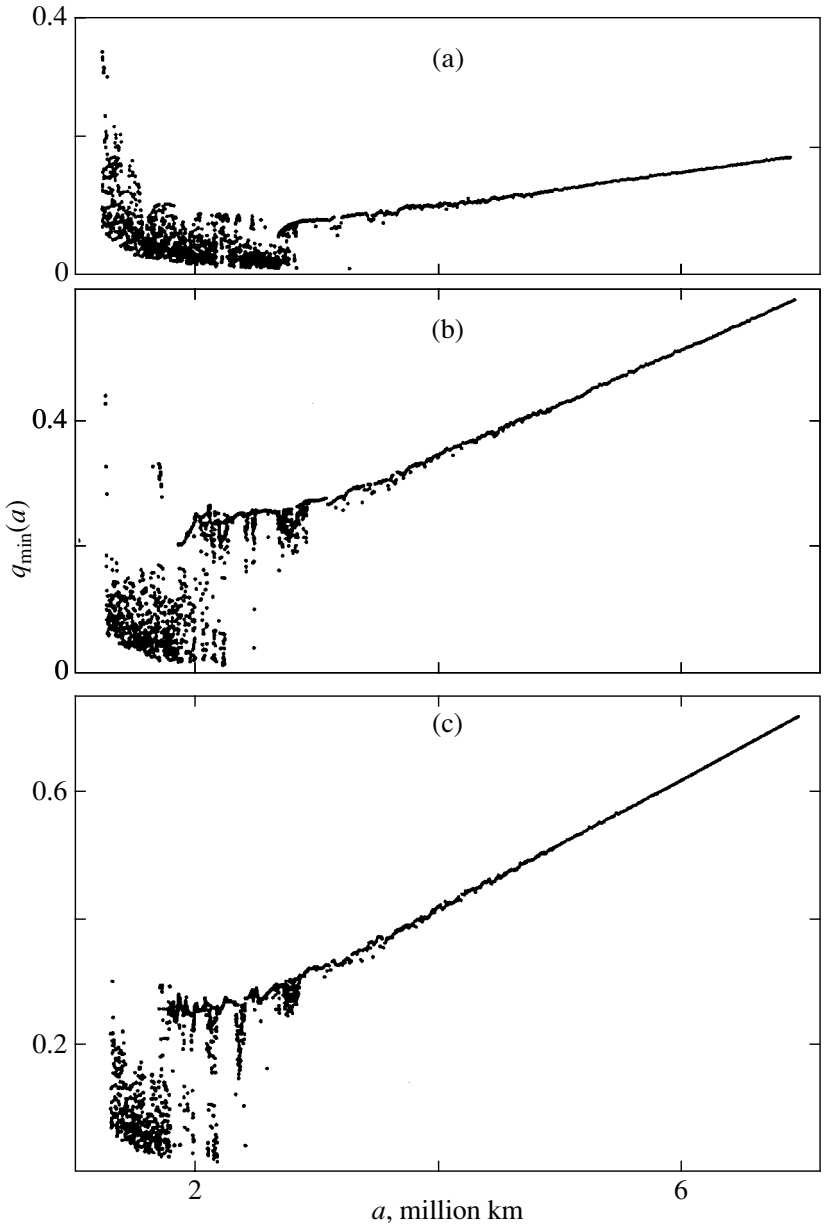


Fig. 5. The dependences $q_{\min}(a)$ on the time interval $T = 400$ thousand years for $e_0 = 0.03$, $\omega_0 = 0^\circ$ and for various i_0: (a) $i_0 = 80^\circ$, (b) $i_0 = 108^\circ$, and (c) $i_0 = 70^\circ$.

lar variation. This is because our evolution problem is nonintegrable for $\alpha/\beta \sim 1$ ($1.3 \leq a \leq 3$ million km). For $a > 3$ million km, when the effect of Uranus's oblateness is marginal ($\alpha/\beta < 10^{-3}$), the problem is similar to the integrable double-averaged Hill problem and the dependences $q_{\min}(a)$ are nearly regular. Nevertheless, even for $\alpha \to 0$, the problem is nonintegrable, because the orbital elements of Uranus are slowly varying functions of time. Note that the boundary values of a that separate the regions of random and regular variations in $q_{\min}$ increase with decreasing e_0, i.e., as the phase trajectory approaches the separatrix.

In contrast to the model problem where the inclination is constant, $i_0 = 90^\circ$, the evolution of initially equatorial orbits in our more complete physical model causes the inclination to vary. The ranges of such variations can be estimated from Fig. 4, which shows the dependences $i_{\min}(a)$ and $i_{\max}(a)$ for $e_0 = 0.03$, $\omega_0 = 0^\circ$. Similarly to Fig. 3, the random variation in extreme inclination with increasing a changes to its regular variation; for the assumed $e_0 = 0.03$, this change occurs at the same $a \approx 2.2$ million km as

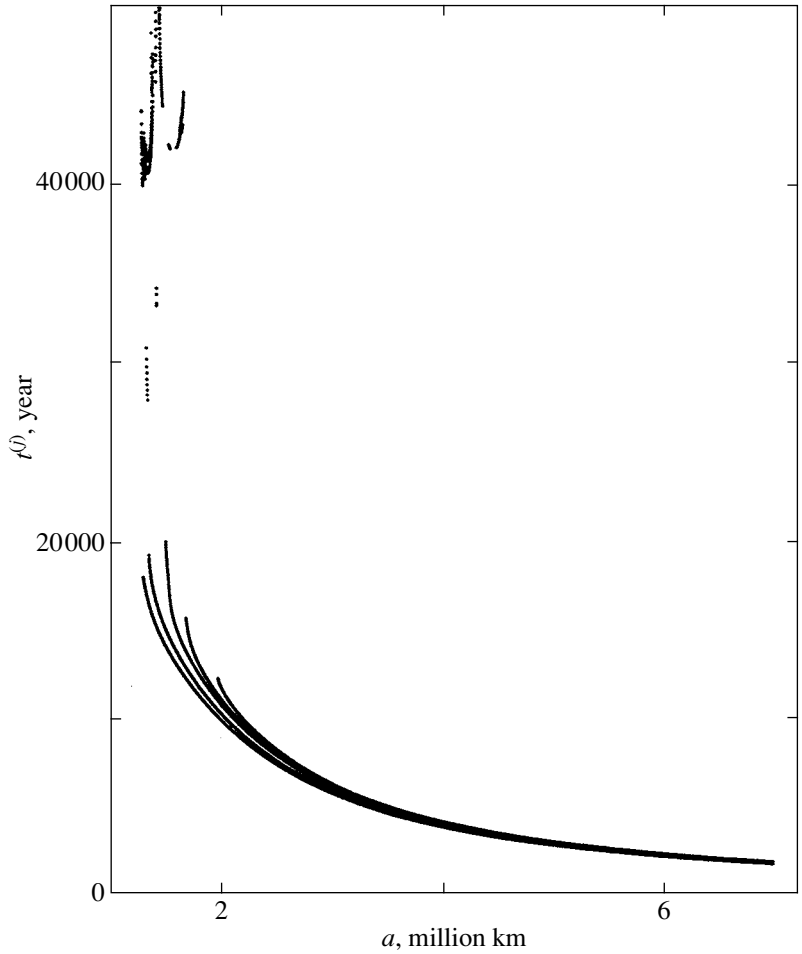


Fig. 6. Time scales of the evolution of hypothetical satellite orbits from initially circular orbits to highly eccentric orbits intersecting the orbits of major inner satellites.

in Fig. 3a. For $\alpha \to 0$, the extreme inclination and eccentricity are roughly related by Eq. (11).

Thus, our analysis of the two series of calculations performed in terms of a realistic model with $I \approx 98°$ and with an evolving orbit of Uranus leads us to the following conclusion.

The orbital evolution of Uranus's hypothetical equatorial satellites with semimajor axes in the range $1.3 \le a \le 7$ million km causes their pericenter distances to decrease to the orbital radii of one or more inner satellites in a time of $\sim 10^4$ yr (Fig. 2). On a time scale of the order of Uranus's orbital evolution period for the range $1.7 \le a \le 3$ million km, the satellite orbits can also intersect with the planetary surface ($q_{\min} \le a_0$, Fig. 3).

The Evolution of Nonequatorial Satellite Orbits

To get an idea of the evolution of Uranus's nonequatorial satellites, we calculated the dependences $q_{\min}(a)$ for some initial inclinations i_0 of the satellite orbit different from I (at the same Ω_0). Figure 5a shows $q_{\min}(a)$ [million km]: (a) for $i_0 = 80°$, (b) for $i_0 = 108°$, and (c) for $i_0 = 70°$. The latter value of i_0 was chosen in such a way that for the orbit with this initial inclination and with a major axis of 7 million km (the orbit of the innermost outer satellite, Caliban), $q_{\min}$ was ≈ 0.6 million km (the orbit of the outermost inner satellite, Oberon). For $i_0 \ne 90°$, $q_{\min}(a \ge 3$ million km) increases proportionally to a. The proportionality coefficient k increases with i_0 differing from 90°. In the limit for $\alpha \to 0$, it is given by the approximate formula $k \approx 1 - \left(1 - \frac{5}{3}c_1\right)^{1/2}$, where $c_1 = (1 - e_0^2)\cos^2 i_0$. It follows from Fig. 5 that for i_0 that do not differ greatly from 90°, the conclusion formulated in the preceding subsection remains valid.

Intersections of Hypothetical and Inner Satellite Orbits

In the model problem where $i(t) = i_0 = 90°$, the intersection of the equatorial orbit of a hypothetical satellite with the orbit of the jth inner satellite directly followed from the condition $q < a^{(j)}$. In our more rigorous evolution problem where the inclination i can differ greatly from its initial value, the following equalities serve as the condition for the intersection of these orbits:

$$(1 - e^2)a/a^{(j)} - e\cos\omega' - 1 = 0, \qquad (12)$$

$$(1 - e^2)a/a^{(j)} + e\cos\omega' - 1 = 0, \qquad (13)$$

where

$$\cos\omega' = \{\cos\omega[\sin i\cos i_0 - \cos i\sin i_0 \qquad (14)$$
$$\times \cos(\Omega - \Omega_0)] + \sin\omega\sin i_0\sin(\Omega - \Omega_0)\}/\sin i',$$
$$\cos i' = \cos i\cos i_0 + \sin i\sin i_0\cos(\Omega - \Omega_0),$$
$$\sin i' = \sqrt{1 - \cos^2 i'},$$

the angle i' is measured from Uranus's equatorial plane, and ω' is measured from the line of nodes of the satellite orbit in the same plane.

At fixed a, conditions (12) and (13) give two values of $t^{(j)}$ for each $j = 1, 2, \ldots, 5$ at which orbit intersections take place at one or another nodal point. In Fig. 6, the minimum values of this quantity are plotted against a. The initial elements e_0, i_0, ω_0, and Ω_0 are the same as those in the case shown in Fig. 3a.

Note that for any a and for conditions (12) and (13), $t^{(j)}$ satisfy the inequalities $t^{(1)} > t^{(2)} > \ldots > t^{(5)}$; i.e., the hypothetical satellite orbits first intersect the orbit of Oberon, then the orbit of Titania, and so on. In the regions of random variations in $q_{\min}$, $i_{\min}$, and $i_{\max}$, $t^{(j)}$ reach 50 thousand years or more (these are not shown in the figure in order not to stretch it vertically too much). As a increases, all five curves virtually merge into a single curve, with the minimum $t^{(j)}$ corresponding to condition (12). At sufficiently large a, i.e., at very small α, the Sun perturbs the satellite orbits so strongly that they intersect the orbits of inner satellites, one by one, in very short time intervals. The intersection times themselves at $a \approx$ 5–7 million km are no further than 3000 years from the initial time.

CONCLUSIONS

The absence of Uranus's equatorial satellites at distances larger than some *critical* value a_{cr} has qualitatively been explained previously. Thus, in the paper by Goldreich (1966), which was included in the well-known collection of papers (Goldreich 1975; paper II), an approximate formula was derived for $a_{\rm cr}$, in planetary radii a_0, from the condition for equality between the average moments of forces exerted on the satellite orbit by the Sun and by an oblate planet. In our notation, it is

$$\frac{a_{\rm cr}}{a_0} \approx \left(-2c_{20}\frac{\mu}{\mu_1}\right)^{1/5}\left(\frac{a_1}{a_0}\right)^{3/5}. \qquad (15)$$

By comparing this formula with Eqs. (2), we can easily verify that it also expresses the equality between the characteristic parameters of perturbations from the Sun and oblateness of the planet; i.e., it is equivalent to the condition

$$\alpha \approx \beta,$$

which yields $a_{\rm cr} \approx 1.36$ million km for the Sun–Uranus system. However, it follows from condition (15) that the moment of solar attractive forces for $a > a_{\rm cr}$ only takes the satellite orbits away from the equatorial plane without forbidding their existence.

Our numerical integration of a fairly complete evolution system in addition to a qualitative analysis of the model problem revealed a specific mechanism that limits the lifetime of Uranus's hypothetical satellites in the intermediate circumplanetary region, where no natural satellites have presently been found, to $\sim 10^4$ yr.

The existence of "empty" ranges of semimajor axes in the satellite systems of Jupiter and Saturn seems much more difficult to explain. In particular, this is true for the separation of outer satellites into two sets of satellites with prograde and retrograde orbits for Jupiter and for the intersection of these sets for Saturn.

ACKNOWLEDGMENTS

We wish to thank V.V. Beletskii for the advice that we used to estimate the evolution time scales of Uranus's rotation axis. This study was supported by by the Council of Grants of the President of Russia and the State Support for Leading Scientific Schools (grant no. 00-15-96036)

REFERENCES

1. V. K. Abalakin, E. P. Aksenov, E. A. Grebenikov, *et al.*, *A Reference Manual on Celestial Mechanics* (Nauka, Moscow, 1976).
2. V. V. Beletskii, *Motion of a Satellite Relative to the Center of Mass in a Gravitational Field* (Mosk. Gos. Univ., Moscow, 1975).
3. V. V. Beletskiĭ, *Tidal Evolution of the Inclinations and Rotations of Celestial Bodies*, Preprint No. 43 (Inst. Prikl. Mekh. Akad. Nauk SSSR, Moscow, 1978).
4. P. Goldreich, Rev. Geophys. **4**, 411 (1966).

5. P. Goldreich, in *Tides and Resonance in the Solar System: Collection of Articles* (Mir, Moscow, 1975), p. 97.
6. M. L. Lidov, Iskusstv. Sputniki Zemli **8**, 5 (1961).
7. M. L. Lidov, *Problems of the Motion of Artificial Celestial Bodies* (Akad. Nauk SSSR, Moscow, 1963a), p. 119.
8. M. L. Lidov, in *Dynamics of Satellites,* Ed. by M. Roy (Springer-Verlag, New York, 1963b), p. 168.
9. M. L. Lidov and M. V. Yarskaya, Kosm. Issled. **12**, 155 (1974).
10. Sh. G. Sharaf and N. L. Budnikova, Byull. Inst. Teor. Astron. Akad. Nauk SSSR **11**, 231 (1967).
11. M. A. Vashkov'yak, Pis'ma Astron. Zh. **27**, 543 (2001) [Astron. Lett. **27**, 464 (2001)].

Translated by V. Astakhov