Astronomy Letters, Vol. 27, No. 7, 2001, pp. 411–415. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 483–488.

Radial Distributions of Gamma-Ray Bursts and Type Ib/c Supernovae in Galaxies

D. Yu. Tsvetkov[1]*, S. I. Blinnikov[1, 2], and N. N. Pavlyuk[1]

[1] *Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119899 Russia*
[2] *Institute of Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia*

Received January 15, 2001

Abstract—Data on the positions of gamma-ray bursts (GRBs) in galaxies are used to construct the radial distributions of their surface density. The gradient in GRB surface density is shown to decrease sharply at a galactocentric distance equal to the effective galactic radius. In central galactic regions, the GRB density distribution agrees with the galactic surface-brightness distribution; in outer regions, the GRB density decreases more slowly than does the surface brightness. Based on improved statistics, we analyze the radial distribution of type Ib/c supernovae. We show that it differs insignificantly from the distributions of other types of supernova and exhibits a much closer similarity to the distribution of star-forming regions than do GRBs. Although the statistics for GRBs is poor, the deviation of their distribution from the distribution of active star-forming regions in nearby galaxies seems to have been firmly established. A correlation of GRBs with the distribution of dark matter in outer galactic regions is not ruled out. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *gamma-ray bursts, supernovae and supernova remnants, theoretical and observational cosmology*

INTRODUCTION

Studies of the spatial distribution of supernovae (SNe) in parent galaxies have provided important information about the probable nature of the objects whose evolution leads to various types of SN explosion (see, e.g., Bartunov *et al.* 1992, 1994; Bartunov and Tsvetkov 1997; van den Bergh 1997).

The currently available observational data on the distribution of gamma-ray bursts (GRBs) are similar to the data for SNe. Of course, it should be remembered that the localization errors for GRBs in parent galaxies are larger than those for SNe; besides the statistics of localized GRBs is much poorer. Nevertheless, the procedure used by Bartunov *et al.* (1992, 1994) yielded important results for SNe, whose statistics was not rich either. It is therefore of interest to apply the procedure used to study SNe to an analysis of the GRB distribution.

The discovery of GRB afterglows and their redshift measurements severely constrain theoretical models for GRBs [see the reviews by Piran (1999) and Postnov (1999)]. Investigating the distribution of GRBs in parent galaxies is of great importance in understanding their nature. Models in which GRBs are produced during the collapse of massive stars must lead to a correlation of GRBs with star-forming regions. In the model of merging binary relativistic stars, GRBs can occur far from formation regions of massive stars. Finally, models that use exotic variants of dark matter have been proposed [see the review by Blinnikov (2000)]. Establishing a correlation of the distribution of GRBs with a particular type of stellar population or with dark matter can shed light on their origin.

ANALYZING THE GRB DISTRIBUTION

The latest observational data on GRB positions in parent galaxies are collected in Bloom *et al.* (2000). The angular distances R from the galactic centers were determined for twenty GRBs. The effective galactic radius $R_{\rm eff}$, the radius within which half the luminosity is emitted, was used to normalize the distances; the relative galactocentric distance $r_{\rm rel} = R/R_{\rm eff}$ was determined for each GRB. The effective radii were determined directly from observational data or estimated from an empirical relation between the magnitude and effective radius. The redshifts were known for fifteen GRBs, and their galactocentric distances were also determined in kiloparsecs. Bloom *et al.* (2000) constructed cumulative distributions for the relative and absolute distances of GRBs from the centers of their parent galaxies as well as the probability distribution of GRBs in relative distance with measurement errors. It was concluded that the observed distributions agreed well with the distribution of the populations belonging to the disks of galaxies with an exponential decline in density from the center to the edge and did not agree with the distribution of merging binary relativistic objects. The latter include binary neutron stars and

* E-mail address for contacts: tsvetkov@sai.msu.su

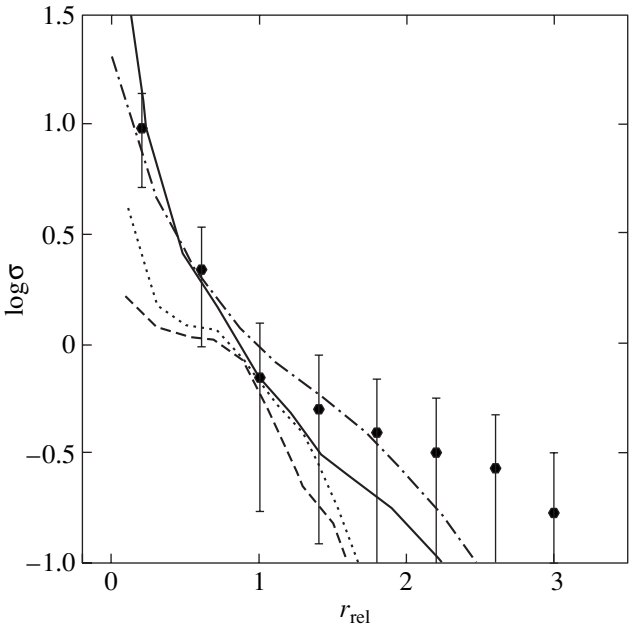


Fig. 1. GRB surface density versus galactocentric distance, in units of effective radius. The GRB surface density referred to the middle of the corresponding bin is denoted by dots, the solid line represents the radial dependence of surface brightness for elliptical galaxies, the dash-dotted line represents the same dependence for spiral galaxies, the dashed line represents the distribution of OB associations in M 33 and of H II regions in NGC 3184, and the dotted line represents the distribution of SNe Ib/c.

pairs of neutron stars with black holes, which can move far from the formation regions of massive stars.

In our view, representing the radial distribution of objects as a dependence of their surface density on galactocentric distance is preferred for comparing the distributions of various galactic populations.

As in Bartunov *et al.* (1992), we determined smoothed GRB surface densities $\sigma_i = N_i/\pi(r_{i+1}^2 - r_i^2)$. Here, r_i is the radius of the ith bin in units of effective galactic radius (r_{rel}) or in kpc; N_i is the smoothed number of objects in the ith bin calculated from the true numbers n_i as $N_i = 0.5n_i + 0.25(n_{i-1} + n_{i+1})$.

In Fig. 1, the logarithm of GRB surface density is plotted against the galactocentric distance. Data for eighteen GRBs were used; we excluded two GRBs with relative distances of 9.7 and 11.0 for which, according to Bloom *et al.* (2000), the parent galaxies were probably not detected, while the above values refer to the nearest field galaxies. The bin was taken to be 0.4; only statistical errors proportional to $N_i^{0.5}$ are shown. The gradient in GRB surface density $d\log\sigma/dr_{rel}$ is seen to change abruptly at $r_{rel} = 1.0$; the steep decline in density gives way to a smoother decline. The dependence of $\log\sigma$ on r_{rel} can be fitted by two straight lines with gradients of -1.4 ± 0.1 and -0.28 ± 0.02, respectively.

Bartunov *et al.* (1992) and Bartunov and Tsvetkov (1997) determined the gradients in the logarithms of surface densities for various types of SN and for various types of galactic populations. Galactic diameters D_{25} from the RC3 Catalog (de Vaucouleurs *et al.* 1991) were used for normalization. To compare these data with the results for GRBs, we calculated the mean ratio $D_{25}/2R_{eff} = 2.7$ for all RC3 galaxies for which these parameters were available and reduced all the previously derived density gradients to the new normalization to R_{eff}.

The gradients in surface density in central galactic regions ($r_{rel} < 1.0$) are: –0.44 for SNe Ia, –0.18 for SNe II, –1.4 for the surface brightness of elliptical galaxies, and –1.0 for the surface brightness of spiral galaxies. In outer galactic regions, these gradients are –0.52, –0.52, –0.56, and –0.81, respectively.

The GRB density gradient in central galactic regions is considerably larger in magnitude than that for the SN samples analyzed and agrees well with the surface-brightness gradients in elliptical galaxies. However, in outer galactic regions, the gradient in GRB surface density is smaller than that for SNe and for the surface brightness of all types of galaxies, although the number of GRBs in this region is too small to draw firm conclusions.

For comparison, Fig. 1 shows the surface-brightness distributions in elliptical and spiral galaxies, as well as the distributions of OB associations in M 33 and of H II regions in NGC 3184, as constructed from the data of Burstein (1979), Humphreys and Sandage (1980), Boroson (1981), and Hodge and Kennicutt (1983). These distributions were displaced along the vertical axis to match best the data for GRBs. The GRB distribution in central galactic regions is seen to agree well with the surface-brightness distribution in elliptical and spiral galaxies. The distribution of star-forming regions differs sharply from the GRB distribution.

In Fig. 2, the logarithm of GRB surface density is plotted against galactocentric distance (in kpc). Data for fourteen GRBs were used, and the bin was 1.2 kpc. The dependence for $r < 5$ kpc is well fitted by a straight line; the gradient in $\log\sigma$ is -0.40 ± 0.03, which corresponds to $r_0 = 1.1$ kpc in the expression $\sigma \sim \exp(-r/r_0)$. As for the distribution in relative radius, the GRB density gradient is considerably larger in magnitude than that for all the SN samples analyzed. The radial dependence of GRB surface density agrees excellently with the surface-brightness distribution in elliptical galaxies and differs from the distribution of star-forming regions.

The GRB distributions we considered differ from the radial distributions of all types of SN (Bartunov *et al.* 1992; Bartunov and Tsvetkov 1997; van den Bergh 1997). The SN distributions in galactic disks are represented satisfactorily by an exponential decline in density with a constant coefficient and agree well with

the surface-brightness distribution in the disks of spiral galaxies.

Analyzing the distribution of SNe Ib/c is of particular interest: first, Bartunov *et al.* (1992) found their unusually strong concentration to the galactic centers; and, second, some GRBs may be associated with SN Ib/c explosions.

For our analysis, we used data on fifty SNe Ib/c from the Catalog of Supernovae, Sternberg Astronomical Institute (Tsvetkov *et al.* 2000). The derived radial distributions are shown in Figs. 1 and 2. We see that the dependences of the logarithm of SN Ib/c surface density on relative radius and radial distance (in kpc) are satisfactorily represented by straight lines. The gradients in $\log\sigma$ relative to $r_{\rm rel}$ and the radius (in kpc) are -0.78 ± 0.04 and -0.18 ± 0.01, respectively; $r_0 = 2.4$ kpc. The SN Ib/c distributions satisfactorily agree with the distributions of OB associations and H II regions and differ markedly from the GRB distributions. The SN Ib/c density gradient essentially matches the surface-brightness gradient of spiral galaxies in their outer parts and is close to the corresponding data for other types of SN.

A comparison of the samples analyzed by using Kolmogorov–Smirnov's test (Press *et al.* 1986) confirms the qualitative conclusions reached above. The GRB distribution in relative radial distance was compared with the distributions shown in Fig. 1. We obtained the following probabilities that a given pair of distributions belongs to the same sample: 68% for the pair "GRB–surface brightness of spirals," 40% for the pair "GRB–surface brightness of ellipticals," 4% for the pair "GRB–OB associations and H II regions," and 9% for the pair "GRB–SN Ib/c."

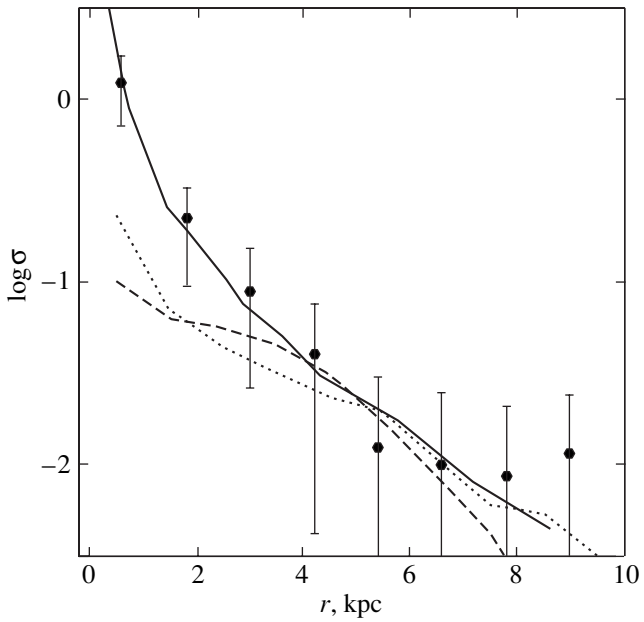


Fig. 2. GRB surface density versus galactocentric distance, in kpc. The GRB surface density is denoted by dots, the solid line represents the radial dependence of surface brightness for elliptical galaxies, the dashed line represents the distribution of OB associations in M 33 and of H II regions in NGC 3184, and the dotted line represents the distribution of SNe Ib/c.

DISCUSSION

Some conclusions can already be drawn from the derived distributions of GRBs in parent galaxies. They are tentative so far, but with the accumulation of observational data and with the improvement of statistics, the procedure considered here will bring us closer to revealing the nature of GRBs.

If GRBs are produced during the collapse of massive stars (Woosley 1993; Popham *et al.* 1999; Gershteĭn 2000; Cherepashchuk and Postnov 2000), then a strong correlation of GRBs with star-forming regions or with Wolf–Rayet stars must be observed.

In the model of merging binary neutron stars (Blinnikov *et al.* 1984) and pairs of neutron stars with black holes (Lattimer and Schramm 1974, 1976; Eichler *et al.* 1989), GRBs can occur far from the formation regions of massive stars, although this is not necessary, and the issue requires a special study.

Finally, because of the numerous difficulties in explaining GRBs in terms of standard physical theories, their models that invoke exotic particles, various variants of dark matter, etc. have been proposed (Loeb 1993; Bertolami 1999; Demir and Mosquera Cuesta 1999; Iwazaki 1999; Blinnikov 1999, 2000). The nature of dark matter is yet to be elucidated. Let us briefly consider the latest results in this field. Until recently, hypothetic weakly interacting particles (for example, superlight axions or, on the contrary, Weakly Interacting Massive Particles, WIMPs, such as neutralinos) were considered to be the most suitable candidates for explaining the entire set of data on dark matter. At present, these particles are nonrelativistic [i.e., they are candidates for Cold Dark Matter (CDM) particles], which is required to account for the formation of galaxies and their clusters. However, the theory that uses such particles ran into several problems. First, the rotation curves of galaxies, which are also determined by dark matter, show that it must have a constant density in central galactic regions, while calculations in CDM models for weakly interacting particles yield a structure with a sharp central peak and a $1/r^2$ decline. Second, such calculations yield too many clouds of dark matter in massive galactic halos, a factor of 10 to 50 more than admissible.

Spergel and Steinhardt (2000) pointed out that these two problems could be solved if the particles of dark matter were assumed to interact strongly and elastically: the central peaks are smoothed out, while the clouds are dispersed.

Meneghetti *et al.* (2000) constrained the strength of CDM self-action. They pointed out that at a too large cross section for particle interaction, dark-matter structures in clusters of galaxies are smoothed out so

strongly that observed phenomena, such as arcs in the images of very distant quasars and galaxies, cannot be explained; these arcs are produced by unseen cluster matter through the effect of a gravitational lens, when the light from distant objects passes through the cluster. A comparison with observations suggests that the cross section for CDM particle interaction per unit mass cannot exceed 0.1 $cm^2\ g^{-1}$. Interestingly, an upper limit on the interaction of dark-matter particles can also be obtained from different considerations. Burkert (2000) gave an overview of the results showing that at a too large interaction cross section, a homogeneous isothermal core of dark matter rapidly collapses. It is important that the cross section cannot be much smaller either, because otherwise, the desired smoothing will not be achieved.

Such a large interaction cross section can give rise to star-like objects composed of dark matter, which may collapse and produce GRBs in various ways (Blinnikov 2000). It should be remembered that dark matter may consist of dissimilar particles and that a hierarchy of stellar populations similar to ordinary matter can form in it (Blinnikov and Khlopov 1983; Berezhiani *et al.* 1996).

Invisible matter in the central regions of many spiral galaxies can be distributed in the same way as neutral hydrogen, while its density in outer regions falls off considerably more slowly (Hoekstra *et al.* 2000). Our analysis of observations indicates that the radial distribution of GRBs in the inner parts of galaxies is similar to the distribution of the population of the galactic spherical component, while in the outer parts, their density decreases more slowly than does the surface brightness of the disks of spiral galaxies, i.e., qualitatively similar to dark matter.

Our results show that the radial distribution of GRBs differs from that of active star-forming regions in nearby galaxies. However, many observers of GRB afterglows insist that they see them mostly in galaxies that are entirely engulfed by a starburst (Sokolov *et al.* 2000). These galaxies are morphologically irregular, and it would be inappropriate to compare them with nearby normal spiral galaxies. Nevertheless, in this case, we cannot say that a GRB was necessarily produced by an ordinary massive star either. There must be a factor that stimulates starbursts in such galaxies. Occasionally, but very rarely, such galaxies have companions; the tides produced by them are responsible for starbursts. However, they are generally single (Taylor *et al.* 1995; Telles and Maddox 2000). As was noted by Trentham *et al.* (2000) (and previously by Blinnikov 1999, 2000), such galaxies probably interact gravitationally with companions that are entirely composed of dark matter. If GRBs are produced by dark matter, then we must see their correlation with starbursts, but, of course, this is not evidence for the genetic relationship between GRBs and ordinary massive stars.

It should also be remembered that there are bright afterglows (for example, the source GRB 0003101C) that have no galaxies brighter than 29^m detected near them. There are also other indications that absolutely dark galaxies exist (Trentham *et al.* 2000).

Let us summarize our conclusions. The distribution of GRBs is similar to the distribution of matter in the bulges of spiral and elliptical galaxies; in outer galactic regions, it is flatter, i.e., differs significantly both from the distribution of active star-forming regions in nearby galaxies and from the distribution of SNe Ib/c. The currently available data do not rule out a correlation of GRBs with dark matter either (despite the strong selection effect, afterglows must be seen more commonly where there is much ordinary matter).

ACKNOWLEDGMENTS

We wish to thank K.A. Postnov and M.E. Prokhorov for stimulating discussions and O.K. Sil'chenko for help. This study was supported by the Russian Foundation for Basic Research (project no. 99-02-16205).

REFERENCES

1. O. S. Bartunov and D. Yu. Tsvetkov, in *Thermonuclear Supernovae*, Ed. by P. Ruiz-Lapuente *et al.* (Kluwer, Dordrecht, 1997), p. 87.
2. O. S. Bartunov, I. N. Makarova, and D. Yu. Tsvetkov, Astron. Astrophys. **264**, 428 (1992).
3. O. S. Bartunov, D. Yu. Tsvetkov, and I. V. Filimonova, Publ. Astron. Soc. Pac. **106**, 1276 (1994).
4. Z. G. Berezhiani, A. D. Dolgov, and R. N. Mohapatra, Phys. Lett. B **375**, 26 (1996).
5. S. van den Bergh, Astron. J. **113**, 197 (1997).
6. O. Bertolami, Astropart. Phys. **11**, 357 (1999).
7. S. I. Blinnikov, astro-ph/9902305 (1999).
8. S. I. Blinnikov, Surv. High Energy Phys. **15**, 37 (2000).
9. S. I. Blinnikov and M. Yu. Khlopov, Astron. Zh. **60**, 632 (1983) [Sov. Astron. **27**, 371 (1983)].
10. S. I. Blinnikov, I. D. Novikov, T. V. Perevodchikova, and A. G. Polnarev, Pis'ma Astron. Zh. **10**, 422 (1984) [Sov. Astron. Lett. **10**, 177 (1984)].
11. J. S. Bloom, S. R. Kulkarni, and S. G. Djorgovski, astro-ph/0010176 (2000).
12. T. Boroson, Astophys. J., Suppl. Ser. **46**, 177 (1981).
13. A. Burkert, astro-ph/0012178 (2000).
14. D. Burstein, Astrophys. J., Suppl. Ser. **41**, 435 (1979).
15. A. M. Cherepashchuk and K. A. Postnov, astro-ph/0012512 (2000).
16. D. A. Demir and H. J. Mosquera Cuesta, astro-ph/9903262 (1999).
17. D. Eichler, M. Livio, T. Piran, and D. N. Schramm, Nature **340**, 126 (1989).
18. S. S. Gershteĭn, Pis'ma Astron. Zh. **26**, 848 (2000) [Astron. Lett. **26**, 730 (2000)].
19. P. W. Hodge and R. C. Kennicutt, Astron. J. **88**, 296 (1983).

20. H. Hoekstra, T. S. van Albada, and R. Sancisi, astro-ph/0010569 (2000).
21. R. M. Humphreys and A. Sandage, Astrophys. J., Suppl. Ser. **44**, 319 (1980).
22. A. Iwazaki, Phys. Lett. B **455**, 192 (1999).
23. J. M. Lattimer and D. N. Schramm, Astrophys. J. Lett. **192**, L145 (1974).
24. J. M. Lattimer and D. N. Schramm, Astrophys. J. **210**, 549 (1976).
25. A. Loeb, Phys. Rev. D **48**, 3419 (1993).
26. M. Meneghetti, N. Yoshida, M. Bartelmann, *et al.*, astro-ph/0011405 (2000).
27. T. Piran, Phys. Rep. **314**, 575 (1999).
28. R. Popham, S. E. Woosley, and C. Fryer, Astrophys. J. **518**, 356 (1999).
29. K. A. Postnov, Usp. Fiz. Nauk **169**, 545 (1999).
30. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1986).
31. V. V. Sokolov, T. A. Fatkhullin, and V. N. Komarova, astro-ph/0006207 (2000).
32. D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).
33. C. L. Taylor, E. Brinks, R. M. Grashuis, and E. D. Skillman, Astrophys. J., Suppl. Ser. **99**, 427 (1995).
34. E. Telles and S. Maddox, Mon. Not. R. Astron. Soc. **311**, 307 (2000).
35. N. Trentham, O. Möller, and E. Ramirez-Ruiz, astro-ph/0010545 (2000).
36. D. Yu. Tsvetkov, N. N. Pavlyuk, O. S. Bartunov, and Yu. P. Pskovskiĭ, *Catalogue of Supernovae*, http://www.sai.msu.su/sn/sncat/ (2000).
37. G. de Vaucouleurs, A. de Vaucouleurs, H. G. Corwin, R. J. Buta, G. Paturel, and P. Fouque, *Third Reference Catalogue of Bright Galaxies* (Springer-Verlag, New York, 1991).
38. S. E. Woosley, Astrophys. J. **405**, 273 (1993).

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 416–420. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 489–494.

Anisotropy in the Sky Distribution of Short Gamma-Ray Bursts

V. F. Litvin[1], S. A. Matveev[1], S. V. Mamedov[1], and V. V. Orlov[2]*

[1] *Research Institute of Physics, St. Petersburg State University, Ul'yanovskaya ul. 1, Petrodvorets, 198904 Russia*
[2] *Astronomical Institute, St. Petersburg State University, Bibliotechnaya pl. 2, Petrodvorets, 198904 Russia*

Received June 19, 2001; in final form, April 24, 2001

Abstract—We analyze the sky distribution of various types of cosmic gamma-ray bursts (GRBs): short, long, and intermediate; they are determined by burst duration T_{90} (T_{90} is the time during which 90% of the burst energy is accumulated). We have found an anisotropy in the distribution of intermediate (2 s < T_{90} < 8 s) and short (T_{90} < 8 s) GRBs in the form of spots with an enhanced GRB concentration near the Galactic coordinates $l = 115°$ and $b = 30°$. Given the BATSE nonuniform exposure function, the statistical significance of the anisotropy is 99.89% for intermediate GRBs and 99.99% for short GRBs. Thus, we suggest that this anisotropy has a natural origin and is not caused by BATSE instrumental effects. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *cosmic gamma-ray bursts*

INTRODUCTION

Here, we analyze the sky distribution of gamma-ray bursts (GRBs) by using the BATSE (Burst and Transient Source Experiment) electronic catalog of GRBs for the end of 2000 (Meegan *et al.* 2000). The catalog contains data on 2702 GRBs, 2037 of which have information about duration T_{90} (the time during which 90% of the burst energy is accumulated; see Paciesas *et al.* 1999, Koshut *et al.* 1996).

Two types of GRB were first identified by T_{90}: type 1, long bursts with T_{90} > 2 s, generally with a higher energy and a soft spectrum; and type 2, short bursts with T_{90} < 2 s with a low energy and a harder spectrum (Kouveliotou *et al.* 1993; Belli 1995; Dezalay *et al.* 1996). Subsequently, three types of GRB were identified: type 1 (long, bright, soft), type 2 (short, weak, hard), and type 3: intermediate bursts (2 s < T_{90} < 8 s) with a moderate energy and a soft spectrum (Mukherjee *et al.* 1998; Horvath 1998; Belousova *et al.* 1999). Figure 1 shows the distribution of BATSE GRBs in duration T_{90}. The intermediate peak with which the third type of GRB is associated is indistinct. Figure 2 shows the hardness ratio-duration T_{90} diagram for BATSE GRBs. The hardness is defined as the ratio of the fluxes in the spectral channels of BATSE discriminators 3 (100–300 keV) and 2 (50–100 keV). In this diagram, the GRBs can be divided into two populations: generally harder short GRBs and generally softer long GRBs. The most recent studies indicate that the longer GRBs can be broken down with a higher confidence into three types rather than two types; however, the very existence of the intermediate class can be methodical in origin (Roiger *et al.* 2000; Hakkila *et al.* 2000a–2000d).

We analyze the sky distribution of various populations of GRBs (short, long, and intermediate); additionally, we consider several subtypes of short bursts to obtain a more reliable estimate for the statistical significance of the detected anisotropy.

To study the sky distribution of GRBs, we map the populations of GRBs under consideration in galactic coordinates (l, b) in the Mercator projection of the celestial sphere and find approximate positions of statistically significant spots with an enhanced GRB concentration. A description of the mapping procedure and the results are given in Section 1.

Since the BATSE exposure is nonuniform (i.e., it depends on the position on the celestial sphere), an artificial anisotropy in the sky distribution of GRBs is expected (Paciesas *et al.* 1999; Briggs *et al.* 1996; Tegmark *et al.* 1996a, 1996b). To quantitatively estimate the anisotropy attributable to the nonuniform exposure, we simulate the sky distribution of GRBs by the Monte Carlo method with allowance for the exposure function. In Section 2, we describe the simulation procedure and present the results. A discussion and conclusions are given in Section 3.

* E-mail address for contacts: vor@astro.spbu.ru

1. MAPPING THE GRB POPULATIONS

The relative surface density of GRBs $n_i = N_i/N_{4\pi}$ is used to map the sky distribution of GRB populations. The celestial sphere is quasi-uniformly covered by a grid of 150 nodes. A circular cell of radius R centered at a given node corresponds to each node. We determine the number N_i of bursts in the ith cell, the mean $N_{4\pi}$ of N_i for all cells, and the rms deviation σ of N_i from $N_{4\pi}$. Using a set of n_i, we construct a family of lines of equal relative GRB density and find regions of enhanced GRB concentration in the form of spots of a light shade (a lighter shade corresponds to a higher relative density). This mapping procedure is applied to all the types of GRB under study for several R. The parameter R was varied to check more thoroughly the nonuniformities found in the sky distribution of various types of bursts. Indeed, if a statistically significant anisotropy is recorded at some R, then it must also be preserved with varying R. Since the characteristic size of the most significant spots was $R_{sp} \sim 40°$, we used R = 35°, 40°, and 45° to determine the statistical significance of the anisotropy. Maps of the distributions of some GRB populations for $R = 40°$ are shown in Figs. 3 and 4.

Having found spots of enhanced GRB concentration on the map, we fitted them by a circle of radius $R_{sp} = R$ roughly centered at the spot under consideration and determined the number of GRBs N_{sp} in it. Knowing $N_{4\pi}$ and σ, we estimated the significance level of this spot as

$$S_{sp} = (N_{sp} - N_{4\pi})/\sigma. \quad (1)$$

Note that the value of S_{sp} determined in this way was obtained without allowance for the anisotropy of the exposure function in the sky (the nonuniform exposure is taken into account in Section 2).

Statistical data for the sky distributions of the types of GRB under study and characteristics of the most significant spots are given in Table 1, where N_{pop} is the number of GRBs in a given population, and l_{sp} and b_{sp} are the Galactic coordinates of the spot center.

As we see from Table 1, the most significant anisotropy is observed for type 2b ($T_{90} < 5$ s), type 2c ($T_{90} <$ 8 s), and type 3 (2 s $< T_{90} <$ 8 s) at all values of R.

2. ALLOWANCE FOR THE EFFECTS OF BATSE NONUNIFORM EXPOSURE ON THE SKY DISTRIBUTION OF GRBs

When studying the sky distribution of GRBs, it should be borne in mind that some anisotropy of GRBs must result from the BATSE nonuniform exposure (Paciesas *et al.* 1999; Briggs *et al.* 1996; Tegmark *et al.* 1996a, 1996b). In equatorial coordinates, the exposure depends on declination δ alone (Paciesas *et al.* 1999; Hakkila *et al.* 1997). The exposure function $F_e(\delta)$ can be represented as the probability of detecting a GRB at some point (α, δ) in the sky. Since this function reaches a maximum near the celestial poles, an increase in GRB density is expected near these poles. Figure 5 shows a topographic map of the normalized exposure function f_e in the (l, b) plane. We normalized the function F_e to its largest maximum, which is reached near the north celestial pole. The most significant anisotropy of short (subtypes 2b and 2c) and intermediate GRBs is observed in the region of the north celestial pole, where the detection probability is highest. Therefore, it is necessary to check if the anisotropy we found results from the nonuniform detection of GRBs by BATSE.

To estimate the significance of the anisotropy with allowance for the nonuniform exposure, we use the Monte Carlo method with carrier function $f_e(\delta)$. Thus, we simulate an anisotropic sky distribution of the GRB

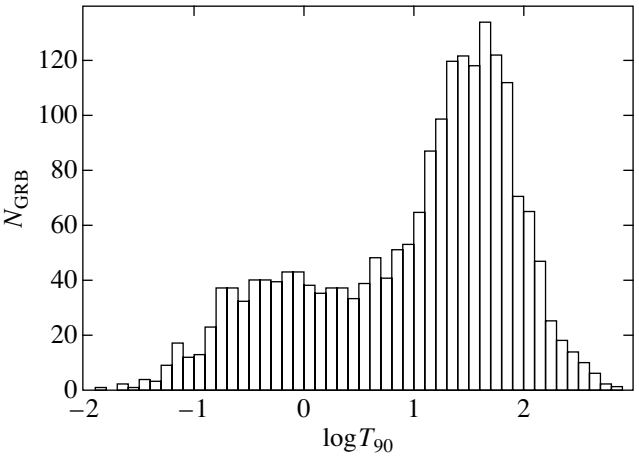


Fig. 1. The distribution of GRBs in duration T_{90}. logT_{90} (T_{90} is in seconds) is plotted along the horizontal axis, and the number of GRBs per 0.1 bin is plotted along the vertical axis.

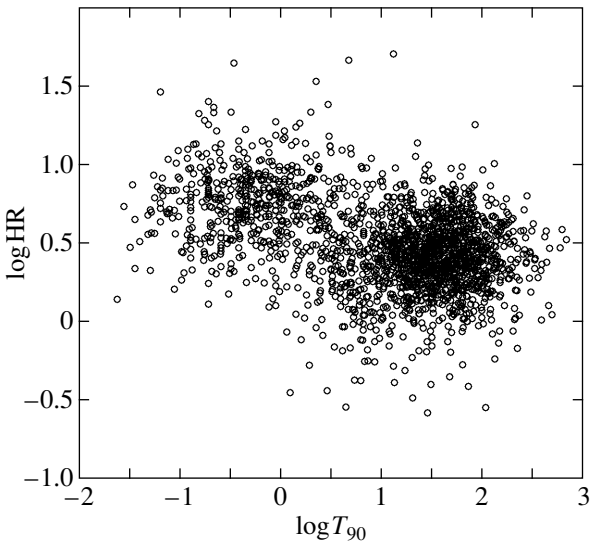


Fig. 2. Hardness ratio (HR) versus duration T_{90} for 1951 GRBs on a logarithmic scale.

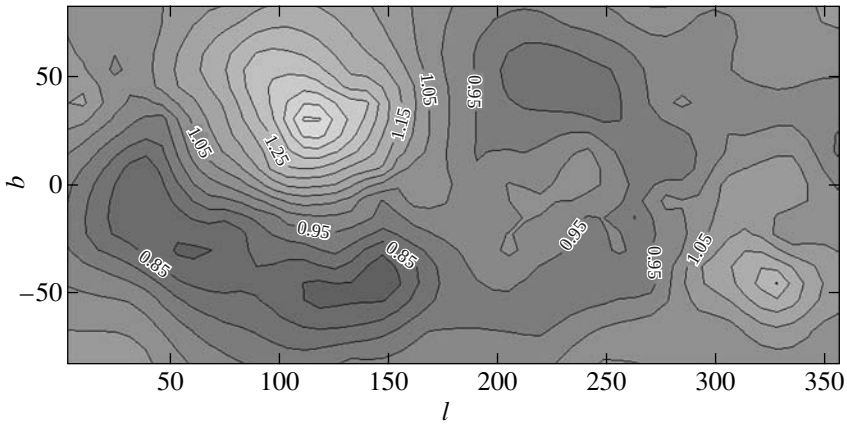


Fig. 3. A topographic map of the distribution of 746 short GRBs with $T_{90} < 8$ s in Galactic coordinates (l, b).

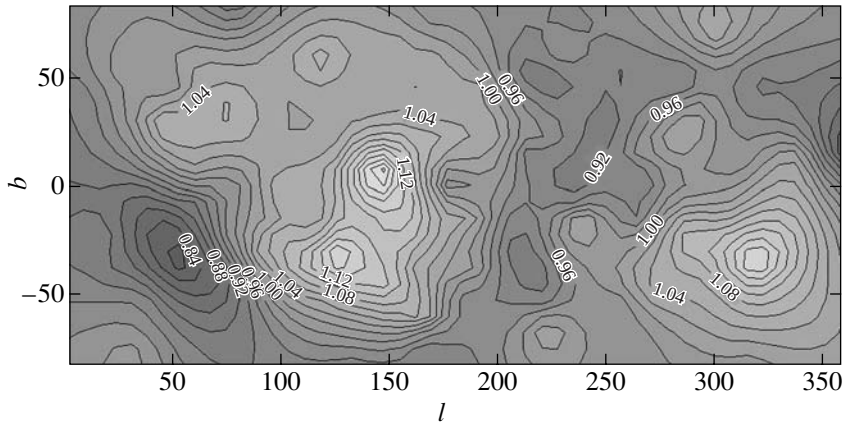


Fig. 4. A topographic map of the distribution of 1291 long GRBs with $T_{90} > 8$ s in Galactic coordinates (l, b).

populations under study, which is expected to result from the nonuniform exposure. The simulations are carried out as follows:

(1) the equatorial coordinates (α, δ) of a model GRB are randomly generated;

(2) with probability $f = f_e(\delta)$, this model GRB is "thrown" on the celestial sphere to the point with coordinates (α, δ);

(3) in this way, $N_{\rm pop}$ objects are thrown on the celestial sphere, where $N_{\rm pop}$ is the number of GRBs in a given population.

Having obtained the sky distribution of model GRBs, we transformed their equatorial coordinates to the galactic coordinates and determined N_m, $N_{4\pi}$, and σ, where N_m is the maximum number of objects falling within one of the grid cells. The nodes and cell size are the same as those used to map the populations of observable GRBs. Based on these data, determine the significance of the simulated spot as

$$S_m = (N_m - N_{4\pi})/\sigma. \quad (2)$$

If $S_m \geq S_{\rm sp}$, where $S_{\rm sp}$ is the spot significance for the population of observable bursts with the same number $N_{\rm pop}$ of bursts in it, then this event is recorded.

We perform this simulation p times and determine the number q of cases in which the condition $S_m \geq S_{\rm sp}$ is satisfied. The quantity

$$P = q/p \quad (3)$$

is then the probability of a random realization of a spot with enhanced GRB concentration with allowance for the nonuniform exposure.

Table 2 gives the results of our simulations for all the burst populations for $p = 20\,000$ tests. Subtypes 2b, 2c, and type 3 exhibit a significant deviation from isotropy at $R = 40°$; by varying R, we find that the deviation from isotropy remains at a high level. Bursts with $T_{90} > 8$ s also exhibit a large deviation from isotropy at $R = 40°$, but the significance of the anisotropy considerably decreases with varying R. Thus, the anisotropy is confirmed for types 2b, 2c, and 3 and is not confirmed for type 1. Bursts with $T_{90} < 2$ s exhibit no statistically significant deviations from anisotropy.

Table 1. Statistical data for the sky distributions of the GRB populations under study and characteristics of spots with enhanced concentration

Population	N_{pop}	R	$N_{4\pi}$	σ	N_{sp}	S_{sp}	$(l_{sp}; b_{sp})$
Short	497	35°	45.0	7.0	68	3.3	(322°; –39°)
Subtype 2a		40°	58.4	8.1	82	2.9	(114°; 29°)
$T_{90} < 2$ s		45°	73.0	8.7	98	2.9	(298°; –43°)
Short	653	35°	58.9	8.5	93	4.0	(119°; 21°)
Subtype 2b		40°	76.6	9.8	116	4.0	(114°; 29°)
$T_{90} < 5$ s		45°	95.7	10.4	134	3.7	(118°; 30°)
Short	746	35°	67.3	9.1	104	4.0	(120°; 17°)
Subtype 2c		40°	87.3	10.7	131	4.1	(114°; 29°)
$T_{90} < 8$ s		45°	109.0	11.8	151	3.6	(112°; 27°)
Intermediate	249	35°	22.4	5.4	44	4.0	(109°; 27°)
Type 3		40°	28.9	6.2	53	3.9	(106°; 31°)
$2 \text{ s} < T_{90} < 8$ s		45°	36.0	7.4	62	3.5	(97°; 24°)
Long	1291	35°	117.8	10.1	151	3.3	(142°; –1°)
Type 1		40°	152.3	11.7	194	3.6	(147°; 7°)
$T_{90} > 8$ s		45°	190.4	12.0	227	3.1	(147°; 1°)

3. DISCUSSION AND CONCLUSIONS

The anisotropy of short GRBs with $T_{90} < 2$ s was reported by Balazs *et al.* (1998a, 1998b). Subsequently, it was reported that intermediate GRBs with $2 \text{ s} < T_{90} < 10$ s (Meszaros *et al.* 1999, 2000a, 2000b) exhibited deviations from isotropy. Analyzing the sky distribution of various GRB populations, we have reached the following conclusions:

(1) If there are three types of GRB, then intermediate bursts with $2 \text{ s} < T_{90} < 8$ s exhibit the largest deviation from isotropy at a statistical significance of 99.89%, while short bursts with $T_{90} < 2$ s and long bursts with $T_{90} > 8$ s exhibit no significant deviations from isotropy.

(2) If there are only two types of GRB, then short GRBs with $T_{90} < 8$ s exhibit the largest deviation from isotropy at a statistical significance of 99.99%.

In addition, our technique allows the localization of anisotropy in the sky approximately at coordinates $l_{sp} = 115°$ and $b_{sp} = 30°$; this enables us to search for other effects in this direction. We suggest that the huge accumulation of matter (Superattractor) in the direction $l_{sa} = 110°$, $b_{sa} = 5°$ with $z = 1$–3, which lenses the emission from quasars (Litvin *et al.* 1999a, 1999b), can be responsible for this anisotropy. Lauer and Postman (1994), Slechta and Meszaros (1997), Meszaros and Vanysek (1997), Coles (1998), and Sylos-Labini *et al.* (1998) discuss the possibility of deviations of the matter distribution in the Universe from uniformity and

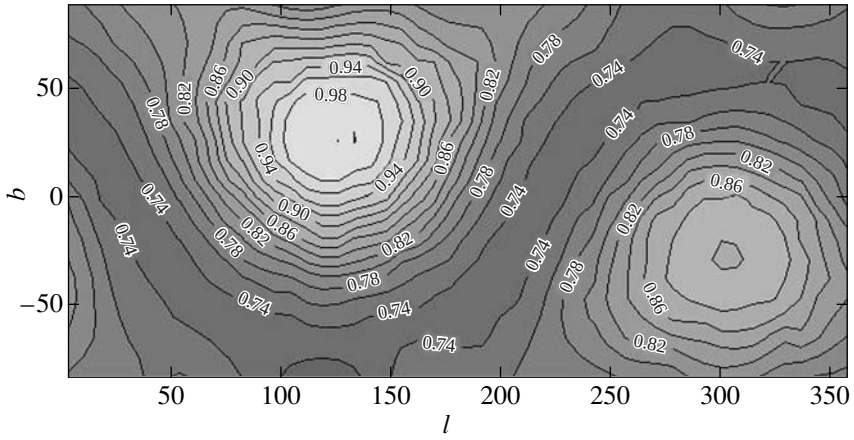


Fig. 5. A topographic map of the distribution of normalized exposure function f_e in the (l, b) plane. Higher values of f_e correspond to lighter shades.

Table 2. Simulations of the GRB populations under study and the probabilities of a chance formation of observable spots with enhanced concentration with allowance for the nonuniform exposure

Population	N_{pop}	R	q	P, %
Short		35°	411	4.11
Subtype 2a	497	40°	1517	15.17
		45°	1235	12.35
Short		35°	6	0.06
Subtype 2b	653	40°	3	0.03
		45°	14	0.14
Short		35°	9	0.09
Subtype 2c	746	40°	1	0.01
		45°	26	0.26
Intermediate		35°	18	0.18
Type 3	249	40°	11	0.11
		45°	60	0.60
Long		35°	368	3.68
Type 1	1291	40°	40	0.40
		45°	586	5.86

isotropy on scales larger than 300 Mpc, which is in agreement with our results.

Note that the spot with a maximum density of short GRBs roughly coincides not only with the direction of the huge accumulation of matter, but also with the direction of the maximum of the exposure function in the region of the north celestial pole (Fig. 5). However, our Monte Carlo simulations confirm the hypothesis that the anisotropy in the sky distribution of short GRBs has a natural origin. We assume that the exposure function does not depend on GRB duration

ACKNOWLEDGMENTS

We are grateful to the referees for critical comments, which helped improve the article.

REFERENCES

1. L. G. Balazs, A. Meszaros, and I. Horvath, Astron. Astrophys. **339**, 1 (1998).
2. L. G. Balazs, A. Meszaros, I. Horvath, and R. Vavrek, Astron. Astrophys., Suppl. Ser. **138**, 417 (1999).
3. B. M. Belli, Astrophys. Space Sci. **231**, 43 (1995).
4. I. V. Belousova, A. Mizaki, T. M. Roganova, and I. L. Rozental', Astron. Zh. **76**, 832 (1999) [Astron. Rep. **43**, 734 (1999)].
5. M. S. Briggs, W. S. Paciesas, G. F. Pendleton, *et al.*, Astrophys. J. **459**, 40 (1996).
6. P. Coles, Nature **391**, L120 (1998).
7. J. P. Dezalay, J. P. Lestrade, C. Barat, *et al.*, Astrophys. J. Lett. **471**, L27 (1996).
8. G. J. Fishman, C. A. Meegan, R. B. Wilson, *et al.*, Astrophys. J., Suppl. Ser. **92**, 229 (1994).
9. J. Hakkila, C. A. Meegan, G. N. Pendleton, *et al.*, astro-ph/9712091 (1997).
10. J. Hakkila, D. J. Haglin, G. N. Pendleton, *et al.*, Astrophys. J. **538**, 165 (2000a).
11. J. Hakkila, D. J. Haglin, R. J. Roiger, *et al.*, astro-ph/0001335 (2000b).
12. J. Hakkila, R. D. Preece, G. N. Pendleton, astro-ph/0001340 (2000c).
13. J. Hakkila, R. J. Roiger, D. J. Haglin, *et al.*, astro-ph/0011583 (2000d).
14. I. Horvath, Astrophys. J. **508**, 757 (1998).
15. T. M. Koshut, W. S. Paciesas, C. Kouveliotou, *et al.*, Astrophys. J. **463**, 570 (1996).
16. C. Kouveliotou, C. A. Meegan, G. J. Fishman, *et al.*, Astrophys. J. Lett. **413**, L101 (1993).
17. T. R. Lauer and M. Postman, Astrophys. J. **425**, 418 (1994).
18. V. F. Litvin, V. V. Orlov, S. A. Matveef, and W. E. Pereira, Astrophys. Space Sci. **266**, 507 (1999a).
19. V. F. Litvin, S. A. Matveev, V. É. Pereĭra, and V. V. Orlov, Vestn. S-Peterb. Univ. **2** (8), 120 (1999b).
20. E. P. Mazets, S. V. Golenetskiĭ, V. N. Il'inskiĭ, *et al.*, Astrophys. Space Sci. **80**, 1 (1981).
21. E. P. Mazets, S. V. Golenetskiĭ, Yu. A. Guryan, and V. N. Ilyinskiĭ, Astrophys. Space Sci. **84**, 173 (1982).
22. C. A. Meegan, G. N. Pendleton, M. S. Briggs, *et al.*, *Current BATSE Gamma-Ray Bursts Catalog*, www.batse.msfc.nasa.gov/grb/catalog/current/ (2000).
23. A. Meszaros, Z. Bagoly, I. Horvath, *et al.*, astro-ph/9912076 (1999).
24. A. Meszaros, Z. Bagoly, I. Horvath, *et al.*, Astrophys. J. **539**, 98 (2000a).
25. A. Meszaros, Z. Bagoly, and R. Vavrek, Astron. Astrophys. **354**, 1 (2000b).
26. A. Meszaros and V. Vanysek, Astron. Astrophys. **319**, 371 (1997).
27. S. Mukherjee, E. D. Feigelson, C. J. Babu, *et al.*, Astrophys. J. **508**, 314 (1998).
28. W. S. Paciesas, C. A. Meegan, G. N. Pendleton, *et al.*, Astrophys. J. Suppl. Ser. **122**, 465 (1999).
29. R. J. Roiger, J. Hakkila, D. J. Haglin, *et al.*, astro-ph/0001381 (2000).
30. M. Slechta and A. Meszaros, Astrophys. Space Sci. **249**, 1 (1997).
31. F. Sylos-Labini, M. Montuori, and L. Pietronero, Phys. Rep. **293**, 61 (1998).
32. M. Tegmark, D. H. Hartmann, M. S. Briggs, *et al.*, Astrophys. J. **466**, 757 (1996a).
33. M. Tegmark, D. H. Hartmann, M. S. Briggs, *et al.*, Astrophys. J. **468**, 214 (1996b).

Translated by G. Rudnitskiĭ

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 421–439. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 495–515.

Formation of the Spiral Structure in SB Galaxies

V. L. Polyachenko and E. V. Polyachenko*

Institute of Astronomy, Russian Academy of Sciences, Pyatnitskaya ul. 48, Moscow, 109017 Russia

Received March 5, 2001

Abstract—We investigate regions near the ends of the principal spiral arms in SB galaxies, where the non-axially symmetric part of the gravitational potential rapidly transforms to its asymptotic quadrupole form. The galactic disk responds to this transformation of the potential by forming nearly circular spirals with an angular extent of the order of $\pi/2$ (quarter-turn spirals). We consider the resonance mechanism for the formation of principal spirals. Expressions are derived for the resonance responses of disks with circular and nearly circular stellar orbits. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *galaxies, intergalactic gas*

1. INTRODUCTION

The ordered system of large-scale spiral arms in galaxies (the so-called grand design) is generally believed to be a single structure with smoothly changing parameters along the spirals; abrupt changes in this structure are admitted only at resonances. In particular, at corotation, where the angular velocities of the spiral wave and the galactic disk are equal, the wave can be reinforced, which is accompanied by its substantial restructuring (Toomre 1981). The ends of spiral arms and the locations of ring structures are usually associated with the Lindblad resonances, at which the spiral wave resonates with free epicyclic stellar oscillations.

Here, we give examples of galaxies in which the regions of abrupt change in spiral structure do not owe their origin to resonances. These regions are clearly associated with the ends of the principal spiral arms, where the gravitational potential of two-arm spirals rapidly transforms to its asymptotic quadrupole form. Below, we show that nearly circular spirals with an azimuthal length of the order of $\pi/2$ emerge under the effect of such a dramatic and rapid change in disk potential. These quarter-turn spirals are analyzed in Section 2. Their presence is demonstrated both in the brightness distribution in the SB galaxy NGC 1365 and in simple calculations in which the linear theory of perturbations is used to determine the response of the galactic disk. We chose the specific galaxy (NGC 1365) as a typical example; quarter-turn spirals are also clearly seen in many other SB galaxies (NGC 157, 1300, etc.) and in many normal SA galaxies (for example, in NGC 3631).

We believe the quarter-turn spirals to be part of a normal mode of an SB galaxy, which also includes a bar and principal spiral arms. The quarter-turn spirals provide a means of lengthening the principal spiral arms. As for the latter, they can have resonance and nonresonance natures. The latter include, for example, the principal spiral arms of the galaxy NGC 1365 considered in Section 2. These arms are open wide, which primarily distinguishes them from tightly wound, nearly circular resonance spirals. In Section 3, where the principal spiral arms of SB galaxies are investigated, we restrict ourselves to describing resonance spirals. They are easiest to describe. At the same time, the resonance responses are highly varied in form, and many of them are very similar to the structures observed in SB galaxies. Here, we consider those cases where the principal spiral arms are mainly the disk response to the bar potential; the self-gravitation of the spiral arms themselves can be easily taken into account by using a rapidly converging iterative procedure.

Note that to prove the existence of quarter-turn spirals in SB galaxies, it essentially does not matter how the bar was formed and to which type it belongs, fast or slow [for the classification of bars, see Pasha and Polyachenko (1994)]. However, this turns out to be of importance in studying the resonance responses.

We consider the responses of disks with circular particle orbits in Subsection 3.1 and the responses of disks with finite (though small) stellar velocity dispersions in Subsection 3.2. Most of the calculations are described in the Appendix.

In conclusion (Section 4), we summarize our main results and note some issues that require a further study.

2. QUARTER-TURN SPIRALS

Figure 1a shows a deprojected image of NGC 1365 (one of the best studied SB galaxies); superimposed on the image is the Fourier harmonic of the brightness dis-

* E-mail address for contacts: evgenii@orc.ru

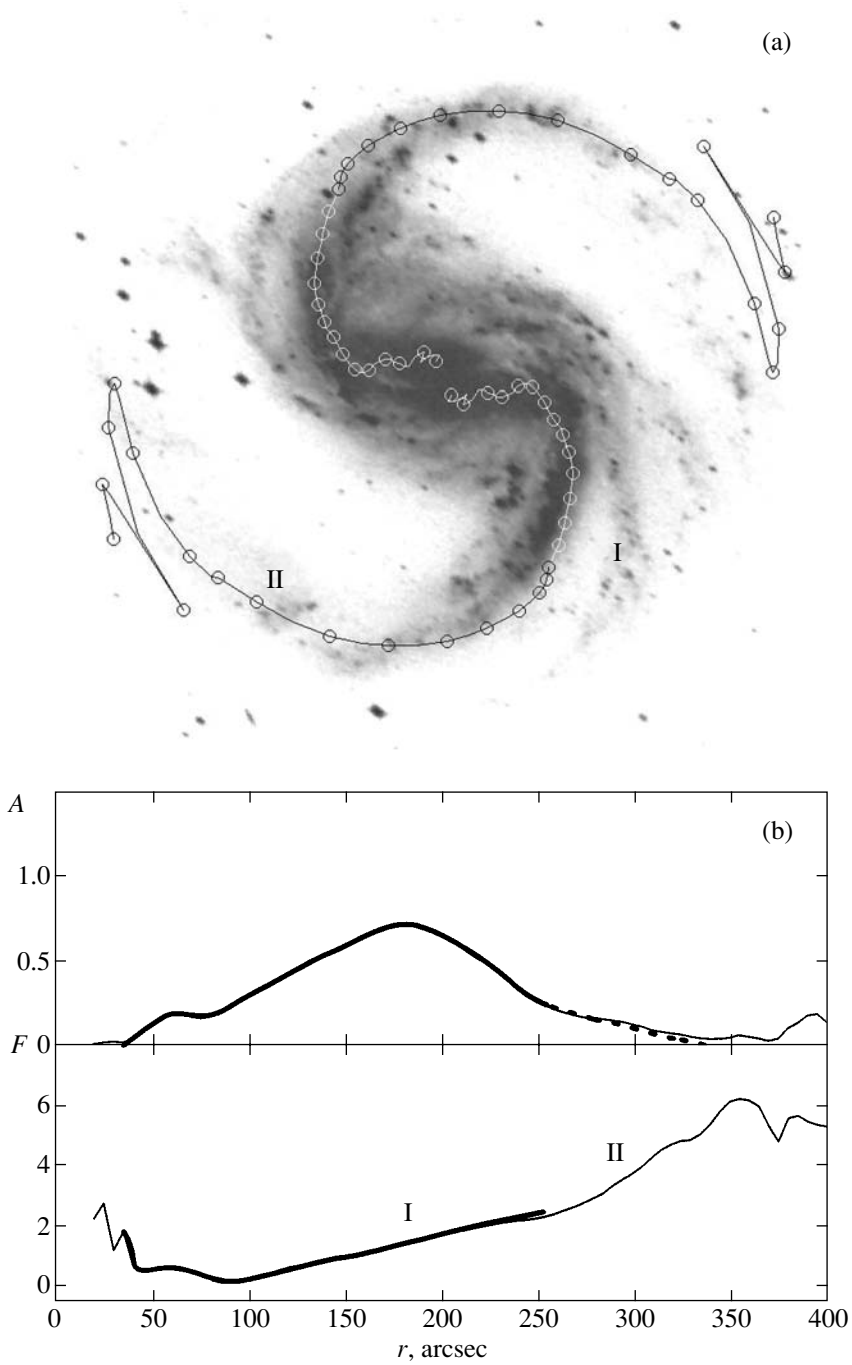


Fig. 1. (a) A deprojected image of NGC 1365 with the $m = 2$ Fourier harmonic of the (B) brightness distribution superimposed on it. The image was retrieved from the NED archive. I, Principal spirals; II, quarter-turn spirals. (b) The amplitude $A(r)$ and phase $F(r)$ of the Fourier harmonic shown in (a). The heavy lines represent the smoothed functions used to calculate the response of a galactic disk (see Fig. 2).

tribution that corresponds to the dominating two-arm symmetry. The surface density σ_2, which is proportional to brightness, can be written in polar coordinates r and φ as

$$\sigma_2(r, \varphi) = \mathrm{Re}[\sigma(r)\exp(2i\varphi)] = A(r)\cos[2\varphi - F(r)],$$

where $\sigma(r) = A(r)\exp[-iF(r)]$ is the complex amplitude, $A(r) > 0$ is the real amplitude, and $F(r)$ is the wave phase. The functions $A(r)$ and $F(r)$ are shown in Fig. 1b [the curves in Fig. 1a are $\varphi(r) = F(r)/2$ and $\varphi(r) = F(r)/2 + \pi$]. We see that the spiral arms consist of two distinct parts: (1) the widely open principal arms branching off from the bar; and (2) the nearly circular quarter-turn spirals adjacent to these arms, each of which ends with radially aligned oscillations made up of short alternating leading and trailing spirals. These

two parts differ greatly both in amplitude and in spiral pitch angle. The same characteristic quarter-turn spirals are available both in other SB galaxies (for example, in NGC 157) and in normal spiral galaxies: for example, these spirals are present in NGC 3631 not only in its outer part (as in NGC 1365), but also in its central region.

In reality, this phenomenon is quite common: the quarter-turn spirals are the response of the galactic disk to the gravitational potential of the principal spiral, which has a characteristic behavior near the end of this spiral (see Fig. 2). This behavior can be described as a transition from the spiral mode to a multipole mode. In the spiral mode, the minima of potential roughly follow the maxima of surface density, i.e., the observed spiral arms. Thus, for example, if the spiral arms are long, tightly wound spirals, so that the phase $F(r) = \int k(r')dr'$ (k is the wave number) changes much faster than the amplitude $A(r)$, then the potential is known (Toomre 1964) to be

$$\Phi_2(r, \varphi) = -[2\pi G/|k(r)|]\sigma_2(r, \varphi),$$

where G is the gravitational constant. Thus, for tightly wound spirals, the maxima of surface density closely coincide with the minima of potential.

Sufficiently far from the spirals, $\Phi_2(r, \varphi) = \bar{\Phi}\cos(2\varphi - \bar{F})$ must clearly approach the quadrupole potential (Landau and Lifshitz 1988): $\Phi_2(r, \varphi) \longrightarrow r^{-3}\cos 2(\varphi - \varphi_0)$ (φ_0 = const). Consequently, Φ_2 must definitely be nonspiral in form at the asymptotics. In reality, however, Φ_2 assumes a nonspiral (multipole) form well before the quadrupole asymptotics is reached: this takes place in a narrow range of radii immediately behind the principal spiral. In this case, we can write $\Phi_2(r, \varphi) \sim r^{-n}\cos 2(\varphi - \varphi_0)$ ($n = -d\ln\bar{\Phi}/d\ln r$, $n \longrightarrow 3$ at sufficiently large radii).

To describe the behavior of the galactic-disk response to such a potential, we take a disk model with circular orbits. We will see that even this simple model yields results that are in qualitatively good agreement with the observations. Clearly, this approximation is completely justified in those cases where the velocity perturbations produced by a nonaxially symmetric potential are larger than the thermal velocities in the subsystem of stars under consideration.

We restrict ourselves to a linear analysis by assuming that all the perturbed quantities are proportional to $\exp(-i\omega t)$, where t is the time and $\omega = 2\Omega_b + i\gamma$ is the complex frequency composed of the bar angular velocity Ω_b and the growth rate γ. Linearizing the Euler hydrodynamic equations with a zero pressure, we can express the perturbed velocities v_r and v_φ in terms of the gravitational potential Φ. Substituting these expressions for v_r and v_φ in the linearized continuity equation, we derive the relation between the perturbations of surface density $\sigma(r)$ and potential $\Phi(r)$

$$\sigma(r) = -\frac{1}{r}\frac{d}{dr}\left(r\varepsilon\frac{d\Phi}{dr}\right) + \frac{4}{r^2}\varepsilon\Phi + \frac{4}{r\omega_*}\Phi\frac{d}{dr}(\varepsilon\Omega). \quad (1)$$

Here, $\varepsilon(r) = \sigma_0(r)/(\omega_*^2 - \kappa^2)$ is a gravitational analog of the permittivity, $\sigma_0(r)$ is the unperturbed density, $\kappa(r)$ is the epicyclic frequency, $\kappa^2 = 4\Omega^2 + rd\Omega^2/dr$, $\Omega(r)$ is the local angular velocity of the galactic disk, and $\omega_*(r) = \omega - 2\Omega(r)$. In the above remarkably compact form, this formula was first given by Bisnovatyĭ-Kogan and Mikhaĭlovskiĭ (1973). The real quantity $\mathrm{Re}[\sigma(r)\exp(2i\varphi)]$ is the density response to potential Φ.

Because of the rapid radial change in potential both on the principal spiral and during the transition from the spiral to multipole mode, we usually have $\sigma(r) \cong -\varepsilon d^2\Phi/dr^2$ instead of Eq. (1). In general (and this is true for the galaxy NGC 1365 considered here), the spiral ends lie in the region where $\omega_*^2 < \kappa^2$, i.e., between Lindblad resonances (Lin *et al.* 1969). Therefore, $\varepsilon < 0$, and we conclude that the surface density $\sigma(r)$ must change in the same way as the second derivative of

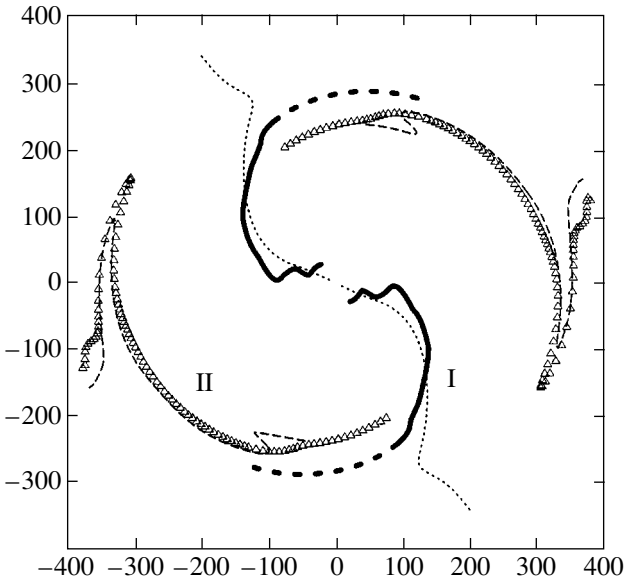


Fig. 2. Positions of the maxima of the density response at each radius (triangles, II) to the gravitational potential of principal spirals (heavy solid lines, I for the galactic disk of NGC 1365). The curves for the positions of response maxima and the quarter-turn spirals in Fig. 1a are seen to be very similar. The dotted lines are used for the curves of the positions of minima of two-arm potential $\Phi_2(r, \varphi)$. The dashed lines indicate the positions of maxima of $\mathrm{Re}[\Phi''\exp(2i\varphi)]$. Qualitatively, the form of quarter-turn spirals does not depend on specific parameters of the disk and the spiral wave. In particular, it is insensitive to the radiation band used (although in other bands, the amplitude of the two-arm harmonic can decrease in the central region more slowly or can even slightly increase).

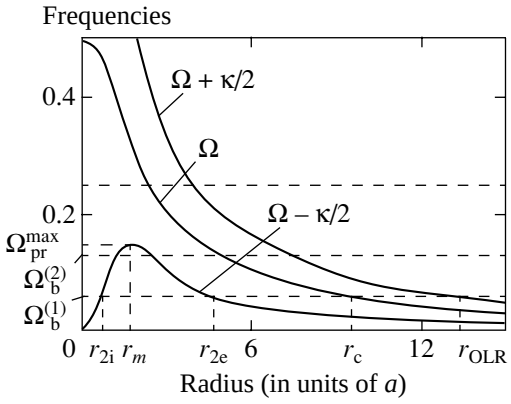


Fig. 3. Principal characteristic frequencies of stellar orbits: $\Omega(r)$ is the angular velocity; $\Omega - \kappa/2 = \Omega_{pr}$ is the precession velocity of nearly circular orbits; r_c is the corotation; r_{2i} and r_{2e} are the inner Lindblad resonances; r_{OLR} is the outer Lindblad resonance; $\Omega_b^{(1)} < \Omega_{pr}^{max}$ is the angular velocity of a slow bar; and $\Omega_b^{(2)} \simeq \Omega_{pr}^{max}$ is the bar velocity in a situation with two close inner Lindblad resonances.

potential, $\sigma(r) \propto \Phi''$. This is precisely observed in our example, as we see from Fig. 2, which shows the positions of the maxima of the response (at each radius) to the potential of the principal spirals in the galaxy under study calculated by using formula (1); the effect of the bar in the region of quarter-turn spirals of interest proves to be negligible. Note that for the spiral behavior of the potential, $\Phi'' \sim -k^2\Phi$, while for the multipole mode, $\Phi'' \sim +n(n+1)\Phi$. We thus see that, if we write $\sigma(r) = B(r)\Phi(r)$, then the phase of the complex function $B(r)$ changes from $-\pi$ at the spiral beginning of the response under consideration to zero at its multipole end. This precisely corresponds to an azimuthal length of the two-arm response equal to $\pi/2$ (because the change in phase ΔF and the turn angle of the two-arm spiral $\Delta\varphi$ are related by $\Delta F = 2\Delta\varphi$; in our case, $\Delta F = \pi$).

3. THE RESONANCE FORMATION OF PRINCIPAL SPIRAL ARMS IN SB GALAXIES

In this section, we discuss the following scenario for the formation of spiral structure in SB galaxies. First, a fast or slow bar is formed at the galactic center. This may result, for example, from the growth of a particular instability. The bar induces primary (resonance) arms, which branch off from the bar ends. In the simplest case considered below, the primary spirals emerge at corotation (CR) for a fast bar and at inner Lindblad resonance (ILR) for a slow bar. Finally, if the primary spirals are sufficiently strong, they trigger the formation of secondary spirals. Nearly circular quarter-turn spirals that are extensions of the primary spirals (in principle, this spiral lengthening can be repeated). However, we restrict our analysis here to the simplest case where the potential of self-gravitating principal spirals is considerably smaller than the bar potential. In that case, quarter-turn spirals are clearly not formed, while the principal spirals are mainly the disk response to the bar potential; allowing for the self-gravitation of principal spirals does not qualitatively change the results.

3.1. The Responses of Cold Disks

To describe the response of a cold galactic disk that produces resonance spirals, we can use relation (1) between surface density and potential as a first approximation. In this case, the potential is generally $\Phi = \Phi_b + \Phi_s$, where $\Phi_b(r, \varphi) \simeq r^{-n}\cos 2\varphi$ is the bar potential ($n > 0$, the bar is assumed to be oriented vertically) and Φ_s is the potential of the spirals themselves.

The case where $|\Phi_s| < |\Phi_b|$ is simplest. When calculating the response using Eq. (1), we can then assume that $\Phi \simeq \Phi_b$ as a first approximation. The surface density $\sigma^{(1)}$ derived in this way can then be substituted in the integral for the potential of a simple layer, whose calculation gives the potential $\Phi_s^{(1)}$ of the self-gravitating spirals in the first approximation. Subsequently, the surface density $\sigma^{(2)}$ is calculated in the second approximation using Eq. (1) with $\Phi = \Phi_b + \Phi_s^{(1)}$ and so on. Thus, a natural iterative process arises; below, we restrict our analysis to those cases where these iterations converge. At the end of this section, we deduce the corresponding approximate conditions and give a typical example of the rapidly converging iterative procedure. First, we will consider the disk responses to the potential of a single bar by completely ignoring the self-gravitation of spirals.

The possible resonance responses of the disk are highly varied. Some of them were previously given by Pasha and Polyachenko (1993, 1994) and Polyachenko (1994). Below, we therefore restrict ourselves to several remarks pertaining to spiral responses.

(1) Near corotation [$\omega_* \simeq 0$, i.e., $\Omega_b = \Omega(r_c)$, $\Omega_b = \omega/2$], we derive from Eq. (1)[1]

$$\sigma(r, \varphi) \simeq \frac{2r^{-n-1}}{\sqrt{\Omega_c'^2(r-r_c)^2 + \gamma^2/4}}\frac{d}{dr}\left(\frac{\Omega\sigma_0}{\kappa^2}\right) \times \cos[2(\varphi - \varphi_c(r))], \quad (2)$$

$$\tan 2\varphi_c(r) = -\frac{\gamma}{2\Omega_c'(r-r_c)}, \quad (3)$$

where $\Omega_c' = d\Omega/dr|_{r_c}$. These formulas describe a two-arm trailing (because $\Omega_c' < 0$; see Fig. 3) spiral. Typical responses are schematically shown in Fig. 4a. In this figure (and in Figs. 4b–8b, 10), the density of points is

[1] Note that Eq. (12) from Polyachenko (1994) holds only for $|\varphi_c(r)| \ll 1$ [in contrast to Eq. (3) in this paper, which is valid for any $\varphi_c(r)$].

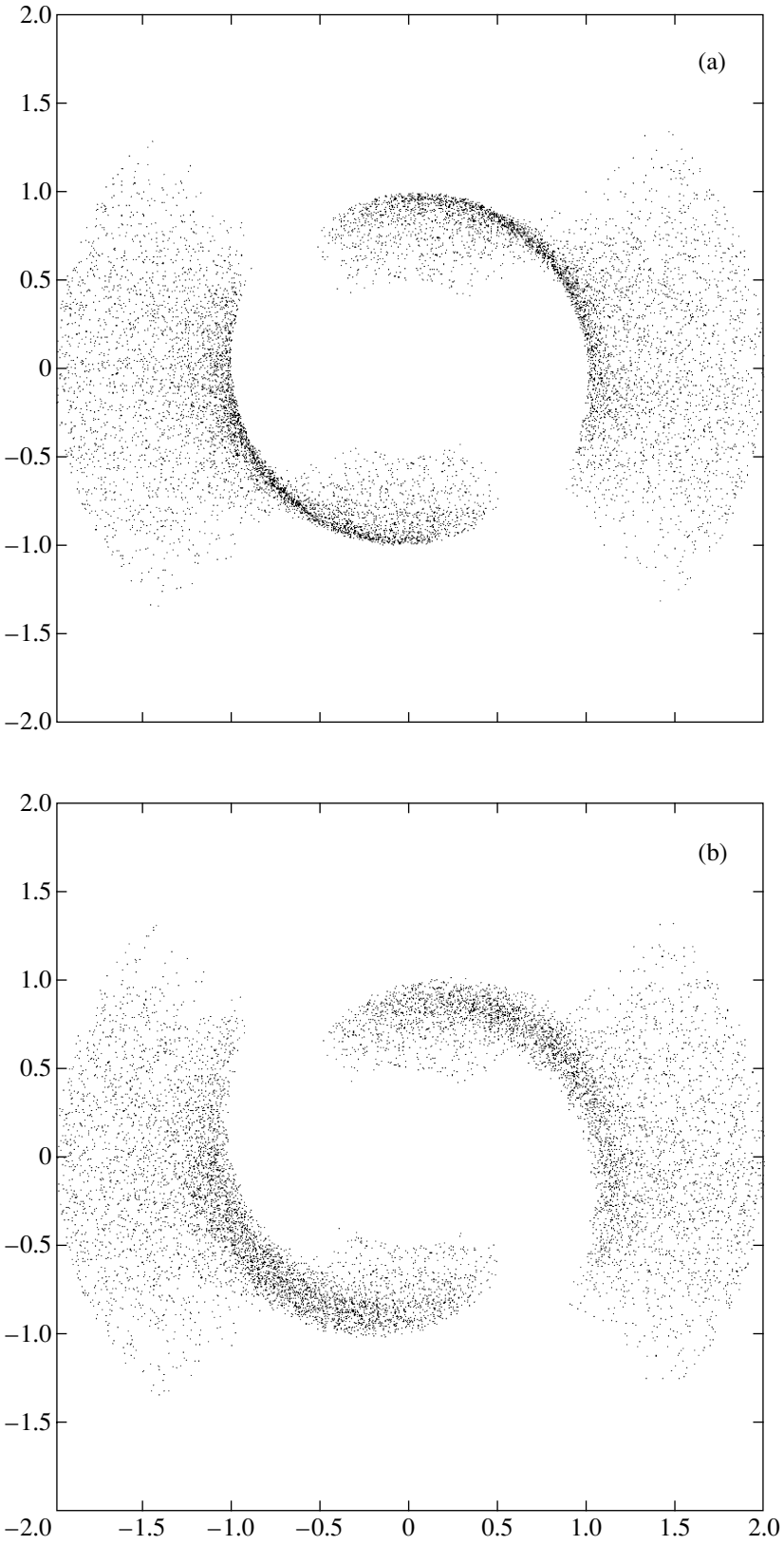


Fig. 4. Resonance response of the disk at corotation (r_c = 1) for γ = 0.05: (a) a cold disk, ρ = 0; (b) a disk with finite stellar velocity dispersion, ρ = 0.1.

proportional to the perturbed surface density σ in the regions where $\sigma > 0$. The maximum length of each arm is obviously $\pi/2$; the arms are longest at low γ. Figure 4a corresponds to the case where $|\sigma_0'/\sigma_0| \gg |\Omega'/\Omega|$ (which apparently always holds). For the inverse inequality, we would have an unobservable arrangement of spirals relative to the bar, which is obtained from that shown in the figure by counterclockwise rotation through $\pi/2$. Intermediate cases, $|\sigma_0'/\sigma_0| > |\Omega'/\Omega|$, describe the realistic pattern of spirals, which slightly go behind the bar and branch off from it by an angle smaller than $\pi/2$.

Note that the resonance responses at corotation are similar. The diversity of resonance responses mentioned above refers to inner Lindblad resonances; the next remark is devoted to them.

(2) For realistic rotation curves (Fig. 3), there is always only one corotation and one outer Lindblad resonance (OLR), at which $\omega_*(r_{\rm OLR}) = \kappa(r_{\rm OLR})$; i.e., $\Omega_b = \Omega(r_{\rm OLR}) + \kappa(r_{\rm OLR})/2$. The inner Lindblad resonances (ILR), where $\omega_*(r_{\rm ILR}) = -\kappa(r_{\rm ILR})$, correspond in Fig. 3 to the intersections of the $\Omega = \Omega_b$ straight line with the curve of precession angular velocities for nearly circular orbits, $\Omega_{\rm pr} = \Omega(r) - \kappa(r)/2$. As we see from the figure, there are several possibilities: (i) if $\Omega_b > \Omega_{\rm pr}^{\max}$, then there are no inner Lindblad resonances at all; (ii) if $\Omega_b < \Omega_{\rm pr}^{\max}$, then two resonances emerge: internal ILR(r_{2i}) and external ILR(r_{2e}); (iii) if Ω_b is only slightly larger than $\Omega_{\rm pr}^{\max}$, then the $\Omega = \Omega_b$ straight line passes near $r = r_m$ closely above the maximum of the $\Omega = \Omega_{\rm pr}(r)$ curve (although it does not intersect the latter); in this case, the disk can strongly respond to the bar potential (almost resonance).

Near one of the ILRs ($r \simeq r_{2i}$ or $r \simeq r_{2e}$), Eq. (1) gives

$$\frac{\sigma(r)}{\sigma_0(r)} \simeq (n_{\rm cr} - n) r\Omega_{\rm pr}' \frac{r^{-n-2}}{\kappa[4(\Omega_b - \Omega_{\rm pr}(r))^2 + \gamma^2]} \times \cos[2(\varphi - \varphi_2(r))], \quad (4)$$

$$\tan\varphi_2(r) = -\frac{\gamma}{2\Omega_{\rm pr}'(r - r_c)}, \quad (5)$$

where $\Omega_{\rm pr}' = d\Omega_{\rm pr}/dr|_{r_c}$ and $n_{\rm cr} = 4\Omega/\kappa|_{r_c}$ (for a flat rotation curve, $n_{\rm cr} = 2\sqrt{2} \simeq 2.82$). It follows from these formulas that at $r \simeq r_{2e}$ and the resonance responses have the form of trailing two-arm spirals, which are schematically shown in Figs. 5a and 6a [for positive and negative ($n_{\rm cr} - n$), respectively]. The maximum length of each spiral arm is π. It is reached at low γ, when the spirals themselves are nearly circular, while the pair of spirals forms a ring-like configuration. At $r \simeq r_{2i}$, the resonance responses (Figs. 7a and 8a) are similar to those just considered but with two important distinctions. The main distinction is that the spiral arms are now leading (because the derivative $\Omega'_{\rm pr} > 0$ at $r = r_{2i}$; see Fig. 3). Thus, if the bar is inside the internal ILR, the resonance spiral structure begins with the leading spiral. The second distinction (the reason is the same) is that for $r \simeq r_{2i}$, the spirals branch off from bar ends at $n > n_{\rm cr}$ (while at $n < n_{\rm cr}$, they fall on the bar ends).

Yet another important conclusion can be drawn from our analysis: the resonance spirals induced by fast bars must probably be approximately a factor of 2 shorter than the same spirals produced by slow bars. It should be emphasized, however, that here we deal with slow bars with widely separated resonances ($\Omega_b^{(1)}$ in Fig. 3). For close resonances ($\Omega_b^{(2)}$ in Fig. 3) or for the almost resonance, the response can have a completely different form (Pasha and Polyachenko 1993, 1994; Polyachenko 1994).

Figures 9 and 10 illustrate the iterative procedure described above by using the corotation resonance as an example. It refers to a disk model with an unperturbed surface density, which falls off as $\exp(-r/r_d)$, $r_d = 1$, at $r > r_c = 6$ and as $\exp(-r/r_{d1})$, $r_{d1} = r_d/3$, at $r < 5$ (the latter exponent describes the decrease in the number of stars in circular orbits inside the bar); the two exponents are joined by a spline. The rotation is assumed to be rigid, $\Omega = \Omega_0 = 0.1$, at $r = 2$; the rotation curve becomes flat at $r = 8$, while at intermediate points, it is complemented by a spline. The mass distribution both in the disk and in the corresponding passive halo provides this rotation law. Figures 9a–9c show the potentials, while Figs. 10a–10c show the spiral responses at three successive iterations. In our example, the iterations are seen to rapidly converge.

Let us now deduce approximate conditions for the convergence of iterations by estimating the spiral potential that corresponds to the response (1) near each of the possible resonances. We will do this separately for CR and ILR.

(1) According to Eq. (2), the amplitude of the resonance response at corotation is

$$\sigma \simeq \frac{2\Phi_b\sigma_0}{\gamma\Omega_c r_c r_d}, \quad (6)$$

where we assume that $\kappa^2 \simeq 2\Omega^2$ (flat rotation curve) and $|\sigma_0'/\sigma_0| \gg |\Omega'/\Omega|$ and introduce the designations r_d for the scale of change in disk surface density ($|\sigma_0'/\sigma_0| = 1/r_d$) and $\Phi_b = r^{-n}$ for the radial part of the dependence of bar potential. The amplitude of the potential that corresponds to the density perturbation under consideration can be estimated from the formula for the WKB approximation

$$\Phi_s = -\frac{2\pi G\sigma}{|k|}, \quad (7)$$

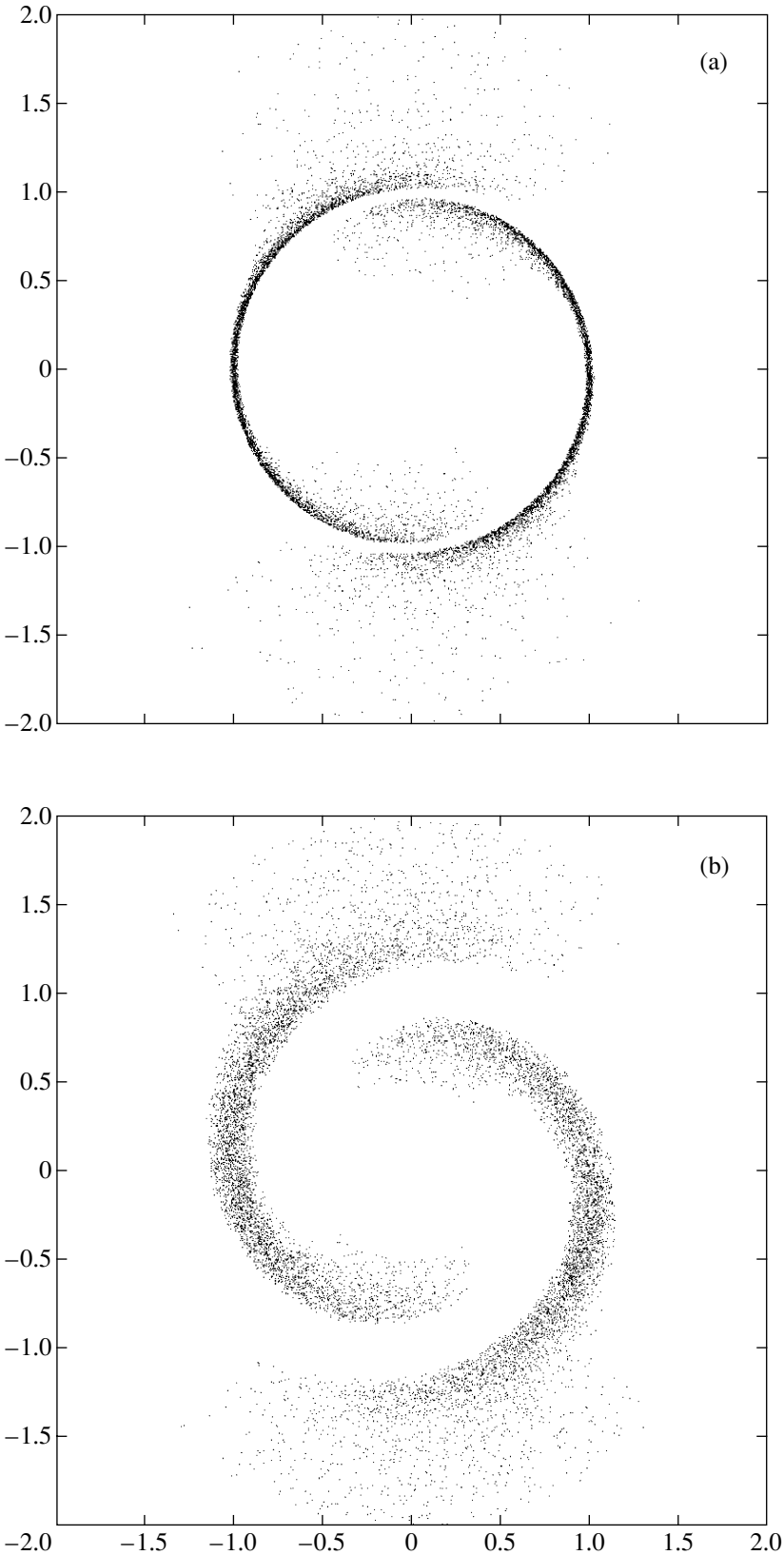


Fig. 5. Resonance response of the disk at external ILR ($r_c = r_{2e} = 1$) for $\gamma = 0.05$ and $n < n_{cr} = 2.82$ ($n = 2$): (a) a cold disk, $\rho = 0$; (b) a disk with finite stellar velocity dispersion, $\rho = 0.1$.

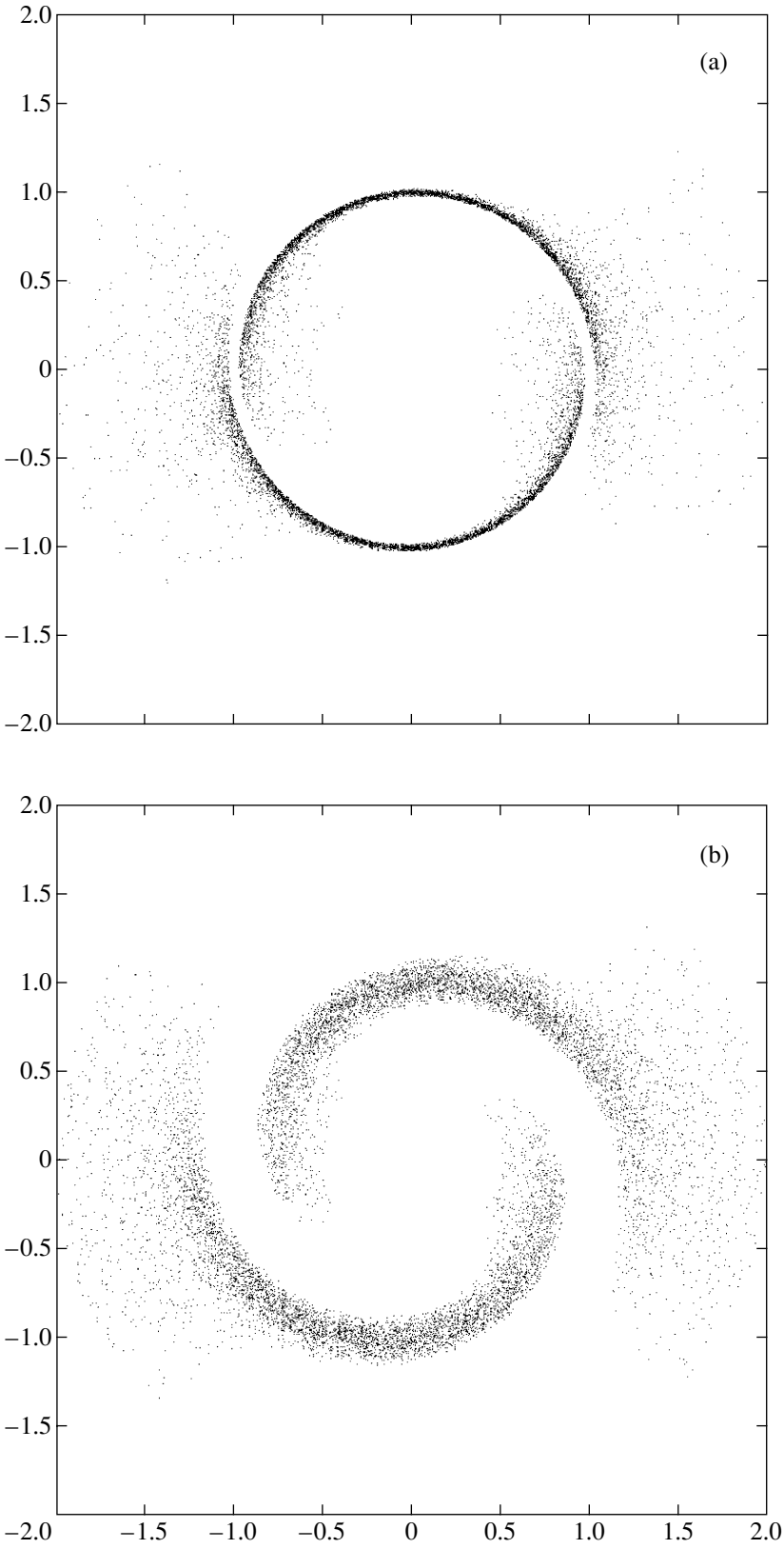


Fig. 6. Same as Fig. 5 for $n > n_{cr} = 2.82$ ($n = 3$).

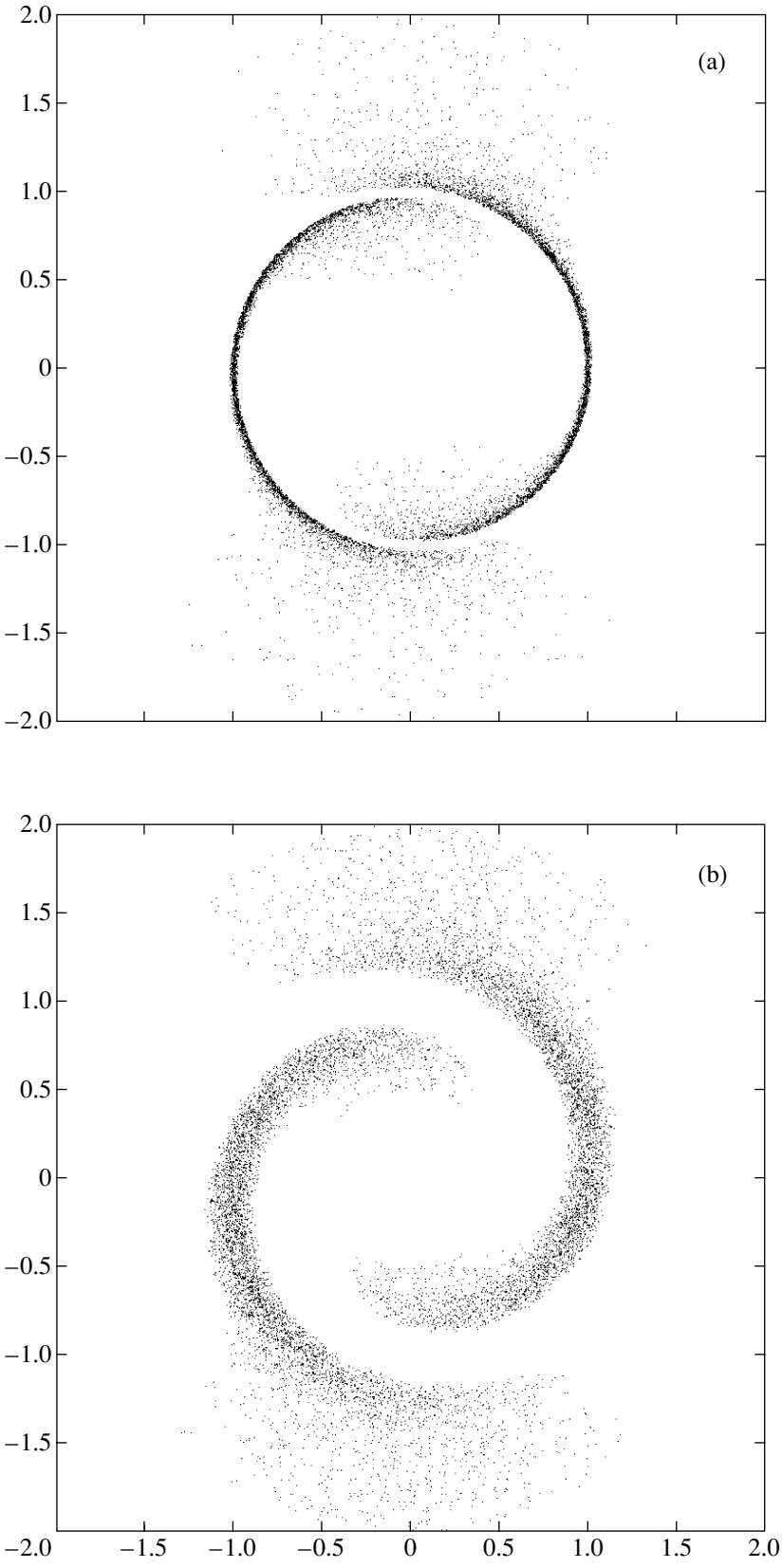


Fig. 7. Same as Fig. 6 at internal ILR ($r_c = r_{2i} = 1$): (a) a cold disk, $\rho = 0$; (b) a disk with finite stellar velocity dispersion, $\rho = 0.1$.

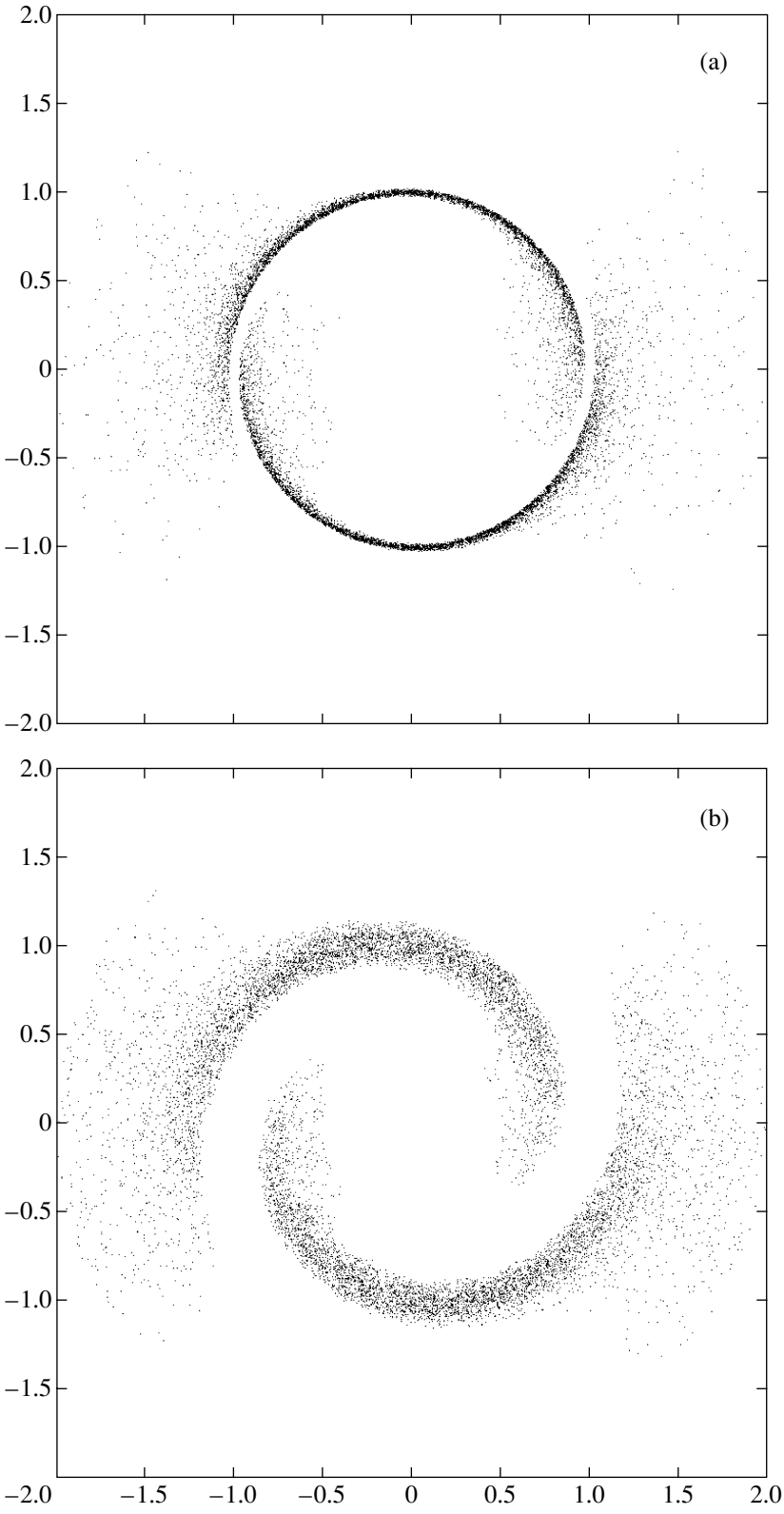


Fig. 8. Same as Fig. 5 at internal ILR ($r_c = r_{2i} = 1$), $\gamma = 0.05$ and $n < n_{cr} = 2.82$ ($n = 2$): (a) a cold disk, $\rho = 0$; (b) a disk with finite stellar velocity dispersion, $\rho = 0.1$.

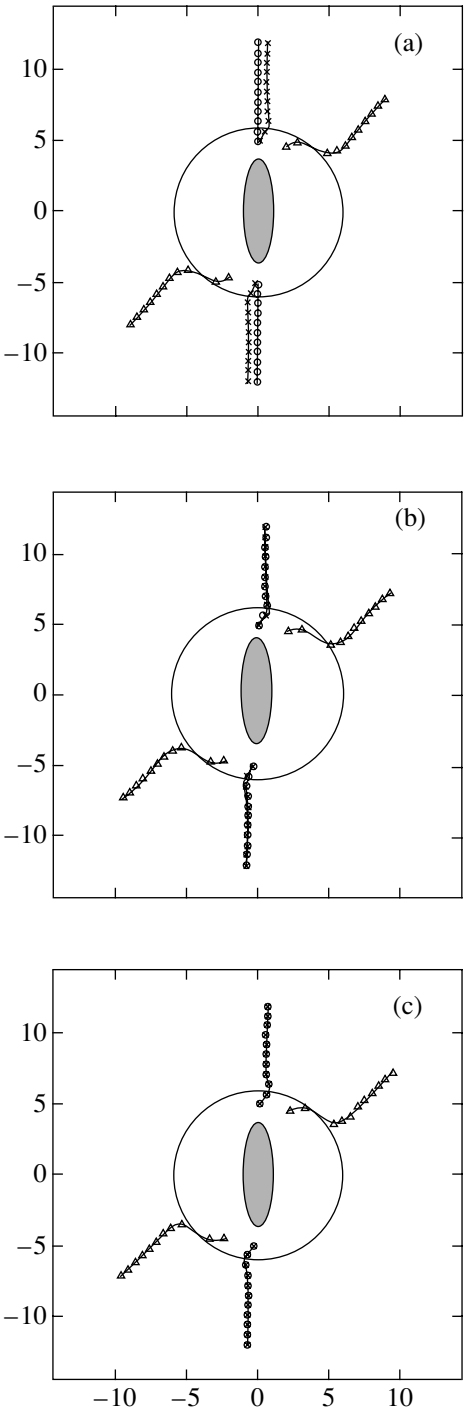


Fig. 9. Curves for the positions of minima of the gravitational potential near corotation (r_c = 6): (a) first iteration; (b) second iteration; and (c) third iteration. The circles mark the positions of minima of the total potential from the preceding iteration (for the first iteration, it degenerates into the vertical straight line corresponding to the positions of minima of the potential for a vertical bar); the triangles mark the positions of minima of the potential for a spiral response; the crosses mark the positions of minima of the total potential after a given iteration. The curves in (b) and (c) essentially coincide, suggesting the convergence of iterations.

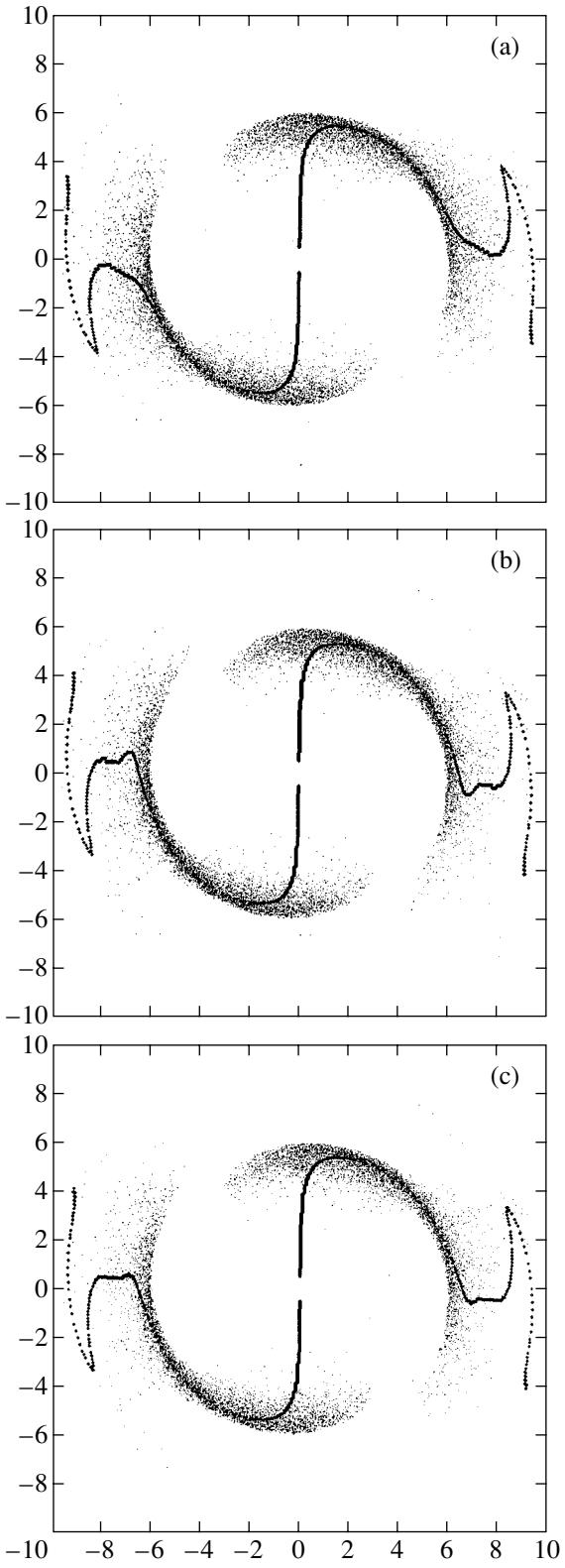


Fig. 10. Disk response near corotation (r_c = 6) after (a) the first, (b) second, and (c) third iterations. The curves indicate the positions of maxima of the density responses at each radius. The patterns of responses essentially cease to differ starting from the second iteration; i.e., in this case, the iterations rapidly converge.

where $k = k(r) = 2\varphi_c'(r)$ is the wave number. Note that direct computations of the potential corresponding to the resonance response σ show that formula (7) underestimates the true potential (by a factor of about 2), but the form of the function $\Phi(r)$ itself is reproduced correctly. Therefore, the inaccuracy that emerges here can be corrected by introducing a constant correction factor in the final expression for the sought-for criterion.

According to Eq. (3),

$$k(r) = -\frac{\gamma\cos^2 2\varphi_c(r)}{2\Omega_c'(r-r_c)^2} = -\frac{2\Omega_c'}{\gamma}, \quad (8)$$

because it follows from the same formula that

$$\cos 2\varphi_c(r) \simeq -\frac{2|\Omega_c'|(r-r_c)}{\gamma}. \quad (9)$$

Hence,

$$|\Phi_s| \simeq \frac{2\pi G\sigma_0|\Phi_b|}{\Omega_c^2 r_d}, \quad (10)$$

so the condition $|\Phi_s| < |\Phi_b|$ means in this case that

$$k_c\frac{2\pi G\sigma_0}{\Omega_c^2 r_d} < 1, \quad (11)$$

where k_c is the constant correction factor mentioned above. Remarkably, the deduced condition (11) has a universal form; for example, it does not depend on the growth rate of perturbation γ. On the other hand, the independence from the bar strength is natural here, because a stronger (weaker) bar induces stronger (weaker) spirals.

Note that for the case corresponding to Figs. 9 and 10, the value of $2\pi G\sigma_0/\Omega_c^2 r_d$, which enters into the deduced criterion for the convergency of iterations, is $0.17 \ll 1$.

(3) Near ILR, similar computations yield a completely different condition:

$$\frac{\gamma}{\Omega(r_2)} > k_I\frac{\pi G\sigma_0(4\Omega/\kappa - n)}{2\kappa\Omega r}\bigg|_{r=r_c}, \quad (12)$$

where k_I is the correction factor.

In conclusion, note that the inequalities that are the inverse of (11) and (12) appear to simultaneously be the necessary conditions for the formation of quarter-turn spirals when the principal spiral arms have a resonance nature. Indeed, as was shown in Section 2 (for nonresonance principal spirals, though), quarter-turn spirals must emerge when the gravitational potential of the primary spiral at its location is larger than the bar potential. The total potential (which is the sum of the bar and spiral potentials) then also assumes a spiral form. The potential again becomes multipole closer to the outer end of the spiral, and one might expect quarter-turn spirals to emerge in the transition region. Unfortunately, this cannot be demonstrated in terms of the iterative procedure described above precisely because such iterations diverge in this case. The physical cause of this divergency is clear: in this case, the spiral potential rather than the bar potential (as assumed) is the correct first approximation; additional remarks on this matter can be found in the last section of this paper.

3.2. The Responses of Disks with Finite Stellar Velocity Dispersions

In general, the effect of finite stellar velocity dispersion (even if it is assumed to be low compared to the circular velocity) on the response of a galactic disk is difficult to take into account. Up until now, this problem has been solved [in the analytic theory of Lin and Shu (1966)] only for tightly wound spirals. In this theory, the potential Φ_1, along with all the other perturbations of an equilibrium axially symmetric background (surface density σ_1, distribution function f_1, etc.), are assumed to be proportional to e^{ikr}, where k is the wave number with $kr \gg 1$ and $d\ln k/d\ln r \ll 1$. Thus, strictly speaking, the theory correctly describes the responses of galactic disks under the effect of only multi-turn, tightly wound spirals.

The expression derived by Lin and Shu (1966) for the local (at point r) response of surface density is generally represented as

$$\frac{\sigma_1}{\sigma_0} = -\frac{k^2\Phi_1}{\kappa^2}\frac{1}{1-\nu^2}\mathcal{F}_\nu(x), \quad (13)$$

where $\nu = \omega_*(r)/\kappa(r)$ is the dimensionless frequency of the spiral perturbation, and $\mathcal{F}_\nu$ is the reduction factor,

$$\mathcal{F}_\nu(x) = \frac{1-\nu^2}{x}\left[1 - \frac{\nu\pi}{\sin\nu\pi}\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-x(1+\cos s)}\cos\nu s\,ds\right], \quad (14)$$

$x = k^2c_r^2/\kappa^2 = k^2\rho^2$ (c_r is the radial stellar velocity dispersion, and $\rho = c_r/\kappa$ is the mean epicycle size). $\mathcal{F}_\nu(x)$ is a monotonically decreasing function of x; $\mathcal{F}_\nu(0) = 1$ in the limit of a cold disk ($c_r = 0$). This justifies the name of $\mathcal{F}_\nu(x)$: its value decreases with increasing stellar velocity dispersion, so, according to Eq. (13), it describes the decrease in the response of surface density with increasing dispersion compared to a cold disk.

Note that the spatial form of the perturbation in the theory under consideration is fixed from the very outset [$\propto\cos(kr + m\varphi)$]. In particular, the thickness of the spiral response is specified: $\delta r = \pi/k = \lambda/2$ (λ is the radial wavelength); as the dispersion increases, only the perturbation amplitude decreases.

Recall that the following standard Lin–Shu dispersion equation is derived from the relation between perturbations of the potential and the surface density that follows from the Poisson equation $\Phi_1 = -2\pi G\sigma_1/|k|$

(which is also local in the approximation under consideration):

$$(\omega - m\Omega)^2 = \kappa^2 - 2\pi G\sigma_0|k|\mathcal{F}_\nu(x), \qquad (15)$$

this is one of the basic relations of the modern theory of galactic spiral structure.

The Lin-Shu theory is definitely not applicable to our problem of the disk response to the bar potential: the bar potential can be described as the limiting case that is opposite to tightly wound spirals, which were considered by Lin and Shu. A general solution of the kinetic equation in our problem seems difficult (if at all possible) to find. However, we can easily determine how the resonance denominators $1/\omega_*$, $1/(\omega_*^2 - \kappa^2)$, and $1/(\omega_*^2 - \kappa^2)^2$ in the response of a cold disk (separately for each of the main resonances) are modified under the effect of finite (although low) velocity dispersion.

Here, we give only the final expression for the responses of surface density (for detailed calculations, see the Appendix).

(1) Near corotation (CR),

$$\sigma_1 = A\Phi_1, \qquad (16)$$

$$A = -\frac{2m\Omega}{r\kappa^2}\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega}\right)\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}}{\omega_*(r_0)} + \frac{m\Omega}{r\kappa^2}\frac{\kappa'}{\kappa}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}(3+y^2)}{\omega_*(r_0)}, \qquad (17)$$

where, for definiteness, the equilibrium distribution function was assumed to have the Schwarzschild form in the epicyclic approximation:

$$f_0 = \frac{2\Omega(r)}{\kappa(r)}\frac{\sigma_0(r)}{2\pi c_0^2}\exp\left(-\frac{v_r^2}{2c_r^2} - \frac{\tilde{v}_\varphi^2}{2c_\varphi^2}\right), \qquad (18)$$

$\int dv_r d\tilde{v}_\varphi f_0 = \sigma_0(r)$, $\tilde{v}_\varphi = v_\varphi - r\Omega(r)$ is the residual azimuthal velocity, $c_r = c_0 = \text{const}$ is the radial velocity dispersion [the possible dependence $c_0(r)$ is of no importance to us], $c_\varphi = c_r\kappa(r)/(2\Omega(r))$ is the dispersion of residual azimuthal velocities, $y = 2\Omega\tilde{v}_\varphi/(\kappa c_0)$, $r_0 = r + \rho y$, $\rho = c_0/\kappa$,

$$\omega_*(r_0) \simeq m|\Omega'(r_c)|[(r - r_c) + \rho y + i\delta], \qquad (19)$$

$\delta = \gamma/(m|\Omega'(r_c)|)$, γ is the imaginary part of the frequency: $\omega = m\Omega_p + i\gamma$, and Ω_p is the wave angular velocity. All the functions that appear in (17) as the coefficients before the integrals are assumed to be calculated at the resonance position $r = r_c$.

Figure 4b shows the responses of a simple model of an exponential disk, $\sigma_0(r) = \sigma_0(0)e^{-r/r_d}$ [for $\sigma_0(0) = 1$ and $r_d = 1/2$), inside an isothermal halo with a flat rotation curve, $\Omega = V_0/r$ ($V_0 = 1$), to the potential $\Phi_1 = r^{-n}\cos 2\varphi$ ($m = 2$) of a bar rotating with angular velocity Ω_p such that $r_c = 1$.

A comparison of Figs. 4a and 4b shows that the main effect of the stellar velocity dispersion is an increase in the thickness of the spiral response and in the degree of its openness (the latter takes place at $\rho > \delta$); in this way, the case under consideration differs fundamentally from previously studied responses to a tightly wound spiral.

(2) Near one of the possible Lindblad resonances (ILR or OLR),

$$\sigma_1 = B\Phi_1'' + C\Phi_1' + D\Phi_1, \qquad (20)$$

$$B = -\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\frac{y^2 e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}, \qquad (21)$$

$$C = C^{(1)} + C^{(2)}, \qquad (22)$$

$$C^{(1)} = -\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{2\Omega}\right)\right], \qquad (23)$$

$$C^{(2)} = -\frac{2\Omega}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}(\alpha y^2 + \beta), \qquad (24)$$

$$D = D^{(1)} + D^{(2)}, \qquad (25)$$

$$D^{(1)} = \frac{2m\Omega\omega_*}{\kappa^2 r}\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0} + \frac{3\Omega'}{2\Omega} - \frac{m\Omega'}{\omega_*} - \frac{2\kappa'}{\kappa} - \frac{1}{r}\right)\right], \qquad (26)$$

$$D^{(2)} = \frac{4\Omega^2}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[(\alpha y^2 + \beta)\frac{m\omega_*}{r\kappa^2} + \frac{m^2}{r^2}\right], \qquad (27)$$

where, for brevity, we designated

$$\alpha = \frac{\mu}{6\Omega} + \frac{1}{2}(r\Omega'' + 2\Omega') - \frac{\Omega}{r}, \qquad (28)$$

$$\beta = \frac{\mu}{3\Omega}, \quad \mu = -\frac{1}{2}\left[(\kappa^2)' - \frac{3\kappa^2}{r}\right]. \qquad (29)$$

For the inner Lindblad resonances (ILR), $\omega_*(r_c) = -\kappa(r_c)$ must be assumed in these formulas, i.e., $\Omega_{pr}(r_c) = \Omega_p$, $\Omega_{pr}(r) \equiv \Omega(r) - \kappa(r)/m$ is the orbital precession velocity; the outer Lindblad resonance (OLR) corresponds to $\omega_*(r_c) = \kappa(r_c)$, i.e., $\Omega_{pr}^{(+)}(r_c) = \Omega_p$, $\Omega_{pr}^{(+)}(r) \equiv \Omega(r) + \kappa(r)/m$. The responses of the simple model described above (an exponential disk with a flat rotation curve) near the inner Lindblad resonances are shown in Figs. 5b and 6b for $r \simeq r_{2e}$ and in Figs. 7b and 8b for $r \simeq r_{2i}$. As in the case of a corotation resonance, the main effect of the velocity dispersion here is also an increase in the thickness of the spiral responses and in the degree of their openness (when $\rho > \delta$). The disk response at the outer Lindblad resonance is similar to the response at the external ILR for $n < n_{cr}$ (Fig. 5)

4. DISCUSSION

Above, we considered new aspects of the formation of spiral structure in SB galaxies.

(1) Previously, spiral perturbations have been studied mainly in a self-consistent case: first, the spiral response σ to a spiral potential Φ was determined from the solution of the kinetic equation for stars or the hydrodynamic equations for gas, and, then, the potential itself was assumed to be produced by a density perturbation σ (i.e., the condition of self-consistency was introduced). This statement of the problem arises, for example, in the theory of galactic spiral density waves; in particular, the Lin–Shu dispersion equation (15) for tightly wound spirals is derived in this way. Naturally, the potential and density perturbations are sought in this case in the same region.

We analyzed (Section 2) disk responses outside the region where the principal spiral arms are located; the region immediately adjacent to the principal spirals turned out to be of greatest interest. In this narrow ring, the spiral behavior of the potential changes to its multipole behavior, which gives rise to quarter-turn spirals as a characteristic response of the disk to this behavior of the potential. The phenomenon of quarter-turn spirals provides a natural means for the successive (generally multiple) lengthening of spirals in galaxies.

(2) We also analyzed (Section 3) the resonance generation of principal spiral arms by a bar and suggested an iterative procedure for consistent allowance for the self-gravitation of spirals. In principle, resonance responses can completely cover the entire variety of the spiral and ring shapes observed in SB galaxies. Therefore, the scenario for the formation of structure in SB galaxies, including bar formation, the resonance generation of principal spirals in the galactic disk, and the nonresonance formation of quarter-turn spirals, seems promising. In this case, a spiral with a length of about $\pi/2$ for a fast bar and about π for a slow bar must be formed (as a resonance response of the disk to the bar potential) during the bar formation. If the primary spiral is strong enough to be able to change the multipole behavior of the bar potential to the spiral behavior (which, in turn, again gives way to the multipole behavior closer to the outer spiral end), then the first quarter-turn spirals that lengthen the primary spiral by $\pi/2$ will emerge. In the example of NGC 1365 we considered, this is the end of the process, because the self-gravitation of the quarter-turn spirals themselves turns out to be insignificant. In general, however, this process can be continued.

A survey of photographs for SB(s) galaxies from available atlases (for example, in the well-known Hubble Atlas) shows that, in general, the spiral arms branching off from the bar are clearly traceable over the length of a half-turn about the center. Similarly, the inner rings in SB(r) galaxies usually consist of a pair of semicircumferences (or segments of tightly wound spirals) slightly displaced from one another. Pasha and Polyachenko (1993), who paid attention to these facts, associated them with resonance responses near (external) ILR: these responses appear exactly as was described.

However, the phenomenon of quarter-turn spirals allows the observed pattern of half-turn spiral to also be obtained in principle for fast bars with a resonance response at corotation; recall that the latter has a maximum length of $\pi/2$. Indeed, the first (nearest the bar) half of the observed spiral can have a resonance nature, while the second half is a quarter-turn spiral. This seems to take place, for example, in NGC 1300 (one of the most beautiful SB galaxies). Note that, in this case, observations in various wavelength ranges give distinctly different spiral patterns. Only short primary spirals are clearly seen in the *K* and *I* bands, while in the *R* and *B* bands, the spiral is clearly traceable over the length of a half-turn. Moreover, currently available photographs (see, e.g., Elmegreen *et al.* 1992) show that there is a weak extension of the half-turn spiral by another $\pi/2$ in the *B* band. In this case, we may have an example of two successive quarter-turn spirals that extend the principal spiral (with a length smaller than $\pi/2$) branching off directly from the bar.

Although the scenario for the formation of spiral structure in SB galaxies described in Section 3 in full, including the resonance generation of principal spiral arms and their nonresonance extension in the form of quarter-turn spirals, seems natural, it cannot be justified in terms of the iteration scheme used. Indeed, this scheme is based on the assumption that the bar potential dominates, whereas the formation of quarter-turn spirals, on the contrary, requires that the potential of principal spirals dominate over the bar potential. In that case, an adequate initial approximation is obviously the spiral potential Φ_s rather than the bar potential Φ_b,

which necessitates determining the self-consistent pair's potential density (Φ_s–σ_s) of the spiral. Since, as can be shown, the nonstationary ($\gamma \neq 0$) resonance spiral response σ_r and the original spiral σ_s that generates this response are necessarily displaced from one another in azimuth by a finite angle ($\pi/4$ for corotation), it will require to find new equilibrium ($\gamma = 0$) solutions that include the resonance spiral. This is possible for the corotation resonance r_c if a minimum at $r = r_c$ appears in the density distribution $\sigma_0(r)$ (Kadomtsev 1981; Polyachenko 1991). The case of Lindblad resonances seems more complex so far. It is clear, however, that these problems require a further detailed study.

(3) It should be emphasized that the main general assumption that we made and that is justified by our results is that the entire complex structure of an SB galaxy, which includes the bar and (both principal and quarter-turn) spiral arms, is the structure of some generally nonlinear normal mode, whose main characteristic is the same (for all mode components) angular velocity Ω_p. In other words, we work in general terms of the modal theory. This approach seems most natural for SB galaxies. In particular, it is difficult to suspect the existence of running spiral waves in these galaxies and, accordingly, to find a place for the mechanism of swing amplification (Toomre 1981). The corotation region most likely can act as a generator, because there are anticyclonic (or, occasionally, cyclonic) vortices in this region [see, e.g., Fridman *et al.* (1999) for the latter].

(4) In conclusion, let us consider in more detail the key structure of SB galaxies, the bars themselves.

Fast bars end near corotation; they are easily obtained in numerical simulations with self-gravitating disks, provided that the latter are sufficiently "cold," i.e., the stellar orbits are nearly circular (see, e.g., Hohl 1971; Ostriker and Peebles 1973).

Slow bars (Lynden-Bell 1979; Polyachenko 1989), which must end near one of the inner Lindblad resonances, can be produced by the mutual attraction and capture of slowly precessing stellar orbits. Two extreme cases of this bar formation are possible. In the first case, the bar grows from a very low (fluctuation, thermal) initial level because of instability similar to the instability of eccentric orbits (see, e.g., Polyachenko and Fridman 1976). However, the initial bar potential can be also induced, for example, by an external tidal effect. At a sufficiently large amplitude of the initial bar potential, we generally cannot reason in terms of the growth of linear instability. To describe the processes taking place in this case, the methods that are used, for example, in plasma physics to investigate nonlinear Landau damping (Artsimovich and Sagdeev 1979) are more suitable. On the other hand, it may be noted that using the idea of interaction between precessing stellar orbits (rather than the stars themselves) to describe the formation of a slow bar is appropriate only when a low-mass unstable group of stars with sufficiently (radially) elongated orbits are within a passive "halo", for example, of stars with an isotropic velocity distribution. Only in this case do the orbits themselves become elementary objects of the theory, because we can then perform averaging over the rapid orbital motions of stars and consider only slow orbital turns (Polyachenko 1992). However, the Jeans nature of the instability of radial orbits is beyond question; some complications can arise only with its accurate interpretation when the system cannot be separated into active and passive components in such a way that the bulk of the total mass is concentrated in the latter.

APPENDIX

DERIVING EXPRESSIONS FOR THE RESONANCE RESPONSES OF DISKS WITH LOW STELLAR VELOCITY DISPERSION

The linearized kinetic equation can be written as (see, e.g., Polyachenko and Fridman 1976)

$$\frac{df_1}{dt} = \frac{\partial \Phi_1}{\partial \mathbf{r}} \cdot \frac{\partial f_0}{\partial \mathbf{v}}, \tag{A1}$$

where d/dt is the time derivative with respect to the unperturbed stellar orbit in phase space. We assume that the equilibrium distribution function is

$$f_0(E, L) = F_0(r_0, \mathscr{E}), \tag{A2}$$

where E and L are, respectively, the energy and angular momentum of the star (per unit mass): $E = \frac{1}{2}(v_r^2 + v_\varphi^2) + \Phi_0(r)$, $L = rv_\varphi$, r_0 and $\mathscr{E}$ are the epicyclic integrals, which can be determined as function of E and L from the equations (Shu 1970)

$$r_0^2\Omega(r_0) = L, \quad \mathscr{E} = E - E_c(r_0), \tag{A3}$$

$$E_c(r_0) = \frac{1}{2}r_0^2\Omega^2(r_0) + \Phi_0(r_0),$$
$$r_0\Omega^2(r_0) = \frac{d\Phi_0}{dr_0}. \tag{A4}$$

The residual azimuthal velocity is

$$\tilde{v}_\varphi = \frac{r_0^2\Omega(r_0)}{r} - r\Omega(r). \tag{A5}$$

We assume the deviations from circular motions to be small, i.e., $|\mathscr{E}| \ll |E_c|$, $|r - r_0| \ll r_0$.

Let us transform the right side of (A1). Since

$$\frac{\partial f_0}{\partial v_r} = \frac{\partial f_0}{\partial E}v_r, \quad \frac{\partial f_0}{\partial v_\varphi} = \frac{\partial f_0}{\partial E}v_\varphi + \frac{\partial f_0}{\partial L}r, \tag{A6}$$

we have

$$\frac{\partial \Phi_1}{\partial \mathbf{r}} \cdot \frac{\partial f_0}{\partial \mathbf{v}} = \frac{\partial f_0}{\partial E}\left(v_r\frac{\partial \Phi_1}{\partial r} + v_\varphi\frac{\partial \Phi_1}{r\partial \varphi}\right) + \frac{\partial \Phi_1}{\partial \varphi}\frac{\partial f_0}{\partial L}. \tag{A7}$$

The total time derivative of the potential is

$$\frac{d\Phi_1}{dt} = \frac{\partial\Phi_1}{\partial t} + v_r\frac{\partial\Phi_1}{\partial r} + v_\varphi\frac{\partial\Phi_1}{r\partial\varphi}. \quad \text{(A8)}$$

Finding $v_r\partial\Phi_1/\partial r + v_\varphi\partial\Phi_1/r\partial\varphi$ from this expression and considering the eigenmodes $\propto\exp(-i\omega t + im\varphi)$, we obtain instead of (A7)

$$\frac{\partial\Phi_1}{\partial\mathbf{r}}\cdot\frac{\partial f_0}{\partial\mathbf{v}} = \frac{\partial f_0}{\partial E}\frac{d\Phi_1}{dt} + i\left(\omega\frac{\partial f_0}{\partial E} + m\frac{\partial f_0}{\partial L}\right)\Phi_1. \quad \text{(A9)}$$

Let us now change to new variables: $(E, L) \longrightarrow (\mathscr{E}, r_0)$. Since the derivatives are transformed according to

$$\frac{\partial}{\partial E} = \frac{\partial}{\partial\mathscr{E}}, \quad \frac{\partial}{\partial L} = \frac{2\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial}{\partial r_0} - \Omega(r_0)\frac{\partial}{\partial\mathscr{E}}, \quad \text{(A10)}$$

we finally derive the right side of the linearized kinetic equation:

$$\frac{\partial\Phi_1}{\partial\mathbf{r}}\cdot\frac{\partial f_0}{\partial\mathbf{v}} = \frac{\partial F_0}{\partial\mathscr{E}}\frac{d\Phi_1}{dt} + i\omega_*(r_0)\frac{\partial F_0}{\partial\mathscr{E}}\Phi_1 + \frac{2im\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial F_0}{\partial r_0}\Phi_1. \quad \text{(A11)}$$

The solution $f_1(r, v_r, v_\varphi)\exp(-i\omega t + im\varphi)$ of Eq. (A1) with the right side (A11) can be written as the "integral over orbits" (see e.g., Polyachenko and Fridman 1976):

$$f_1(r, v_r, v_\varphi) = \frac{\partial F_0}{\partial\mathscr{E}}\Phi_1 + i\omega_*(r_0)\frac{\partial F_0}{\partial\mathscr{E}} \times\int_{-\infty}^{0}\Phi_1(r')e^{-i\omega t + im(\varphi'-\varphi)}dt + \frac{2im\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial F_0}{\partial r_0}\int_{-\infty}^{0}\Phi_1(r')e^{-i\omega t + im(\varphi'-\varphi)}dt, \quad \text{(A12)}$$

where $r' = r'(t)$ and $\varphi' = \varphi'(t)$ is the unperturbed stellar orbit with the coordinates (r, φ) and velocity (v_r, v_φ) at time $t = 0$.

For definiteness, we take an equilibrium distribution function in the form of a generalized Schwarzschild distribution (Shu 1970):

$$F_0 = P(r_0)\exp\left[-\frac{\mathscr{E}}{c_0^2(r_0)}\right]. \quad \text{(A13)}$$

The functions $P(r_0)$ and $c_0(r_0)$ can be expressed, if desired, in terms of surface density and stellar velocity dispersion. However, we will not need these general expressions; the required simpler formulas are given below. We also assume, for simplicity, that $c_0 = \text{const}$.

For this distribution function, the perturbation of surface density $\sigma_1(r)\exp(-i\omega t + im\varphi)$ is

$$\sigma_1(r) = \int dv_r d\tilde{v}_\varphi f_1 = -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0\left\{\Phi_1(r) + i\omega_*(r_0)\int_{-\infty}^{0}\Phi_1(r')e^{-i\omega t + im(\varphi'-\varphi)}dt\right\} \quad \text{(A14)}$$

$$+\int dv_r d\tilde{v}_\varphi\frac{2im\Omega(r_0)}{r_0\kappa^2(r_0)}\frac{\partial F_0}{\partial r_0}\int_{-\infty}^{0}\Phi_1(r')e^{-i\omega t + im(\varphi'-\varphi)}dt.$$

Hence, it is clear that the orbits for r' and $(\varphi' - \varphi)$ calculated in the post-epicyclic approximation are required for the proper transition to the limiting case of a cold disk ($c_0 \longrightarrow 0$). Below, we also make this transition, but without expanding the resonance denominators. The equations of motion for a star in an axially symmetric potential $\Phi_0(r)$ are

$$\ddot{r} = -\frac{d\Phi_0}{dr} + \frac{L^2}{r^3},$$

$$\frac{d}{dt}(r^2\dot{\varphi}) = 0,$$

where $L = r^2\dot{\varphi} = \text{const}$ is the conserved angular momentum of the star. As was pointed out above, we need the solutions of these equations for nearly circular orbits in the postepicyclic approximation. The sought-for equations of the orbit are

$$r' = r + \delta r, \quad \text{(A15)}$$

$$\delta r = \sin\kappa_0 t\frac{v_r}{\kappa_0} + (1-\cos\kappa_0 t)\frac{1}{\kappa_0}\left(\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right) + (1-\cos\kappa t)\frac{2\Omega\alpha}{\kappa_4}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 + (1-\cos\kappa t)\frac{2\Omega\beta}{\kappa^4}v_r^2, \quad \text{(A16)}$$

$$\varphi' - \varphi = (\Omega_0 + \delta\Omega)t + \delta\varphi, \quad \text{(A17)}$$

$$\delta\varphi = -\frac{2\Omega_0}{r_0\kappa_0^2}(1-\cos\kappa_0 t)v_r + \frac{2\Omega_0}{r_0\kappa_0^2}\sin\kappa_0 t\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0} + \sin\kappa t\frac{4\Omega^2\alpha}{\kappa^5 r}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 - \sin\kappa t\frac{4\Omega^2\beta}{\kappa^5 r}v_r^2, \quad \text{(A18)}$$

where we designated: $\Omega_0 = \Omega(r_0)$, $\kappa_0 = \kappa(r_0)$, and

$$\alpha = \frac{\mu}{6\Omega} + \frac{1}{2}(r\Omega'' + 2\Omega') - \frac{\Omega}{r},$$

$$\beta = \frac{\mu}{3\Omega}, \quad \mu = -\frac{1}{2}\left[(\kappa^2)' - \frac{3\kappa^2}{r}\right],$$

$$\delta\Omega = \left(\frac{3}{2r^2} - \frac{\mu}{r\kappa^2}\right)\frac{\Omega}{\kappa^2}\left[v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right] = \frac{\kappa'}{\kappa}\cdot\frac{\Omega}{r\kappa^2}\left[v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right].$$

Note that the postepicyclic approximation for the radius $r' = r + \delta r$ was previously used by Shu (1970) in his theory of tightly wound spiral density waves, which generalizes the theory of Lin and Shu (1966); however, there are errors in the formulas for δr and the relation between r_0 and r given by Shu (1970).

Let us expand the function $\Phi_1(r')$ that enters into the integral over orbits in a Taylor power series of $\delta r = r' - r$ to within terms $\sim(\delta r)^2$

$$\Phi_1(r') \simeq \Phi_1(r) + \delta r \cdot \Phi_1'(r) + \frac{1}{2}(\delta r)^2\Phi_1''(r). \quad \text{(A19)}$$

Let us also expand the exponent in terms of powers of $\delta\varphi = \varphi' - \varphi - (\Omega_0 + \delta\Omega)t$:

$$e^{-i\omega t + im(\varphi' - \varphi)} = e^{-i[\omega_*(r_0) - m\delta\Omega]t}e^{im\delta\varphi}$$
$$\simeq e^{-i[\omega_*(r_0) - m\delta\Omega]t}\left\{1 + im\delta\varphi - \frac{m^2}{2}(\delta\varphi)^2\right\}. \quad \text{(A20)}$$

It is convenient to represent the resulting expression for the perturbation of surface density as

$$\sigma_1 = \Sigma_1\Phi_1''(r) + \Sigma_2\Phi_1'(r) + \Sigma_3\Phi_1(r), \quad \text{(A21)}$$

where

$$\Sigma_1 = -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 \frac{1}{2}\omega_*(r)$$
$$\times \int_{-\infty}^{0} dt e^{-i\omega_*(r)t}\left[\frac{1}{2}(1 - \cos 2\kappa t)\frac{v_r^2}{\kappa^2}\right. \quad \text{(A22)}$$
$$\left. + \frac{1}{2}(3 - 4\cos\kappa t + \cos 2\kappa t)\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right],$$

$$\Sigma_2 = -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 i\omega_*(r_0)\int_{-\infty}^{0} dt e^{-i\omega_*(r_0)t}$$
$$\times\left[(1 - \cos\kappa_0 t)\frac{1}{\kappa_0}\left(\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right) + \sin\kappa_0 t\frac{v_r}{\kappa_0}\right. \quad \text{(A23)}$$
$$\left. + \frac{2\Omega\alpha}{\kappa^4}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2(1 - \cos\kappa t) + \frac{2\Omega\beta}{\kappa^4}v_r^2(1 - \cos\kappa t)\right]$$
$$\times\left\{1 + im\left[\sin\kappa_0 t\frac{2\Omega_0}{\kappa_0^2 r_0}\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0} - (1 - \cos\kappa_0 t)\frac{2\Omega_0}{\kappa_0^2 r_0}v_r\right]\right\},$$

$$\Sigma_3 = -\frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0 i\omega_*(r_0)$$
$$\times\int_{-\infty}^{0} dt e^{-i\omega_*(r_0)t}\left\{im\left[\sin\kappa_0 t\frac{2\Omega_0}{\kappa_0^2 r_0}\left(\frac{2\Omega_0\tilde{v}_\varphi}{\kappa_0}\right)\right.\right.$$
$$\left. - \frac{2\Omega_0 v_r}{\kappa_0^2 r_0}(1 - \cos\kappa_0 t) + \sin\kappa t\left(\frac{4\Omega^2}{\kappa^2}\right)\frac{\alpha}{\kappa^3 r}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right.$$
$$\left. - \sin\kappa t\left(\frac{4\Omega^2}{\kappa^2}\right)\frac{\beta}{\kappa^3 r}v_r^2\right] \quad \text{(A24)}$$
$$- \frac{m^2\Omega^2}{\kappa^4 r^2}\left[(1 - \cos 2\kappa t)\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2\right.$$
$$\left.\left. + (3 - 4\cos\kappa t + \cos 2\kappa t)v_r^2\right]\right\}$$
$$+ \frac{1}{c_0^2}\int dv_r d\tilde{v}_\varphi F_0\frac{m\delta\Omega}{\omega_*(r_0)}$$
$$+ \frac{2im\Omega}{r\kappa^2}\int dv_r d\tilde{v}_\varphi\left.\frac{\partial F_0}{\partial r_0}\right|_{r_0 = r}\int_{-\infty}^{0} dt e^{-i\omega_*(r)t}.$$

The integrals over t in (A22)–(A24) are calculated according to the formulas

$$\int_{-\infty}^{0} dt e^{-i\omega_*(r_0)t} = \frac{1}{-i\omega_*(r_0)},$$

$$\int_{-\infty}^{0} e^{-i\omega_*(r_0)t}\sin\kappa_0 t dt = \frac{\kappa_0}{\omega_*^2(r_0) - \kappa_0^2},$$

$$\int_{-\infty}^{0} e^{-i\omega_*(r_0)t}\cos\kappa_0 t dt = -\frac{\omega_*(r_0)}{i(\omega_*^2(r_0) - \kappa_0^2)}.$$

When integrating over velocities, one should first bear in mind that the equilibrium distribution function of a disk with nearly circular orbits must satisfy the conditions

$$\int v_r F_0 dv_r d\tilde{v}_\varphi = 0, \quad \int v_r\tilde{v}_\varphi F_0 dv_r d\tilde{v}_\varphi = 0, \quad \text{(A25)}$$

$$\int v_r^2 F_0 dv_r d\tilde{v}_\varphi = \int\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2 F_0 dv_r d\tilde{v}_\varphi. \quad \text{(A26)}$$

It is also important to take into account the fact that the azimuthal flux is

$$\Pi_\varphi = \int\tilde{v}_\varphi F_0 dv_r d\tilde{v}_\varphi \simeq \sigma_0\frac{c_0^2}{\Omega r} \neq 0. \quad \text{(A27)}$$

The properties (A25)–(A27) and the specific expression for flux Π_φ can be derived most easily by representing the distribution function of a stellar system with nearly circular orbits as a formal series of δ-functions of v_r, $\tilde{v}_\varphi$, and their derivatives (Polyachenko and Fridman 1976):

$$f_0 = a_1\delta(v_r)\delta(\tilde{v}_\varphi) + b_1\delta'(v_r)\delta(\tilde{v}_\varphi) + b_2\delta(v_r)\delta'(\tilde{v}_\varphi) + c_1\delta''(v_r)\delta(\tilde{v}_\varphi) + c_2\delta'(v_r)\delta'(\tilde{v}_\varphi) + c_3\delta(v_r)\delta''(\tilde{v}_\varphi). \quad \text{(A28)}$$

Substituting the expansion (A28) in the equilibrium kinetic equation

$$v_r\frac{\partial f_0}{\partial r} - \frac{\partial\Phi_0}{\partial r}\frac{\partial f_0}{\partial v_r} + \frac{v_\varphi^2}{r}\frac{\partial f_0}{\partial v_r} - \frac{v_r v_\varphi}{r}\frac{\partial f_0}{\partial v_\varphi} = 0 \quad \text{(A29)}$$

and setting the coefficients of various combinations of the derivatives of δ functions equal to zero, we derive the equalities $b_1 = c_2 = 0$ equivalent to (A25), $c_3 = c_1\kappa^2/4\Omega^2$, which corresponds to the well-known Lindblad relation (A26) between radial (c_r) and azimuthal (c_φ) stellar velocity dispersions, as well as the following general expression for the azimuthal stellar flux:

$$\Pi_\varphi = \frac{1}{2\Omega r}\left[(r\sigma_0 c_r^2)' - \frac{\kappa^2\sigma_0}{4\Omega^2}c_r^2\right]. \quad \text{(A30)}$$

Accordingly, instead of Eq. (A28), we have a general expression for the equilibrium distribution function of stars with nearly circular orbits in the form[2]:

$$f_0 = \sigma_0(r)\left\{\delta(v_r)\delta(\tilde{v}_\varphi) - \frac{1}{2\Omega r\sigma_0}\left[(r\sigma_0 c_r^2)' - \frac{\kappa^2\sigma_0}{4\Omega^2}c_r^2\right]\delta(v_r)\delta'(\tilde{v}_\varphi) + \frac{1}{2}c_r^2(r)\delta''(v_r)\delta(\tilde{v}_\varphi) + \frac{1}{2}\frac{\kappa^2}{4\Omega^2}c_r^2(r)\delta(v_r)\delta''(\tilde{v}_\varphi)\right\}. \quad \text{(A31)}$$

In general, we can arbitrarily specify the system's surface density $\sigma_0(r)$ and radial velocity dispersion $c_r(r)$ (which must be low). The function $\Omega(r)$ can be calculated from the equilibrium condition for a star in a circular orbit: $\Omega^2 r = d\Phi_0/dr$, where the potential $\Phi_0(r)$ is determined by the distribution of all masses in the galaxy one of the components of which we consider.

The Schwarzschild-type "spread" distribution function corresponding to (A31) can be written as

$$F_0(r, v_r, \tilde{v}_\varphi) = \frac{2\Omega}{\kappa}\frac{\sigma_0}{2\pi c_0^2}\exp\left[-\frac{v_r^2 + \left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)^2}{c_0^2}\right] \times\left\{1 + \frac{1}{\kappa}\left(\frac{2\Omega\tilde{v}_\varphi}{\kappa}\right)\left(\frac{\sigma_0'}{\sigma_0} - \frac{\Omega'}{2\Omega}\right)\right\}, \quad \text{(A32)}$$

where we took into account the fact that the azimuthal flux at $c_r = c_0 \neq c_0(r)$ is

$$\Pi_\varphi = \frac{\sigma_0 c_0^2}{2\Omega}\left(\frac{\sigma_0'}{\sigma_0} - \frac{\Omega'}{2\Omega}\right). \quad \text{(A33)}$$

If we make the symmetrizing change

$$v_r \longrightarrow x = v_r/c_0, \quad \tilde{v}_\varphi \longrightarrow y = 2\Omega\tilde{v}_\varphi/(\kappa c_0), \quad \text{(A34)}$$

we obtain

$$F_0(r, v_r, \tilde{v}_\varphi)dv_r d\tilde{v}_\varphi = f_0(r, x, y)dxdy, \quad \text{(A35)}$$

$$f_0(r, x, y) = \frac{\sigma_0}{2\pi}\exp\left(-\frac{x^2 + y^2}{2}\right) \times\left\{1 + \rho y\left(\frac{\sigma_0'}{\sigma_0} - \frac{\Omega'}{2\Omega}\right)\right\}. \quad \text{(A36)}$$

Simple calculations reduce expressions (A22)–(A24) for Σ_1, Σ_2, and Σ_3 to

$$\Sigma_1 = -\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dy\frac{y^2 e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}, \quad \text{(A37)}$$

$$\Sigma_2 = C^{(1)} + C^{(2)}, \quad \text{(A38)}$$

$$C^{(1)} = -\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty}dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{2\Omega}\right)\right], \quad \text{(A39)}$$

$$C^{(2)} = -\frac{2\Omega}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dy\frac{e^{-y^2/2}}{\omega_*^2(r_0) - \kappa^2(r_0)}(\alpha y^2 + \beta), \quad \text{(A40)}$$

$$\Sigma_3 = A + D^{(1)} + D^{(2)}, \quad \text{(A41)}$$

$$A = -\frac{2m\Omega}{r\kappa^2}\left(\frac{\sigma_0'}{\sigma_0} + \frac{\Omega'}{\Omega}\right)\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dy\frac{e^{-y^2/2}}{\omega_*(r_0)} + \frac{m\Omega}{r\kappa^2}\frac{\kappa'}{\kappa}\frac{\sigma_0}{\sqrt{2\pi}}\int_{-\infty}^{\infty}dy\frac{e^{-y^2/2}(3 + y^2)}{\omega_*(r_0)}, \quad \text{(A42)}$$

[2] In Eq. (A31), we corrected the error in the similar formula (4) from the monograph of Polyachenko and Fridman (1976); the above formula (4) is valid only for a homogeneous (in density) disk when $\sigma_0' = 0$.

$$D^{(1)} = \frac{2m\Omega\omega_*}{\kappa^2 r}\frac{\sigma_0}{\sqrt{2\pi}\rho}\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0)-\kappa^2(r_0)} \times\left[y + y^2\rho\left(\frac{\sigma_0'}{\sigma_0}+\frac{3\Omega'}{2\Omega}-\frac{m\Omega'}{\omega_*}-\frac{2\kappa'}{\kappa}-\frac{1}{r}\right)\right], \quad \text{(A43)}$$

$$D^{(2)} = \frac{4\Omega^2}{\kappa^2}\frac{\sigma_0}{\sqrt{2\pi}} \times\int_{-\infty}^{\infty} dy \frac{e^{-y^2/2}}{\omega_*^2(r_0)-\kappa^2(r_0)}\left[(\alpha y^2+\beta)\frac{m\omega_*}{r\kappa^2}+\frac{m^2}{r^2}\right]. \quad \text{(A44)}$$

In order to pass to the limit of a cold disk, we must expand the resonance denominators in (A37)–(A44):

$$\frac{1}{\omega_*(r_0)} \longrightarrow \frac{1}{\omega_*(r)},$$

$$\frac{1}{\omega_*^2(r_0)-\kappa_0^2} \longrightarrow \frac{1}{\omega_*^2(r)-\kappa^2}+\left(\frac{1}{\omega_*^2(r)-\kappa^2}\right)'\rho y,$$

take into account the fact that $\alpha + \beta = \kappa^2(1 - r\Omega'/2\Omega)/2\Omega r$, and calculate the emerging integrals over y. It is easy to verify that the resulting expression for $\sigma_1(r)$ matches Eq. (1); recall that the latter is derived directly from the hydrodynamic equations with a zero pressure. For a convenient comparison, it is useful to transform Eq. (1) to the form

$$\sigma_1 = -\Phi''\frac{\sigma_0}{\omega_*^2-\kappa^2}-\frac{\Phi'}{r}\left\{\frac{\sigma_0}{\omega_*^2-\kappa^2}+\frac{r\sigma_0'}{\omega_*^2-\kappa^2} + \sigma_0 r\left(\frac{1}{\omega_*^2-\kappa^2}\right)'\right\}+\Phi\left\{\frac{2m\Omega\sigma_0'}{r\kappa^2}\left(\frac{\omega_*}{\omega_*^2-\kappa^2}-\frac{1}{\omega_*}\right)\right.$$
$$+\frac{m^2}{r^2}\frac{\sigma_0}{\omega_*^2-\kappa^2}+\frac{2m\Omega'\sigma_0}{r\kappa^2}\left(\frac{\omega_*}{\omega_*^2-\kappa^2}-\frac{1}{\omega_*}\right) \quad \text{(A45)}$$
$$-\frac{2m\Omega\sigma_0}{r\kappa^2}\frac{\omega_*(\omega_*^2-\kappa^2)'}{(\omega_*^2-\kappa^2)^2}-\frac{4m^2\Omega\sigma_0}{r^2\kappa^2}\frac{r\Omega'}{\omega_*^2-\kappa^2}$$
$$\left.-\frac{4m\Omega\sigma_0}{r(\omega_*^2-\kappa^2)\omega_*}\frac{\kappa'}{\kappa}\right\}.$$

Note that, surprisingly, nobody has previously made this passage to the limit. However, we are interested in the effect of a low velocity dispersion that modifies the resonance denominators. Being interested in the disk regions near a particular resonance, we must separate the dominating terms from (A37)–(A44).

The expression for the response of surface density is particularly simple in the region of corotation resonance:

$$\sigma_1 = A\Phi_1. \quad \text{(A46)}$$

Near Lindblad resonances, the expression is much more cumbersome:

$$\sigma_1 \simeq B\Phi_1'' + C\Phi_1' + D\Phi_1, \quad \text{(A47)}$$

where we designated: $B = \Sigma_1$, $C = \Sigma_2$, $D = \Sigma_3 - A$.

ACKNOWLEDGMENTS

We made use of the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. The study was partially based on data from the ING Archive. This work was supported in part by the Russian Foundation for Basic Research (project nos. 99-02-18432 and 00-15-96528).

REFERENCES

1. L. A. Artsimovich and R. Z. Sagdeev, *Plasma Physics for Physicist* (Atomizdat, Moscow, 1979).
2. G. S. Bisnovatyĭ-Kogan and A. B. Mikhaĭlovskiĭ, Astron. Zh. **50**, 312 (1973) [Sov. Astron. **17**, 205 (1973)].
3. B. G. Elmegreen, D. M. Elmegreen, and L. Montenegro, Astrophys. J., Suppl. Ser. **79**, 37 (1992).
4. A. M. Fridman, O. V. Khoruzhii, E. V. Polyachenko, *et al.*, Phys. Lett. A **264**, 85 (1999).
5. F. Hohl, Astrophys. J. **168**, 343 (1971).
6. B. B. Kadomtsev, Pis'ma Zh. Éksp. Teor. Fiz. **33**, 361 (1981) [JETP Lett. **33**, 344 (1981)].
7. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975; Nauka, Moscow, 1988).
8. C. C. Lin and F. H. Shu, Proc. Nat. Acad. Sci. USA **55**, 229 (1966).
9. C. C. Lin, C. Yuan, and F. H. Shu, Astrophys. J. **155**, 721 (1969).
10. D. Lynden-Bell, Mon. Not. R. Astron. Soc. **187**, 101 (1979).
11. J. P. Ostriker and P. J. E. Peebles, Astrophys. J. **186**, 467 (1973).
12. I. I. Pasha and V. L. Polyachenko, Pis'ma Astron. Zh. **19**, 3 (1993) [Astron. Lett. **19**, 1 (1993)].
13. V. L. Polyachenko and I. I. Pasha, Mon. Not. R. Astron. Soc. **266**, 92 (1994).
14. V. L. Polyachenko, Pis'ma Astron. Zh. **15**, 890 (1989) [Sov. Astron. Lett. **15**, 385 (1989)].
15. V. L. Polyachenko, Astrofizika **34**, 4 (1991).
16. V. L. Polyachenko, Astron. Zh. **69**, 10 (1992) [Sov. Astron. **36**, 5 (1992)].
17. V. L. Polyachenko, in *Physics of Gaseous and Stellar Disks of the Galaxy*, Ed. by I. R. King; Astron. Soc. Pac. Conf. Ser. **66**, 103 (1994).
18. A. M. Fridman and V. L. Polyachenko, *Physics of Gravitating Systems* (Nauka, Moscow, 1976; Springer-Verlag, New York, 1984).
19. A. Toomre, Astrophys. J. **139**, 1217 (1964).
20. A. Toomre, *Structure and Evolution of Normal Galaxies*, Eds. by S. M. Fall and D. Lynden-Bell (Cambridge Univ. Press, Cambridge, 1981), p. 111.
21. F. H. Shu, Astrophys. J. **160**, 99 (1970).

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 440–444. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 516–520.

An Iterative Method for Simultaneously Computing the Synthesis of Light and Heavy Elements

D. K. Nadyozhin and I. V. Panov*

Institute for Theoretical and Experimental Physics, Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia

Received February 5, 2001

Abstract—We propose an iterative algorithm for computing the synthesis of heavy elements through the rapid capture of neutrons (*r*-process) and, at sufficiently high temperatures, protons by simultaneously using two distinct computer codes. One of the codes describes the kinetics of nuclear reactions between light and intermediate chemical elements, which are the source of free neutrons and protons used by the second code to synthesize heavy elements from seed nuclides (isotopes near the iron peak of the cosmic abundance curve). The two codes interact through the neutron and proton reaction channels. We demonstrate the efficiency of our method with an example of the nucleosynthesis in a supernova's helium shell triggered by the evaporation of neutrons and protons from α particles exposed to the neutrino flux from a collapsing stellar core. In this case, three or four iterations are enough to obtain an almost exact self-consistent solution. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *nuclear astrophysics, supernovae and supernova remnants*

INTRODUCTION

The synthesis of heavy elements in stars through the rapid capture of neutrons and protons largely depends on the attainable neutron and proton fluxes. Free neutrons are generally produced in reactions between light and intermediate nuclei, for example, in the $^{13}C(\alpha, n)^{16}O$ and $^{22}Ne(\alpha, n)^{25}Mg$ reactions. If the temperature is sufficiently high ($\gtrsim 1.5 \times 10^9$ K), then free protons also emerge in considerable quantities for the nucleosynthesis via (γ, p) and other reactions. Free neutrons and protons can also be produced during the interaction of neutrinos with a supernova's envelope matter. In this case, the fraction of the neutrons and protons involved in the synthesis of heavy elements depends on numerous neutrino-induced reactions between light and intermediate elements, from He to Mg (Nadyozhin *et al.* 1998).

Thus, we essentially have to deal with two distinct systems of nuclear-kinetic equations. One describes the reactions between light and intermediate elements (L-code), and the other describes the production of heavy elements from primary seed nuclei (H-code). The two codes are related to each other through the neutron and proton reaction channels. The proposed iterative method allows the neutron and proton exchange between the two codes to be computed in a mutually consistent way. It also allows us to compute the production of heavy elements (H-code) triggered by the neutron source whose physical model was realized in the L-code. Below, we illustrate the efficiency of our method with computations of nucleosynthesis in the helium shell of a collapsing supernova when free neutrons and protons are produced by inelastic scattering of neutrinos by helium nuclei (an excited $^4He^*$ nucleus decays predominantly with the emission of a neutron or a proton).

Our method can also be used to compute the nucleosynthesis in thermonuclear flashes as hydrogen and helium are accreted onto the surfaces of degenerate stars (white dwarfs and neutron stars). In general, the idea of iterations between different nuclear-kinetic codes is not new, particularly in nucleosynthesis calculations of elements beyond the iron peak of the cosmic abundance curve. Since intermediate nuclei and iron-peak elements, on the one hand, and elements heavier than iron, on the other hand, are produced under completely different conditions determined by different scenarios, their calculation requires different mathematical approaches and models. Elements heavier than iron are synthesized under conditions when the main element-producing processes are interactions with neutrons and β decay. However, seed nuclei for the *r*-process and free neutrons are produced in reactions between light and intermediate nuclides at high temperatures. Of course, the entire evolution of chemical elements during explosive nucleosynthesis, which transforms into the *r*-process, can be analyzed in principle in terms of a single nuclear-kinetic model. However, as was shown by Panov *et al.* (2001), it takes much computer time to compute the explosive nucleosynthesis accompanied by heavy-element synthesis in a supernova's envelope or during the merging of neutron stars. Therefore, these two processes—the production of free

* E-mail address for contacts: panov@vitep5.itep.ru

neutrons and iron-peak elements (as well as several nuclides with masses $A < 100$ beyond the iron peak) in nuclear reactions between light and intermediate elements and the synthesis of heavy elements in the *r*-process—are modeled on the basis of mathematical schemes that differ from each other and that are adapted to specific physical scenarios. In this case, the final elemental abundances derived in one scenario are used as seed abundances in another scenario. In particular, Rauscher *et al.* (1994) considered a two-code method for computing the synthesis of heavy elements where data between their L- and H-codes were exchanged at each computational step in time by applying corrections to the number density of free neutrons common to the two codes; however, these authors did not aim at obtaining a strictly consistent solution.

Our method makes it possible to completely reconcile calculations based on the two nuclear-kinetic codes in the case where the physical scenario allows the nucleosynthesis calculation to be naturally separated into two ranges of nuclei that interact only through the neutron and proton channels. Using different codes has several advantages over synthesis calculations of light, intermediate, and heavy elements in terms of a single code. The most important advantage seems to be the following: we can analyze in detail the nucleosynthesis kinetics separately for light and heavy elements and can combine the available codes into a single computer code that provides a coordinated interaction between specialists in the synthesis of light elements and those who are interested in the synthesis of heavy elements.

ITERATIVE SCHEME

We begin iterations by computing the fractions of free neutrons and protons as functions of time, $Y_n(t)$ and $Y_p(t)$, using the L-code. Then, we numerically solve the system of ordinary differential equations (H-code) that describes the synthesis of heavy elements by assuming that the neutron and proton fractions are determined by the derived functions $Y_n(t)$ and $Y_p(t)$ rather than by the corresponding kinetic equations for neutrons and protons. During the H-code calculation, we compute the rates of change in neutron and proton fractions, $\left(\frac{dY_n}{dt}\right)_H$ and $\left(\frac{dY_p}{dt}\right)_H$, that would take place if we took into account the interaction of neutrons and protons with heavy elements alone.

Next, we perform calculations with the L-code, in which the kinetics of neutrons and protons is described by modified equations,

$$\frac{dY_n}{dt} = S_n - Y_n F_n, \quad \frac{dY_p}{dt} = S_p - Y_p F_p, \qquad (1)$$

where S_n and S_p are the sums over all neutron- and proton-emission and absorption reactions in the system of kinetic equations described by the L-code; F_n and F_p are given by

$$F_n = -\frac{1}{Y_n(t)}\left(\frac{dY_n}{dt}\right)_H, \quad F_p = -\frac{1}{Y_p(t)}\left(\frac{dY_p}{dt}\right)_H, \qquad (2)$$

where $Y_{n,p}(t)$ and $\left(\frac{dY_{n,p}}{dt}\right)_H$ are taken from the previous L- and H-code calculations, respectively.

The H-code [developed for *r*-process calculations and described in detail by Blinnikov and Panov (1996) and Nadyozhin *et al.* (1998)] uses the method of Gear (1971) to solve stiff systems of differential equations. This method with improved matrix-inversion algorithms (Blinnikov and Bartunov 1993) allows the system of nucleosynthesis equations that determines changes in the fractions of all the isotopes under consideration to be effectively solved. The number of equations in these calculations of heavy-nuclei production is equal to the number of isotopes under consideration. We determined the boundary conditions for the range of nuclei of the H-code in which the nucleosynthesis was modeled as follows: all isotopes of the chemical elements from calcium ($Z_{min} = 20$) to lead ($Z_{max} = 82$) were included in the analysis. Thus, the nucleosynthesis range was wide enough even for an extremely low metallicity ($Z = 0.0001Z_\odot$), when the nucleosynthesis wave reached the heaviest nuclei ($A \approx 196$, $Z \leq 78$) during the synthesis.

The procedure for reconciling the L- and H-code calculations (iterations) described above is repeated until the fractions of neutrons and protons (along with all the remaining nuclides) converge with the required accuracy in the entire time interval of interest. An important remark should be made here. If the functions $F_{n,p}(t)$ become positive in some time intervals (for example, due to the emission of delayed neutrons at the cooling stage after shock passage), then $\left(\frac{dY_{n,p}}{dt}\right)_H$ should be used instead of $-Y_{n,p}F_{n,p}$ to avoid numerical instability in Eqs. (1) in these time intervals.

The initial approximation for $F_{n,p}(t)$ in the very first calculation with the L-code is discussed in the next section.

APPLICATION OF THE METHOD

Figures 1–4 show examples of calculations based on the two-code technique for neutrino-induced nucleosynthesis in a helium shell. Free neutrons and protons are produced through the decay of helium nuclei when mostly μ and τ neutrinos and antineutrinos are scattered inelastically. The initial temperature and density in the helium shell were chosen to be the following: $T = 0.8 \times 10^9$ K and $\rho = 10^4$ g cm^{-3}. The distance of the helium shell from the stellar center was assumed to be $R = 1 \times 10^9$ cm, which is a factor of 4 smaller than that in stan-

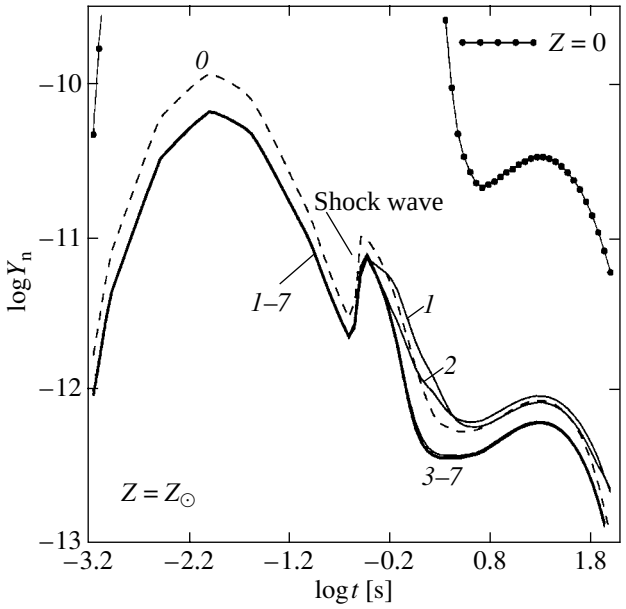


Fig. 1. The fraction of free neutrons versus time ($t = 0$ corresponds to the onset of neutrino emission) for solar metallicity ($Z = Z_\odot$). The curves with numbers correspond to different iterations.

dard evolutionary models (Woosley and Weaver 1995). This underestimated value of R was specially chosen in order that the temperature at the shock front exceed 1.5×10^9 K and that it become possible to test our method not only on the neutron channel, but also on the proton channel. However, the possibility that helium is considerably closer to the neutrino-emitting collapsing core than it is in simple spherically symmetric models should not be overlooked. For example, helium-enriched jets could penetrate deep into the star immediately before the supernova explosion (see, e.g., Bazan and Arnett 1994). In the accepted presupernova model, the shock wave reaches the helium shell $t = 0.26$ s after the onset of collapse and heats it up to a temperature $T \approx 2.2 \times 10^9$ K. We calculated these parameters and the matter cooling law after shock passage by using analytic formulas that approximated detailed hydrodynamic calculations (Nadyozhin *et al.* 2000). The supernova explosion energy was taken to be 2×10^{51} erg.

In Figs. 1–4, the fractions of neutrons and protons Y_n and Y_p (the volume neutron and proton densities are $n_{n,p} = N_A \rho Y_{n,p}$, where N_A is the Avogadro number) are plotted against time for solar metallicity ($Z = Z_\odot$, Figs. 1 and 2) and for a considerably lower metallicity ($Z = 0.001 Z_\odot$, Figs. 3 and 4), which is typical of very old stars (Sneden *et al.* 2000). The intermediate iterations are indicated by thin lines with iteration numbers. The final solution that converged with an accuracy better than 1% in the entire time interval 0–100 s under consideration is represented by heavy lines. This accuracy is achieved at iterations 7 (Figs. 1 and 2) and 9 (Figs. 3 and 4). The line with dots corresponds to $Y_n(t)$, which was computed by using the L-code for zero metallicity ($Z = 0$). In the case of zero metallicity (curves with dots), Y_n reaches a maximum of $\approx 10^{-5}$ at $\log t = -0.36$.

The very first calculation of $Y_{n,p}(t)$ based on the L-code (zero approximation) is indicated in Figs. 1–4 by dashed lines with number 0. In this case, the functions F in Eqs. (1) are given by

$$F_n = Y^0_{Fe56} N_A \rho \langle \sigma_{n,\gamma} v \rangle, \quad F_p = Y^0_{Fe56} N_A \rho \langle \sigma_{p,\gamma}, v \rangle, \quad (3)$$

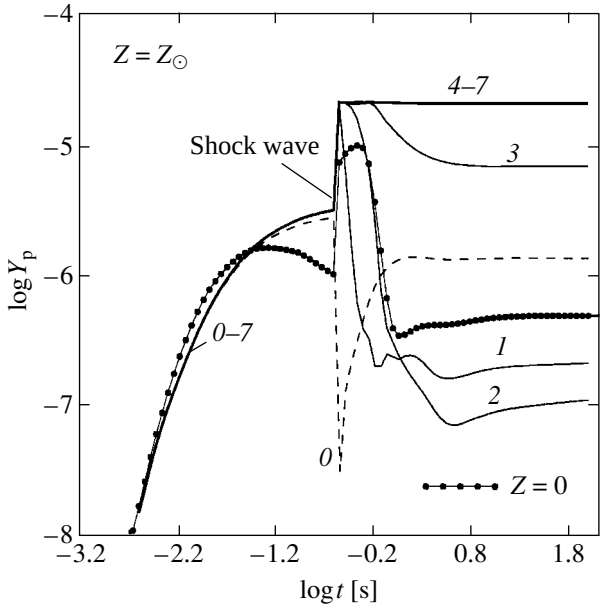


Fig. 2. Same as Fig. 1 for the fraction of protons.

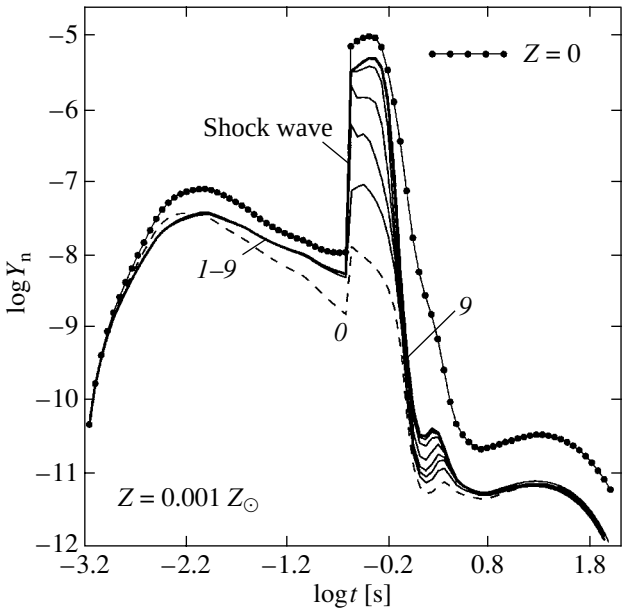


Fig. 3. Same as Fig. 1 for metallicity $Z = 0.001 Z_\odot$.

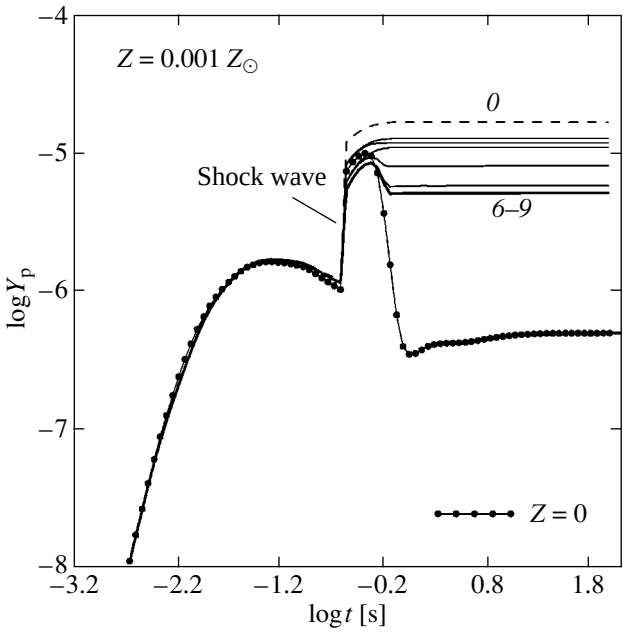


Fig. 4. Same as Fig. 2 for metallicity $Z = 0.001Z_\odot$.

where Y^0_{Fe56} is the initial fraction of seed ^{56}Fe nuclei, which was constant during the entire period under consideration ($0 \le t \le 100$ s). In all calculations, Y^0_{Fe56} was chosen to correspond to the cosmic elemental abundance: $Y^0_{Fe56} = 1.27 \times 10^{-3}Z$ (Anders and Grevesse 1989). In addition, the zero approximation for $F_{n,p}$ is that the cross-sections for radiative neutron ($\sigma_{n,\gamma}$) and proton ($\sigma_{p,\gamma}$) captures by heavy elements are always the same as those for seed ^{56}Fe nuclei.

In the case of solar metallicity, the last five iterations [(3–7) for neutrons (Fig. 1) and four iterations (4–7) for protons (Fig. 2)] differ by no more than a few percent, and they are almost indistinguishable from the corresponding final solutions in the figures. This is also true for the last four iterations (6–9) in Figs. 3 and 4.

At low metallicity ($Z \lesssim 0.01Z_\odot$), the iterations also rapidly converge if the functions $Y_{n,p}(t)$ computed with the L-code at $Z = 0$, i.e., when $F_{n,p} = 0$ in Eqs. (1), are used as the zero approximation.

In this example, the L-code controls the kinetics of ~120 reactions between 30 light and intermediate nuclides (n, p, D, T, ^{3,4}He, …, ^{24}Mg), whereas the H-code includes almost 3000 equations. These equations describe the conversion of seed ^{56}Fe nuclei into heavier elements through multiple neutron and proton captures and β decay [for more details, see Nadyozhin *et al.* (1998) and references therein].

CONCLUSION

Our main conclusions can be formulated as follows:

(1) A mere three or four iterations are required to achieve an accuracy of no less than a few percent.

(2) For metallicity $Z \gtrsim 0.01Z_\odot$, even the first approximation defined by Eqs. (3) ensures a satisfactory accuracy for $Y_n(t)$ (compare the dashed and heavy lines in Fig. 1). At not-too-low metallicities, the neutrino-induced nucleosynthesis of light and intermediate elements in a helium shell can therefore be studied by using the simple procedure defined by Eqs. (3) to estimate the absorption of neutrons by iron-peak elements. Thus, there is no need to resort to time-consuming calculations of heavy-element kinetics with the sole aim to estimate their effect on the fraction of free neutrons, on which the yields of several astrophysically important light isotopes (^{7}Li, ^{11}B, ^{15}N, etc.) strongly depend.

(3) A considerable number of iron nuclei are converted into heavier nuclei only at very low metallicities ($Z \lesssim 0.01Z_\odot$); the number of captured neutrons in the calculation per iron nucleus, *n*/Fe, is 0.3, 34, 162, and 185 for $Z = Z_\odot$, $10^{-2}Z_\odot$, $10^{-3}Z_\odot$, and $10^{-4}Z_\odot$, respectively. At $Z \gtrsim 0.01$, *n*/Fe is not large enough for the full-scale *r*-process to proceed; in addition to iron-peak elements, a number of heavy nuclei with $A < 130$ are produced, while the peak with $A \sim 130$ begins to form only at smaller *Z*. This result can be of interest in connection with the prediction of a weak *r*-process component (Wasserburg *et al.* 1996) and with its recent confirmation in chemical-composition observations of low-metallicity stars (Sneden *et al.* 2000; Truran and Cowan 2000).

This paper was presented at the International Conference "Nuclei in Cosmos" and published in Nadyozhin and Panov (2001).

ACKNOWLEDGMENTS

We are grateful to the Russian Foundation for Basic Research (project nos. 00-02-17230 and 00-15-96572) and the Swiss National Science Foundation (SNSF) for support.

REFERENCES

1. E. Anders and N. Grevesse, Geochim. Cosmochim. Acta **53**, 197 (1989).
2. G. Bazan and D. Arnett, Astrophys. J. Lett. **433**, L41 (1994).
3. S. I. Blinnikov and O. S. Bartunov, Astron. Astrophys. **273**, 106 (1993).
4. S. I. Blinnikov and I. V. Panov, Pis'ma Astron. Zh. **22**, 45 (1996) [Astron. Lett. **22**, 39 (1996)].
5. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, New York, 1971).
6. D. K. Nadyozhin and I. V. Panov, in *Proceedings of the NIC-2000, Aarhus, Denmark, 2000*; Nucl. Phys. A (2001) (in press).
7. D. K. Nadyozhin, I. V. Panov, and S. I. Blinnikov, Astron. Astrophys. **335**, 207 (1998).

8. D. K. Nadyozhin, I. V. Panov, and A. Yu. Deputovich, in *Proceedings of 10th Workshop on "Nuclear Astrophysics," 2000*, Ed. by W. Hillebrandt and E. Müller, MPA/P12, 2000, p. 99.
9. I. V. Panov, S. I. Blinnikov, and F.-K. Thielemann, Pis'ma Astron. Zh. **27** (4), 279 (2001) [Astron. Lett. **27**, 239 (2001)].
10. T. Rauscher, J. H. Applegate, J. J. Cowan, *et al.*, Astrophys. J. **429**, 499 (1994).
11. C. Sneden, J. J. Cowan, I. I. Ivans, *et al.*, Astrophys. J. Lett. **533**, L139 (2000).
12. J. Truran and J. J. Cowan, in *Proceedings of 10th Workshop on "Nuclear Astrophysics," 2000*, Ed. by W. Hillebrandt and E. Müller, MPA/P12, 2000, p. 64.
13. G. J. Wasserburg, M. Busso, and R. Gallinio, Astrophys. J. Lett. **466**, L109 (1996).
14. S. E. Woosley and T. A. Weaver, Astrophys. J., Suppl. Ser. **101**, 181 (1995).

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 445–450. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 521–527.

A Polarization–Intensity Anticorrelation Diagram for the Solar Coronal Green Line

O. G. Badalyan[1]* and J. Sýkora[2]

[1] *Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation, Russian Academy of Sciences, Troitsk, 142092 Russia*

[2] *Astronomical Institute, Slovak Academy of Sciences, Tatranska Lomnica, 05960 Slovakia*

Received May 3, 2000; in final form, January 22, 2001

Abstract—We analyze the complex pattern of anticorrelation between the degree of polarization p in the green λ530.3-nm line and its intensity I_λ, which was revealed by coronal observations during the total solar eclipse of July 11, 1991. For coronal points located at approximately the same distance from the disk center, the anticorrelation diagram breaks up into two branches with a zone of avoidance between them. High-latitude streamers form the upper branch, while the lower branch belongs to active equatorial regions. The arrangement of large-scale coronal structures in the p–$\log I_\lambda$ diagram is considered for a distance of $1.2R_\odot$. The change in anticorrelation diagram with distance is analyzed in detail for the giant high-latitude coronal streamers observed on July 11, 1991. Our results contain important information about the scattering of photospheric radiation at line frequencies in the presence of a coronal magnetic field. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *Sun*

INTRODUCTION

In recent years, a wealth of information about the solar corona has come from the modern Yohkoh, SoHO, and TRACE spacecraft. Long series of coronal observations were obtained in white light (coronal shape, streamers, and coronal mass ejections) and in permitted X-ray and ultraviolet lines, both on the disk and above the limb (coronal holes, various kinds of brightening on the disk, and loops of the inner corona). Extra-atmospheric observations in visible forbidden coronal lines are not yet successful. Nevertheless, since forbidden lines carry different information about physical conditions in the corona and about the origin of its radiation, including polarized radiation, a continuation of the corresponding observations is an important problem of current interest. A study of coronal magnetic fields by analyzing polarized radiation in forbidden coronal lines is particularly promising.

Polarization observations of the solar corona in emission lines are a complex experimental problem. Up until now, only a very limited amount of reliable observational data has been available. Therefore, after high-quality polarization filtergrams of the corona were obtained by one of us (J.S.) in the green λ530.3-nm line during the total solar eclipse of July 11, 1991, considerable progress has been in studying the properties of this forbidden line, and several regularities have been found in the distribution of polarization characteristics in various large-scale coronal structures.

This work is an elaboration of our studies based on polarization observations in the green coronal line on July 11, 1991 (Badalyan and Sýkora 1997; Badalyan *et al.* 1997a, 1999a). Here, we continue to consider the anticorrelation between polarization p in the line and its intensity I_λ. We analyze peculiarities of the arrangement of large-scale coronal structures of various types in the p–$\log I_\lambda$ anticorrelation diagram. The change in polarization and intensity with height in the two most characteristic coronal structures, huge NE and SW streamers, on July 11, 1991, is considered separately. We also discuss the statistical scatter of data on the green-line polarization.

OBSERVATIONAL DATA AND A BRIEF DESCRIPTION OF PREVIOUS RESULTS

A characteristic feature of the white-light corona on July 11, 1991, (see Fig. 1 in Badalyan *et al.* 1999a) was its deviation from a spherically symmetric shape, which is commonly observed during the maxima of solar cycles. On the eclipse day, a system of huge streamers was located at high latitudes, which was attributable to a large inclination of the solar magnetic equator (see, e.g., Sýkora *et al.* 1999). Another peculiar feature of the corona was that the coronal condensations in equatorial regions were not very bright. The condensations mostly did not overlap with streamers, which allowed these coronal structures to be studied

* E-mail address for contacts: badalyan@izmiran.troitsk.ru

separately. A short, but bright streamer above a region of X-ray brightening lay in the east near the equator. A small coronal hole was observed slightly westward of the north pole. There was a region of reduced brightness at a position angle of about 150°, which proved to be similar in radiation characteristics to coronal holes.

High-quality green-line images of the corona were obtained with a narrow-band filter ($\Delta\lambda$ = 0.17 nm) at four polaroid positions. Using the same instrument (but without a narrow-band filter), three series of polarization images were also obtained for the corona in white light with various exposures. These images allowed us to analyze the polarization distribution in white-light coronal structures (Badalyan *et al.* 1997b) and to subtract the contribution of white light that passed through the narrow-band filter when obtaining green-line filtergrams. The details of data reduction and a justification of the method of subtracting the contribution of white light can be found in Badalyan and Sýkora (1997) and Badalyan *et al.* (1997a).

As a result, we constructed polarization maps for the entire inner corona of July 11, 1991. These maps allowed us to analyze the green-line polarization distribution in various coronal regions. Of particular interest was the anticorrelation between the degree of polarization in the line and its intensity. The following factors are mainly responsible for this effect. The green coronal line is excited by electron collisions and by the absorption of photospheric photons. The linear polarization is produced by the anisotropic scattering of photospheric radiation. In dense coronal layers, where the role of electron collisions considerably increases, the relative contribution of the polarized component is reduced, which causes the polarization p to decrease with increasing total line intensity.

The total anticorrelation between line intensity I_λ and polarization p includes both the dependence of p on distance from the disk center and its change along the limb (the dependence of polarization on activity and structural peculiarities of the corona in a given direction). To exclude the distance dependence of polarization, we analyzed the anticorrelations for points inside a narrow ring, i.e., for points at approximately the same distance from the disk center (Badalyan *et al.* 1999a). As might be expected, the anticorrelation in the ring actually proved to be distinct.

The fact that the anticorrelation effect turned out to be much more informative was unexpected. We found the anticorrelation for points inside the ring to break up into two branches. A remarkable feature of the anticorrelation diagram was the zone of avoidance: an empty zone between the two branches. When analyzing the correlation between green-line polarization and the intensity for the points inside a narrow ring at a fixed distance from the disk center, we noticed that the points fell on the $p-\log I_\lambda$ diagram not randomly, but followed some pattern associated with the change in position angle. This led us to assume that the position of a point in the diagram depends on which large-scale coronal structure this point belongs to. Subsequent studies confirmed this assumption.

It emerged that the upper and lower branches belonged mainly to huge high-latitude streamers and equatorial regions, respectively. Each of the branches may be considered as a set of separate clouds of points (clusters), which describe large-scale structures of the same type within each of the branches. At low intensities and high polarizations, the points that correspond to the regions near coronal holes lie in the diagram (there are no central regions of coronal holes because of the adopted method of subtracting the contribution of white light). Bright coronal knots lie at low p and high I_λ.

The cluster pattern of arrangement of the points in all the other diagrams was subsequently confirmed when analyzing the polarization angles and the correlation of line polarization characteristics with the coronal magnetic fields calculated from photospheric observations (Badalyan *et al.* 1999b). This suggests that each of the large-scale structures is characterized by its set of physical parameters (including the magnetic field), which describe the coronal plasma in the coronal region under consideration.

LARGE-SCALE STRUCTURES IN THE ANTICORRELATION DIAGRAM

This study was carried out with small, averaged arrays with a pixel size of ≈15″ to reduce the statistical scatter of points. Our main analysis was performed for points at a mean distance of $1.2R_\odot$ within a $0.03R_\odot$-wide ring. We used our system of position angles corresponding to the position of the north pole in Fig. 1 from Badalyan *et al.* (1999a).

In Fig. 1, green-line polarization p and intensity $\log I_\lambda$ (in relative units) are plotted against position angle for points at a mean distance of $1.2R_\odot$ from the disk center. The two quantities exhibit a smooth large-scale change along the limb. A comparison of Fig. 1 with a schematic image of the corona allows us to trace the changes in p and I_λ from one large-scale structure to another. The total anticorrelation between p and I_λ clearly shows up in Fig. 1.

Eleven large-scale structures (listed in the table) were analyzed in more detail. We singled out four bright equatorial regions above the coronal condensations [the X-ray image obtained by Golub (Golub and Pasachoff 1997) during an eclipse shows characteristic brightenings in these regions], three moderately active equatorial regions (weak X-ray brightenings), two high-latitude streamers, and two regions near the coronal holes. Since no data are available for the central regions of the coronal holes because of the adopted method of subtracting the contribution of white light, the name coronal hole in the table is conditional. The second column lists the position angles within which a

given structure is located, and the third column gives the symbol that corresponds to the structure in Fig. 2.

Figure 2 shows the arrangement of the selected structures in the anticorrelation diagram. The points on the upper branch belong to giant high-latitude streamers. The compact cloud of points (open down triangles) corresponds to the isolated NE streamer at $1.1 < \log I_\lambda < 1.3$ and $12\% < p < 20\%$. The system of southern streamers (filled down triangles) covers the entire upper branch, which is attributable to the more complex structure of SW streamers. The uppermost cloud of points belongs to the southern streamer proper located near the plane of the sky at $164° \le P \le 208°$. At higher I_λ, the points that correspond to the region where the two streamers are superimposed lie further; at $P \ge 217°$, a transition to the second streamer, which is probably located outside the plane of the sky, occurs. In this second streamer, a brightening was observed in the soft X-ray image (Golub and Pasachoff 1997) at $P \sim 225°$; in the anticorrelation diagram, the upper branch joins the lower branch here.

All four coronal condensations lie at high I_λ and low p, i.e., in the lower right corner of the diagram. Interestingly, the bright condensation at the western limb (the first row in the table, crosses), above which a bright short equatorial streamer was observed, may be considered as an extension of the upper branch. This condensation and the part of the SW streamer at $P \ge 217°$ may be considered as transition structures between the high-latitude streamers and the lower-branch structures. Note that such a feature of these two coronal regions is also traceable in all the other diagrams (Badalyan *et al.* 1999b).

The three moderately active equatorial regions lie on the lower branch in Fig. 2. Each of them forms a compact cloud of points (sequentially from right to left, open up triangles, filled squares, and open squares), with these clusters overlapping only slightly. As the

A list of the coronal structures considered

Structure	Position angle, deg	Symbol in Fig. 2
Coronal condensation	67–80	Open circle
Coronal condensation	108–115	Asterisk (*)
Coronal condensation	241–258	Cross (×)
Coronal condensation	272–287	Filled circle
Equatorial region	84–107	Open up triangle
Equatorial region	119–129	Open square
Equatorial region	259–268	Filled square
NE streamer	20–56	Open down triangle
SW streamers	164–226	Filled down triangle
S coronal hole	133–138, 152–158	Filled up triangle
N coronal hole	0–10, 357–360	Plus in an open circle

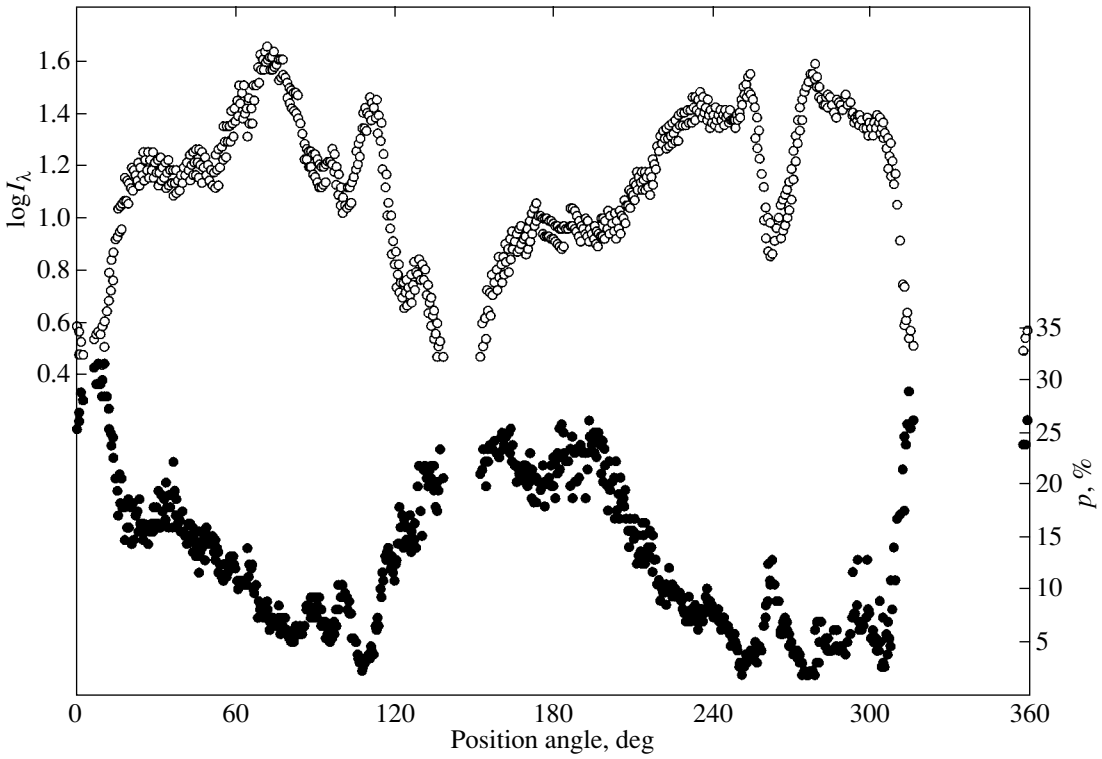


Fig. 1. Green-line polarization (lower curve, right scale) and intensity (upper curve, left scale) versus the position angle for a set of points in a $0.03R_\odot$-wide ring at a distance of $1.2R_\odot$.

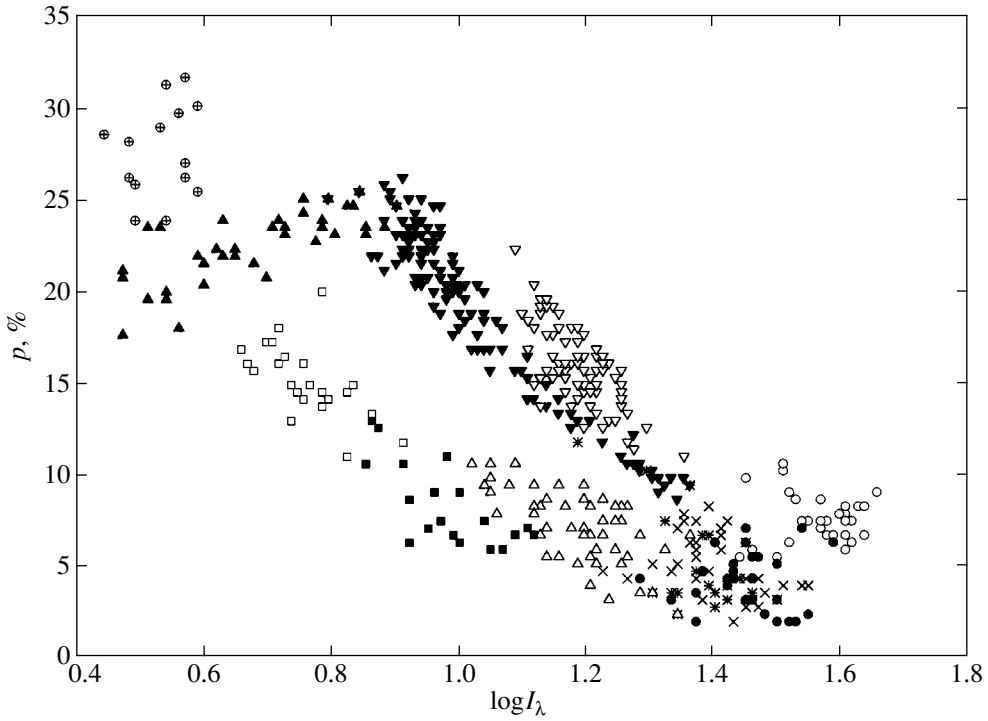


Fig. 2. Anticorrelation diagram for various large-scale structures (for a list of structures and a description of the corresponding symbols, see the table).

intensity decreases further, the vicinities of the coronal holes show up. The filled up triangles correspond to a small southeastern coronal hole; the points on the lower branch belong to its lower latitude part, while the points on the extension of the upper branch belong to the higher latitude region adjacent to the system of southern streamers. Finally, the points in the northeastern vicinity of the N hole lie in the lower left corner of the diagram. In this region, the coronal brightness in the green line is very low, and these points should be viewed with caution.

CHANGES OF THE p–I_λ DIAGRAM IN HIGH-LATITUDE STREAMERS WITH DISTANCE FROM THE DISK CENTER

We constructed similar diagrams for various distances from the disk center, which enabled us to trace the changes in anticorrelation with height for individual structures. Here, we consider the diagrams for the NE and SW streamers at distances of 1.15, 1.20, 1.25, and $1.30R_\odot$. Figures 3a and 3b show the polarizations and intensities for the entire NE streamer and for the subpolar part of the SW streamer, where no superposition of streamers was observed (the upper cloud of points in Fig. 2). The general pattern of the diagrams in Fig. 3 corresponds to the well-known variation in polarization with distance from the limb [see Picat *et al.* (1979) and references therein]. The symbols in Fig. 3 are unrelated to the structures shown in Fig. 2; they are used here only for a convenient identification of the points that refer to different distances. All circles, up triangles, squares, and down triangles (irrespective of whether they are open or filled) refer to distances of 1.15, 1.20, 1.25, and $1.30R_\odot$, respectively. The open and filled symbols in Fig. 3b are also used only for convenience.

A comparison of Figs. 3a and 3b reveals both the common properties of the two giant streamers under considerations and their differences. Thus, we can see that the clouds of points that refer to different distances of the corresponding narrow rings from the disk center move in the diagram with changing $R_\odot$ in approximately the same way; in other words, the slopes of the curves in both panels of Fig. 3 are approximately the same. It may also be noted that the values of p and $\log I_\lambda$ for the two streamers are similar at a given distance. At the same time, we may note that the northern streamer is generally slightly brighter than the southern streamer, and, according to the anticorrelation law, the polarizations in the former are slightly lower than in the latter.

Formally, as the distance from the limb increases, the monotonic increase in the scatter of points along the p axis for the NE streamer is appreciably larger than that for the SW streamer. However, a more detailed analysis indicates that the spread of the cloud of points belonging to the NE streamer with increasing distance is attributable to a larger change in the degree of polarization in the streamer within the corresponding range of position angles (i.e., when moving along the limb). This is illustrated by Fig. 3a, where the open and filled

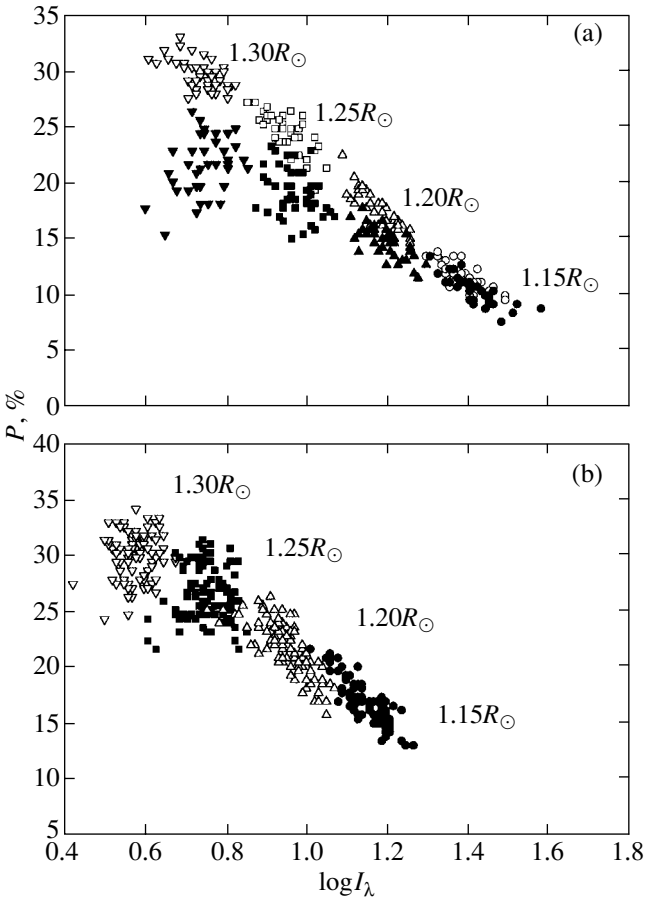


Fig. 3. Green-line polarizations and intensities for various distances (a) in the northern streamer and (b) in the subpolar part of the southern streamer. The open and filled symbols in the upper panel refer to the higher and lower latitude (relative to the axis) parts of the NE streamer, respectively.

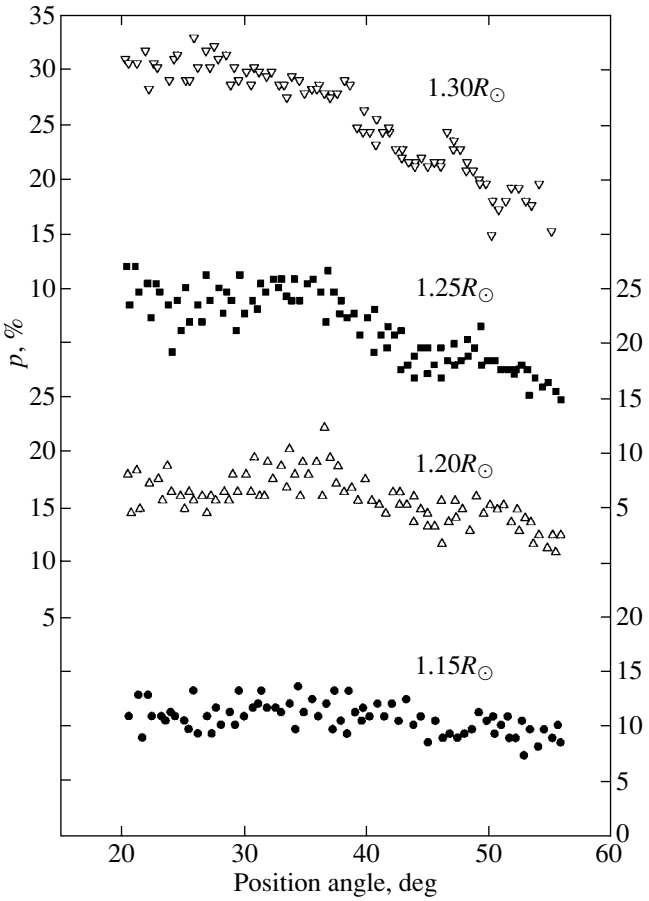


Fig. 4. Variations of the green-line polarization in the NE streamer with position angle for various distances from the disk center. The right scale is for distances of 1.15 and 1.25$R_\odot$, and the left scale is for 1.20 and 1.30$R_\odot$.

symbols correspond to the higher ($20° \le P \le 38°$) and lower ($38° \le P \le 56°$) latitude parts of the streamer, respectively. We can see that the values of p in these two parts of the NE streamer are actually different and vary differently with height. The scatter of points in the high-latitude part of the streamer changes only slightly with height, in contrast to its low-latitude part.

Figure 4 shows variations in the degree of polarization in the NE streamer with position angle for the four distances chosen. We see how the pattern of polarization distribution in the streamer changes with distance from the disk center: the total p range in the same range of position angles increases. The position angle at which p reaches a maximum gradually decreases. The statistical scatter of points at all distances is approximately the same, and the rms deviation is ~1.5%. In a similar figure for the SW streamer, we can see characteristic turning points in the polarization plot at position angles of about 200° and 220°, which divide the entire curve into three segments that correspond to three different parts of the system of SW streamers described above.

CONCLUSION

Our analysis has revealed a complex pattern of the anticorrelation diagram. We found a distinct dependence of the pattern of polarization distribution on the type of large-scale coronal structure. The smooth change in p and I_λ along the limb with a small statistical scatter of points suggests good internal agreement of the data. We traced the locations of eleven large-scale coronal structures in the anticorrelation diagram; they belong to four main types: bright coronal condensations, moderately active equatorial regions, giant high-latitude streamers, and the vicinities of the coronal holes.

The discovery of two branches in the anticorrelation diagram is the development of previously known conclusions about the behavior of polarization in the green line. It follows from Fig. 7 in Picat *et al.* (1979), where virtually all the data available by that time are shown, that the polarizations in streamers are generally higher than in active equatorial regions. The above figure also clearly shows a hint at a difference between the gradi-

ents of increase in p with height for these two main types of large-scale coronal structure.

We analyzed in detail the distributions of green-line polarization and intensity in the high-latitude helmet streamers observed on July 11, 1991. These streamers formed the characteristic shape of the corona under consideration. As our measurements show, the line polarizations in the lower coronal layers are high, suggesting a significant role of the scattering in forming the λ530.3-nm line emission.

Our derived characteristics of the green-line polarization and their comparison with the magnetic-field strength and its components allowed us to make progress in studying helmet streamers. In the July 11, 1991 corona, helmet streamers were observed at high latitudes separately from active regions. It was therefore fortunate that various streamer structures (a streamer with low activity at the base, a large system of streamers, and a NE streamer in the plane of the sky) are represented on the upper branch of the anticorrelation diagram. These data can be used for detailed model calculations.

Polarization observations in the forbidden λ530.3-nm line allow the problem of measuring the magnetic field strength directly in the solar corona to be addressed. The magnetic field plays a double role in producing the green-line polarization. On the one hand, it largely governs the formation of coronal structures and their density and temperature distributions, particularly in coronal condensations. On the other hand, our results, in particular the discovery of the zone of avoidance in the anticorrelation diagram (a jump-like transition from one branch to the other at the same intensity) and the relationship of polarization direction to the type of coronal structure (Badalyan *et al.* 1999b), suggest that the magnetic field directly affects the formation of polarized radiation in the λ530.3-nm line.

Unfortunately, this is currently the only series of data that allows such a detailed analysis of polarization characteristics to be performed in various large-scale coronal structures. It is imperative to receive a further observational confirmation of the detected effects and to perform new theoretical analyses of the issues related to the formation of polarized radiation in forbidden lines of the solar corona with allowance for real magnetic-field configurations.

ACKNOWLEDGMENTS

We wish to thank M.A. Livshits for the discussions. This study was supported by the Russian Foundation for Basic Research (project no. 99-02-18346) and the Slovak Academy of Sciences (VEGA 2/1022/21 grant).

REFERENCES

1. O. G. Badalyan and J. Sýkora, Astron. Astrophys. **319**, 664 (1997).
2. O. G. Badalyan, M. A. Livshits, and J. Sýkora, Astron. Zh. **74**, 767 (1997a) [Astron. Rep. **41**, 682 (1997a)].
3. O. G. Badalyan, M. A. Livshits, and J. Sýkora, Sol. Phys. **173**, 67 (1997b).
4. O. G. Badalyan, M. A. Livshits, and J. Sýkora, Astron. Astrophys. **349**, 295 (1999a).
5. O. G. Badalyan, V. N. Obridko, and J. Sýkora, Astron. Zh. **76**, 869 (1999b) [Astron. Rep. **43**, 767 (1999b)].
6. L. Golub and J. M. Pasachoff, *The Solar Corona* (Cambridge Univ. Press, Cambridge, 1997).
7. J. P. Picat, P. Felenbok, B. Fort, *et al.*, Astron. Astrophys. **75**, 176 (1979).
8. J. Sýkora, O. G. Badalyan, V. N. Obridko, and T. Pintér, Contrib. Astron. Obs. Skalnaté Pleso **29**, 89 (1999).

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 451–454. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 528–532.

Effects of Activity Complexes and Active Longitudes on Variations in Total Solar Irradiance

A. V. Mordvinov[1]* and R. C. Willson[2]

[1] *Institute for Solar-Terrestrial Physics, P.O. Box 4026, Irkutsk, 664033 Russia*
[2] *Research Center for Climatic Systems, Columbia University, Coronado, CA 92118, USA*

Received January 31, 2001

Abstract—A numerical technique of time–longitude analysis has been developed by studying the fine structure of temporal variations in total solar irradiance (TSI). This analysis produces maps of large-scale thermal inhomogeneities on the Sun and reveals corresponding patterns of radiative excess and deficit relative to the unperturbed solar photosphere. These patterns are organized in two- and four-sector structures and exhibit the effects of both activity complexes and the active longitudes. Large-scale patterns with radiative excess show a facular macrostructure caused by the relaxation of large-scale thermo-magnetic perturbations and/or energy output due to very large-scale solar convection. These thermal patterns are related to long-lived magnetic fields that are characterized by rigid rotation. The patterns with radiative excess tend to concentrate around the active longitudes and are centered at 103° and 277° in the Carrington system when averaged over the time-longitude distribution of thermal inhomogeneities during activity cycles 21–23. © *2001 MAIK "Nauka/Interperiodica"*.

Key words: *Sun, irradiance variations, large-scale magnetic fields*

INTRODUCTION

A continuous monitoring of basic solar parameters in recent years and earlier long-term observations have shown that the Sun is a magnetic variable star in the full sense of the word: its luminosity, diameter, rotation parameters, and magnetic fields vary with an 11-year cycle of activity (Willson and Hudson 1991; Laclare *et al.* 1996; Howard and LaBonte 1980). Since these global variations exhibit a complex interrelation, thermo-magnetic phenomena span over the bulk of the Sun (Pipin and Kichatinov 2000).

The solar magnetic fields have a multi-scale and hierarchical structure. Sunspots form structures that are seen as sunspot groups and activity complexes. Large sunspot groups concentrate within narrow intervals of heliographic longitudes, which are known as active longitudes. Long-lived longitudinal structures exist for many years, passing from one cycle of activity to another (Vitinskiĭ *et al.* 1986; Bumba and Hejna 1991). These large-scale magnetic structures are characterized by rigid rotation with an angular velocity different from the Carrington velocity. The longitudinal organization of activity mapped by using various indices occasionally differs in details; nevertheless, general regularities can be found in the distribution of activity for its various manifestations (Vitinskiĭ *et al.* 1986; Bumba and Hejna 1991). The active longitudes are often located at opposite heliographic longitudes, thus generating a 13.5-day periodicity of activity indices (Dodson and Hedeman 1968). Some active longitudes disappear for several years, and, subsequently, they are restored near their previous positions (Vitinskiĭ 1997). This behavior of active longitudes can be explained as a manifestation of the fossil solar magnetic field (Kitchatinov *et al.* 2000). Occasionally, abrupt changes occur in the spatial distribution of activity: the zone of maximum activity passes from one interval of active longitudes to another, or comparatively short-lived active structures emerge.

Since major solar flares also concentrate within the intervals of active longitudes, Bai (1988) called such regions on the Sun zones of superactivity or hot spots. He found that the 154-day periodicity played an important role in their evolution. A continuous wavelet analysis of variations in total solar irradiance (TSI) has revealed cascades of spectral energy toward larger scales, which take place during the primary and secondary maxima of cycle 22 in a time scale of 155 days (Willson and Mordvinov 1999). Here, we propose a numerical technique for time-longitude analysis of TSI variations; we have found large-scale thermal perturbations associated with long-lived magnetic structures.

SPACIOTEMPORAL ANALYSIS OF TSI

High-accuracy TSI measurements have been carried out during long-term space experiments since 1978,

* E-mail address for contacts: avm@iszf.irk.ru

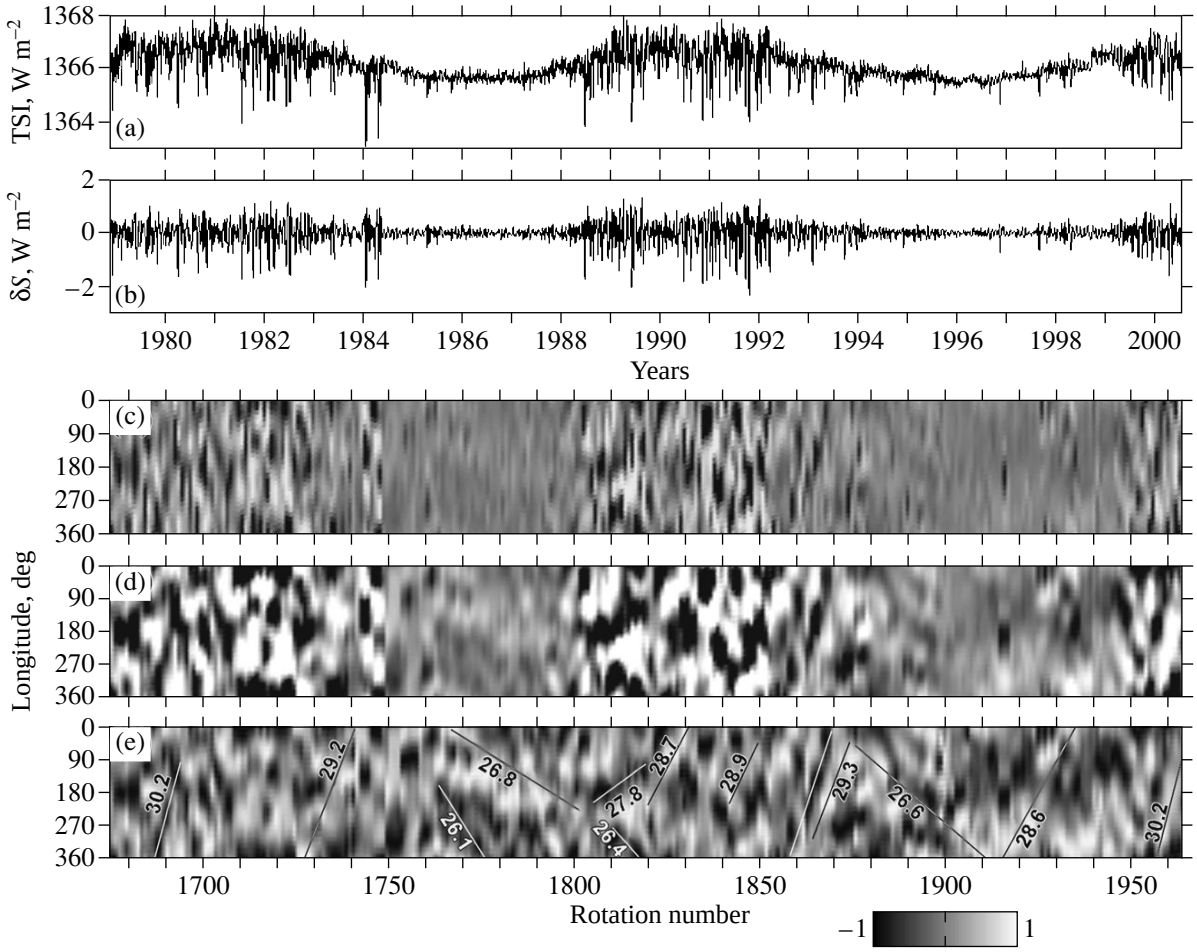


Fig. 1. (a) A series of TSI measurements and (b) its filtered component. The (c) nonsmoothed and (d) smoothed time-longitude distributions of brightness inhomogeneities. (e) The diagram of normalized brightness inhomogeneities. The maximum and minimum values of the gray scale for figures (c) and (d) should be multiplied by 0.75 and 0.25 W/m^2, respectively. The periods of rigidly rotating structures are given in days.

starting from the Nimbus-7 satellite. Subsequently, the measurements were continued, and they are currently being carried out with the radiometers mounted on four satellites. Various radiometer calibration techniques were used to construct different versions of the composite TSI time series (Fröhlich and Lean 1998; Willson 2000). Despite some differences in the fine structure of these versions, their behaviors are in good agreement with each other on a longer time scale (Willson and Mordvinov 1999). Here, we perform a spatiotemporal analysis of the nineteenth version of the composite TSI time series (Fröhlich and Lean 1998) accessible via the Internet (http://www.pmodwrc.ch). A plot of the composite time series of TSI measurements is shown in Fig. 1a.

The TSI signal from the entire Sun is formed by the convolution of the distribution of brightness inhomogeneities over the solar disk with the weighting function of limb darkening. Despite the fact that TSI is a global index, information about the spatial distribution of brightness inhomogeneities can be extracted from temporal TSI variations due to solar rotation. Here, based on wavelet filtering, we develop a numerical technique, which has allowed us to study the spatial distribution of brightness inhomogeneities and its changes by using temporal variations in TSI fine structure. By numerically performing deconvolution, an inverse operation for the convolution of measurements, we can map long-lived brightness inhomogeneities in heliographic longitude.

The main idea behind wavelet deconvolution is to filter out the signal component associated with the TSI rotational modulation and to map this component as a two-dimensional time–longitude diagram. A similar diagram for the mean solar magnetic field exhibits multimode solar rotation and its variations with an 11-year cycle (Mordvinov and Plyusnina 2000). Orthogonal wavelet transformation is an efficient tool for filtering complex nonstationary processes. Here, we used the technique of decomposition into Daubechies orthogonal wavelets, which have good localization both in time

and in time scales. A discrete wavelet transformation can be written for TSI as a function of time as

$$S(t) = \sum_{j,k=-\infty}^{\infty} c_{jk}\psi_{jk}(t), \quad (1)$$

where $\psi_{jk}(t)$ are the wavelet functions that form an orthonormal basis, and the wavelet decomposition coefficients are given by $c_{jk} = \langle f\psi_{jk}\rangle$ (Astaf'eva 1996). The main effects of rotational modulation concentrate in the range of time scales 13–30 days. Therefore, wavelet filtering should be performed by retaining only the coefficients that correspond to discrete time scales of 8, 16, and 32 days when making inverse discrete wavelet transformation. The filtered component that contains the main rotational effects is then

$$\delta S = \sum_{j=3}^{j=5}\sum_{k=-\infty}^{\infty} c_{jk}\psi_{jk}, \quad (2)$$

where 2^j correspond to the time scales (in days) on which the rotational modulation shows up. The filtered TSI component is shown in Fig. 1b. The nature and statistical properties of the negative and positive TSI fluctuations are markedly different (Mordvinov 1996). The maximum amplitude of the positive TSI fluctuations reaches 1 W/m^2, whereas the negative fluctuations reach 2 W/m^2.

Next, the filtered component was divided into subsets corresponding in time to Carrington rotations by using interpolation and mapped rotation by rotation in the form of a two-dimensional diagram. The time of observations within a Carrington rotation corresponds to a heliographic longitude. The time–longitude diagram constructed by using TSI deconvolution is shown on a gray scale in Fig. 1c. This diagram displays spatial brightness inhomogeneities as deviations of the filtered component from a mean TSI level, which slowly varies with time. Positive deviations are shown in light tones relative to the 50% gray background, and negative deviations are shown in dark tones. In the time-longitude brightness distribution, we can see a more or less regular pattern, which is formed by long-lived structures with reduced and excess radiation with respect to the mean TSI level varying with an 11-year cycle. These structures are associated with long-lived magnetic structures, activity complexes, and the macrostructure of facular fields. To study the large-scale organization of such thermal structures and to reveal the effect of activity complexes (Obridko 1985), we smoothed the time-longitude diagram with a window of three rotations by 40° in size. Figure 1d shows the smoothed distribution of TSI deviations relative to the slowly varying TSI component. In this diagram, regions whose activity complexes have lifetimes of 0.5–1 year show stand out as dark spots. Between these regions, we can see areas with excess radiation; they are shown in light tones and are apparently associated with the macrostructure of facular fields. These hot areas are clearly seen at the epochs of minimum activity, when there are no sunspots.

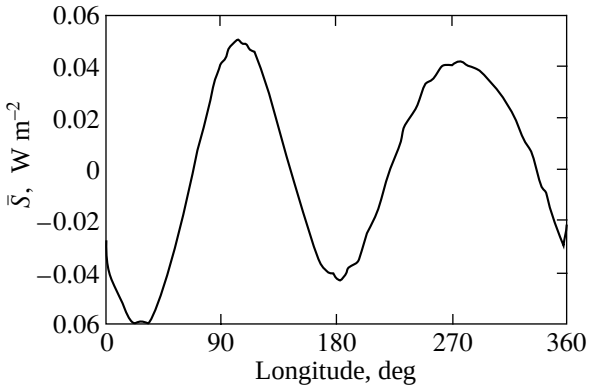


Fig. 2. The distribution of brightness inhomogeneities averaged for 1978–2000.

In cycles 21 and 22, alternating two- and four-sector structures are seen in the longitude distribution of thermal inhomogeneities. The four-sector structures generate a 13.5-day periodicity in TSI variations. For example, in the first half of 1984, when two deep dips in TSI were observed, there was a distinct four-sector structure in the time-longitude distribution of thermal inhomogeneities; precisely at this time, an appreciable 13-day variation appeared in the TSI wavelet spectrum (Willson and Mordvinov 1999). A continuous wavelet analysis of TSI variations has also shown that during the primary and secondary maxima of solar activity, a transition of spectral energy to longer time scales occurs (Willson and Mordvinov 1999). Interestingly, these cascades take place when the four-sector structure is observed in the distribution of brightness inhomogeneities, and they may be associated with cells with a size of 90° in longitude.

To study the behavior of thermal inhomogeneities in heliographic longitude on a long time scale, we averaged the distribution of brightness inhomogeneities (see Fig. 1c) over the entire interval of observations. Figure 2 shows the result of this averaging, which demonstrates the existence of two sectors with excessive radiation in the distribution of brightness in heliographic longitude. The TSI variations due to spatial brightness inhomogeneities do not exceed ±0.06 W/m^2. Remarkably, the maxima in the brightness distribution occur at longitudes of 103° and 277°, which are located within or in the immediate vicinity of active longitudes. During cycles 21 and 22, the most stable active longitudes existed in the intervals 80°–120° and 280°–320° (Vitinskiĭ 1997). Accordingly, in the time-longitude distribution of TSI variations, long-lived structures with excessive and reduced radiation are often located within or near these intervals, but, in general, the contribution of the regions with excessive radiation overcomponents for the deficit of radiation produced by sunspots.

THE ROTATION OF MAGNETO-THERMAL STRUCTURES

The structures that are horizontally arranged in the time-longitude diagrams are characterized by the Carrington rotation velocity. The regions of excessive and reduced radiation sometimes form inclined structures in the time-longitude diagrams, pointing to the presence of various modes of rigid rotation with periods different from the Carrington period. At the epochs of minimum activity, the regions with excess radiation exhibit a systematic inclination, suggesting that their rotation is slower than the Carrington rotation. By contrast, the oppositely inclined diagonal structures observed at the beginning of cycle 22 suggest that the regions of reduced radiation rotate more slowly.

When the inclination of the structures and their rotation velocities are estimated, it does not matter which events (in amplitude) recurred; therefore, to represent the time–longitude distribution of brightness inhomogeneities at different activity levels in a comparable form, we normalized the filtered TSI component to its maximum value within each rotation. Figure 1e shows the time–longitude diagram for the normalized TSI component. This distribution clearly reveals the rotational effects, irrespective of the phase of activity cycle (Mordvinov and Plyusnina 2000). At the epoch of growing activity in 1988–1989, the regions with reduced radiation form inclined structures associated with activity complexes, whose rotation periods are 27.8 and 26.4 days. At the epoch of minimum activity in 1985–1987 and 1995–1996, the regions with excess radiation dominate and exhibit a faster rotation with periods of 26.1, 26.6, and 26.8 days.

CONCLUSION

Our time–longitude analysis of variations in total solar irradiance has revealed large-scale structures on the Sun with reduced and excess radiation relative to the mean level, which varies with an 11-year cycle. During cycles 21–23, giant thermal spots, which were separated in heliographic longitude and associated with long-lived magnetic structures, existed on the Sun. These structures were associated with active longitudes, activity complexes, and the macrostructure of facular fields. The large-scale magneto-thermal perturbations are characterized by a two- and four-sector structure in heliographic longitude. Although the distribution of brightness inhomogeneities is complex and nonstationary in pattern, averaging this distribution over the entire interval of observations clearly reveals a concentration of the regions with excessive radiation within active longitude intervals; the maxima of this distribution occur at longitudes of 103° and 277°. The thermal perturbations associated with activity complexes exhibit rigid rotation with an angular velocity different from the Carrington velocity. The regions of excessive radiation trace the macrostructure of facular fields and may be attributable to the relaxation of large-scale thermo-magnetic perturbations and/or to the heat flux through giant subphotospheric convection.

ACKNOWLEDGMENTS

We are grateful to the staff of the World Data Center in Switzerland for the opportunity to work with the TSI data retrieved via the Internet (http://www.pmodwrc.ch). We thank V.M. Grigor'ev and L.L. Kitchatinov for a discussion and helpful remarks. This work was supported by the Russian Foundation for Basic Research (project no. 99-02-16088) and the Program of State Support for Leading Scientific Schools (project no. 00-15-96659). The NASA provided support for Dr. Willson at Columbia University under contact NAS5-97164.

REFERENCES

1. N. M. Astaf'eva, Usp. Fiz. Nauk **166**, 1145 (1996) [Phys. Usp. **39**, 1085 (1996)].
2. T. Bai, Astrophys. J. **328**, 860 (1988).
3. V. Bumba and L. Hejna, Bull. Astron. Inst. Czech. **42**, 76 (1991).
4. H. W. Dodson and E. R. Hedeman, in *Structure and Development of Solar Active Regions*, Ed. by K. O. Kipenheuer (Reidel, Dordrecht, 1968), p. 56.
5. C. Fröhlich and J. Lean, Geophys. Res. Lett. **25**, 4377 (1998).
6. R. Howard and B. J. LaBonte, Astrophys. J. Lett. **239**, 33 (1980).
7. L. L. Kitchatinov, M. Jardine, and A. C. Cameron, Astron. Astrophys. (in press).
8. F. Laclare, C. Delmas, J. P. Coin, and A. Irbah, Sol. Phys. **166**, 211 (1996).
9. A. V. Mordvinov, Sol. Phys. **163**, 309 (1996).
10. A. V. Mordvinov and L. A. Plyusnina, Sol. Phys. **197**, 1 (2000).
11. V. N. Obridko, *Sunspots and Complex of Activity* (Nauka, Moscow, 1985).
12. V. V. Pipin and L. L. Kichatinov, Astron. Zh. **77**, 872 (2000) [Astron. Rep. **44**, 771 (2000)].
13. Yu. I. Vitinskiĭ, in *Modern Problems of Solar Cyclicity*, Ed. by V. I. Makarov and V. N. Obridko (Glavn. Astron. Obs. Ross. Akad. Nauk, St. Petersburg, 1997), p. 33.
14. Yu. I. Vitinskiĭ, M. Kopetskiĭ, and G. V. Kuklin, *Statistics of Sunspot Activity* (Nauka, Moscow, 1986).
15. R. C. Willson, http://www.acrim.com; http://acrim.jpl.nasa.gov (2000).
16. R. C. Willson and H. S. Hudson, Nature **351**, 42 (1991).
17. R. C. Willson and A. V. Mordvinov, Geophys. Res. Lett. **26**, 3613 (1999).

Translated by G. Rudnitskiĭ

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 455–463. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 533–542.

Orbital Evolution of Saturn's New Outer Satellites and Their Classification

M. A. Vashkov'yak*

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

Received February 22, 2001

Abstract—Based on data for twelve recently discovered outer satellites of Saturn, we investigate their orbital evolution on long time scales. For our analysis, we use the previously obtained general solution of Hill's double-averaged problem, which was refined for libration orbits, and numerical integration of the averaged system of equations in elements with allowance for Saturn's orbital evolution. The following basic quantitative parameters of evolving orbits are determined: extreme eccentricities and inclinations, as well as circulation periods of the pericenter arguments and of the longitudes of the ascending nodes. For four new satellite orbits, we have revealed the libration pattern of variations in pericenter arguments and determined the ranges and periods of their variations. Based on characteristic features of the orbits of Saturn's new satellites, we propose their natural classification. *© 2001 MAIK "Nauka/Interperiodica".*

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

1. INTRODUCTION. INPUT DATA

Reports on the discovery of Saturn's four new satellites appeared in late October 2000. Their authors are B. Gladman, J. Kavelaars, J.-M. Petit, H. Scholl, M. Holman, B.G. Marsden, P. Nicholson, and J. A. Burns (Marsden 2000a, 2000b). The same team of scientists headed by B. Gladman reported the discovery of another two satellites in mid-November 2000 (Marsden 2000c) and four more satellites in early December (Marsden 2000d, 2000e). Finally, data on the orbits of the last pair of Saturn's new satellites appeared in late December of the same year (Marsden 2000f, 2000g). Thus, the total number of satellites discovered in the last three months of 2000 was twelve; they were tentatively designated as S/2000 S1, S2, …, S12.

The orbital parameters of all twelve satellites were refined immediately after their discovery (Marsden 2000f–2000i). Table 1 gives the orbital elements of Saturn's new satellites required to compute the orbital evolution; they were taken from the Internet source at http://cfa-www.harvard.edu/iau/mpc.html with MPEC (Minor Planet Electronic Circular) numbers 2000-Y13, 14, 15, and 33. The semimajor axis a is given in astronomical units (AU), and the angular Keplerian elements in the standard notation i, ω, and Ω (in degrees) refer to the ecliptic and equinox 2000.0. For all twelve satellites, the initial epoch is the same, 2001 Apr. 1.0 TT = JDT 2 452 000.5. For brevity, the satellites are designated as S1, S2, …, S12.

Thus, Saturn turned out to possess a large family of outer satellites. This family can naturally be divided into two groups:

I—Satellites with direct motion ($i < 90°$):

S2, S3, S4, S5, S6, S10, S11;

II—Satellites with retrograde motion ($i > 90°$):

S1, S7, S8, S9, S12, Phoebe.

As we show below in Section 3, group I can be divided by the pattern of orbital evolution into two subgroups:

IC—Satellites moving in orbits with circulation variations in pericenter argument ω: S4, S10, S11;

IL—Satellites moving in orbits with libration variations in ω: S2, S3, S5, S6.

Below, we give numerical values of Saturn's basic parameters (Jacobson 1998):

$\mu = 37\,940\,629.764$ km^3 s^{-2}—Saturn's gravitational constant;

$a_0 = 60\,330$ km—Saturn's mean equatorial radius;

$c_{20} = -0.016298$—coefficient of the second zonal harmonic of its attraction potential.

In addition, we take the following parameters:

$\mu' = 132\,712\,442\,007$ km^3 s^{-2}—solar gravitational constant;

$a' = 1\,427\,014\,000$ km—semimajor axis of Saturn's orbit;

and the parameters that specify Saturn's orbital orientation relative to the ecliptic plane:

inclination $i' = 2\overset{\circ}{.}4871$ and

* E-mail address for contacts: vashkov@spp.keldysh.ru

Table 1. Orbital elements for Saturn's new satellites S/2000 S1–S12

Satellite	a	e	i, deg	ω, deg	Ω, deg
S1	0.1526229	0.3665803	172.79616	39.90853	206.33370
S2	0.1006928	0.4585252	46.17993	239.33838	352.04333
S3	0.1102672	0.2934525	48.65572	60.80548	63.73841
S4	0.1192466	0.6345960	34.96993	284.71111	94.51408
S5	0.0754791	0.1578260	48.45127	91.35673	352.00016
S6	0.0759049	0.3672667	49.32856	70.52638	151.27029
S7	0.1320295	0.5439151	174.95400	89.16654	246.53051
S8	0.1045683	0.2140774	148.56133	208.62059	285.02467
S9	0.1240056	0.2539976	169.59340	298.20537	79.78166
S10	0.1225551	0.6143584	33.28509	287.75284	141.94939
S11	0.1185613	0.3870307	34.88605	73.00111	107.27236
S12	0.1191037	0.0866045	174.76580	8.12139	252.71210

longitude of the ascending node $\Omega' = 113°.35$ for initial epoch t_0.

Recall that only one outer (or distant) satellite of Saturn, Phoebe, which was discovered by W.H. Pickering slightly more than a hundred years ago, was known before October 2000. For comparison, we point out that the semimajor axis of its orbit is ≈0.0865 AU, the eccentricity is 0.163, and the inclination to the ecliptic plane is 175°. Based on long-term ground-based observations and on Voyager-2 data, Jacobson (1998) deduced a system of mean elements that represented Phoebe's orbit. Previous studies of the motion of Phoebe and Saturn's other satellites were comprehensively reviewed by Nasonova (1991).

2. STATEMENT OF THE EVOLUTIONAL PROBLEM AND METHODS OF ITS SOLUTION

Let us consider the motion of Saturn's distant satellites, S1–S12. Based on the estimates made when Phoebe's motion was investigated, we assume that for the newly discovered satellites, the principal perturbing factor that far exceeds all the remaining factors is solar attraction. Exceptions are the most distant satellites S1 and S7, for which the perturbing effect of Jupiter can be appreciable and can account for a few percent of the solar perturbation. The perturbing effect of other planets, Saturn's inner satellites, and its oblateness on satellites S1–S12 is several orders of magnitude weaker than the effect of solar attraction.

Thus, we arrive at the model that is most suitable for analyzing the orbital evolution of distant planetary satellites under the effect of secular perturbations from the Sun. This is Hill's problem, in which the perturbing function is averaged over all fast variables, the mean anomalies of the satellite and the perturbing body. In this way, all the short-period inequalities related to the orbital motions of satellites and the perturbing body are excluded from the analysis. In the averaged statement, this problem has the three first integrals

$$\begin{aligned} a &= c_0, \quad (1-e^2)\cos^2\tilde{i} = c_1, \\ e^2&(2/5 - \sin^2\tilde{i}\sin^2\tilde{\omega}) = c_2, \end{aligned} \tag{1}$$

in which the satellite orbital inclination $\tilde{i}$ and pericenter argument $\tilde{\omega}$ refer to the orbital plane of the perturbing body. Integrals (1), which were derived by Lidov (1961), allow a comprehensive qualitative analysis of this integrable problem to be performed. Its general solution yields dependences of the satellite orbital elements on time and on four arbitrary constants, which are initial values of the elements at $t = t_0$ (Vashkov'yak 1999). However, this solution should be supplemented with two formulas that must be used for $c_2 < 0$ (in the case of libration variations in $\tilde{\omega}$). More specifically, for $c_2 < 0$, in addition to the variations described in subsection 3.4.2 of our previous paper (Vashkov'yak 1999), the following formulas should be used to calculate integrals **I** and J:

$$\begin{aligned} \mathbf{I}(l^2, k^2) &= \mathbf{\Pi}(l^2, k^2), \\ J(x, l^2, k^2) &= \Pi(\varphi, l^2, k^2), \end{aligned} \tag{2}$$

where **Π** and Π are, respectively, the complete and incomplete elliptic integrals of the third kind with modulus k and parameter $l > k$, $\sin\varphi = \mathrm{sn}x$.

For the general solution of Hill's double-averaged problem to be properly used, we must first transform the system of ecliptic elements i, ω, and Ω to the system of elements $\tilde{i}$, $\tilde{\omega}$, and $\tilde{\Omega}$; the latter are related to the orbital plane of the perturbing body (Sun) and to the direction of the line of its intersection with the ecliptic plane. This transformation is performed in a standard way for specified angles i' and Ω', which determine the orientation of Saturn's orbital plane with respect to the ecliptic plane. Subsequently, the elements of the perturbed satellite orbit are calculated for an arbitrary instant of time in an interval of the order of a thousand years by using the analytic formulas of Hill's double-

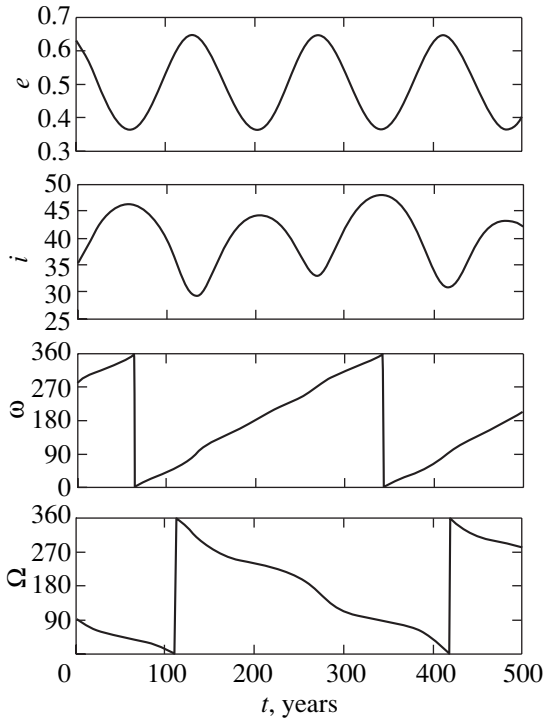


Fig. 1. Orbital evolution of S/2000 S4.

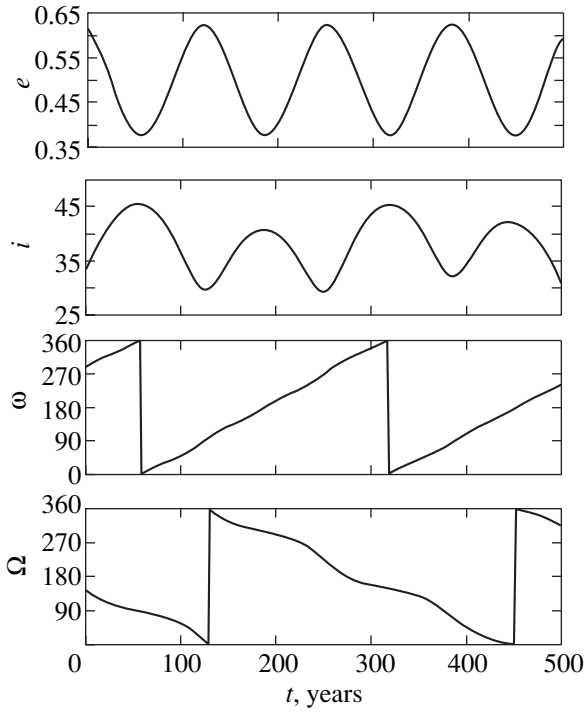


Fig. 2. Orbital evolution of S/2000 S10.

averaged problem. The inverse transformation from $\tilde{i}$, $\tilde{\omega}$, and $\tilde{\Omega}$ to the ecliptic elements i, ω, and Ω is performed for the same initial values of i' and Ω'. Since Saturn's orbital elements vary slowly (the characteristic period is about 50 000 years), we may roughly assume that i' = const and Ω' = const when analyzing the evolution of satellite orbits on time scales no longer than a thousand years.

To check the results, we numerically integrated the fuller averaged system of secular equations by taking into account the ellipticity of Saturn's orbit, its noncoplanarity with the ecliptic plane, and its evolution because of the secular perturbations described by the Lagrange–Brower–Wurkom theory (Sharaf and Budnikova 1967). In contrast to the analytic solution, the next term [after Hill's term $\sim(a/a')^2$] in the expansion of the perturbing function in terms of Legendre polynomial $\sim(a/a')^3$ is also taken into account in numerical integration. Since the secular part of this term also contains $e'\cos(\omega' + \Omega')$ and $e'\sin(\omega' + \Omega')$ as the factors, it simultaneously includes Saturn's orbital ellipticity. As we show in subsection 3.3, this additional term of the perturbing function significantly affects the change in orbital eccentricities of the group II satellites even on time scales of ~1000 years. Numerical integration on a considerably longer time scale of 50 000 years for the orbits of the group IL satellites makes it possible to estimate the effect of spatial evolution of Saturn's orbit and its ellipticity (compared to Hill's averaged problem) on the amplitudes of variations in elements.

3. THE EVOLUTION OF SATELLITE ORBITS

3.1. Group IC

This group includes satellites S4, S10, and S11. Their orbital evolution is characterized by a monotonic increase in pericenter argument, by a monotonic decrease in longitude of the ascending node, and by long-period variations in eccentricity and inclination.

Figures 1–3 show the time dependences of orbital elements for satellites S4, S10, and S11 that were derived by analytically solving Hill's double-averaged problem (the ecliptic angular variables i, ω, and Ω are given in degrees). For the satellites of this group, the results of numerical integration of the fuller evolutional system in a 500-year interval match the analytic results with a graphical accuracy; therefore, they are not highlighted in the figures.

Table 2 gives basic parameters of evolving orbits: constants c_1 and c_2 of the integrals, extreme eccentricities and inclinations, circulation periods of the pericenter arguments, $T_{\tilde{\omega}}$, and of the longitudes of the ascending nodes, $T_{\tilde{\Omega}}$. In contrast to Figs. 1–3, all the angular variables used to calculate these parameters refer to Saturn's orbital plane. The periods of variations in e and $\tilde{i}$ are half the circulation period of $\tilde{\omega}$.

3.2. Group IL

This (apparently) most interesting group includes satellites S2, S3, S5, and S6. Their orbital evolution is

Table 2. Parameters of evolving orbits for the group IC satellites

Satellite	c_1	c_2	$e_{\min}$	$e_{\max}$	$\tilde{i}_{\min}$, deg	$\tilde{i}_{\max}$, deg	$T_{\tilde{\omega}}$, years	$T_{\tilde{\Omega}}$, years
S4	0.424	0.053	0.365	0.646	31.5	45.7	283	329
S10	0.456	0.057	0.379	0.624	30.2	43.1	263	317
S11	0.606	0.020	0.226	0.404	31.7	37.0	351	434

Table 3. Parameters of evolving orbits for the group IL satellites

Satellite	c_1	$-c_2$	$e_{\min}$	$e_{\max}$	$\tilde{i}_{\min}$, deg	$\tilde{i}_{\max}$, deg	$T_{\tilde{\omega}}$, years	$T_{\tilde{\Omega}}$, years	$\Delta\tilde{\omega}$, deg
S2	0.360	0.0054	0.153	0.621	40.0	52.6	330	528	29.8
S3	0.424	0.0025	0.120	0.532	39.7	49.0	353	510	26.7
S5	0.407	0.0045	0.156	0.553	40.1	49.8	550	849	25.8
S6	0.397	0.0092	0.225	0.551	41.0	49.7	468	775	21.6

characterized by libration variations in pericenter argument, by a monotonic decrease in longitude of the ascending node, and by long-period variations in eccentricity and inclination.

In Fig. 4–7, which are similar to Figs. 1–3, the numerical results match the analytic results with a graphical accuracy in a 1000-year interval.

Basic parameters of evolving orbits for the satellites of this group are given in Table 3; it is similar to Table 2, but it is supplemented with the amplitudes of variations in pericenter arguments $\tilde{\omega}$. The periods of variations in e and $\tilde{i}$ are equal to the libration period of $\tilde{\omega}$.

Note that all orbits of the satellites from this group are characterized by a nearly zero constant c_2. In the integrable problem, this means that the phase trajectories are located in the libration regions that are "not very deep," i.e., near the separatrixes, the lines that separate these regions from the circulation regions of $\tilde{\omega}$. It is of interest to establish whether the qualitative behavior of the solution will change if we take into account the spatial evolution of Saturn's orbit and additional terms of the perturbing function $\sim(a/a')^3e'$ on a long time scale. To this end, we numerically integrated the averaged system on a time scale of 50 000 years, an approximate period of variations in Saturn's orbital elements.

In Figs. 8 and 9, the solid lines indicate the projections of phase trajectories of the integrable problem onto the pericenter argument—eccentricity and pericenter argument—inclination planes, respectively, and

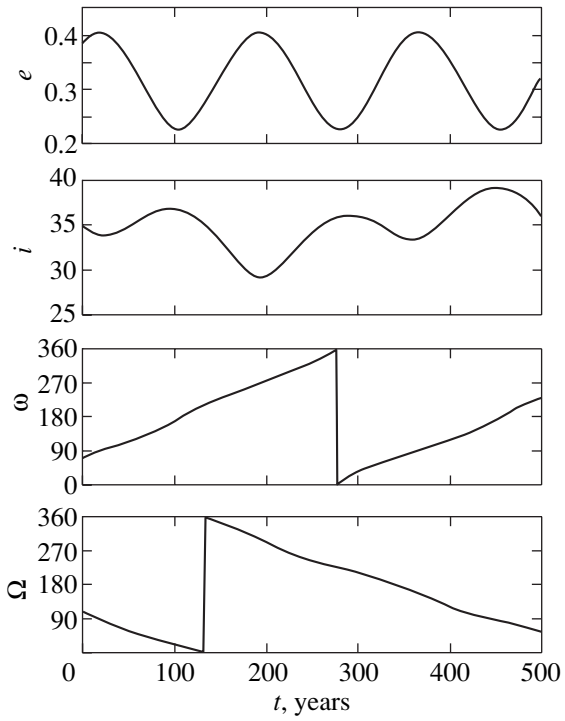


Fig. 3. Orbital evolution of S/2000 S11.

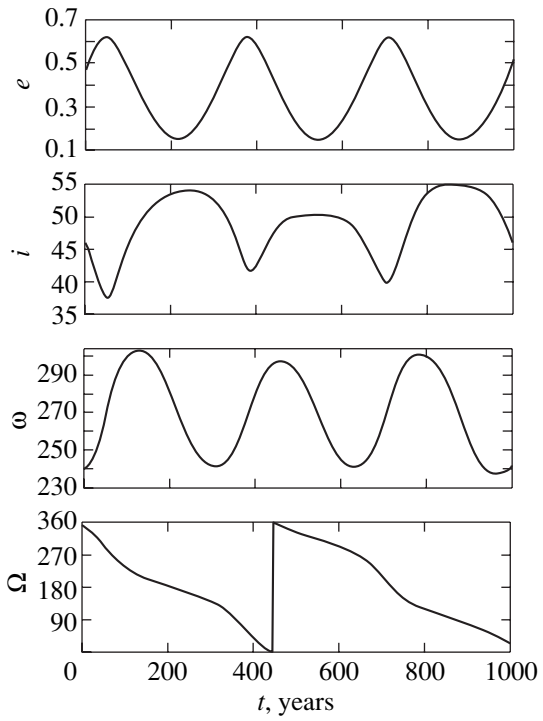


Fig. 4. Orbital evolution of S/2000 S2.

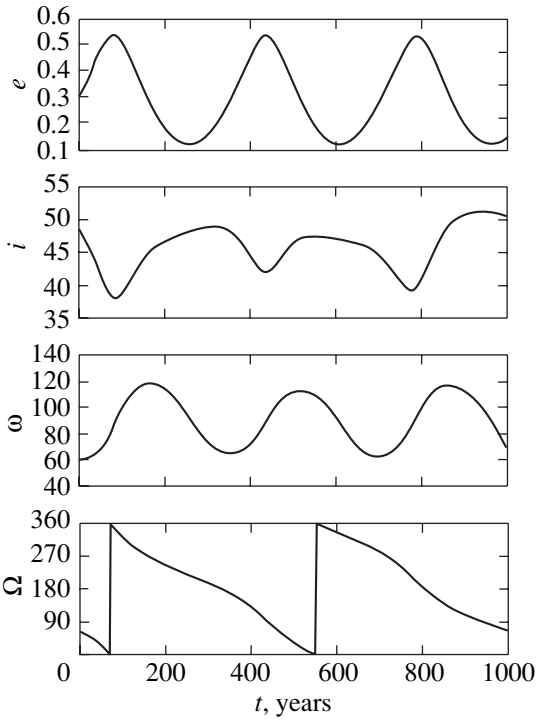


Fig. 5. Orbital evolution of S/2000 S3.

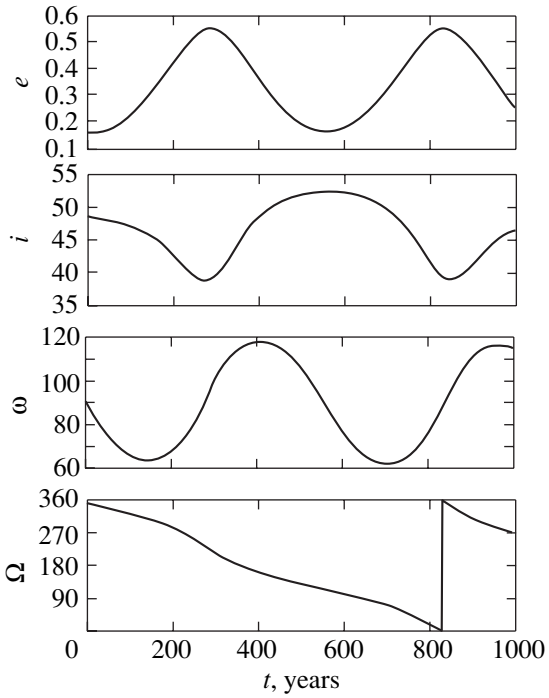


Fig. 6. Orbital evolution of S/2000 S5.

the dashed lines represent the separatrixes. On these curves, the inclination $\tilde{i}$ and the pericenter argument $\tilde{\omega}$ refer to Saturn's orbital plane. The small crosses mark the values of $\tilde{i}$ and $\tilde{\omega}$ obtained by numerical integration. The crosses are plotted at time steps of 50 years, with $\tilde{i}$ and $\tilde{\omega}$ being associated with the ecliptic plane. All angular variables are given in degrees. We see from Figs. 8 and 9 that the libration pattern of variations in pericenter arguments is also preserved in the ecliptic coordinate system with allowance for the above additional perturbing factors. It would therefore be natural to call the orbits of this group, by standard terminology, ω-librators. It should be noted that libration orbits are encountered rarely even among the many thousands of asteroid orbits. Therefore, the fact that four of the thirteen orbits of Saturn's outer satellites (including Phoebe) proved to be libration orbits is surprising enough.

Of course, the libration pattern of satellite orbits from this group is entirely determined by initial values of the elements from Table 1 and $\tilde{i}$, and $\tilde{\omega}$. Therefore, their further refinement can transfer any group IL satellite into group IC with $c_2 > 0$. However, this is possible when i and ω were refined significantly, by ~10^{-2}–10^{-3}, because c_2 for the group IL satellites varies over the range –0.0092 to –0.0025 (see Table 3).

Note: The change from ω libration to circulation at a nearly zero c_2 can also be caused in principle by secular resonances (commensurability of the periods of variations in slowly evolving orbits of the perturbed and perturbing bodies). Such cases are known among asteroid orbits. The chaotic orbit of asteroid (2335) James, which changes from one mode of variations in ω to the other and back on time scales of several hundred thousand years (Vashkov'yak 1986; Froeschle *et al.* 1991) can serve as an example. However, in the satellite problem under consideration, there are appar-

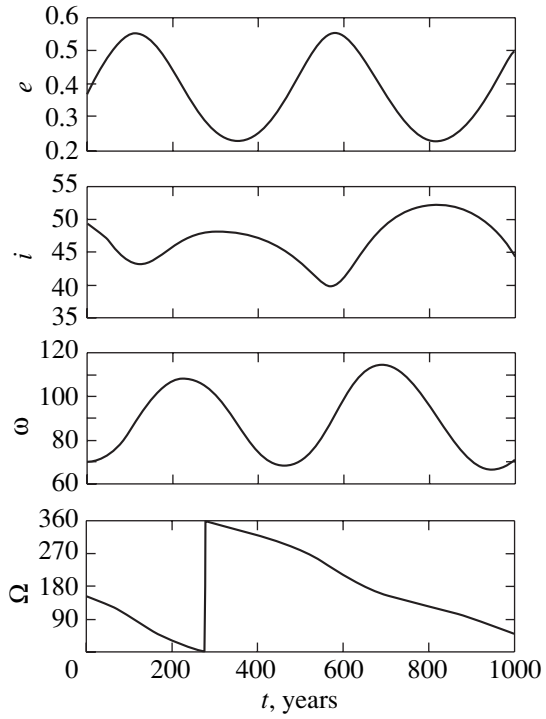


Fig. 7. Orbital evolution of S/2000 S6.

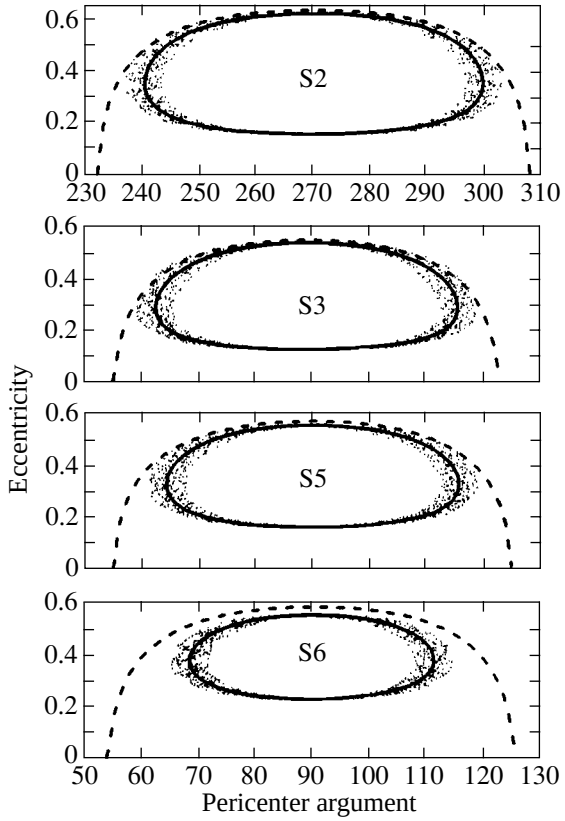


Fig. 8. Projections of libration phase trajectories onto the pericenter argument–eccentricity plane.

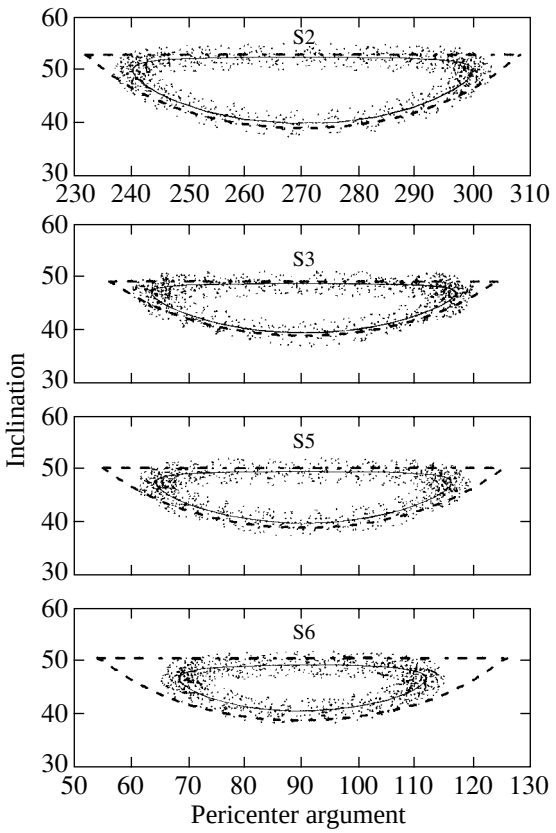


Fig. 9. Projections of libration phase trajectories onto the pericenter argument–inclination plane.

ently no such resonant perturbations with periods of the order of the libration periods of ω for the orbits of S2, S3, S5, and S6.

3.3. Group II

This group includes satellites S1, S7, S8, S9, S12, and Phoebe. Their orbital evolution is characterized by a monotonic increase in pericenter arguments and in longitudes of the ascending nodes. The periods of variations in eccentricities are approximately half the circulation periods of the pericenter arguments, except for the nearly circular orbit of satellite S12. The principal harmonic of the variations in eccentricity of this orbit is the harmonic with the circulation period of the ascending node. It also dominates for the variations in inclinations of all orbits from this group of satellites.

Figures 10–15 show time dependences of the orbital elements on a time scale of 500 years (1000 years for Phoebe). The solid lines correspond to our numerical solution of the full evolutional system, and the filled circles denote the elements derived by analytically solving Hill's double-averaged problem. For satellite S8, these dependences coincide with a graphical accuracy (Fig. 12). For the remaining satellites of this group, this coincidence takes place in all elements except the eccentricity. This is because the effect of the perturbing factors that distort the solution of Hill's double-averaged problem is stronger for evolving orbits with small differences $e_{max} - e_{min}$ (Figs. 10, 11, 13, and 15). The

Table 4. Parameters of evolving orbits for the group II satellites

Satellite	c_1	c_2	e_{min}	e_{max}	$\tilde{i}_{min}$, deg	$\tilde{i}_{max}$, deg	$T_{\tilde{\omega}}$, years	$T_{\tilde{\Omega}}$, years
S1	0.851	0.0535	0.365	0.373	172.3	173.5	152	260
S7	0.701	0.1173	0.541	0.546	174.9	176.4	170	261
S8	0.730	0.0160	0.200	0.286	150.7	153.1	380	560
S9	0.891	0.0237	0.243	0.260	166.8	167.8	226	405
S12	0.988	0.0030	0.085	0.087	176.2	176.3	238	470
Phoebe	0.966	0.0110	0.165	0.167	175.2	175.4	379	730

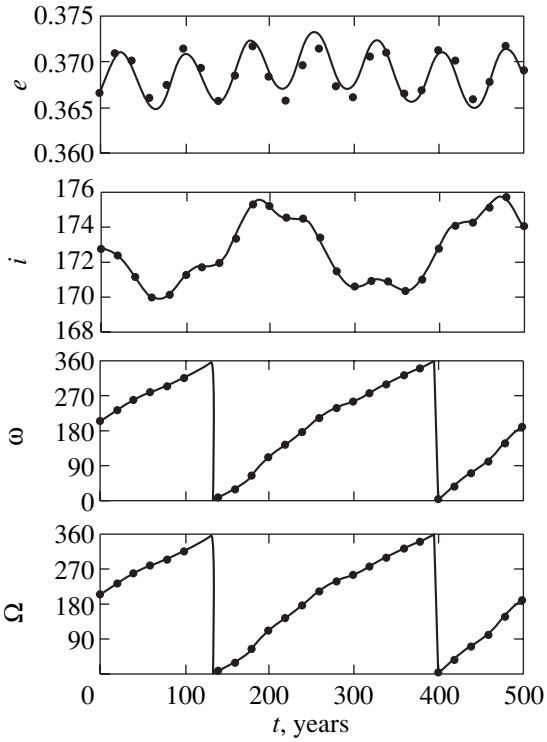


Fig. 10. Orbital evolution of S/2000 S1.

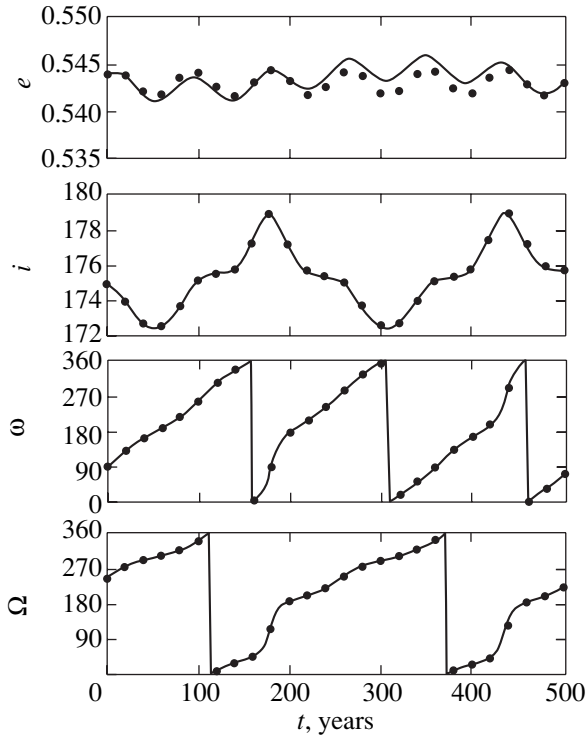


Fig. 11. Orbital evolution of S/2000 S7.

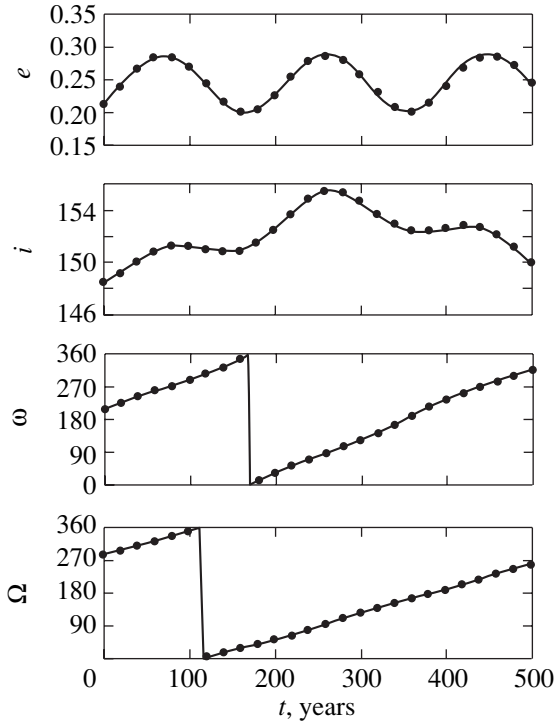


Fig. 12. Orbital evolution of S/2000 S8.

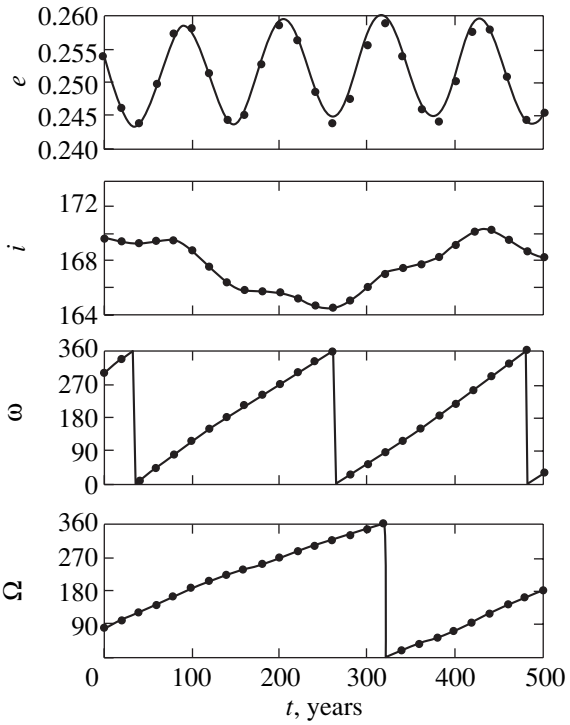


Fig. 13. Orbital evolution of S/2000 S9.

discrepancy between the results is at a maximum for the nearly circular orbit of satellite S12 (Fig. 14).

The basic parameters of evolving orbits for the satellites of this group, which is naturally called the Phoebe group, are given in Table 4. However, the extreme eccentricities (in contrast to Tables 2 and 3) correspond to the results of the numerical integration of the fuller system of secular equations.

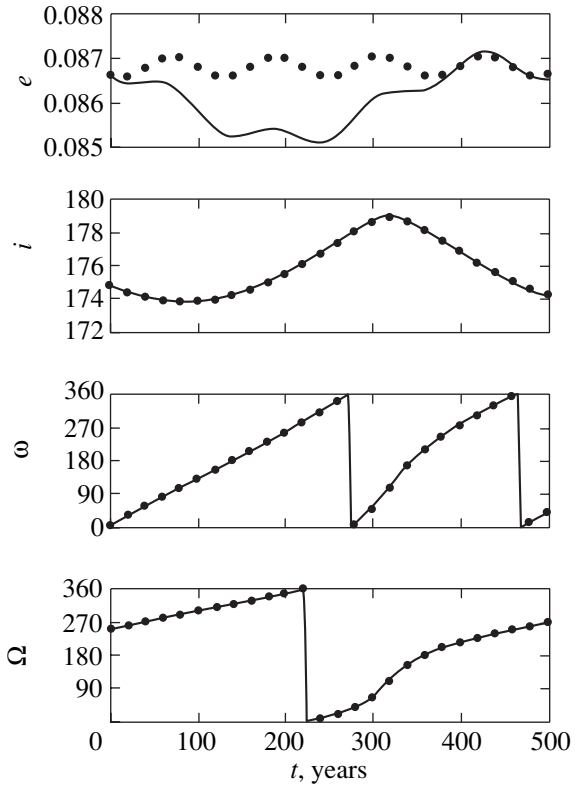


Fig. 14. Orbital evolution of S/2000 S12.

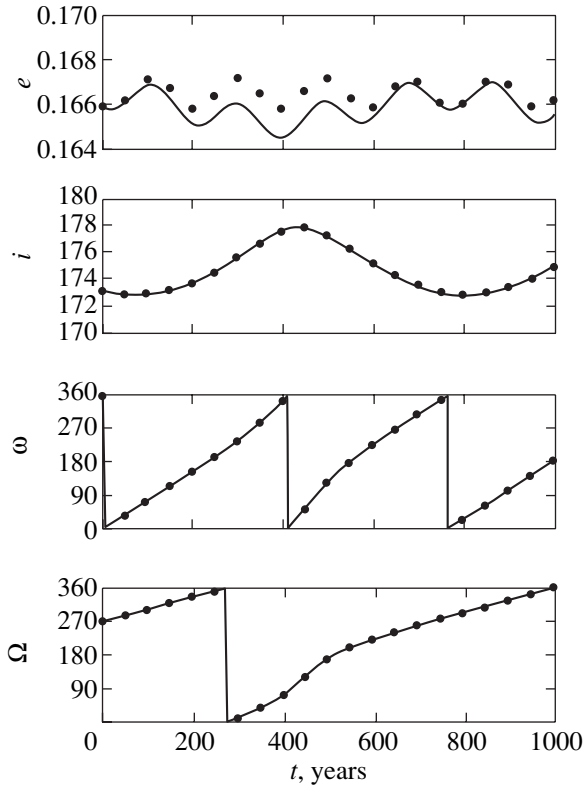


Fig. 15. Orbital evolution of Phoebe.

As was already noted in the introduction, the orbits of satellites S1 and S7 are subject to appreciable perturbations from Jupiter, which we ignore in the model used. Therefore, the numerical data for these satellites (see Table 4) give only approximate parameters of their orbital evolution.

4. CONCLUSION

We have described the principal qualitative features of the orbital evolution of Saturn's recently discovered outer satellites. The approximate quantitative evolutional parameters were derived in terms of the theory of secular perturbations by analytically solving and numerically integrating the system of equations averaged over fast variables of the problem: the mean anomalies of the satellite and the perturbing body (Sun). The largest of the disregarded secular perturbations are those from Jupiter, which reach several percent of the solar perturbations for the two most distant satellites considered. Next in magnitude are the secular perturbations from Saturn's largest satellite, Titan. They are similar in structure to the perturbations of planetary oblateness and can be taken into account in the main approximation by a formal (fourfold) increase in the coefficient of the second zonal harmonic of Saturn's gravitational field. The subsequent application of standard formulas for the secular perturbations from oblateness with a changed coefficient c_{20} will result only in small (~0.001 relative to the solar perturbations) corrections to the evolution of the pericenter arguments and longitudes of the ascending nodes referred to Saturn's equator.

Based on our analysis, we have proposed a simple and natural classification of the orbits of Saturn's outer satellites by the direction of their motion and by the pattern of their orbital evolution. Although the discovered ω-librators are located near the separatrixes of the integrable problem, they apparently retain their qualitative properties even when the disregarded perturbations are added to the model. However, these orbits can pass into the group of circulation orbits when their elements will be refined in the future.

ACKNOWLEDGMENTS

I wish to thank G.R. Kastel for the assistance in obtaining information about Saturn's new satellites and N.M. Teslenko for help. This work was supported 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. Ch. Froeschle, A. M. Morbidelli, and H. Scholl, Astron. Astrophys. **249**, 553 (1991).
2. R. A. Jacobson, Astron. Astrophys., Suppl. Ser. **128**, 7 (1998).

3. M. L. Lidov, Iskusstv. Sputniki Zemli **8**, 5 (1961).
4. B. G. Marsden, IAU Circ., No. 7512 (2000a).
5. B. G. Marsden, IAU Circ., No. 7513 (2000b).
6. B. G. Marsden, IAU Circ., No. 7521 (2000c).
7. B. G. Marsden, IAU Circ., No. 7538 (2000d).
8. B. G. Marsden, IAU Circ., No. 7539 (2000e).
9. B. G. Marsden, Minor Planet Electronic Circ., No. 2000-Y13 (2000f).
10. B. G. Marsden, Minor Planet Electronic Circ., No. 2000-Y33 (2000g).
11. B. G. Marsden, Minor Planet Electronic Circ., No. 2000-Y14 (2000h).
12. B. G. Marsden, Minor Planet Electronic Circ., No. 2000-Y15 (2000i).
13. L. P. Nasonova, Itogi Nauki Tekh., Ser. Issled. Kosmich. Prostranstva **35**, 70 (1991).
14. Sh. G. Sharaf and N. L. Budnikova, Byull. Inst. Teor. Astron. Akad. Nauk SSSR **11**, 231 (1967).
15. M. A. Vashkov'yak, Kosm. Issled. **24**, 323 (1986).
16. M. A. Vashkov'yak, Pis'ma Astron. Zh. **25**, 554 (1999) [Astron. Lett. **25**, 476 (1999)].

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 464–469. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 543–548.

Celestial-Mechanical Peculiarities of Uranus's Satellite System

M. A. Vashkov'yak*

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

Received January 10, 2001

Abstract—We consider the structural peculiarities of Uranus's satellite system associated with its separation into two groups: inner equatorial satellites moving in nearly circular orbits and distant irregular satellites with retrograde motion in highly elliptical orbits. The intermediate region is free from satellites in a wide range of semimajor axes. By analyzing the evolution of satellite orbits under the combined effect of solar attraction and Uranus's oblateness, we offer a celestial-mechanical explanation for the absence of equatorial satellites in this region. M.L. Lidov's studies during 1961–1963 have served as a basis for our analysis. *© 2001 MAIK "Nauka/Interperiodica".*

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

1. INTRODUCTION

In recent years, the discovery of Uranus's five distant satellites—Caliban, Sycorax, Prospero, Setebos, and Stephano—has considerably extended our knowledge of its satellite system. With their discovery, Uranus joined the ranks of the remaining giant planets, which have a subsystem of close (or inner) satellites with nearly circular equatorial orbits and distant (or outer) satellites with highly elliptical orbits significantly inclined to the ecliptic plane and to the equatorial planes of the planets. Uranus's rings and the orbits of its inner satellites are located at distances from about 40 000 to 600 000 km from the center of this planet with a radius of 26 200 km. The orbits of distant satellites have semimajor axes from about 7 to 18 million km. Thus, the range of semimajor axes 0.6 million km $< a <$ 7 million km is free from satellite orbits. At the same time, when using currently available observational facilities and when conducting thorough systematic searches, which led to the discovery of Uranus's five distant satellites, the probability that satellites of appreciable sizes remain undiscovered in this range seems fairly low.

The possible mechanism of the avoidance of this intermediate region by satellites is therefore of considerable interest. Here, we elucidate this question by using the theory of secular perturbations of satellite orbits.

The numerous studies devoted to the dynamics of Uranus's rings and to the motion of its inner satellites were comprehensively reviewed, in particular, by Semenova (1991).

The orbits of inner satellites ($a < 600\,000$ km) evolve mainly under the effect of Uranus's oblateness and their mutual attraction. The resonant motion of Uranus's largest satellites also plays a major role in the evolution. By contrast, the orbital evolution of distant satellites with $a > 7$ million km is mainly governed by the influence of solar attraction (Vashkov'yak 1999, 2001). In the intermediate region, the effects of Uranus's oblateness and solar attraction may be comparable in magnitude, and the orbital evolution is determined by their combined secular perturbations.

The basic parameter that characterizes the ratio of perturbing accelerations from Uranus's oblateness and solar attraction is

$$\gamma(a) = 0.5\alpha(a)/\beta(a), \tag{1}$$

where

$$\alpha(a) = -\frac{3c_{20}a_0^2}{8a^2}, \tag{2}$$

$$\beta(a) = \frac{3\mu_S a^3}{16\mu_U a_U^3}, \tag{3}$$

c_{20} is the coefficient of the second zonal harmonic of Uranus's gravitational field; a_0 is its mean equatorial radius; μ_U and μ_S are the products of the gravitational constant by Uranus's and solar masses, respectively; and a_U is the semimajor axis of Uranus's heliocentric orbit. It follows from Fig. 1, in which $\log\gamma$ is plotted against a, that γ changes in the intermediate range 0.6 million km $< a <$ 7 million km from about 10^{-4} to 30.

* E-mail address for contacts: vashkov@spp.keldysh.ru

1063-7737/01/2707-0464$21.00 © 2001 MAIK "Nauka/Interperiodica"

To investigate the evolution of satellite orbits in this range, it would be natural to use Hill's standard averaged problem with allowance for oblateness of the central planet (Lidov 1963a, 1963b). Under the assumption of $a \ll a_U$, the averaging is performed over all fast variables—the mean longitudes of the satellite and the perturbing body. Apart from γ, the inclination I of the equator of an oblate planet to the orbital plane of the perturbing body is another important parameter in this problem. In general, when I is arbitrary, the problem is not integrable. Various special cases were qualitatively analyzed by Lidov and Yarskaya (1974). This problem was also used to investigate the orbital evolution of artificial satellites of the Moon (Kozai 1963) and the Earth (Kudielka 1994). Stationary and periodic solutions of the evolutional problem and their stability have been analyzed recently (Vashkov'yak 1998; Vashkov'yak and Teslenko 2001).

In the next section, we consider one of the integrable cases of the problem as applied to Uranus's satellite system with its peculiar features. Since the inclination of Uranus's equator to the plane of its heliocentric orbit (or to the plane of the Uranus-centric orbit of the Sun) is $\approx 97\overset{\circ}{.}8$, we can idealize the problem by setting $I = 90°$ for a qualitative analysis. In addition, for an equatorial satellite, we may assume with the same degree of roughness that its orbital inclination to Uranus's equatorial plane is zero and that, accordingly, its inclination to Uranus's orbital plane is 90°. Under these assumptions, Lidov (1963a, 1963b) explained the existence of Uranus's inner satellites in nearly circular orbits, which in the absence of planetary oblateness ($c_{20} = \gamma = 0$) must have turned into highly elongated ellipses with increasing eccentricities during their evolution (Lidov 1961). At a constant semimajor axis (a = const is the integral of the averaged problem), the pericenter distance $q = a(1 - e)$ would inevitably become equal to Uranus's radius, and, having fallen to its surface, the satellite would cease to exist. Since $a < 600\,000$ km for the actually oblate Uranus and $\gamma > 30$ (Fig. 1), the orbital eccentricity of Oberon, the inner satellite that is most perturbed by the Sun, undergoes only modest long-period variations (Lidov 1963a, 1963b). To put it simply, Uranus's oblateness saves the inner satellites from falling to its surface by retaining them in nearly circular orbits.

In the next section, we consider in more detail the problem of the evolution of Uranus's satellite orbits that lie in its equatorial plane by assuming that $I = 90°$.

2. MODEL PROBLEM (I = 90°)

2.1. The Results of a General Study

Let us consider the problem of orbital evolution for a satellite of an oblate planet whose rotation axis lies in its orbital plane, and, consequently, its equator is inclined to this plane at angle $I = 90°$. The evolutional

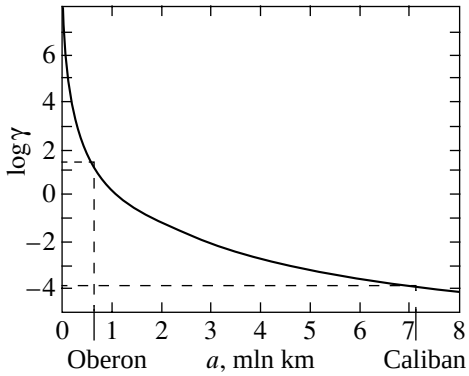


Fig. 1. Logarithmic dependence of γ on semimajor axis of Uranus's satellite orbit.

system that describes long-period variations in elements of the satellite orbit then takes the form

$$\frac{de}{d\tau} = 10e\sqrt{1-e^2}\sin^2 i \sin 2\omega,$$

$$\frac{di}{d\tau} = -10\frac{e^2}{\sqrt{1-e^2}}\sin i\cos i\sin 2\omega + \frac{8\gamma}{(1-e^2)^2}\sin i\sin\Omega\cos\Omega, \quad (4)$$

$$\frac{d\omega}{d\tau} = \frac{2\gamma}{(1-e^2)^2}[1-5\cos^2 i + (3-5\cos^2 i)\cos 2\Omega] + \frac{2}{\sqrt{1-e^2}}[5\cos^2 i - 1 + e^2 + 5(\sin^2 i - e^2)\cos 2\omega],$$

$$\frac{d\Omega}{d\tau} = \left[\frac{4\gamma}{(1-e^2)^2}(1+\cos 2\Omega) + \frac{2}{\sqrt{1-e^2}}(5e^2\cos 2\omega - 3e^2 - 2)\right]\cos i.$$

Here, e, i, ω, and Ω are the standard designations for the elements of a Keplerian orbit (eccentricity, inclination, argument of the pericenter latitude, and longitude of the ascending node); the angular elements refer to the orbital plane of the perturbing body and its line of intersection with the planet's equatorial plane. The normalized independent variable τ is related to time t by

$$\tau = \beta(a)n(a)(t - t_0), \quad (5)$$

where $n(a)$ is the mean motion of the satellite, which is a known function of the semimajor axis of its orbit, and $\beta(a)$ is given by Eq. (3).

Eqs. (4) can be taken from Lidov (1963a, 1963b) for I = 90°. The following averaged perturbing function serves as the only first integral of these equations:

$$W(e, i, \omega, \Omega) = 2(e^2 - \sin^2 i) + e^2(5\cos 2\omega - 3)\sin^2 i + \frac{2\gamma}{(1-e^2)^{3/2}}[1/3 - \cos 2i - 2\sin^2 i \sin^2 \Omega] = \text{const}. \quad (6)$$

Being not integrable, Eqs. (4) have the stationary solution

$$i^* = 90°, \quad \sin\Omega^* = 0, \quad \sin 2\omega^* = 0 \quad \text{(Lidov and Yarskaya 1974)}. \quad (7)$$

In this case, $\frac{de}{d\tau} = \frac{di}{d\tau} = \frac{d\Omega}{d\tau} = 0$, and the satellite's orbital plane lies in the planet's equatorial plane. The stationary eccentricity e^* can be determined from the condition $\frac{d\omega}{d\tau} = 0$ by using the formula

$$e^* = \sqrt{1 - \left(\frac{2\gamma}{5\sin^2\omega - 2}\right)^{2/5}}. \quad (8)$$

At $\omega^* = 0, 180°$ and $\gamma > 0$, no stationary solution exists. At $\omega^* = \pm 90°$ and $0 < \gamma < 3/2$, we have

$$e^* = \sqrt{1 - \left(\frac{2\gamma}{3}\right)^{2/5}}. \quad (9)$$

Linearizing system (4) in the vicinity of this stationary solution causes the equations to split into two independent pairs.

Since the system that describes the change $x_1 = e - e^*$ and $x_3 = \omega - \omega^*$ has the characteristic equation

$$\lambda^2 + 1200e^{*2} = 0, \quad (10)$$

the motion in the (ω, e) plane is stable in the linear approximation for $e^* > 0$.

For the system that describes the change $x_2 = i - i^*$ and $x_4 = \Omega - \Omega^*$, we have

$$\lambda^2 + 48(2 - 7e^{*2}) = 0. \quad (11)$$

Therefore, for $0 < e^* < e_b = \sqrt{2/7} \approx 0.5345$, the motion in the (Ω, i) plane is stable in the linear approximation, while for

$$e^* > \sqrt{2/7} \quad \left(\text{or} \quad \gamma < \gamma_b = \frac{3}{2}\left(\frac{5}{7}\right)^{5/2} \approx 0.6468\right) \quad (12)$$

it is unstable just as the solution in general. The cases with $e^* = 0$ and $e^* = \sqrt{2/7}$ are critical.

Note. Conditions (12) supplement our analysis of the stability of stationary solutions for Hill's averaged problem with allowance for planetary oblateness with the $i_0 = 90°$ vertical straight lines that pass through the bifurcation points $e_0 = \sqrt{2/7}$ and $\gamma = \frac{3}{2}\left(\frac{5}{7}\right)^{5/2}$ (Vashkov'yak 1998; Figs. 3a and 3b).

Next, we consider a particular (but not only stationary) solution of Eqs. (4)

At $i = i^* = 90°$ and $\Omega = \Omega^* = 0, 180°$, Eqs. (4), which describe the change in e and ω, and the first integral (6) take a simpler form:

$$\frac{de}{d\tau} = 10e\sqrt{1-e^2}\sin 2\omega,$$
$$\frac{d\omega}{d\tau} = \frac{8\gamma}{(1-e^2)^2} + 4\sqrt{1-e^2}(2 - 5\sin^2\omega), \quad (13)$$
$$e^2(2/5 - \sin^2\omega) + \frac{4\gamma}{15(1-e^2)^{3/2}} = C = \text{const}.$$

This integrable case was qualitatively analyzed by Lidov (1963a, 1963b) and Lidov and Yarskaya (1974; case V). For coherence, we present here the results of the above authors and derive some quantitative evolutional characteristics.

For $3/2 < \gamma < \infty$, the satellite's orbital evolution reduces to a monotonic increase in ω when the eccentricity varies in the range $e_{min} = e(\omega = 0, 180°) \le e \le e(\omega = \pm 90°) = e_{max}$. In this case of circulation variations in ω, the extreme e can be determined from the equation

$$\rho e^2 + \frac{4\gamma}{15(1-e^2)^{3/2}} = C_{circ}, \quad (14)$$

where $\rho = 2/5$ for e_{min} and $\rho = -3/5$ for e_{max}.

For $0 \le \gamma \le 3/2$, singularities emerge in the (ω, e) phase plane. The center-type points have $\omega^* = \pm 90°$, and e^* is given by Eq. (9). The saddle-type points have

$$e = 0, \quad \omega_s = \arcsin\sqrt{\frac{2}{5}(1+\gamma)}. \quad (15)$$

The singular trajectory (separatrix), which corresponds to the constant of the integral

$$C_s = 4\gamma/15, \quad (16)$$

separates the ω circulation ranges

$$C_s < C < \infty \quad (17)$$

from the ω libration range

$$C^* < C < C_s. \quad (18)$$

C^* corresponds to the stationary point (ω^*, e^*) and is given by

$$C^* = -\frac{3}{5} + \left(\frac{2\gamma}{3}\right)^{2/5}. \quad (19)$$

At $C = C_s$, the limiting motion takes place, for which $e_{min} = 0$ and $e_{max} = e(\omega = \pm 90°) = e_s$ is given by the equation

$$\left(\gamma + \frac{9}{4}e_s^2\right)(1 - e_s^2)^{3/2} = \gamma. \quad (20)$$

Whereas the extreme values of e in the circulation range (17) are given by Eq. (14) for the two values of ρ in the libration range (18), these are two different roots of the same equation for $\rho = -3/5$. For any $C^* \le C < \infty$, the phase trajectories in the (ω, e) plane are symmetric about the $\omega = 0, 180°$ and $\omega = \pm 90°$ straight lines.

The first integral yields a dependence of ω on e; substituting this dependence in the first equation (13) reduces the solution of the entire system to calculating the quadrature

$$\tau = \frac{3}{4}\mathrm{sgn}(\sin 2\omega_0)\int_{e_0}^{e}\frac{e(1-e^2)de}{\sqrt{f(e)}}, \quad (21)$$

where

$$f(e) = (1-e^2)^3(54e^4 - 45Ce^2 - 225C^2) + 12\gamma(e^2 + 10C)(1-e^2)^{3/2} - 16\gamma^2, \quad (22)$$

and e_0 and ω_0 are the initial elements at $\tau = 0$.

Since the integrand is complex, there is apparently no possibility for analytically representing $\tau(e)$ let alone the possibility for finding the inverse dependence $e(\tau)$ (even in the usually simpler case of limiting motion). Nevertheless, the periods of circulation,

$$T_c = 3\int_{e_{\min}}^{e_{\max}}\frac{e(1-e^2)de}{\sqrt{f(e)}} \quad (23)$$

and libration

$$T_l = \frac{3}{2}\int_{e_{\min}}^{e_{\max}}\frac{e(1-e^2)de}{\sqrt{f(e)}} \quad (24)$$

can be calculated numerically, for example, by the Gauss method.

Note. A slight simplification is achieved by introducing the variable

$$G = \sqrt{1-e^2}, \quad (25)$$

which is a canonical conjugate of ω. This allows us to reduce the radicand in Eq. (21) to rational form and the quadrature itself to the form

$$\tau = -\frac{3}{2}\mathrm{sgn}(\sin 2\omega_0)\int_{G_0}^{G}\frac{G^3 dG}{\sqrt{R(G)}}, \quad (26)$$

where $R(G)$ is a polynomial of the 10th degree for G, which is given by

$$R(G) = 216(1-G^2)^2G^6 + 12(1-G^2)(4\gamma - 15CG^3)G^3 - 4(4\gamma - 15CG^3)^2. \quad (27)$$

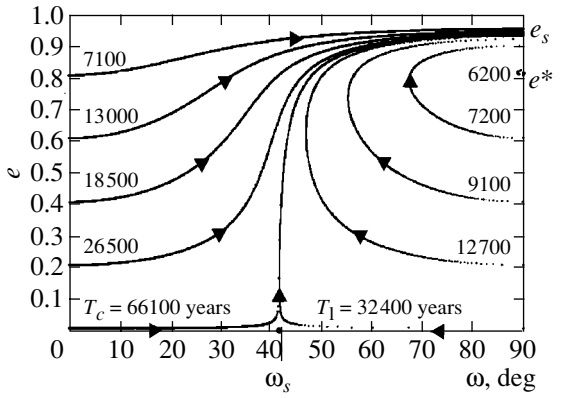


Fig. 2. Family of phase trajectories in the (ω, e) plane for $\gamma = 0.1$ (a = 1.88423 mln km).

2.2. Application to Uranus's Satellite System

Figure 2 illustrates the results described in the previous section. It shows a family of trajectories in the (ω, e) phase plane for $\gamma = 0.1$, which corresponds to the semimajor axis of the satellite orbit a = 1.88423 million km in the model system (Sun–Uranus-satellite). Since the trajectories are symmetric, only the range $0° \le \omega \le 90°$ is shown. The libration and circulation periods of ω (in years) with allowance for the normalizing relation (5) are given alongside the trajectories. For $C = C^*$ and $C \longrightarrow \infty$, these periods approach the periods of small variations, while for $C \longrightarrow C_s$, $T_l \longrightarrow T_c \longrightarrow \infty$. The arrows indicate the direction in which the phase point moves along the trajectories. Note that the family of phase trajectories shown in the figure corresponds to $e^* > e_b \approx 0.5345$, $\gamma < \gamma_b \approx 0.6468$, and $a > a_b = a(\gamma_b) \approx 1.297$ million km. According to conditions (12), it represents an unstable structure. In the linear approximation, however, small deviations of i and Ω from their equilibrium values i^* and Ω^* on time scales of the order of the periods of small variations in e and ω do not affect this family, because the equations in variations split into two independent systems.

Note that during the evolution, the shape of an equatorial satellite orbit significantly changes, while the distance of its pericenter $q = a(1-e)$ inevitably becomes smaller than

$$q^* = a[1 - e^*(a)], \quad (28)$$

in the case of ω libration and smaller than

$$q_s = a[1 - e_s(a)] \quad (29)$$

in the case of ω circulation.

The functions $q^*(a)$ and $q_s(a)$ are nonlinear, because e^* and e_s depend on γ and, consequently, on a via relations (1)–(3), (9), and (20). The corresponding solid curves shown in Fig. 3 start at point $a = q \approx 1.1$ million km ($\gamma = 3/2$) and approach the x axis as $a \longrightarrow \infty$ ($\gamma \longrightarrow 0$). Since there are no singularities in the (ω, e) plane for

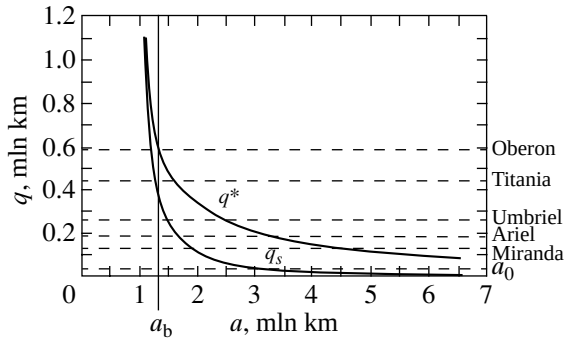


Fig. 3. Pericenter distance versus semimajor axis. The dashed lines indicate Uranus's radius and the orbital radii of its largest inner satellites.

$\gamma > 3/2$, equatorial satellite orbits with $a < 1.1$ million km could theoretically exist beyond the orbit of Oberon.

The horizontal dashed lines correspond to Uranus's radius and to the orbital radii of its five largest inner satellites.

The table gives approximate semimajor axes a of the equatorial satellite orbits that satisfy the conditions

$$q_s(a) = a_j \quad \text{and} \quad q^*(a) = a_j, \tag{30}$$

where $j = 0$ corresponds to Uranus's radius, and j from 1 to 5 corresponds to the orbital radii of its five satellites from Miranda to Oberon.

The satellites fall to Uranus's surface at $q_s < a_0$ and $q^* < a_0$ for orbits with circulation and libration variations in ω, respectively. Therefore, if Uranus's inner satellites are ignored, then circulation and libration orbits could exist up to $a \approx 3.3$ and 15 million km (Fig. 3), respectively. However, these boundary values of the semimajor axis significantly decrease when the actually existing inner satellites are taken into account.

An analysis of the averaged evolutional problem makes the following simplification natural: modeling the orbits of Uranus's massive inner satellites by a coplanar system of Gaussian mass rings. This assumption, introducing effective radii a_j instead of planetary radius a_0, increases it to $\max\limits_j a_j \approx 0.583$ million km.

Semimajor axes a (in million km) satisfying the conditions $q_s(a) = a_j$ and $q^*(a) = a_j$

j	$q_s(a) = a_j$	$q^*(a) = a_j$
0	3.3	>15
1	1.9	4.8
2	1.7	3.3
3	1.5	2.5
4	1.3	1.6
5	1.2	1.3

An analysis of Fig. 3 and the second column of the table then leads us to conclude that only satellite orbits with semimajor axes no larger than about 1.2 million km can exist in Uranus's equatorial plane in the case of circulation variations in ω. At $a > 1.2$ million km, the orbital evolution results in such a decrease in pericenter distance q that the orbit inevitably intersects with one of the Gaussian mass rings that model the orbits of Uranus's inner satellites. Physically, this means either the fall of an outer satellite to an inner satellite or their close encounter (highly probable when orbits intersect) and an abrupt random change in the size and shape of the outer satellite's orbit (in the restricted problem, its mass is assumed to be negligible).

A similar comparison of Fig. 3 and the third column of the table shows that only satellite orbits with $a < 1.3$ million km can exist in Uranus's equatorial plane in the case of libration variations in ω.

Interestingly, this value of a essentially matches the boundary value a_b (the solid vertical straight line in Fig. 3). It is given by condition (12), which corresponds to the loss (at $a > a_b$) of stability of the stationary solution (7) and (9). This means that at $a > a_b$, the regular structure of phase trajectories in the (ω, e) plane will change with time; the largest changes will take place near the separatrix, where the libration and circulation periods of ω tend to infinity. The chaotic state that arises in this case will introduce a random component into the variations in satellite orbital eccentricity and pericenter distance.

3. CONCLUSION

The results of our analysis of the model problem ($I = 90°$) with an increased effective radius of Uranus can explain the absence of equatorial orbits with semimajor axes larger than ≈1.3 million km in the satellite system of this planet. The pattern of orbital evolution shows that in the course of time, the outer satellites (if they existed in the above range) must either have fallen to the inner satellites, thereby significantly increasing their masses (an appreciable massiveness of the most distant inner satellites can serve as circumstantial evidence for this scenario), or greatly change their orbits through close encounters with inner satellites.

Of course, the model problem we considered can serve only as a qualitative guide. To obtain reliable quantitative estimates requires taking into account such factors as a deviation of Uranus's equatorial inclination to its orbital plane from 90° and changes in ecliptic elements of Uranus's orbit with time because of secular planetary perturbations.

ACKNOWLEDGMENTS

This work was supported 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. Y. Kozai, Publ. Astron. Soc. Jpn. **15**, 301 (1963).
2. V. Kudielka, Celest. Mech. Dyn. Astron. **60**, 455 (1994).
3. M. L. Lidov, Iskusstv. Sputniki Zemli **8**, 5 (1961).
4. M. L. Lidov, in *Problems of the Motion of Artificial Celestial Bodies* (Akad. Nauk SSSR, Moscow, 1963a), p. 119.
5. M. L. Lidov, in *Dynamics of Satellites*, Ed. by M. Roy (Springer-Verlag, Berlin, 1963b), p. 168.
6. M. L. Lidov and M.V. Yarskaya, Kosm. Issled. **12**, 155 (1974).
7. S. L. Semenova, Itogi Nauki Tekh., Ser. Issled. Kosm. Prostranstva **35**, 134 (1991).
8. M. A. Vashkov'yak, Pis'ma Astron. Zh. **24**, 791 (1998) [Astron. Lett. **24**, 682 (1998)].
9. M. A. Vashkov'yak, Pis'ma Astron. Zh. **25**, 554 (1999) [Astron. Lett. **25**, 476 (1999)].
10. M. A. Vashkov'yak, Pis'ma Astron. Zh. **27**, 470 (2001) [Astron. Lett. **27** (6), 404 (2001)].
11. M. A. Vashkov'yak and N. M. Teslenko, Pis'ma Astron. Zh. **27**, 232 (2001) [Astron. Lett. **27**, 198 (2001)].

Translated by V. Astakhov

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 470–474. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 549–553.

Approach–Ejection Correlation in the General Three-Body Problem

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

[1] *Astronomical Institute, St. Petersburg State University, Bibliotechnaya pl. 2, Petrodvorets, 198904 Russia*
[2] *St. Petersburg Academy of Forestry Engineering, Institutskiĭ per. 5, St. Petersburg, 194021 Russia*

Received December 25, 2000

Abstract—The relations between parameters of triple approaches and the lengths of subsequent ejections are analyzed for the general three-body problem with components of equal masses and zero initial velocities. A statistically significant correlation is shown to exist between the closeness of approaches and the lengths of subsequent ejections: closer approaches generally result in longer ejections. We have found several systems that evolve to a temporary quasi-stable chain-like configuration. *© 2001 MAIK "Nauka/Interperiodica".*

Key words: *celestial mechanics*

INTRODUCTION

The general three-body problem is one of the problems of current interest in analytical mechanics, celestial mechanics, and stellar dynamics. Despite the large number of works on this problem, an analytic solution that could be effectively used in practice has not yet been obtained.

Numerical simulations (see, e.g., Anosova and Orlov 1985; Valtonen and Mikkola 1991) are a more promising approach to analyzing the three-body problem. Based on their numerical simulations, Szebehely (1971) and Agekyan and Martynova (1973) proposed a classification of the states in the general three-body problem with negative total energy. According to this classification, an unstable triple system can be in one of the following three states in the course of its dynamical evolution:

(0) Triple approach;

(1) Simple interaction;

(2) Ejection without escape.

During the motion, a changeover of states 0–2 occurs, and the final evolutionary state of the system is

(3) Escape.

Agekyan and Martynova (1973) suggested the criteria of states 0–2 for plane nonrotating triple systems. Martynova and Orlov (2000) generalized these criteria to rotating and three-dimensional systems by. Several escape criteria were described in the literature [see Orlov (1986) for a review]. Here, we use the criterion of Standish (1971).

* E-mail address for contacts: vor@astro.spbu.ru

We consider successive transitions between the states and analyze the correlation between several parameters of triple approaches and the length of the subsequent ejection.

FORMULATION OF THE PROBLEM

We consider the dynamical evolution of triple systems with components of equal masses. The initial velocities of all three bodies are zero. The initial configurations of triple systems are chosen uniformly randomly within the domain of all possible configurations of triple systems (Agekyan and Anosova 1967). The initial positions of component M_3 (Fig. 1) change at steps of $\Delta\xi = \Delta\eta = 0.02$.

For each set of initial conditions (ξ, η), we trace the evolution of a triple system until one of the three times:

(1) The escape criterion for one of the bodies is satisfied;

(2) One of the bodies is ejected to a distance larger than $100d$ from the center of mass of the triple system, where d is the mean size of the triple system,

$$d = \frac{G\sum_{i<j} m_i m_j}{2|E|}; \quad (1)$$

(3) The evolution time exceeds 1000τ, where τ is the mean time it takes for a component to cross the system,

$$\tau = \frac{G\sqrt{\sum_{i=1}^{3} m_i}\sum_{i<j} m_i m_j}{(2|E|)^{3/2}}. \quad (2)$$

1063-7737/01/2707-0470$21.00 © 2001 MAIK "Nauka/Interperiodica"

Here, G is the gravitational constant, E is the total energy of the triple system, and m_i are the masses of the bodies.

In our calculations, we recorded the minimum and maximum distances R of the most distant body from the center of mass of the triple system. At the times of minimum R, we calculated the following parameters:

(1) The perimeter p of the configuration triangle;

(2) The static moment

$$I_1 = \sum_{i=1}^{3} m_i r_i, \quad (3)$$

where r_i is the distance from the ith body to the center of mass of the triple system;

(3) The moment of inertia

$$I_2 = \sum_{i=1}^{3} m_i r_i^2; \quad (4)$$

(4) The geometric parameters

$$\alpha = \frac{|r_{13} - r_{23}|}{r_{12}}, \quad \beta = \frac{\min r_{ij}}{\max r_{ij}}, \quad (5)$$

where r_{ij} are the mutual separations between the bodies, which are numbered as follows:

$$r_3 \le r_2 \le r_1; \quad (6)$$

(5) The cosine of angle ϕ between the direction of velocity $\mathbf{V}_3$ and distance $\mathbf{r}_{12}$:

$$\cos\phi = \frac{\mathbf{V}_3 \cdot \mathbf{r}_{12}}{|\mathbf{V}_3| \cdot |\mathbf{r}_{12}|}; \quad (7)$$

(6) The relative energy of body 3

$$\varepsilon = \frac{1}{U}\left[\frac{1}{2} m_3 |\mathbf{V}_3|^2 - \frac{G m_1 m_3}{r_{13}} - \frac{G m_2 m_3}{r_{23}}\right], \quad (8)$$

where U is the potential energy of the triple system;

(7) The potential energy U of the triple system.

RESULTS

To analyze the relations between parameters 1–7 of a triple approach and length R of the subsequent ejection, we calculate the correlation coefficients r between these parameters and R. The values of r and their errors σ_r are given in the table. Figures 2a–2h show plots of ejection length R against parameters 1–7 of the prior approach.

Column 1 gives the notation of the triple-approach parameters under consideration. Columns 2–7 list the correlation coefficients and their errors for all the approach–ejection pairs, for the approach–ejection pairs with the so-called W-approaches excluded (see Anosova and Zavalov 1981), for the ejections with $R > R_{ejc} = G\sqrt{3}/|E| = 2d/\sqrt{3}$ (Agekyan and Martynova 1973), for simple interactions with $d/\sqrt{3} = G\sqrt{3}/(2|E|) = R_{app} < R < R_{ejc} = 2d/\sqrt{3}$ (Agekyan and Martynova 1973), for the approach–ejection pairs with $R < R_{app} = d/\sqrt{3}$, and for the approach–ejection pairs without excluding W-approaches, respectively. We excluded W-approaches as follows: if the time interval Δt between two successive minima of R did not exceed $\Delta t_{cr} = 0.5\tau$, then the two approaches were treated as a single approach with the parameters corresponding to the deepest minimum of R. The first row of the table gives the corresponding number N of approach–ejection pairs. The last rows gives the correlation coefficients between ejection length R and ejection time t_{ejc}. The relation between R and t_{ejc} is shown in Fig. 2i.

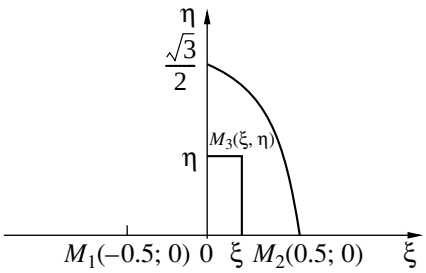


Fig. 1. Domain D of all possible configurations of triple systems.

An examination of the table and Fig. 2 leads us to the following conclusions. There is a statistically significant correlation between the closeness of the approach (parameters p, I_1, I_2, ε, and U) and the length of the subsequent ejection: in general, the closer the approach, the farther the ejection. There is no statistically significant correlation with the other approach parameters considered. The dependence $R(\cos\varphi)$ (Fig. 2f) shows an increase in the concentration of points toward the $\cos\varphi = \pm 1$ straight lines, suggesting the dominance of central-body passages that are nearly orthogonal relative to the line connecting the other two bodies. As expected, there is a strong positive correlation between R and t_{ejc}.

For approach–simple interaction pairs, apart from the preservation of statistically significant correlation between R and the closeness of approach, there is a correlation between R and approach parameters α, β, $\cos\varphi$. In particular, the ejection length R increases with increasing difference between the distances from the central body to the two other bodies. In addition, the length R decreases as the triangle becomes equilateral at the approach time.

The situation changes radically for approach–approach pairs (see the last two columns of the table and Figures 2a–2c). The length R of the subsequent ejection decreases with increasing closeness of approach (parameters p, I_1, I_2, and U). There is a distinct linear correlation between p and R as well as

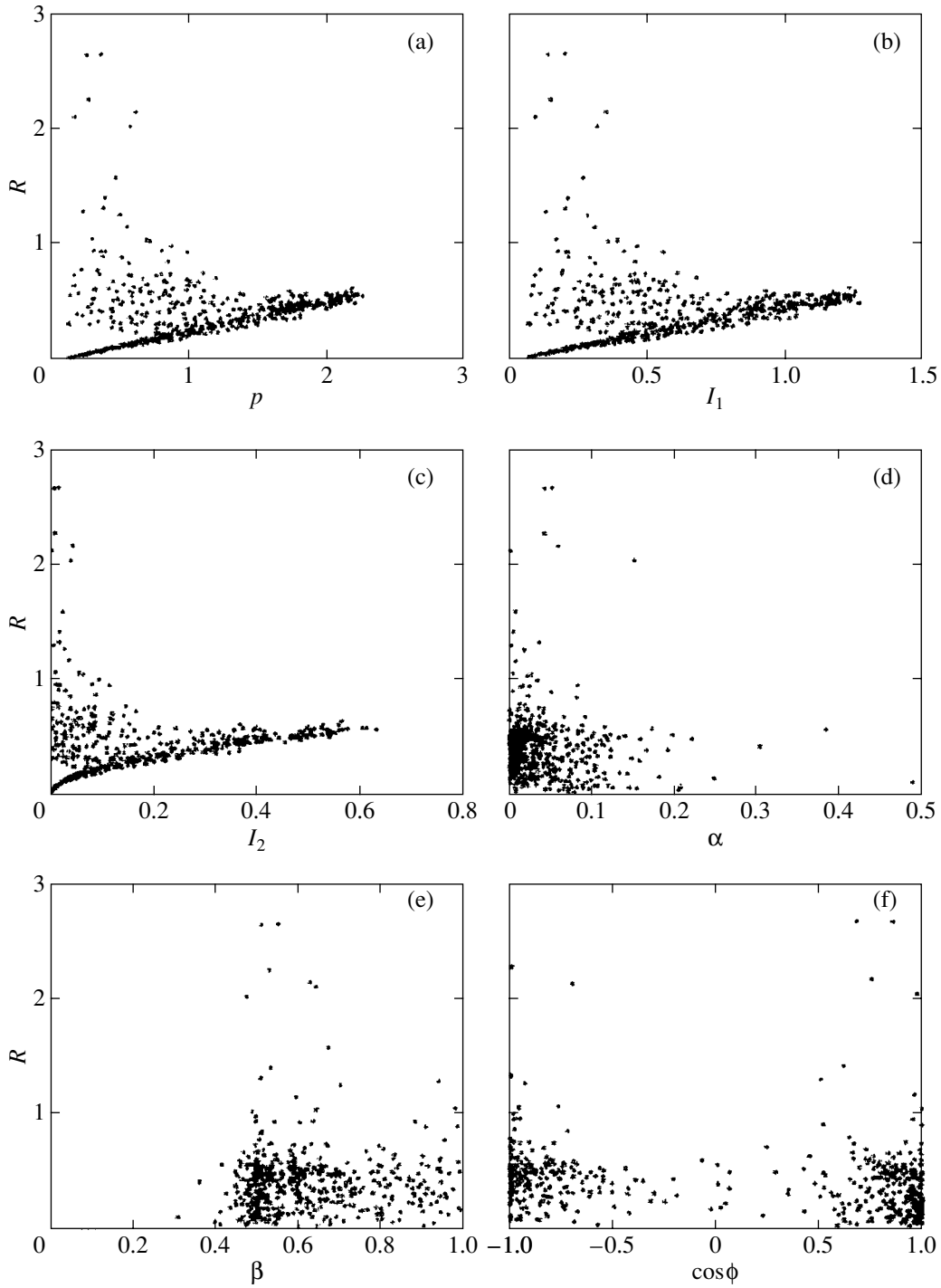


Fig. 2. Ejection length R versus (a) perimeter p, (b) static moment I_1, (c) moment of inertia I_2, (d) parameter $\alpha = \frac{|r_{13} - r_{12}|}{r_{12}}$, (e) parameter $\beta = \frac{\min r_{ij}}{\max r_{ij}}$, (f) $\cos\phi = \frac{\mathbf{V}_3 \cdot \mathbf{r}_{12}}{|\mathbf{V}_3| \cdot |\mathbf{r}_{12}|}$, (g) parameter ε, (h) potential energy U, and (i) ejection time t_{ejc}.

between I_1 and R (see Fig. 2a and 2b) with correlation coefficients >0.9. This correlation can also be seen in the I_2–R, U–R plots, but it is nonlinear. This correlation corresponds to *W*-approaches.

Note that for approach–approach pairs with *W*-approaches excluded, there is no correlation between R and relative energy ε of the central body. At the same time, this correlation is strong for *W*-approaches. We

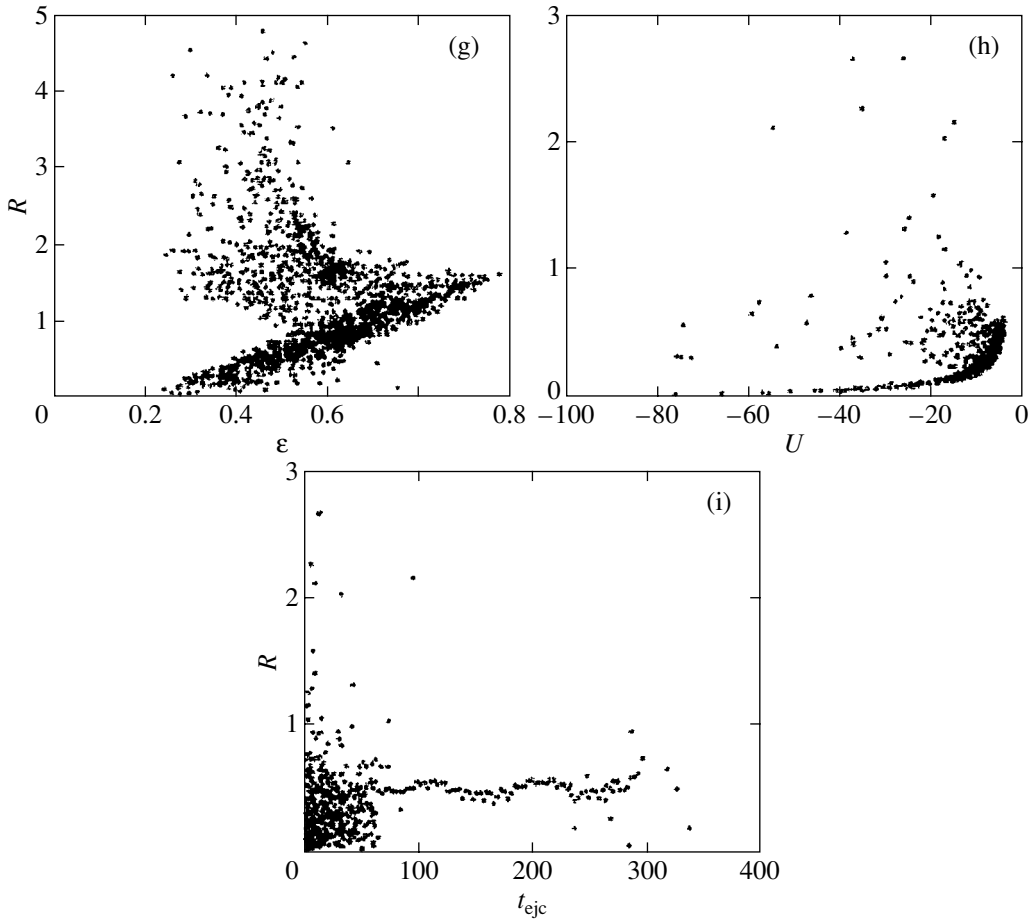


Fig. 2. (Contd.)

can thus conclude that parameter ε can be used as an indicator of *W*-approaches.

Note also the horizontal chain of points in Fig. 2i. They correspond to the dynamical evolution of a triple system becoming quasi-stable, when one of the bodies oscillates between the other two bodies (chain-like configuration). We plan to study such systems in more detail in the future.

Correlation coefficients

Parameter	All	Without *W*-approaches	Ejections	Interactions	Approaches	Approaches with *W*
N	18826	10 771	4063	6400	308	6411
p	−0.116	−0.204	−0.207	−0.272	0.431	0.956
	±0.007	±0.009	±0.015	±0.012	±0.046	±0.001
I_1	−0.116	−0.204	−0.207	−0.287	0.347	0.936
	0.007	0.009	0.015	0.012	0.050	0.002
I_2	−0.073	−0.161	−0.178	−0.188	0.472	0.923
	0.007	0.009	0.015	0.012	0.044	0.002
α	0.007	0.026	−0.013	0.128	0.051	−0.098
	0.007	0.010	0.016	0.012	0.057	0.012
β	−0.081	0.004	0.052	−0.115	−0.256	0.026
	0.007	0.010	0.016	0.012	0.053	0.013
cosϕ	−0.081	0.008	0.014	−0.122	−0.129	−0.102
	0.007	0.010	0.016	0.012	0.056	0.012
ε	−0.107	−0.178	−0.148	−0.189	0.045	0.769
	0.007	0.009	0.015	0.012	0.057	0.005
U	−0.103	−0.246	−0.198	−0.306	0.333	0.643
	0.007	0.009	0.015	0.011	0.051	0.007
t_{ejc}	0.205	0.263	0.437	−0.166	0.156	0.031
	0.007	0.009	0.013	0.012	0.056	0.013

CONCLUSION

The results obtained have led us to the following conclusions:

—There is a significant correlation between the closeness of approach and the length of the subsequent ejection: in general, closer approaches result in longer ejections;

—An inverse correlation was found for *W*-approaches with a strong linear correlation between the approach perimeter and the length of the subsequent ejection;

—Several systems were found to evolve to a temporary quasi-stable chain-like configuration.

REFERENCES

1. T. A. Agekyan and Zh. P. Anosova, Astron. Zh. **44**, 1261 (1967) [Sov. Astron. **11**, 1006 (1967)].
2. T. A. Agekyan and A. I. Martynova, Vestn. Leningr. Univ., Ser. Mat., Mekh., Astron., No. 1, 122 (1973).
3. Zh. P. Anosova and V. V. Orlov, Tr. Astron. Obs. Leningr. Gos. Univ. **40**, 66 (1985).
4. Zh. P. Anosova and N. N. Zavalov, Tr. Astron. Obs. Leningr. Gos. Univ. **36**, 109 (1981).
5. A. I. Martynova and V. V. Orlov, Vestn. St. Peterb. Univ., Ser. Mat., Mekh., Astron., No. 2, 132 (2000).
6. V. V. Orlov, Vestn. St. Peterb. Univ., Ser. Mat., Mekh., Astron., No. 2, 82 (1986).
7. E. M. Standish, Celest. Mech. **4**, 44 (1971).
8. V. Szebehely, Celest. Mech. **4**, 116 (1971).
9. M. Valtonen and S. Mikkola, Ann. Rev. Astron. Astrophys. **29**, 9 (1991).

Translated by A. Dambis

Astronomy Letters, Vol. 27, No. 7, 2001, pp. 475–479. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 27, No. 7, 2001, pp. 554–559.

Dynamical Causes of Asymmetry in the Arrangement of Gaps in the Asteroid Belt

A. P. Markeev*

Institute for Problems of Mechanics, Russian Academy of Sciences, pr. Vernadskogo 101, Moscow, 117526 Russia

Received January 12, 2001

Abstract—The centers of the gaps observed in the asteroid belt are displaced toward Jupiter from their positions that correspond to the exact commensurability between the mean motions of an asteroid and Jupiter. Using the current theory of stability and nonlinear oscillations of Hamiltonian systems, we point out the dynamical causes of this asymmetry. Our analysis is performed in terms of the plane circular restricted three-body problem. The orbits that correspond to Poincaré periodic solutions of the first kind are taken as unperturbed asteroid orbits. © *2001 MAIK "Nauka/Interperiodica".*

Key words: *planets and asteroids, Solar system, periodic solutions, resonance, stability*

INTRODUCTION

More than 130 years ago, Kirkwood (1867) discovered a remarkable property of the distribution of asteroids in mean motion within their main belt located between the orbits of Mars and Jupiter. This distribution turned out to be distinctly nonuniform: there are distribution intervals in which few or no asteroids are observed. These intervals are now called voids or Kirkwood gaps. The discovery of gaps aroused great interest among astronomers and gave rise to many studies devoted to the motion of asteroids. In their monographs and reviews, Putilin (1953), Hagihara (1961), Samoĭlova-Yakhontova (1973), and Demin and Zhuravlev (1979) analyzed the results of several hundred publications on the origin, statistics, and evolution of asteroid orbits.

Kirkwood believed that gaps resulted from the perturbing effect of Jupiter on asteroid orbits and attributed the nonuniformity in the distribution of asteroids to commensurability between an asteroid's mean motion n and Jupiter's mean motion n'. Indeed, observations show that $n : n'$ is close to a rational number $l : s$. The number $l - s$ is called the order of commensurability or the order of resonance.

In the early twentieth century, Brown (1911) noticed an asymmetry in the arrangement of gaps: the centers of the observed gaps are displaced toward Jupiter from exact commensurability. This asymmetry had been repeatedly confirmed until recent decades by a statistical analysis of observations based on the restricted three-body problem and by numerical simulations (Batrakov 1958; Message 1966; Lecar and Franklin 1973; Gerasimov 1975, 1976; Wiesel 1976).

Here, our goal is to establish the dynamical causes of the asymmetry in the arrangement of Kirkwood gaps by using rigorous methods of the theory of stability and nonlinear oscillations of Hamiltonian systems. The orbits that correspond to Poincaré periodic solutions of the first kind for the plane circular restricted three-body problem are taken as unperturbed asteroid orbits.

CONSTRUCTING PERIODIC SOLUTIONS

Let two mass points, S (Sun) and J (Jupiter), move in circular orbits about their common center of mass O with angular velocity n'. Point P of negligible mass moves in the orbital plane of points S and J and is attracted by them according to the Newton law. We will analyze its motion in a synodic coordinate system. The polar coordinates r and φ, where r is the distance OP of point P from the center of mass of points S and J and φ is the angle between vector $\overline{OP}$ and vector $\overline{SJ}$ rotating with angular velocity n', are taken as the generalized coordinates that specify the position of point P. The corresponding generalized momenta are denoted by p_r and p_φ. We choose the units of measurement in such a way that the sum of the masses of points S and J and the separation between them are equal to unity, denote the mass of point J by μ $(0 < \mu \ll 1)$, and take the dimensionless quantity $\nu = n't$ as an independent variable. The Hamiltonian is

$$H = \frac{1}{2}\left(p_r^2 + \frac{p_\varphi^2}{r^2}\right) - p_\varphi - \frac{1-\mu}{r_1} - \frac{\mu}{r_2},$$
$$r_1 = \sqrt{r^2 + 2\mu r\cos\varphi + \mu^2}, \qquad (1)$$
$$r_2 = \sqrt{r^2 - 2(1-\mu)r\cos\varphi + (1-\mu)^2}.$$

* E-mail address for contacts: markeev@ipmnet.ru

1063-7737/01/2707-0475$21.00 © 2001 MAIK "Nauka/Interperiodica"

At $\mu = 0$, the equations of motion allow a particular solution that corresponds to circular motions of point P about point S:

$$r = R, \quad \varphi(\nu) = \Omega\nu + \varphi(0),$$
$$p_r = 0, \quad p_\varphi = \sqrt{R}, \tag{2}$$
$$\Omega = R^{-3/2} - 1, \quad R = (n/n')^{-2/3}.$$

Here, R is constant. We assume that $0 < R < 1$. At solution (2), we have

$$H = h_0 = \text{const}, \quad h_0 = -(2R^{3/2} + 1)/(2R).$$

For low, but nonzero μ, we take solution (2) as the generating solution. Many studies are devoted to constructing periodic solutions of the restricted three-body problem, which transform to a generating circular solution at $\mu = 0$, and to their use for analyzing the dynamics of asteroids (Szebehely 1967; Bruno 1990). We construct these solutions (Poincaré solutions of the first kind) as follows. Writing the constant of the energy integral $H = h$ = const as a series $h = h_0 + \mu h_1 + \mu^2 h_2 + \ldots$ and solving the equality $H = h$ for p_φ yields

$$p_\varphi = -K = r^2 + r\sqrt{F} + \mu r\frac{V(r, \varphi) + h_1}{\sqrt{F}} + \ldots, \tag{3}$$

$$F = 2h_0 + r^2 - p_r^2 + \frac{2}{r},$$
$$V(r, \varphi) = \frac{1}{\sqrt{1 - 2r\cos\varphi + r^2}} - \frac{1}{r} - \frac{\cos\varphi}{r^2}. \tag{4}$$

At the isoenergetic level $H = h$, the motion is described by Whittaker's equations in Hamiltonian form (Markeev 1999):

$$\frac{dr}{d\varphi} = \frac{\partial K}{\partial p_r}, \quad \frac{dp_r}{d\varphi} = -\frac{\partial K}{\partial r}. \tag{5}$$

Let us find the analytic (in μ), 2π-periodic (in φ) solution $r^*(\varphi)$ and $p_r^*(\varphi)$ of system (5) by the Poincaré method (Malkin 1956). We seek the solution in the form of a series

$$r = r^* = R + \mu f(\varphi) + \ldots,$$
$$p_r = p_r^* = \mu g(\varphi) + \ldots. \tag{6}$$

The functions f and g satisfy the system of equations

$$\frac{df}{d\varphi} = \frac{1}{\Omega}g,$$
$$\frac{dg}{d\varphi} = -\Omega\omega^2 f + \frac{2(\Omega + 1)(V_0 + h_1)}{\Omega^2 R} + \frac{1}{\Omega}\frac{\partial V_0}{\partial R}, \tag{7}$$

where

$$\omega = \omega(R) = \frac{\Omega + 1}{\Omega} = \frac{1}{1 - R^{3/2}}, \tag{8}$$

and V_0 is the function $V(r, \varphi)$ from (4) calculated at $r = R$. It can be represented as a Fourier series of the form (Aksenov 1986)

$$V_0 = \left(\frac{1}{2}L_1^{(0)} - \frac{1}{R}\right) + \left(L_1^{(1)} - \frac{1}{R^2}\right)\cos\varphi + \sum_{n=2}^{\infty} L_1^{(n)}\cos n\varphi,$$

where $L_1^{(n)} = L_1^{(n)}(R)$ are the Laplace coefficients.

If ω is not an integer (i.e., there is no resonance of the first order), then the periodic solution of Eqs. (7) can be written as the following Fourier series:

$$f = \sum_{n=0}^{\infty} f_n\cos n\varphi, \quad g = -\Omega\sum_{n=1}^{\infty} n f_n \sin n\varphi,$$
$$f_0 = -\frac{(\Omega + 2)}{\Omega(\Omega + 1)^2 R^2}$$
$$+ \frac{1}{(\Omega + 1)^2}\left[\frac{(\Omega + 1)(L_1^{(0)} + 2h_1)}{\Omega R} + \frac{1}{2}\frac{dL_1^{(0)}}{dR}\right], \tag{9}$$
$$f_1 = -\frac{2}{(2\Omega + 1)\Omega R^3}$$
$$+ \frac{1}{(2\Omega + 1)}\left[\frac{2(\Omega + 1)L_1^{(1)}}{\Omega R} + \frac{dL_1^{(1)}}{dR}\right],$$
$$f_n = \frac{1}{\Omega^2(\omega^2 - n^2)}\left[\frac{2(\Omega + 1)L_1^{(n)}}{\Omega R} + \frac{dL_1^{(n)}}{dR}\right].$$

The dependence of momentum p_φ on φ is found from (3), (4), and (6) to be

$$p_\varphi = p_\varphi^* = \sqrt{R} + \mu\frac{V_0 + h_1}{\Omega} + \ldots. \tag{10}$$

Functions (6) and (10) are 2π-periodic in φ. In turn, the dependence of φ on dimensionless time ν can be derived from the equation

$$\frac{d\varphi}{d\nu} = \frac{\partial H}{\partial p_\varphi} = \frac{p_\varphi}{r^2} - 1.$$

Substituting r and p_φ from (6) and (10) in this equation yields

$$\frac{d\varphi}{d\nu} = \Omega + \mu G(\varphi) + \ldots, \tag{11}$$

where

$$G(\varphi) = -\frac{2(\Omega + 1)}{R}f + \frac{V_0 + h_1}{\Omega R^2}. \tag{12}$$

The period in ν is equal to the time interval T in which φ increases by 2π. We find from (9), (11), and (12) that $T = 2\pi/\Omega_1$, where $\Omega_1 = \Omega + \mu\bar{G} + O(\mu^2)$ and

$$\bar{G} = \frac{1}{2\pi}\int_0^{2\pi} G(\varphi)d\varphi = \frac{(\Omega+1)(\Omega+3)}{\Omega}$$
$$-\frac{3(L_1^{(0)}+2h_1)}{2\Omega R^2} - \frac{1}{(\Omega+1)R}\frac{dL_1^{(0)}}{dR}.$$

The constant of the energy integral is an important parameter on which the constructed periodic solution depends. In particular, h_1, h_2, … can be chosen in such a way that its period is equal to the period of the generating circular motion.

THE HAMILTONIAN OF PERTURBED MOTION

Let us introduce perturbations q, p, and I in the vicinity of the periodic solution by using the canonical transformation

$$r = r^* + q, \quad p_r = p_r^* + p,$$
$$\varphi = w, \quad p_\varphi = p_\varphi^* + q\frac{dp_r^*}{d\varphi} - p\frac{dr^*}{d\varphi} + I.$$

In the expansion of Hamiltonian (1) in a power series of q, p, and I, the coefficient of the first power of I coincides with the right-hand side of Eq. (11). Let us introduce new canonically conjugate variables φ_1 and I_1 instead of φ and I by using the generating function

$$S(\varphi, I_1) = I_1\left(\varphi - \frac{\mu}{\Omega}\int\tilde{G}(\varphi)d\varphi - \ldots\right),$$
$$\tilde{G}(\varphi) = G(\varphi) - \bar{G}.$$

The coefficient of I_1 in the expansion of the Hamiltonian for perturbed motion in a series will then coincide with the frequency Ω_1 of unperturbed periodic motion. If we also make the substitution $q = (\Omega+1)^{-1/2}q_2$ and $p = (\Omega+1)^{1/2}p_2$ and change to a new independent variable $\tau = \Omega_1\nu$, then the Hamiltonian of perturbed motion will be written as the following series (we use the old designation φ for the new angular variable φ_1):

$$H = H_2 + H_3 + H_4 + \ldots + H_m + \ldots, \qquad (13)$$

where H_m is the form of power m with respect to $|I_1|^{1/2}$, q_2, and p_2, with

$$H_2 = I_1 + \frac{1}{2}\omega(q_2^2 + p_2^2)$$
$$+\mu(h_{20}q_2^2 + h_{11}q_2p_2 + h_{02}p_2^2) + O(\mu^2),$$
$$h_{20} = \frac{3(V_0+h_1)}{\Omega^2R^2} - \frac{3(\Omega+1)}{\Omega R}f - \frac{2}{\Omega R}\frac{dg}{d\varphi}$$
$$-\frac{\Omega+1}{2\Omega^2}\bar{G} - \frac{1}{2\Omega(\Omega+1)}\frac{\partial^2V_0}{\partial R^2},$$
$$h_{11} = \frac{2(\Omega+1)}{\Omega^2R}g, \quad h_{02} = -\frac{\Omega+1}{2\Omega^2}\bar{G},$$
$$H_3 = -2\frac{\sqrt{\Omega+1}}{\Omega R}I_1q_2 - \frac{\sqrt{\Omega+1}}{\Omega R}q_2^3 + O(\mu),$$
$$H_4 = \frac{1}{2\Omega R^2}I_1^2 + \frac{3}{\Omega R^2}I_1q_2^2 + \frac{3}{2\Omega R^2}q_2^4 + O(\mu).$$

CHARACTERISTIC EXPONENTS AND THE DISPLACEMENT OF KIRKWOOD GAP CENTERS

The Hamiltonian H_2 from expansion (13) corresponds to a linearized system of equations of perturbed motion. Two characteristic indices of this system are zero, while the other two, $\pm i2\pi\lambda$, at a sufficiently low μ and in the absence of resonance of the second order (parametric resonance) are purely imaginary. The quantity λ can be treated (when $\tau = \Omega_1 n' t$ plays the role of time) as one of the frequencies of the small oscillations in the vicinity of the periodic solution under study.

If μ is small enough, then λ is an analytic function of μ and can be represented as the series

$$\lambda = \lambda_0 + \mu\lambda_1 + \ldots, \qquad (14)$$

where $\lambda_0 = \omega$ and λ_1 is given by

$$\lambda_1 = \frac{1}{2\pi}\int_0^{2\pi}(h_{20}+h_{02})d\varphi = -\frac{(\Omega+1)R^3}{4\Omega}\frac{d^2L_1^{(0)}}{dR^2}$$
$$+\frac{(\Omega+1)(2-\Omega)R^2}{2\Omega^2}\frac{dL_1^{(0)}}{dR} \qquad (15)$$
$$+\frac{3(\Omega+1)^2R}{2\Omega^3}(L_1^{(0)}+2h_1) - \frac{(\Omega+1)(\Omega+3)}{\Omega^3}.$$

Let the $n : n' = k : (k-m)$ commensurability be exact at $R = R_0$. The resonance of order m, i.e., $m\omega_0 = k$, where $\omega_0 = \omega(R_0)$, is then realized.

If, however, $\mu \neq 0$, then $R = R_\mu$, which differs from R_0 and corresponds to the resonance of order m. This value must satisfy the resonance relation

$$m\lambda = k, \qquad (16)$$

where λ is given by Eq. (14). We derive the following expression for $\delta = R_\mu - R_0$ from relations (8) and (16):

$$\delta = -\mu\frac{\lambda_1(R_0, h_1)}{\omega_0'} + O(\mu^2)$$
$$\left(\omega_0' = \frac{d\omega_0}{dR_0} = \frac{3\omega_0^2\sqrt{R_0}}{2}\right). \qquad (17)$$

For a known μ and specified $n : n'$ commensurability, Eq. (17) describes the displacement of Kirkwood gap centers. To be specific, let us consider periodic

motions with a period equal to the period of the generating circular motion. We then have

$$h_1 = \frac{1}{3}(\Omega+1)(\Omega+3)R^2 - \frac{1}{2}L_1^{(0)} - \frac{\Omega R}{3(\Omega+1)}\frac{dL_1^{(0)}}{dR}$$

and derive the following expression for δ from (15) and (17) using the equality $L_1^{(0)} = 4K(R)/\pi$, where $K(R)$ is the complete elliptic integral of the first kind:

$$\delta = \mu\frac{2R_0^{3/2}}{3\pi\omega_0}\left[R_0\frac{d^2K(R_0)}{dR_0^2} + 2\frac{dK(R_0)}{dR_0}\right] + O(\mu^2).$$

At sufficiently low μ, δ is positive; i.e., the gap centers are displaced from R_0 to larger R (toward Jupiter) at low μ.

THE STABILITY OF PERIODIC MOTIONS AT THE BOUNDARIES OF THE REGIONS OF PARAMETRIC RESONANCE

Let us assume that 2ω is close to an odd number $2N+1$; i.e., the ratio $n:n'$ is close to a rational number $(2N+1):(2N-1)$ ($N = 1, 2, 3, \ldots$). An analysis of the stability of periodic solutions for the restricted three-body problem for such commensurability was begun long ago (Heinrich 1912; Zeipel 1915); the corresponding review is given in the monograph of Bruno (1990).

Let the equality $2\omega_0 = 2\omega(R_0) = 2N+1$ holds at $R = R_0$. For R close to R_0, we assume that $R = R_0 + \Delta$, $|\Delta| \ll 1$. We then have

$$2N + 1 - 2\omega = 2\mu\alpha, \tag{18}$$

where, according to (18) and (8),

$$\alpha = -\frac{\omega_0'}{\mu}\Delta + O(\Delta^2). \tag{19}$$

Using the algorithm from Markeev (2000), we can make a real, canonical, analytic (with respect to μ, ξ_2, η_2, η_1), and 2π-periodic (in ξ_1) change of variables, q_2, p_2, φ_1, $I_1 \longrightarrow \xi_2$, η_2, ξ_1, η_1, which reduces the Hamiltonian of perturbed motion to

$$\Gamma = \Gamma_2 + \Gamma_4 + \ldots + \Gamma_m + \ldots,$$

where Γ_m is the form of m with respect to $|\eta_1|^{1/2}$, ξ_2, η_2 with 2π-periodic (in ξ_1) coefficients;

$$\Gamma_2 = \eta_1 + \frac{1}{2}\lambda(\xi_2^2 + \eta_2^2)$$

$$+ \sigma\left[-\frac{1}{2}(\xi_2^2 - \eta_2^2)\cos k\xi_1 + \xi_2\eta_2\sin k\xi_1\right],$$

$$\Gamma_4 = c_{20}\eta_1^2 + \frac{1}{2}c_{11}(\xi_2^2 + \eta_2^2)\eta_1$$

$$+ \frac{1}{4}c_{02}(\xi_2^2 + \eta_2^2)^2 + O(\mu).$$

Here, $k = 2N+1$, $\lambda = N + 1/2 - \mu(\alpha - \lambda_1) + O(\mu^2)$, and $\sigma = \mu\chi + O(\mu^2)$, where λ_1 and α are given by formulas (15) and (19), while

$$\chi = \frac{1}{2\pi}\int_0^{2\pi}[h_{11}\sin k\varphi + (h_{02} - h_{20})\cos k\varphi]d\varphi$$

$$= \frac{1}{4\Omega(\Omega+1)}\frac{d^2L_1^{(k)}}{dR^2}$$

$$+ \frac{1}{\Omega^2 R}\left\{1 - \frac{(\Omega+1)[(2k-1)\Omega+2]}{2[(\Omega+1)^2 - k^2\Omega^2]}\right\}\frac{dL_1^{(k)}}{dR}$$

$$+ \frac{1}{2\Omega^3 R^2}\left\{\Omega + 4 - \frac{2(\Omega+1)^2[(2k-1)\Omega+2]}{(\Omega+1)^2 - k^2\Omega^2}\right\}L_1^{(k)}.$$

The coefficients c_{ij} in the function Γ_4 are given by the equalities

$$c_{20} = \frac{1}{2}c_{11} = c_{02} = -\frac{3}{2\Omega R^2}. \tag{20}$$

At low μ, the regions of parametric resonance (the regions of orbital instability of the periodic solution under study) are specified (Markeev 2000) by the inequality $|\alpha - \lambda_1| < |\chi|$. In the R, μ plane, they are wedge-shaped regions that emerge from points $R = R_0$ of the $\mu = 0$ axis. The boundaries γ_+ and γ_- of these regions are specified, respectively, by the equations

$$R = R_0 + \mu a + O(\mu^2) \quad \text{and} \quad R = R_0 + \mu b + O(\mu^2), \tag{21}$$

where $a = (|\chi| - \lambda_1)/\omega_0'$ and $b = -(|\chi| + \lambda_1)/\omega_0'$.

At the boundary of the region of parametric resonance for sufficiently low μ, there is orbital stability if the signs of $a_{02} = c_{20}(N+1/2)^2 - c_{11}(N+1/2) + c_{02}$ and $\alpha - \lambda_1$ are opposite and orbital instability if their signs are the same (Markeev 2000).

However, it follows from (20) that

$$a_{02} = -\frac{3(2N-1)^2}{8\Omega R^2} < 0,$$

while from (19) and (21), we can find that the inequalities $\alpha < \lambda_1$ and $\alpha > \lambda_1$ hold at γ_+ and γ_-, respectively. Therefore, at sufficiently low μ, the Poincaré periodic solution under study is orbitally stable at boundary γ_- and unstable at boundary γ_+. This is yet another dynamical cause of the asymmetry in the arrangement of gap centers [which correspond to $n:n' = (2N+1):(2N-1)$ commensurability] in the asteroid belt, because the points of boundaries γ_+ lie closer to Jupiter than do the points of boundaries γ_- for a given μ.

CONCLUSION

We have established the dynamical causes of the asymmetry in the arrangement of Kirkwood gap centers in the asteroid belt. The orbits corresponding to Poincaré periodic solutions of the first kind were taken as unperturbed asteroid orbits. The first cause is that even to a first approximation in μ, one of the frequencies of small oscillations in the vicinity of an asteroid's periodic motion differs from the oscillation frequency in the vicinity of the corresponding generating circular motion. For $n : n' = (2N + 1) : (2N - 1)$ commensurability, we additionally pointed out the second cause of the asymmetry: the Poincaré solutions are orbitally unstable at one of the two boundaries of the regions of parametric resonance or, more specifically, at the boundary that lies closer to Jupiter.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-01-00405) and the Program of State Support for Leading Scientific Schools of the Russian Federation (project no. 00-15-96088).

REFERENCES

1. E. P. Aksenov, *Special Functions in Celestial Mechanics* (Nauka, Moscow, 1986).
2. Yu. V. Batrakov, Byull. Inst. Teor. Astron. Akad. Nauk SSSR **6**, 577 (1958).
3. E. W. Brown, Science **33**, 79 (1911).
4. A. D. Bruno, *The Restricted Problem of Tree Bodies* (Nauka, Moscow, 1990).
5. V. G. Demin and S. G. Zhuravlev, Itogi Nauki Tekh., Ser. Astron. **15**, 3 (1979).
6. I. A. Gerasimov, Vestn. Mosk. Gos. Univ., Ser. 3: Fiz., Astron., No. 6, 749 (1975).
7. I. A. Gerasimov, Vestn. Mosk. Gos. Univ., Ser. 3: Fiz., Astron., No. 5, 569 (1976).
8. Y. Hagihara, Smithson. Contrib. Astrophys. **5**, 59 (1961).
9. W. W. Heinrich, Astron. Nachr. **192** (20), 325 (1912).
10. D. Kirkwood, *Meteoric Astronomy* (Lippencott, Philadelphia, 1867).
11. M. Lecar and F. Franklin, Icarus **20**, 422 (1973).
12. I. G. Malkin, *Problems of the Theory of Nonlinear Oscillations* (Gostekhizdat, Moscow, 1956).
13. A. P. Markeev, *Theoretical Mechanics* (CheRo, Moscow, 1999).
14. A. P. Markeev, Prikl. Mat. Mekh. **64**, 833 (2000).
15. P. J. Message, in *Proceedings of the 25th IAU Symposium on Theory Orbits Solar System and Stellar Systems, 1966*, Ed. by G. Contopulos (Academic, 1966), p. 197.
16. I. I. Putilin, *Minor Planets* (Gostekhizdat, Moscow, 1953).
17. *Minor Planets*, Ed. by N. S. Samoĭlova-Yakhontova (Nauka, Moscow, 1973).
18. V. Szebehely, *Theory of Orbits: The Restricted Problem of Three Bodies* (Academic, Orlando, 1967; Nauka, Moscow, 1982).
19. W. H. Wiesel, Celest. Mech. **13**, 3 (1976).
20. H. Zeipel, Ark. Mat., Astron. Fys. **10** (30), 1 (1915).

Translated by V. Astakhov

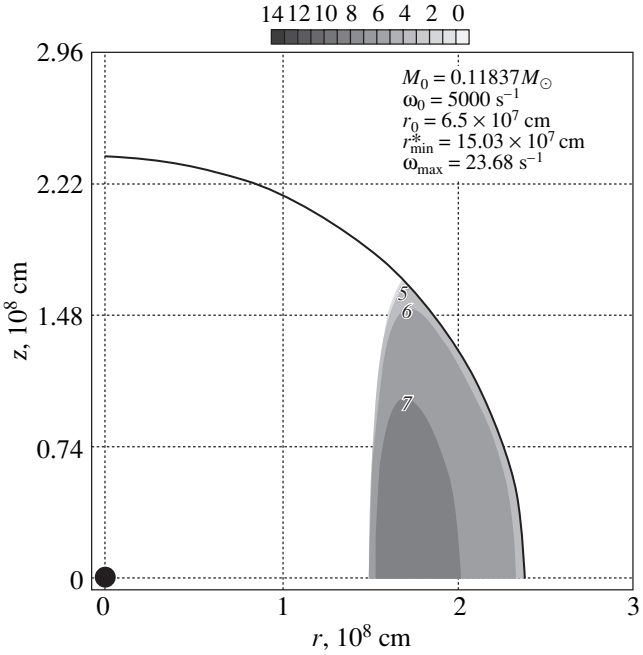
Astronomy Letters, Vol. 27, No. 7, 2001, p. 480.

ERRATA

Erratum: "A Two-Dimensional Hydrostatically Equilibrium Atmosphere of a Neutron Star with Given Differential Rotation" [*Pis'ma Astron. Zh.* 26, 917 (2000); *Astronomy Letters* 26, 788 (2000)]

V. S. Imshennik and K. V. Manukovskiĭ

In the paper by V.S. Imshennik and K.V. Manukovskiĭ [Astron. Lett. **26**, 788 (2000)], a misprint was made on page 794: Fig. 2 was given instead of Fig. 4 for the second time. Here, we present the correct Fig. 4.



1063-7737/00/2601-0480$20.00