Astronomy Reports, Vol. 49, No. 12, 2005, pp. 1001–1008. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1119–1126.

Magnetic Field Models for HD 116458 and HD 126515

Yu. V. Glagolevskiĭ

Special Astrophysical Observatory, Russian Academy of Sciences, Nizhniĭ Arkhyz, Karachaĭ-Cherkessian Republic, 357147 Russia

Received October 25, 2004; in final form, July 6, 2005

Abstract—We have modeled the magnetic fields of the slowly rotating stars HD 116458 and HD 126515 using the "magnetic charge" technique. HD 116458 has a small angle between its rotation axis and dipole axis ($\beta = 12°$), whereas this angle is large for HD 126515 ($\beta = 86°$). Both stars can be described with a decentered-dipole model, with the respective displacements being $r = 0.07$ and $r = 0.24$ in units of the stellar radius. The decentered-dipole model is able to satisfactorily explain the phase relations for the effective field, $B_e(P)$, and the mean surface field, $B_s(P)$, for both stars, along with the fact that the $B_e(P)$ phase relation for HD 126515 is anharmonic. We discuss the role of systematic measurement errors possibly resulting from instrumental or methodical effects in one or both of the phase relations. The displacement of the dipole probably reflects real asymmetry of the stellar field structure, and is not due to measurement errors. Using both phase relations, $B_e(P)$ and $B_s(P)$, in the modeling considerably reduces the influence of the nonuniform distribution of chemical elements on the stellar surface. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

This paper continues our studies of slowly rotating ($P > 25^d$), magnetic, chemically peculiar (CP) stars [1, 2]. Using literature data on the mean effective (B_e) and mean surface (B_s) magnetic fields, we model the magnetic fields of selected stars by applying the "magnetic-charge" technique described in detail in [3–6]. The main goals of our study are the following:

(1) to compare the properties of fast and slow rotators;

(2) to check the correctness of the predictions made by Stępień [7] that angular-momentum loss is more efficient for CP stars with magnetic fields in the case of small angles between the rotation and dipole axes, β;

(3) to compare magnetic-field modeling results based on the "magnetic charge" technique and other methods.

Since the studies of H.W. Babcock, the magnetic-field configuration for chemically peculiar stars has been thought to be predominantly dipolar [8]. Our modeling technique assumes the presence of several dipoles in the star (depending on the complexity of the structure), with magnetic moment $M = Ql$ (where Q is the magnetic charge of a monopole and l the distance between the charges). Each of the monopoles has the coordinates λ (longitude) and δ (latitude). The modeling parameters are the charge Q, the star's inclination to the line of sight i, λ and δ, and the distance from the dipole's center to the center of the star, r, in units of the stellar radius. By calculating B_e and B_s averaged over the visible hemisphere with limb darkening taken into account for various rotational phases, P, we obtained calculated $B_e(P)$ and $B_s(P)$ relations, to be compared with the observed relations. We selecting the parameters iteratively, based on coincidence of the calculated and observed relations. We already noted earlier [2] that, to first approximation, two models with different i can be derived for the same B_s. The more realistic version should be adopted.

2. MODEL OF HD 116458

We took the observational data for the mean effective magnetic field from [9, 10] and for the mean surface field from [11]. The phase relations based on these data are presented in Figs. 1a and 1b (dots).

The parameters were selected in the following order. First, assuming a centered dipole, we chose the coordinates of the monopoles and the magnetic moment M, for a preliminary value of the inclination i, in order to select a calculated phase relation $B_e(P)$. We then found the calculated $B_s(P)$ relation via iterative correction of the chosen parameters, and so on. The best solutions were chosen according to a least-squares criterion.

Our first version assumed a small inclination, $\beta < 45°$, and yielded the calculated $B_e(P)$ and $B_s(P)$ phase relations shown in Figs. 1a and 1b by the solid curves. All the derived parameters are collected in Table 1A. These relations indicate that, whereas the

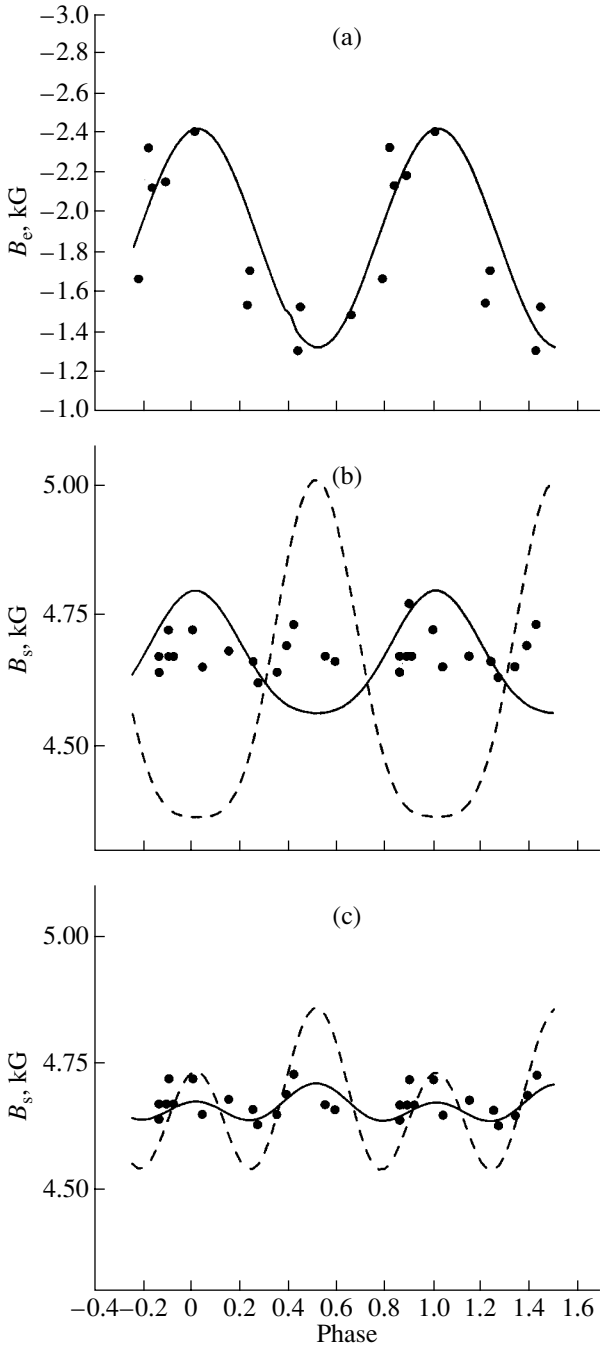

Fig. 1. Phase relations for HD 116458. The points show the observations, curve (a) is the model relation for all the versions, curves (b) are the centered-dipole models for small (solid) and large (dashed) angles β, and curves (c) are the decentered-dipole models for small (solid) and large (dashed) angles β.

$B_e(P)$ phase relation is approximated relatively well, the observed and calculated $B_s(P)$ relations do not coincide. The observations suggest that the mean surface magnetic field virtually does not vary with the rotational phase, and that the calculated relation's amplitude, about 500 G, is too large. We conclude that the model assumptions of a centered dipole and small β angle do not agree with the observations.

The second version assumed a large inclination, $\beta > 45°$. The calculated $B_e(P)$ relation is the same as for the previous version; the $B_s(P)$ relation is plotted by the dashed curve in Fig. 1b. The calculated parameters are collected in Table 1B. It is clear that the $B_s(P)$ relation deviates from the observations even more for this version. We conclude that both versions of the centered-dipole model, with small and large β, are inconsistent with the observations.

The next step is to assume a decentered dipole. Let us first select a model with a small β. The best agreement between the calculated and observed phase relations is achieved when the dipole is displaced by $r = 0.07$ towards the positive monopole (in units of the star's radius). The $B_s(P)$ relation corresponding to this case is shown in Fig. 1c by the solid curve, and the parameters for this model are given in Table 1C. The assumption of large β gives a worse result. The relation corresponding to this case is shown in Fig. 1c by the dashed curve, and the model parameters are presented in Table 1D.

Our final conclusion is that the best agreement with the observations is achieved when the dipole is displaced by 0.07 of the stellar radius towards the positive charge, with the dipole axis close to the rotational axis (the angle between them is $\beta = 12°$). This distinguishes HD 116458 from the three stars considered in [2], which have large angles β. The fact that the dipole is decentered leads to a difference in the magnetic intensities at the poles by about 3 kG. Figure 2 is a Mercator-projection map showing the distribution of the magnetic intensity on the stellar surface, demonstrating the concentration of the intensity contours toward the positive magnetic pole.

Because of the large inclination of the rotational axis to the line of sight ($i = 75°$), the star is observed virtually directly along the magnetic equator. For this reason, the amplitude of B_e is small, whereas the magnetic field at the poles, B_p, is fairly large.

The magnetic field of HD 116458 was modeled in [12] using a dipole-quadrupole-octupole model, with $\beta = 10°$ and $i = 52°$; i.e., close to our values, especially the value for β.

Chemical anomalies are known to be primarily concentrated near the magnetic poles. These are precisely the regions on surfaces of CP stars that give rise to photometric variability. Since the chemically anomalous regions on HD 116458 are near its rotational poles ($\beta = 12°$) and at the edges of the disk, its photometric variability is relatively small. This explains why several authors have claimed an absence of variability for HD 116458 [13].

Table 1. Model magnetic-field parameters for HD 116458

Monopole sign	Longitude λ, deg	Latitude δ, deg	Field at the pole, G
A. First case, centered-dipole model with small $\beta = 10°$ and $i = 108°$ (82°)			
+	180	80	+7290
−	0	−80	−7290
B. Second case, centered-dipole model with large $\beta = 66°$ and $i = 158°$ (32°)			
+	0	24	+7570
−	180	−24	−7570
C. Third case, decentered-dipole model with small $\beta = 12°$, $i = 105°$ (75°), and $r = 0.07$			
+	180	78	+9510
−	0	−78	−6220
D. Fourth case, decentered-dipole model with large $\beta = 66°$, $i = 158°$ (32°), and $r = 0.07$			
+	0	22	+9880
−	180	22	−6470

The concentration of the chemical elements toward the magnetic poles distorts the phase relations, which some authors think could result in a fictitious displacement of the dipole relative to the stellar center. However, as can be seen from an inspection of Figs. 1a and 1b, our model calculations show that the amplitude of the $B_e(P)$ relation must be a factor of a few lower than is observed in order to explain the narrow range of the data scatter in Fig. 1b. Thus, the magnetic field structure for HD 116458 does indeed correspond to a decentered-dipole model.

3. MODEL OF HD 126515

A preliminary study of HD 126515 was presented in [14]. The star exhibits an anharmonic $B_e(P)$ phase relation, and we had to assume a model with a strongly decentered magnetic dipole to achieve the best agreement between the calculated and observed phase relations. The displacement of the negative monopole is $r = 0.55$, while that of the positive one is $r = 0.37$. Thus, we find that the field structure is determined by a barlike dipole with a considerable distance between the monopoles. In the model of Preston [15], the dipole is displaced by $r = 0.36$ along the dipole axis. Preston assumed that the anharmonic phase relation was probably due to a nonuniform distribution of chemical elements and the star's strongly distorted magnetic field. In our study, we decided to investigate the model in more detail. Since the mean surface magnetic field, B_s, has a weak dependence on the inhomogeneity of the surface abundance distributions, we tried to find agreement between the calculated and observed relations only for $B_s(P)$, whereas we selected the $B_e(\max)$ relation solely based on its amplitude.

We based the $B_e(P)$ phase relation on the data from [9, 15–19] and the $B_s(P)$ relation on the data from [9, 15, 20]. The ephemeris for $B_s(\max)$ was taken from [15]:

$$\mathrm{JD}(B_s(\max)) = 2\,437\,015.0 + 129.9474^{\mathrm{d}}.$$

This value for the rotational period, which is the most accurate available, was given in [18].

The $B_e(P)$ and $B_s(P)$ phase relations are shown in Figs. 3a and 3b. We can easily see that the $B_s(P)$ relation is symmetric and its extrema are at phases 0 and 0.5, whereas the negative extremum for the $B_e(P)$ relation coincides with $P = 0$ and the positive extremum with $P \approx 0.4$.

As for HD 116458, we consider two models, for small and large angles between the rotational axis and the dipole axis β, and try to identify which option is most realistic.

We iteratively determined that the centered-dipole model does not agree with the observations. The solid curve in Fig. 3a is the calculated relation for the parameters presented in Tables 2A and 2B (the two curves coincide). The calculated $B_s(P)$ phase relations in Fig. 3b (solid and dashed thin curves) virtually coincide, and are very different from the observed relation. This situation can be improved only by supposing a decentered dipole model.

Using the same scheme, we iteratively obtained two versions of the model, for small and large β, with the parameters presented in Tables 2C and 2D. Under

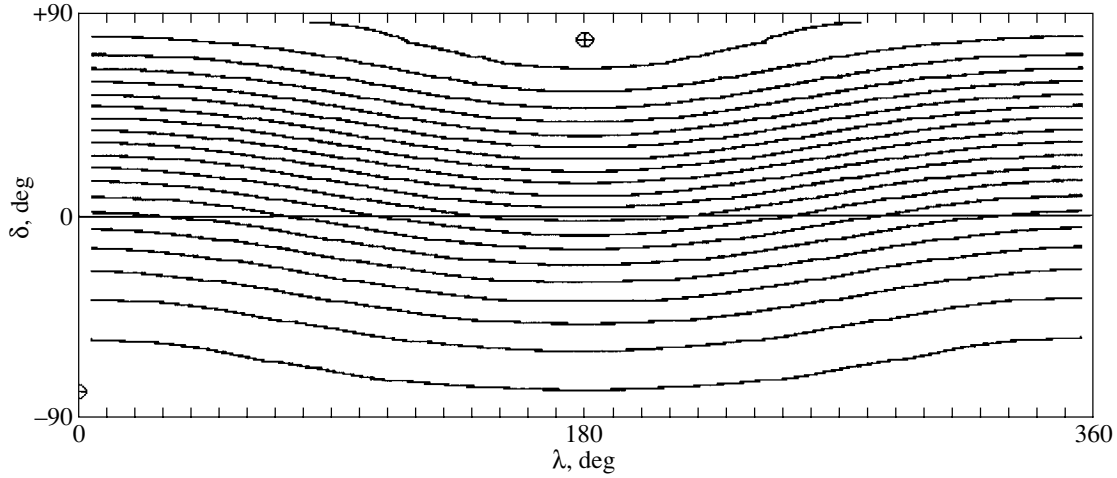

Fig. 2. Distribution of the magnetic intensity on the surface of HD 116458.

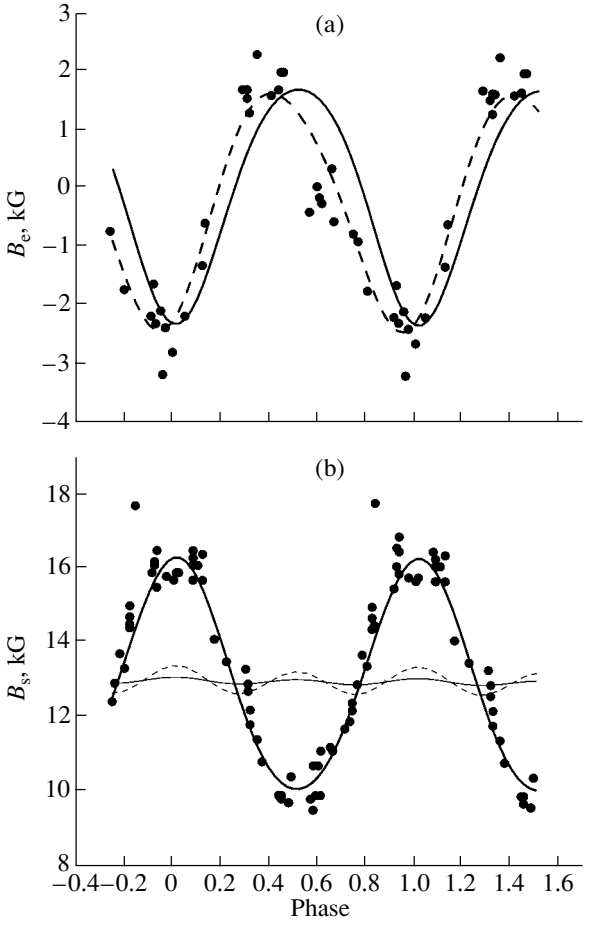

Fig. 3. Phase relations for HD 126515. The points represent the observations, while the curves show (a) the model for all cases (solid curve) and the model taking into account the anharmonic nature of the phase curve (dashed curve), (b) the decentered-dipole model for large β (bold solid curve) and the centered dipole models for large (thin solid curve) and small (thin dashed curve) β.

these conditions, the calculated and observed $B_e(P)$ and $B_s(P)$ phase relations agree well (with the exception of a phase shift near $P = 0.4$, which we decided to ignore); however, in the case of small $\beta = 12°$, we were forced to displace the dipole by 0.45 of the stellar radius, while the required displacement of the dipole for large $\beta = 86°$ is half this value, 0.24 of the stellar radius. Note that the possibility of a dipole-quadrupole-octupole model with small β was also considered in [12], and the derived parameters were close to our values ($\beta = 20°$, $i = 78°$). The version with large β suggested by Preston [15] gives angles close to ours: $\beta = 85°$ and $i = 17°$. Thus, the models obtained by different authors give fairly similar results, but, in our opinion, the model with the smaller displacement of the dipole from the center, and thereby the smaller magnetic-field asymmetry, is more realistic.

The parameters of our preferred model in Table 2D suggest that the dipole is not only displaced along its axis, but also slightly across it. The dipole axis is $r = 0.08$ below the equatorial plane, with its inclination to the plane being 4°. The Mercator-projection map of the surface magnetic-field distribution for this model is shown in Fig. 4a.

We also attempted to develop a model best describing both phase relations. The best agreement was achieved when the dipole was displaced perpendicular to its axis by $r = 0.4$, with the dipole axis being in a plane parallel to the equatorial plane (the latitude for both monopoles is $\delta = +30°$). The dipole parameters are presented in Table 3. It is characteristic that the distance between the monopoles is 0.4 of the stellar radius; i.e., we are dealing with a field resembling that of a bar magnet. Only this position of the monopoles can imitate the surface field structure and give phase relations similar to the observed ones.

Table 2. Model parameters for HD 126515

Monopole sign	Longitude λ, deg	Latitude δ, deg	Field at the pole, G
A. First case, centered-dipole model with small $\beta = 12°$ and $i = 89°$			
+	0	78	−20 770
−	180	−78	+20 770
B. Second case, centered-dipole model with large $\beta = 88.5°$ and $i = 25°$			
+	0	1.5	−22 600
−	180	−1.5	+22 600
C. Third case, decentered-dipole model with small angle $\beta = 12°$, $i = 75°$, and $r = 0.45$			
+	180	78	+91 600
−	0	−78	−6000
D. Fourth case, decentered-dipole model with large $\beta = 86°$, $i = 22°$, and $r = 0.24$			
+	0	−9	−45 800
−	180	−1	+11 100

The star's inclination is $i = 14°$. The $B_s(P)$ phase relation is virtually identical to the previous one (Fig. 3b), and we do not plot it to avoid cluttering the figure. The $B_e(P)$ relation is shown in Fig. 3a by the dashed curve, which clearly coincides with the observations within the scatter in the data points. The Mercator-projection map of the surface magnetic-field distribution is shown in Fig. 4b. Thus far, the data are insufficient to constrain the derived field configuration. We do not insist that this particular, strongly distorted magnetic-field configuration must exist. Currently, we can only say that this result follows from the relative phase shift of the extrema of the $B_e(P)$ and $B_s(P)$ phase relations, based on our modeling technique. The actual situation can be elucidated only after further studies.

4. CONCLUSION

The magnetic-field structures of chemically peculiar stars is a key issue in studies of such stars. Babcock [8] had argued that the magnetic-field configuration of CP stars was predominantly dipolar, but wrote, "...many of the field patterns are probably complex, with considerable fine structure...the results are reported in terms of the effective field on the simplifying assumption that this is homogeneous" [8]. In fact, harmonic phase relations are observed for many stars. Maximum field strengths during passage of the magnetic pole across the apparent meridian and the crossover effect during meridian passage of the magnetic equator are readily observed. However, the accumulating data show that many stars display anharmonic magnetic-field variability. The behavior of the magnetic field resembles that of a decentered dipole [21]. This raises the question of whether this anharmonic behavior is due to asymmetry of the magnetic field relative to the stellar center or differences in the field structure at different magnetic poles. The magnetic-charge technique we have used enables us to obtain a good fit for both the $B_e(P)$ and $B_s(P)$ phase relations for the same initial conditions. As a rule, the iterative fitting indicates the presence of a magnetic-field source consisting of one positive and one negative magnetic charge; i.e., of a dipole. Thus, we can consider the solution correct to first approximation. Further evidence for the predominant role of the dipolar component is provided by the fact that the FeII 6149 lines used to estimate the mean surface field are quite clearly resolved [22]. In the presence of strong deviations of the field from a dipolar character, unshifted or differently shifted components would be observable between the split lines, and such components are completely absent. We also note that the mean surface field is strongest at phases 0 and 0.5, when the longitudinal component is strongest. The magnetic poles pass the meridian at this time. The dipole's displacement from the stellar center, Δr, can characterize the deviation from dipolar structure. The fact that the field structures of many stars differ from a purely dipolar field is well known. Recently found stars displaying this property especially prominently include HD 37776 [23], β CrB [24], 53 Cam [25], HD 126515 (this paper), and others. Our technique could lead to large errors for such stars. For other stars, however, we try to estimate some "effective" dipole parameters, quite similar to the determination

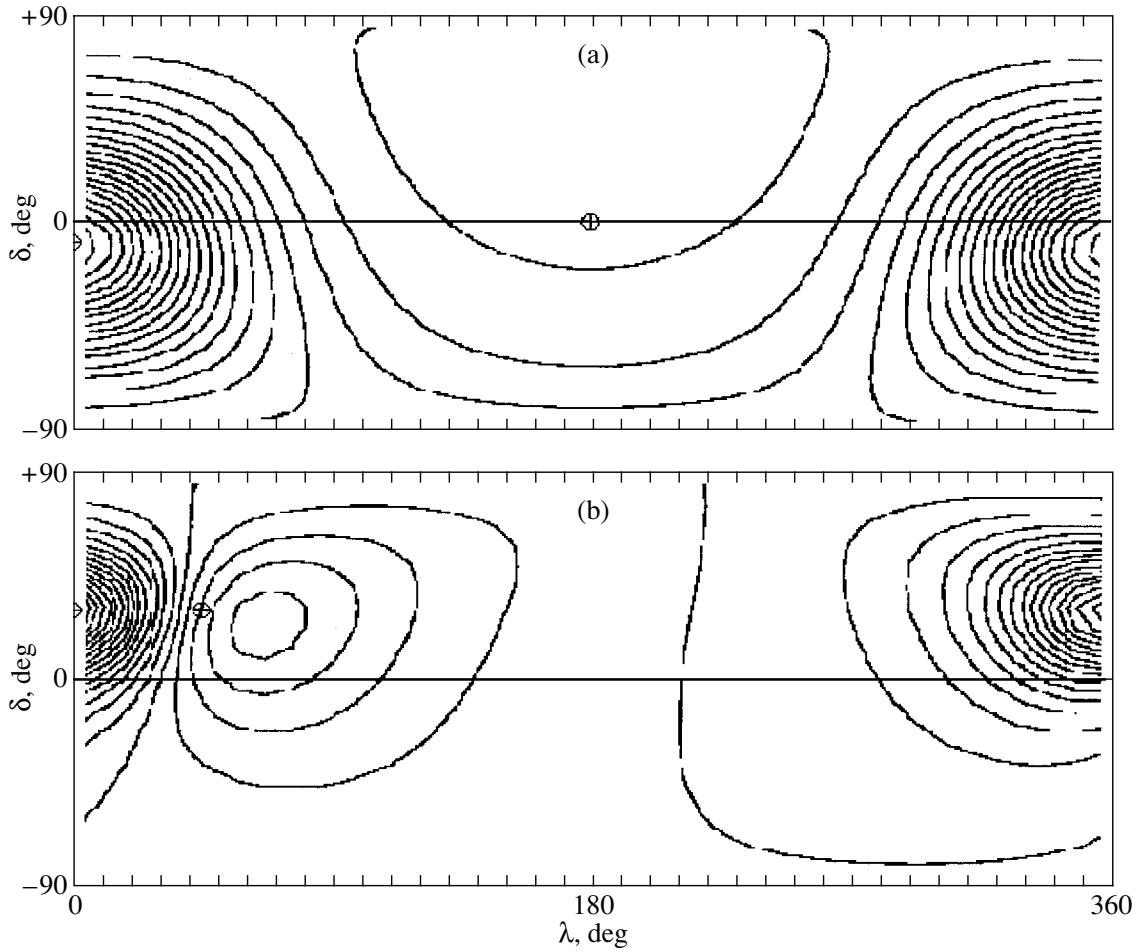

Fig. 4. Distribution of the magnetic intensity on the surface of HD 126515, (a) without and (b) with taking into account the anharmonic nature of the $B_e(P)$ relation.

of mean chemical compositions, despite the fact that the composition may be different at different points of the star. Such data are badly needed to solve a number of problems.

Thus, we currently do not possess sufficient data to judge how correctly the decentered-dipole models describe the real situation, and to what extent they are influenced by measurement errors. We can only assert that the decentered-dipole model is able to satisfactorily explain both phase relations for all the studied stars, as well as the anharmonic nature of the $B_e(P)$ phase relation for HD 126515. In the latter case, we must assume a very strong asymmetry for the magnetic field, which has un unstable nature, so that the magnetic-field configuration should rapidly be transformed into the simplest symmetric structure of a centered dipole. Recall that the anharmonic character of the phase relations for some CP stars, such as β CrB, are even more strongly expressed; for this star, it was necessary to suggest a model with two dipoles on either side of the stellar center [24].

The role of the nonuniform distribution of chemical elements on the surface is obvious. In the case of a dipolar structure, the strengths of the dipolar field at the pole and equator differ by a factor of two, and the field integrated over the surface varies by a factor of about 1.5. In the extreme case when spectral lines used to measure the magnetic field are formed near the magnetic poles, where the corresponding elements are concentrated, the $B_e(P)$ and $B_s(P)$ phase relations will have a larger amplitude, while this amplitude will be smaller if these elements are concentrated near the magnetic equator. In the case of HD 126515, the difference in the magnetic intensities of the poles intensity is more than a factor of two. Hence, there is a real asymmetry of the magnetic field. It is also claimed in [15] that the observations of B_e and B_s require that the star has a hemisphere-asymmetric distribution of its magnetic field, unlike the field of a centered dipole. The properties of the decentered-dipole magnetic field are also discussed in [25]. Of interest in this connection are the results

Table 3. Model parameters for HD 126515, with the anharmonic character of the $B_e(P)$ relation taken into account. Third version of the decentered-dipole model

Monopole sign	Longitude λ, deg	Latitude δ, deg	Field at the pole, G
+	0	30	−60 000
−	45	30	+20 000

Table 4. Magnetic field parameters for slowly rotating CP stars

Star	Type	P, days	i, deg	β, deg	B_p, G
HD 2453	SrCrEu	521	14	80	±6560
HD 12288	CrSi	34.5	24	66	±13 400
HD 116458	SiEuCr	148	75	12	+9510 −6220
HD 126515	CrSr	130	22:	86:	+11 100: −45 800:
HD 200311	SiCrHg	52	30	86	+18 520 −11 420

of Doppler imaging for 53 Cam derived from polarization profiles in the four Stokes parameters [26], which demonstrate that the field structure is strongly nondipolar.

It is striking that the models for many stars suggest that the dipoles are displaced specifically along the dipole axis; i.e., towards the magnetic poles. What physical processes are capable of distorting the magnetic field on a star in this way? This is one of the problems to be solved by future modeling.

If one or both phase relations are determined with a systematic error of any sign, this will always result in errors in all the parameters. Such a bias could be due to either instrumental or methodical effects. For example, geometric effects can easily lead to underestimation of B_e [5]. The model calculated for a centered dipole with $\delta = \pm 45°$ demonstrates that reducing B_e by 10% reduces δ by 5° and i by 5°, and also displaces the dipole from the center by approximately 0.012. These changes are relatively minor, and the above values of r are unlikely to be due solely to measurement errors. The distortion of the field corresponding to the dipole's displacement probably reflects a real asymmetry in the field structure.

We summarize modeling results for five stars with large rotational periods in Table 4. It is difficult to judge at this time if the magnetic fields of these stars display any peculiar properties. Concerning the prediction of Stępień [7] that slowly rotating CP stars should have small angles β, we can already conclude that only one of the five stars we have studied has a small value of this parameter.

REFERENCES

1. Yu. V. Glagolevskiĭ, Astrofizika **47**, 575 (2004) [Astrophys. **47**, 487 (2004)].
2. Yu. V. Glagolevskij and E. Gerth, Bull. Spec. Astrophys. Obs. **58**, 31 (2004).
3. E. Gerth, Yu. V. Glagolevskij, and G. Scholz, in *Stellar Magnetic Fields*, Ed. by Yu. V. Glagolevskij and I. I. Romanyuk (Moscow, 1997), p. 67.
4. E. Gerth and Yu. V. Glagolevskij, *Magnetic Fields of Chemically Peculiar and Related Stars*, Ed. by Yu. V. Glagolevskij and I. I. Romanyuk (Moscow, 2000), p. 151.
5. E. Gerth and Yu. V. Glagolevskij, Bull. Spec. Astrophys. Obs. **55** (2004) (in press).
6. V. P. Khalak, Yu. N. Khalak, A. V. Shavrina, and N. S. Polosukhina, Astron. Zh. **78**, 655 (2001) [Astron. Rep. **45**, 564 (2001)].
7. R. Stępień, Astron. Astrophys. **353**, 227 (2000).
8. H. W. Babcock, in *Stellar Atmospheres*, Ed. by J. Greenstein (University of Chicago, Chicago, 1960), p. 282.
9. G. Mathys and S. Hubrig, Astron. Astrophys., Suppl. Ser. **124**, 475 (1997).
10. G. Mathys, Astron. Astrophys., Suppl. Ser. **89**, 121 (1991).
11. G. Mathys, S. Hubrig, J. D. Landstreet, T. Lanc, and J. Manfroid, Astron. Astrophys., Suppl. Ser. **123**, 353 (1997).
12. J. D. Landstreet and G. Mathys, Astron. Astrophys. **359**, 213 (2000).
13. H. M. Maitzen and H. J. Wood, Astron. Astrophys. **58**, 389 (1977).
14. Yu. V. Glagolevskij and E. Gerth, in *Magnetic Fields of Chemically Peculiar and Related Stars*, Ed. by Yu. V. Glagolevskij and I. I. Romanyuk (Moscow, 2000), p. 161.

15. G. Prestom, Astrophys. J. **160**, 1059 (1970).
16. G. A. Wade, J.-F. Donati, J. D. Landstreet, and S. L. S. Shorlin, Mon. Not. R. Astron. Soc. **313**, 851 (2000).
17. G. Mathys, Astron. Astrophys., Suppl. Ser. **108**, 547 (1994).
18. F. Leone and G. Catanzaro, Astron. Astrophys. **356**, 118 (2001).
19. van den Heuvel, Astron. Astrophys. **11**, 4612 (1971).
20. G. Mathys and N. Lanc, Astron. Astrophys. **256**, 169 (1992).
21. J. D. Landstreet, Astrophys. J. **159**, 1001 (1970).
22. G. Mathys and T. Lanc, Astron. Astrophys. **256**, 169 (1992).
23. Yu. V. Glagolevskij and E. Gerth, Bull. Spec. Astrophys. Obs. **51**, 84 (2001).
24. Yu. V. Glagolevskij and E. Gerth, *Magnetic Stars*, Ed. by Yu. V. Glagolevskij, D. O. Kudryavtsev, and I. I. Romanyuk (Moscow, 2004), p. 142.
25. E. Gerth and Yu. V. Glagolevskij, in *IAU Symp. No. 224,* Ed. by J. Zverko, W. W. Weiss, J. Ziznovsky, and S. J. Adelman (Cambridge Univ. Press, 2004), p. EP6.
26. O. Kochukhov, S. Bagnulo, G. A. Wade, *et al.*, Astron. Astrophys. **414**, 613 (2004).

Translated by N. Samus'

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 1009–1017. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1127–1136.

Dissipation of Diamagnetic Currents and Plasma Heating in Coronal Magnetic Loops

V. V. Zaĭtsev[1] and K. Shibasaki[2]

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

[2]*Nobeyama Observatory, National Astronomical Observatory, Mitaka, Tokyo, 181-8588 Japan*

Received March 13, 2005; in final form, May 18, 2005

Abstract—SOHO and TRACE data have shown that the coronal plasma is heated most actively near sunspots, in magnetic loops that issue from the penumbral region. The source of heating is nonuniform in height, and its power is maximum near the footpoints of the magnetic loops. The heating process is typically accompanied by the injection of dense chromospheric plasma into the coronal parts of the magnetic loops. It is important that the radiative losses cannot be compensated for via electron thermal conduction in the loops, which have temperatures of 1.0–1.5 MK; therefore, some heating source must operate throughout the entire length of the loop, balancing radiative losses and maintaining a quasi-steady state of the loop over at least several hours. As observations show, the plasma density inside the loops exceeds the density of the ambient plasma by more than an order of magnitude. It is supposed that the enhanced plasma density inside the loops results from the development of the ballooning mode of a flute-type instability in the sunspot penumbra, where the plasma of the inner sunspot region, with $\beta_i \ll 1$, comes into contact with the dense chromospheric plasma, which has $\beta_e \gg \beta_i$ (β is the gas-to-magnetic pressure ratio). As the chromospheric plasma penetrates into the potential field of the sunspot, the generated diamagnetic currents balance the excess gas pressure. These currents efficiently decay due to the Cowling conductivity. Even if neutrals are few in number in the plasma (accounting for less than 10^{-5} of the total mass density), this conductivity ensures a heating rate that exceeds the rate of the normal Joule dissipation of diamagnetic currents by 7–8 orders of magnitude. Helium is an important factor in the context of plasma heating in magnetic loops. Its relatively high ionization potential, while not forbidding dielectronic recombination, ensures a sufficiently high number of neutrals in the coronal plasma and maintains a high heating rate due to the Cowling conductivity, even at coronal temperatures. The heating results from the "burning-out" of the nonpotential component of the magnetic field of the coronal magnetic loops. This mechanism provides the necessary heating rate for the plasma inside the loops if the loops are thin enough (with thickness of the order of 10^5–10^6 cm). This may imply that the observed $(1–5) \times 10^8$-cm-thick loops consist of numerous hot, thin threads. For magnetic loops in hydrostatic equilibrium, the calculated heating function exponentially decreases with height on characteristic scales a factor of 1.8 smaller than the total-pressure scale height, since the scale heights for the total pressure and for the ^{4}He partial pressure are different. The heating rate is proportional to the square of the plasma pressure in the loop, in agreement with observational data.

1. INTRODUCTION

The heating of the coronal plasma remains an unresolved problem of solar physics. The heating mechanisms that have been suggested for coronal magnetic loops, which have a structuring effect on the regions surrounding sunspots, can be divided into internal and external mechanisms. Internal mechanisms, which convert the free energy of a magnetic loop into the thermal energy of the plasma, include the ohmic dissipation of magnetic-field-aligned currents inside the loop [1], the tearing instability [2], and the dissipation of magnetic energy due to numerous microflares generated by reconnection process inside the loop [3, 4]. The ohmic dissipation of field-aligned currents requires the presence of an anomalous resistance, and, therefore, the strong filamentation of currents with current-sheet thicknesses of the order of 5.5×10^2 cm. External heating mechanisms, which deliver free energy from outside the loop, include the evaporation of hot chromospheric plasma into the corona [5, 6], the resonant dissipation of Alfvén waves [7–11], and the cyclotron absorption of electromagnetic waves generated during flares [12]. Several years ago, Aschwanden *et al.* [13, 14] carried out a stereoscopic analysis of coronal magnetic loops observed using the *Extreme Ultraviolet Imaging Telescope* on board the SOHO satellite in the tem-

1063-7729/05/4912-1009$26.00

perature range (1.0–2.5) MK. These data showed that the temperature in these loops varies little over several scale heights. A similar result was obtained by Reale and Pares [15] from their analysis of observational data obtained using the TRACE satellite.

Since the temperature gradients are small, the radiative losses cannot be balanced by the electronic thermal conduction in the loop, with the possible exception of segments located near the footpoints and in the transition region. Therefore, as observations suggest, the heating source must be balanced by radiative losses over a considerable part of the loop length, ensuring a quasi-steady state for the loop over at least several hours. The plasma density inside such loops declines exponentially with altitude, with the scale height corresponding to the temperature of the plasma. This testifies that the loops remain nearly in hydrostatic equilibrium during observations covering several (one to three) scale heights. In this case, the heating function (the power per unit volume that is spent on plasma heating) should also exponentially decrease with altitude, with a scale height that is half the barometric scale height for the density. This results from the fact that the radiative losses are proportional to the square of the plasma density. Observational data have revealed no correlation between the magnetic field and the plasma-heating function near the loop footpoints, but a quadratic dependence of the heating function on the plasma pressure.

We point out here that the quadratic pressure dependence of the heating rate is due to the dissipation of diamagnetic currents in the plasma. We consider the dissipation of diamagnetic currents in magnetic-flux tubes under the proviso that a small fraction of neutral atoms is present in the plasma. In this case, Cowling dissipation [16, 17], which results from ion–atom collisions, plays an important role. The rate of this dissipation is many orders of magnitude higher than the dissipation rate for classical conduction due to electron–ion collisions. This is true because the Ampère force associated with the emerging electrical (in our case, diamagnetic) currents drives the plasma relative to the neutral gas, and the speed of this motion can be much larger than the relative speed of the electrons and ions responsible for the current. If we also take into account the fact that the ion mass is substantially larger than the electron mass, it becomes clear that the energy of the plasma (ion) motion exceeds the energy of motion of the electrons relative to the ions by many orders of magnitude. Since the ionic and atomic masses are comparable, even one collision converts a considerable fraction of the directional energy into turbulent energy. In this case, the plasma is heated due to the slow dissipation of the diamagnetic-current-produced, nonpotential component of the loop magnetic field. Then, if the gas-to-magnetic pressure ratio is small ($\beta = 8\pi p/B^2 \ll 1$), the magnetic-field dissipation rate will be lower than the plasma-heating rate by a factor of $\beta^{-1} \gg 1$; i.e., the heating does not considerably change the magnetic field.

At high (coronal) temperatures, a small relative fraction of neutrals is maintained by the presence of helium. Its ionization potential is higher than that of hydrogen, and, in contrast to hydrogen, dielectronic recombination is not forbidden for helium; this keeps the mass fraction of neutral helium atoms at a level of 10^{-5} at a temperature of 2 MK, which is sufficient to maintain the necessary plasma-heating rate in coronal magnetic loops. In this case, the heating function exhibits an altitude dependence with a scale height approximately half that for the mean pressure, since, in hydrostatic equilibrium, the scale heights for the helium partial pressure and the mean plasma pressure differ by about a factor of two. We consider the ballooning mode of the flute instability, which develops at the chromospheric footpoints of coronal magnetic loops [18], as the mechanism that fills the coronal loops with dense plasma. This instability results in additional filamentation of the plasma in the magnetic loop, which makes the heating more efficient.

Section 2 presents the governing equations used to analyze the plasma heating in magnetic loops. In Section 3, we study plasma heating due to the dissipation of diamagnetic currents in a flux tube, including the effect of the Cowling conductivity. In Section 4, we calculate the plasma-heating function and its altitude dependence in the corona taking into account the role of helium in maintaining the necessary number of neutrals at high coronal temperatures. Section 5 considers one possible mechanism for filling the magnetic fields around sunspots with dense chromospheric plasma—the development of the ballooning mode of the flute instability; we also discuss the problem of plasma filamentation in magnetic tubes, which enhances the efficiency of plasma heating. The results are discussed and the basic conclusions formulated in Section 6.

2. GOVERNING EQUATIONS

Let us consider a vertical, cylindrically symmetric magnetic tube in magnetohydrostatic equilibrium with the solar atmosphere. The equation of magnetohydrostatic balance between the pressure gradient, Ampère force, and gravitational force is

$$-\nabla p + \frac{1}{c}\mathbf{j} \times \mathbf{B} + \rho \mathbf{g} = 0, \qquad (1)$$

where p is the plasma pressure,

$$\mathbf{j} = \frac{c}{4\pi}\nabla \times \mathbf{B} \tag{2}$$

is the electrical current, $\mathbf{B}$ the magnetic-field induction, ρ the plasma density, and $\mathbf{g}$ the gravitational acceleration (equal to $g = 2.74 \times 10^4$ cm/s^2 at the solar surface). Equation (1) is valid if the plasma speed V is much less than the sound speed, Alfvén speed, and free-fall speed.

We will obtain the plasma heating and magnetic-field dissipation in the tube from the generalized Ohm's law, which, if neutrals are present in the plasma, can be written [16, 17]

$$\mathbf{E} + \frac{1}{c}\mathbf{V} \times \mathbf{B} = \frac{c}{4\pi\sigma}\nabla \times \mathbf{B} \tag{3}$$
$$- \frac{1}{4\pi en}\mathbf{B} \times (\nabla \times \mathbf{B}) - \frac{\nabla p_\mathrm{e}}{en} + \frac{F^2}{cnm_\mathrm{i}\nu'_\mathrm{ia}}\mathbf{B}$$
$$\times \left[\frac{1}{8\pi}\frac{m_\mathrm{i}}{m_\mathrm{a}}\mathbf{B} \times (\nabla \times \mathbf{B}) - \left(1 - \frac{m_\mathrm{i}}{2m_\mathrm{a}}\right)\rho\mathbf{g}\right],$$

where $\mathbf{E}$ is the electric field;

$$\sigma = \frac{e^2 n}{m_\mathrm{e}(\nu_\mathrm{ei} + \nu_\mathrm{ea})} \tag{4}$$

is the classical conductivity; n is the electron number density; $m_\mathrm{e,i,a}$ are the electron, ion, and atom masses, respectively; ν_ei and ν_ea are the effective rates of collisions of electrons with ions and atoms, respectively; $\nu'_\mathrm{ia} = m_\mathrm{a}\nu_\mathrm{ia}/(m_\mathrm{a} + m_\mathrm{i})$, ν_ia is the effective rate of ion–atom collisions; $p_\mathrm{e} = n\kappa_\mathrm{B}T$ is the electronic pressure; $\kappa_\mathrm{B} = 1.38 \times 10^{-16}$ erg/K is Boltzmann's constant; T is the temperature, which is assumed to be the same for the electrons, ions, and atoms; and

$$F = \frac{n_\mathrm{a}m_\mathrm{a}}{n_\mathrm{a}m_\mathrm{a} + nm_\mathrm{i}} \tag{5}$$

is the mass fraction of the neutral component with the number density n_a. Equation (3) was derived from the equations of three-fluid magnetohydrodynamics for electrons, ions, and neutral atoms, assuming that $F \ll 1$ and $\partial\mathbf{V}/\partial t = 0$.

Variations in the induction $\mathbf{B}$ due to the dissipation of currents in the plasma can be described by the equation [19–23]

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V} \times \mathbf{B}) + \frac{c^2}{4\pi\sigma}\Delta\mathbf{B} \tag{6}$$
$$+ \frac{F^2}{8\pi m_\mathrm{a}}\nabla \times \left\{\frac{1}{n\nu'_\mathrm{ia}}\mathbf{B} \times [\mathbf{B} \times (\nabla \times \mathbf{B})]\right\}.$$

The system of equations (1), (3), (6) should be supplemented with the pressure equation [24]

$$\frac{1}{\gamma - 1}\frac{\partial p}{\partial t} = E_\mathrm{H} - E_\mathrm{r} - \frac{\partial}{\partial s}\left(kT^{5/2}\frac{\partial T}{\partial s}\right), \tag{7}$$

where $\gamma = c_\mathrm{p}/c_\mathrm{v}$ is the ratio of specific heats, E_H the plasma-heating function, and E_r the radiative-loss function [25–27]. The last term on the right-hand side of (7) describes the electronic heat conduction, where

$$k = 0.92 \times 10^{-6}\ \mathrm{erg\ s^{-1}\ K^{-7/2}} \tag{8}$$

is the heat conductivity and s the coordinate measured along the axis of the tube.

In principle, two types of magnetic tubes are possible. The first are tubes that form due to the sweeping of the background magnetic field by convective flows of the photospheric plasma. The footpoints of such tubes are normally located at the meeting points of several supergranulation cells, where the horizontal convective flows converge. Such tubes can be situated far from sunspots, and large electrical currents (up to 10^{12} A) can flow along them; these currents result from the interaction of the convective plasma flow with the intrinsic magnetic field of the tube [28]. For a vertical tube with a magnetic field $\mathbf{B}(0, B_\varphi, B_z)$, the components B_φ and B_z are frequently specified in the form suggested by Gold and Hoyle [29]. Magnetic tubes with field-aligned currents can carry high nonpotential energies and can be sources of large flares.

The second type of tubes are the numerous magnetic tubes that emerge near sunspots due to the filling of thin, threadlike volumes stretched along the magnetic field lines of the sunspot with dense chromospheric plasma. As a rule, no electrical currents flow along such tubes, and only a diamagnetic current develops, so as to reduce the magnetic field inside the tube by a quantity corresponding to the excess gas pressure in the tube. To all appearances, numerous tubes of precisely this type were observed around sunspots by the TRACE satellite. The lack of any notable twisting of the magnetic fields in magnetic tubes of this type suggests an absence of appreciable field-aligned currents in such tubes. It is this type of magnetic tube that we will consider below.

To be specific, we will assume that the plasma pressure in a cylindrically symmetric, vertical magnetic tube varies with the radial coordinate r as

$$p(r) = p_\infty + p_0 \exp\left(-\frac{r^2}{a^2}\right), \tag{9}$$

where p_∞ is the plasma pressure outside the tube, $(p_\infty + p_0)$ the pressure at the axis of the tube, and a the tube's radius. We project (1) onto the radial direction to obtain the radial dependence of the magnetic field in the tube:

$$B_z^2(r) = B_\infty^2 - 8\pi p_0 \exp\left(-\frac{r^2}{a^2}\right), \tag{10}$$

where B_∞ is the z component of the magnetic field outside the tube.

The total electrical current in the tube coincides in this case with the diamagnetic current, and is equal to

$$j_\phi = j = \frac{c}{B_z}(\nabla p)_\mathrm{r} = -\frac{2cp_0 r}{B_z a^2}\exp\left(-\frac{r^2}{a^2}\right), \quad (11)$$

where $(\nabla p)_\mathrm{r}$ is the radial component of the plasma-pressure gradient. The SOHO and TRACE data [13, 14, 30] indicate that the pressure inside loops in hydrostatic equilibrium is of the order of 0.5–1.0 dyne/cm^2 at the loop footpoints. This far exceeds the value for the upper chromosphere (6×10^{-2}–2×10^{-1} dyne/cm^2 [31]). Here and below, we will assume for simplicity that $p_\infty \ll p(r)$. We project (1) onto the z axis, taking into account the fact that the density and pressure are linked by the relationship

$$\rho = \frac{\mu m_\mathrm{H} p}{2\kappa_\mathrm{B} T}, \quad (12)$$

where μm_H is the mean ion mass and m_H the hydrogen-atom mass (in the solar atmosphere, $\mu \approx 1.4$ for H : He = 10 : 1). This yields the altitude dependence of the pressure

$$p = p_0 \exp\left(-\int_0^z \frac{dz}{\Lambda}\right), \quad (13)$$

where $\Lambda = 2\kappa_\mathrm{B}T(z)/\mu m_\mathrm{H} g$ is the pressure scale height.

3. PLASMA HEATING

We obtain from the generalized Ohm's law (3) the rate of plasma heating due to the dissipation of electrical currents in the magnetic tube:

$$q_j(r) = (\mathbf{E} + \frac{1}{c}\mathbf{V}\times\mathbf{B})\mathbf{j} = \frac{c^2}{(4\pi)^2\sigma}(\nabla\times\mathbf{B})^2 + \frac{F^2}{32\pi^2 n m_\mathrm{a}\nu'_\mathrm{ia}}(\mathbf{B}\times(\nabla\times\mathbf{B}))^2. \quad (14)$$

The first term on the right-hand side of (14) describes the contribution of the classical conductivity (4), which is determined by collisions of electrons with ions and atoms, to the dissipation of the current $\mathbf{j} = (4\pi/c)\nabla\times\mathbf{B}$ in the magnetic tube. The second term arises only if neutral atoms are present in the plasma, and describes the dissipation of the diamagnetic currents due to ion–atom collisions. In view of (1), (2), and (11), the heating rate can be written

$$q_j(r) = \frac{c^2}{B_z^2\sigma}(\nabla p)_\mathrm{r}^2 + \frac{F^2}{2nm_\mathrm{a}\nu'_\mathrm{ia}}(\nabla p)_\mathrm{r}^2. \quad (15)$$

We assume that $p_\infty \ll p(r)$ and introduce the mean pressure inside the tube. Using (9), we obtain

$$\overline{p} = \frac{1}{\pi a^2}\int_0^\infty p(r)2\pi r dr = p_0. \quad (16)$$

Similarly, we introduce the plasma-heating rate averaged over the cross section of the tube,

$$q_j = \frac{1}{\pi a^2}\int_0^\infty q_j(r)2\pi r dr. \quad (17)$$

Substituting the expression for $(\nabla p)_\mathrm{r}$ into (15) and using (9) yields for the mean heating rate

$$q_j = \left(\frac{c^2}{B_z^2 a^2\sigma} + \frac{F^2}{2nm_\mathrm{a}\nu'_\mathrm{ia}a^2}\right)p_0^2. \quad (18)$$

Formula (18) seems to reflect the actual dissipation scheme, since, according to the TRACE data [13], the rate of heating of coronal magnetic loops is proportional to the square of the gas pressure inside the loops.

Let us compare the plasma-heating rates associated with the ion–atom collisional dissipation of diamagnetic currents (the Cowling conductivity) and the classical conductivity, determined by electron–ion and electron–atom collisions. We will assume that $\nu'_\mathrm{ia} = m_\mathrm{a} n_\mathrm{a}\sigma_\mathrm{ia}V_\mathrm{Ti}/(m_\mathrm{a}+m_\mathrm{i})$, $V_\mathrm{Ti} = (\kappa_\mathrm{B}T/m_\mathrm{i})^{1/2}$, $n_\mathrm{a} = nm_\mathrm{i}F/m_\mathrm{a}$ and that the cross section for ion–atom collisions at temperatures $T \geq 5\times10^5$ K is given by the formula [32]

$$\sigma_\mathrm{ia} = \frac{3.3\times10^{-10}}{T}\ \mathrm{cm}^2. \quad (19)$$

This yields for the ratio R of the second and first terms on the right-hand side of (18)

$$R = \frac{F^2\omega_\mathrm{e}\omega_\mathrm{i}}{8\nu_\mathrm{ei}\nu'_\mathrm{ia}} \simeq 4\times10^{14}\frac{FB_z^2T^2}{n^2}, \quad (20)$$

where $\omega_\mathrm{e} = eB_z/m_\mathrm{e}c$ and $\omega_\mathrm{i} = eB_z/m_\mathrm{i}c$ are the electron and ion gyrofrequencies, respectively. In addition, we have assumed that ^{4}He makes the main contribution to the neutral component of the plasma at high (coronal) temperatures (this will become clear below); for ^{4}He, $m_\mathrm{a} = 4m_\mathrm{i}$, where m_i is the mass of a hydrogen ion. For loops with temperatures of $T = (1$–$2.5)\times10^6$ K, as were observed by TRACE [13, 14], and assuming $n = 2\times10^9$ cm^{-3} and $B_z = 200$ G, we obtain $R = 4\times10^{12}F$. For hot SXT loops with $T = 5\times10^6$ K and $n = 3\times10^9$ cm^{-3} at the footpoints, we find heating-rate ratios of $R = 2.8\times10^{13}F$.

As we will show below, the mass fraction of neutral helium atoms at 1.5×10^6 K is $F = 2.6\times10^{-5}$,

which yields a heating-rate ratio of $R = 10^8$. When $T = 5 \times 10^6$ K, we have $F = 0.5 \times 10^{-5}$, and $R = 1.4 \times 10^8$. In other words, even at coronal temperatures of $(1-5) \times 10^6$ K, the rate of plasma heating due to ion–atom collisions is eight orders of magnitude higher than that due to the classical electronic conduction. In the Introduction, we explained why allowing for ion–atom collisions substantially increases the plasma-heating rate. The ionized gas must move relative to the neutral gas in order for ion–atom collisional dissipation to operate. The presence of a relative velocity between the ions and neutrals in a stationary tube is also dictated by the equilibrium conditions. It follows from (1) that the radial component of the Ampère force in a stationary tube balances the radial component of the total pressure gradient (of both the ionized and neutral components). On the other hand, the Ampère force is directly exerted only on the ionized component, while it acts on the neutrals indirectly, through ion–neutral collisions (this can easily be checked by writing the equilibrium conditions for the neutrals). In other words, the neutral-pressure gradient inside the tube is balanced by the ion–neutral friction force, whose action implies the presence of a relative motion between the ionized and neutral components of the plasma inside the tube. The primary reason for such motion is the Ampère force, which acts to accelerate the ionized component.

We neglect the first term on the right-hand side of (18) and write for the plasma-heating function (i.e., the amount of heat released due to the dissipation of diamagnetic currents per cubic centimeter of plasma per second)

$$E_{\rm H} = \frac{F^2}{2nm_{\rm a}\nu'_{\rm ia}a^2}p_0^2. \qquad (21)$$

Under chromospheric conditions ($T \le 10^5$ K), the number of neutrals in the plasma is basically determined by the ionization and recombination of hydrogen; therefore, $m_{\rm a} = m_{\rm i}$, and we find that the effective ion–atom collision rate is $\nu'_{\rm ia} = 4.5 \times 10^{-11} nFT^{1/2}$ s^{-1} [32]. Under coronal conditions ($T \ge 5 \times 10^5$ K), as we will show below, the number of neutral atoms is determined by the ionization and recombination of helium, so that $m_{\rm a} = 4m_{\rm i}$, and the effective collision rate is $\nu'_{\rm ia} = 0.75 \times 10^{-6} nFT^{-1/2}$ s^{-1} [32].

4. HEATING FUNCTION

4.1. The Role of Helium

Let us calculate the heating function $E_{\rm H}$ for the specific conditions of the solar corona. As we already note, at sufficiently high temperatures ($T \ge 10^5$ K), the relative density of neutrals in the corona, $F = n_{\rm a}m_{\rm a}/(n_{\rm a}m_{\rm a} + n_{\rm i}m_{\rm i}) \approx n_{\rm a}m_{\rm a}/n_{\rm i}m_{\rm i}$, is mainly governed by the ionization and recombination of helium, for the following reasons. First, the ionization potential of helium ($\varepsilon_{\rm i}({\rm HeI}) = 24.6$ eV) is almost twice that of hydrogen, ($\varepsilon_{\rm i}({\rm H}) = 13.6$ eV). Second, in contrast to hydrogen, dielectronic recombination is not forbidden for helium; at coronal temperatures, this recombination rate is two orders of magnitude higher than the rate of radiative recombination of hydrogen ions [33].

We find from the ionization-balance equation [34]

$$S(T)n_{\rm a}({\rm HeI})n_{\rm e} = \alpha(T)n_{\rm i}({\rm HeII})n_{\rm e}, \qquad (22)$$

where $n_{\rm a}$(HeI) is the number density of neutral helium atoms, $n_{\rm i}$(HeII) the number density of singly ionized helium, $S(T)$ the ionization rate, $\alpha(T)$ the recombination rate, and $n_{\rm e}$ the electron number density. For the ionization rate in the temperature range $0.02 \le \kappa_{\rm B}T_{\rm e}/\epsilon_{\rm i} \le 100$, we can use the formula [35]

$$S(T) \simeq 2.56 \times 10^{-11} \frac{T^{1/2}}{1 + \dfrac{T}{1.71 \times 10^6}} \times \exp\left(-\frac{2.85 \times 10^5}{T}\right) \text{ cm}^3 \text{ s}^{-1}. \qquad (23)$$

We have for the dielectronic-recombination rate of singly ionized helium [33, 36, 37]

$$\alpha(T) \simeq \frac{3.05 \times 10^{-3}}{T^{3/2}} \exp\left(-\frac{4.74 \times 10^5}{T}\right) \text{ cm}^3 \text{ s}^{-1}. \qquad (24)$$

Equations (22)–(24) yield

$$F = \frac{m_{\rm a}\alpha(T)n_{\rm i}({\rm HeII})}{m_{\rm i}S(T)n_{\rm e}} \simeq \frac{m_{\rm a}\alpha(T)p({\rm HeII})}{m_{\rm i}S(T)p}; \qquad (25)$$

when passing to the last equality, we used the fact that the partial pressure of helium is $p({\rm HeII}) \approx 2n_{\rm i}({\rm HeII})\kappa_{\rm B}T$, whereas the total pressure is $p \approx 2n_{\rm e}\kappa_{\rm B}T$, since the degree of ionization is high. The partial pressure of helium decreases exponentially with altitude with the scale height $\Lambda = 2\kappa_{\rm B}T/4m_{\rm i}g$, while the scale height for the total pressure is $\Lambda = 2\kappa_{\rm B}T/\mu m_{\rm i}g$, with $\mu \approx 1.4$. Thus, we finally obtain for the temperature and altitude dependence of the mass fraction of neutrals in the corona

$$F(T,z) \simeq 4 \times 10^7 \frac{1 + \dfrac{T}{1.71 \times 10^6}}{T^2} \times \exp\left(-\frac{1.89 \times 10^5}{T}\right) \exp\left(-\frac{1.86z}{\Lambda}\right). \qquad (26)$$

We assumed in (26) that the ratio of the helium partial pressure to the total pressure is $p({\rm HeII})/p \approx 0.1$ at

the base of the corona (at $z = 0$), consistent with observational data.

It follows from (26) that the fractional content of neutral helium atoms in the corona at temperatures of $T = (1.5–5.0) \times 10^6$ K is $F_{\rm HeI}(z = 0) \approx (0.5–2.6) \times 10^{-5}$, while the fractional content of neutral hydrogen atoms at the same temperatures is $F_{\rm H}(z = 0) \approx (0.56–1.7) \times 10^{-7}$, two orders of magnitude lower. Thus, helium plays the main role in the dissipation of diamagnetic currents in the coronal plasma at high temperatures, and the most efficient dissipation channel is provided by ion–atom collisions, which assign the determining role to the Cowling conductivity (see Section 3).

4.2. Heating Function for Coronal Loops

Substituting the value of $\nu'_{\rm ia}$ at $T \geq 5 \times 10^5$ K into (21), assuming that $p_0 = 2n_{\rm e}\kappa_{\rm B}T$, and using (26), we obtain for the plasma-heating function for the dissipation of diamagnetic currents in coronal magnetic loops

$$E_{\rm H}(T, z) = Q_{\rm ia}(T) \exp\left(-\frac{1.86z}{\Lambda(T)}\right), \qquad (27)$$

where

$$Q_{\rm ia}(T) \simeq \frac{2.8 \times 10^5\, T^{1/2}}{a^2}\left(1 + \frac{T}{1.71 \times 10^6}\right) \times \exp\left(-\frac{1.79 \times 10^5}{T}\right). \qquad (28)$$

The heating function does not depend on the plasma density in the tube, but does depend on the temperature, tube radius, and height. The scale of the height variations of the heating function is approximately half the scale of the pressure variations; therefore, if the temperature inside the loop is nearly constant (as TRACE data indicate for loops with temperatures of $(1.0–2.5) \times 10^6$ K near sunspots [13, 14, 30]), the height variation of the heating function is similar to that for the radiative-loss function, which is proportional to the square of the electron density. Thus, if the electronic thermal conduction is not important, the heating function can be in balance with the radiative-loss function over a wide range of heights, ensuring a quasi-steady state in the loop. The small difference in the height dependences of the heating function and radiative-loss function can be smoothed by the electronic thermal conduction. Taking into account the electronic heat conduction may also be important when analyzing hot flare loops with temperatures of the order of 10^7 K.

In the temperature range $10^{5.75} \leq T \leq 10^{6.3}$ K, the radiative cooling of the plasma due to optical emission can be approximated by the function [25]

$$E_R = n_{\rm e}^2 Q(T)\ {\rm erg\ cm^3\ s^{-1}}, \qquad (29)$$

where $Q(T) \simeq 10^{-21.94}$. The balance condition between the heating and radiative losses,

$$Q_{\rm ia}(T) \exp\left(-\frac{1.86z}{\Lambda(T)}\right) \approx Q(T) n_{\rm e}^2(0) \exp\left(-\frac{2z}{\Lambda(T)}\right), \qquad (30)$$

can then be satisfied if the tube is sufficiently thin, i.e., if

$$a^2 \leq \frac{2.4 \times 10^5\, T^{1/2}}{n_{\rm e}^2 Q(T)}\left(1 + \frac{T}{1.76 \times 10^6}\right) \times \exp\left(-\frac{1.89 \times 10^5}{T}\right). \qquad (31)$$

For the mean values $T = 1.2 \times 10^6$ K and $n_{\rm e} = 2 \times 10^9\ {\rm cm^{-3}}$, typical of the loops observed by TRACE in the UV [13], the condition (31) yields $a \leq 10^6$ cm. If $T = 5 \times 10^6$ K and $n_{\rm e} = 3 \times 10^9\ {\rm cm^{-3}}$, as is characteristic of the hot loops observed by Yohkoh in soft X-rays, we find from (31) that $a \leq 2 \times 10^6$ cm.

Thus, the dissipation of diamagnetic currents can provide the heating of loops with observed widths of $(2–5) \times 10^8$ cm if the loops consist of numerous thin, hot threads with thicknesses of $(1–2) \times 10^6$ cm. It is noteworthy that attempts to interpret the TRACE data led some investigators to the similar conclusion that numerous hot, thin threads may be present inside the observed coronal magnetic loops [15, 30].

As equation (7) implies, the characteristic plasma-heating time $t_{\rm H}$ can be estimated as follows:

$$t_{\rm H} = \frac{p}{(\gamma - 1)E_{\rm H}}. \qquad (32)$$

Setting $\gamma = 5/3$, $T = 1.2 \times 10^6$ K, $n_{\rm e} = 2 \times 10^9\ {\rm cm^{-3}}$, and $a = 10^6$ cm yields $p = 0.8\ {\rm dyne/cm^2}$ and $E_{\rm H} = 0.3 \times 10^{-3}\ {\rm erg\ cm^{-3}\ s^{-1}}$, and a heating time of $t_{\rm H} \simeq 1.9 \times 10^3$ s, or about 32 min, which coincides with the average radiative-loss time, estimated by Aschwanden *et al.* [13] to be about 40 min for the coronal magnetic loops observed by TRACE.

4.3. Magnetic-Field Dissipation

Plasma heating due to the dissipation of diamagnetic currents in a coronal magnetic loop ultimately results from the dissipation of the loop magnetic field. This can be checked by taking the dot product of (6) with $\mathbf{B}$ and integrating over the volume of the tube, using (9) and (10). Setting for our case $\mathbf{V} = 0$, we obtain

$$\int_0^\infty \frac{\partial}{\partial t}\frac{B_z^2}{8\pi} 2\pi r dr = -\int_0^\infty q_{\rm j}(r) 2\pi r dr, \qquad (33)$$

where $q_j(r)$ is determined by (15). In other words, the rate of dissipation of the magnetic field in the tube is equal in magnitude to the rate of plasma heating in the tube. Therefore, although the magnitudes of the dissipation and heating rate coincide, the time of magnetic-field dissipation when $\beta = 8\pi p/B^2 \ll 1$ is a factor of β^{-1} longer than the plasma-heating time. As follows from (33), the time for the decay of the magnetic field due to the dissipation of the diamagnetic currents in the loop is

$$t_B = \frac{\gamma - 1}{\beta} t_H. \quad (34)$$

When $\beta^{-1} = 10^{-2}$, we find from (34) that the magnetic-field dissipation time is $t_B = 35.5$ h if the heating time is $t_H = 32$ min.

5. FLUTE-TYPE INSTABILITY AND MAGNETIC-LOOP FILAMENTATION

The filling of the potential magnetic field near sunspots with high-density plasma, which gives rise to magnetic loops with an enhanced plasma density inside, can result from the development of the ballooning mode of the flute instability at the chromospheric footpoints of magnetic spots [18]. This instability develops at the boundary between the sunspot penumbra, where $\beta \ll 1$, and the ambient chromosphere, where $\beta \leq 1$. The magnetic field expands in this region, and acquires a curvature directed from the ambient chromosphere into the spot. The radius of curvature of the magnetic-field lines is then of the order of the scale height of the inhomogeneous atmosphere:

$$R_c \approx \frac{\kappa_B T}{\mu m_i g}. \quad (35)$$

The curvature of the magnetic field gives rise to the centrifugal force

$$\mathbf{f}_c = \frac{2n\kappa_B T}{R_c^2}\mathbf{R}_c, \quad (36)$$

per cubic centimeter of plasma with density $\rho \approx (n + n_a)m_i$. Therefore, the effective centrifugal acceleration that acts on the chromospheric plasma surrounding the spot is

$$\mathbf{g}_c = \frac{\mathbf{f}_c}{\rho} = \frac{2\kappa_B T}{m_i R_c^2}\frac{n}{n + n_a}\mathbf{R}_c. \quad (37)$$

Substituting (35) into (37) yields

$$g_c = 2g\frac{\mu n}{n + n_a}. \quad (38)$$

The condition for the ballooning instability has the form

$$g_c - g\cos\theta > 0, \quad (39)$$

where θ is the angle between the direction of the radius of curvature $\mathbf{R}_c$ and the vertical.

If $\mathbf{R}_c$ is nearly perpendicular to $\mathbf{g}$, the growth time of the ballooning mode of the flute instability is

$$\tau_b = \frac{1}{2}\left(\frac{\lambda}{\pi g}\right)^{1/2}\left(\frac{\mu n}{n + n_a}\right)^{-1/2}, \quad (40)$$

where λ is the perturbation wavelength. The estimate (40) corresponds to the most unstable situation in which the perturbation wavevector is normal to the magnetic field (see, e.g., [24]). If $n_a \ll n$,

$$\tau_b \approx 2 \times 10^{-3}\lambda\,(\mathrm{cm})^{1/2}\ \mathrm{s}, \quad (41)$$

and perturbations with wavelengths of $\lambda \approx \Lambda \approx 3 \times 10^8$ cm grow at $T = 10^5$ K within a time of the order of 35 s. The perturbation wavelength $\lambda \approx \Lambda$ can be assumed to be the maximum wavelength for the ballooning mode of the flute instability. Magnetic loops with thicknesses of $(3–6) \times 10^8$ cm, as were observed by SOHO/EUT and TRACE, could be associated with this mode. However, we can see from (41) that short-wavelength perturbations grow more rapidly. This means that an original long-wavelength perturbation should break down into smaller-scale perturbations.

The minimum instability scale can be governed by the heating of the plasma inside the flutes due to the dissipation of diamagnetic currents on the scale λ_{min}. If the heating time t_H of the plasma in a flute with a thickness λ_{min} is considerably shorter than the growth time for a flute instability with a wavelength λ_{min}, the enhanced-pressure plasma will have time to leave the flute along the magnetic-field lines under the action of the pressure gradient produced by the heating. As a result, the instability disappears on this scale. In this case, we can assume that the minimum thickness of the flutes can be approximately determined from the condition that the heating time and the time for development of the flute instability are equal; thus, we obtain

$$\lambda_{min} = \left(\frac{F^2 p}{n m_a \nu'_{ia}\sqrt{g}}\right)^{2/3}. \quad (42)$$

For chromospheric regions with $T = 8 \times 10^3$ K, $n = 1.25 \times 10^{10}$ cm^{-3}, $p = 0.17$ dyne/cm^2, $F = 0.75$, and $n_a = 4 \times 10^{10}$ cm^{-3} [38], we find that $\lambda_{min} \geq 3 \times 10^5$ cm.

Comparing the maximum and minimum thicknesses of the flute perturbations $\lambda_{max} \approx 3 \times 10^8$ cm and $\lambda_{min} \approx 3 \times 10^5$ cm, and taking into account the fact that efficient plasma heating occurs in flutes of thickness λ_{min}, we can suppose that observed coronal magnetic loops with thicknesses of $(3–6) \times 10^8$ cm are filled with thin, hot magnetic threads of thickness

$3 \times 10^5 - 3 \times 10^6$ cm. Such structures could be observed with spatial resolutions no worse than $0.1''$.

6. DISCUSSION

We have considered the dissipation of diamagnetic currents that result from the filling of the magnetic field near sunspots with additional plasma, e.g., due to the flute instability at the boundary between the sunspot penumbra and the ambient chromosphere, as a possible mechanism heating the plasma in coronal magnetic loops. Due to the instability, chromospheric plasma with $\beta \leq 1$ finds its way into the regions of the sunspot with $\beta \ll 1$, filling some portion of the magnetic-field lines. To compensate for the excess gas pressure, diamagnetic currents develop in the tube, which are dissipated in part via the usual, classical conduction, but mainly via the Cowling conductivity determined by ion–atom collisions. The rate of ion–atom collisional dissipation in the chromosphere and corona is approximately seven orders of magnitude higher than the rate of dissipation due to the usual conduction. The number of neutrals necessary for the dissipation, $F = n_a m_a / n m_i \approx 10^{-5}$, is maintained in the corona by helium, whose ionization potential is higher than that of hydrogen. In addition, in contrast to hydrogen, dielectronic recombination is not forbidden for helium; the rate of dielectronic recombination at temperatures of (0.5–5) MK is more than two orders of magnitude higher than the rate of radiative recombination [33].

The rate of plasma heating due to diamagnetic currents is proportional to the square of the plasma pressure inside the tube [see (18)]. The heating time (32) in the chromosphere is shorter than the radiative-loss time if the thickness of the tube is $a \leq 10^5$ cm, and comprises $t_H \approx 2$ s when $T = 2 \times 10^4$ K, $p = 1.42 \times 10^{-1}$ dyne/cm^2, $n = 2.4 \times 10^{10}$ cm^{-3}, $n_a = 1.7 \times 10^9$ cm^{-3}, and $F = 6.8 \times 10^{-2}$, which corresponds to a height of 2×10^3 km above the photosphere in the model of the transition region [38]. In the corona, when $T = 1.2 \times 10^6$ K, $n_e = 2 \times 10^9$ cm^{-3}, and $p = 0.8$ dyne/cm^2, the heating function is in balance with the radiative-loss function at a tube radius of $a = 10^6$ cm. In this case, the heating time is $t_H = 1.9 \times 10^3$ s, or about 32 min.

If the tube is in hydrostatic equilibrium, the heating rate exponentially decreases with height with a characteristic scale that is a factor of 1.86 smaller than the scale for the total-pressure variations, which stems from the difference between the scale heights for hydrogen and helium. As SOHO observational data [13] show, the losses due to electronic thermal conduction for coronal magnetic loops with temperatures not exceeding 2.5 MK and located near sunspots are two orders of magnitude lower than the radiative losses. Therefore, for quasi-stationary loops, the heating rate must be in balance with the radiative-loss rate. In the case of heating due to the dissipation of diamagnetic currents, this balance will be achieved if the heating occurs in thin magnetic threads with thicknesses of about 10^5 cm in the chromosphere and about 10^6 cm in the corona. Therefore, it cannot be ruled out that the hot coronal magnetic loops of thickness $(3–5) \times 10^8$ cm that were observed by SOHO and TRACE actually consist of numerous hot, thin threads. The suggestion that the filamentation of magnetic structures plays an important role in the heating of solar and stellar coronas was put forward in [39]. Moreover, it was shown in [15] that the nearly uniform temperature distributions along coronal magnetic loops observed by TRACE [40] can be attributed to the presence of subsecond structures stretched along the magnetic field inside the loops. In their analysis of 41 loops observed by the TRACE satellite, Aschwanden *et al.* [30] noted a good agreement between the observed isothermal state of the loops and their internal threadlike structure.

Thermal instability has been examined as a possible origin for the filamentation of currents in a magnetic tube [41], which requires currents of up to 10^{12} A in a tube of diameter 10^8 cm and turbulent conduction; this instability is characterized by a filamentation time of 2×10^3 s for these parameters. Note that such currents are only typical of flare loops, while no appreciable azimuthal component of the magnetic field is present in the numerous magnetic loops surrounding sunspots; this provides evidence for small values of field-aligned electrical currents in such loops. Another possible means of filamentation is associated with the tearing instability [2], but this has mainly been studied for plane layers.

We have considered the ballooning mode of the flute instability in the chromosphere as a filamentation mechanism for the field in the tube; the growth rate of this mode increases with decreasing instability scale. The flute instability also provides a means for filling coronal magnetic loops with dense chromospheric plasma, so that the plasma density is an order of magnitude higher inside the loops than in the ambient corona. We estimate the minimum scale of the instability to be of the order of $10^5–10^6$ cm, assuming that the growth rate of the ballooning mode is equal to the heating time. Shorter scales are stabilized by heating, since the heating on such scales results in an ejection of plasma from the unstable region of the chromosphere to the coronal part of the loop.

Plasma heating in coronal magnetic loops is due to the dissipation of the loop magnetic field. If $\beta \ll 1$, only a small fraction of the magnetic-field energy

is spent on heating. In this case, the time for the magnetic-field dissipation is a factor of β^{-1} longer than the plasma-heating time. The plasma-heating time is substantially shorter in the chromospheric part of the loop than in the corona. Therefore, the heating of the coronal part of the loop can be accompanied by additional ejections of hot plasma from the loop footpoints.

ACKNOWLEDGMENTS

This work was supported by the National Astronomical Observatory of Japan, the Russian Foundation for Basic Research (project nos. 03-02-20009 (BNTS) and 05-02-16252a), the Basic Research Programs of the Presidium of the Russian Academy of Sciences "Nonstationary Phenomena in Astronomy" and "Solar Activity and Physical Processes in the Sun–Earth System," the Basic Research Program of the Division of Physical Sciences of the Russian Academy of Sciences "Plasma Processes in the Solar System," and the Program for the Support of Leading Scientific Schools of Russia (grant no. NSh-1744.2003.2).

REFERENCES

1. D. C. Spicer, in *Mechanisms of Chromospheric and Coronal Heating*, Ed. by P. Ulmschneider, E. R. Prist, and R. Rosner (Springer-Verlag, Berlin, 1991), p. 547.
2. A. A. Galeev, R. Rosner, S. Serio, and G. S. Vaiana, Astrophys. J. **243**, 131 (1981).
3. E. N. Parker, Astrophys. J. **330**, 447 (1988).
4. E. N. Parker, in *Mechanisms of Chromospheric and Coronal Heating*, Ed. by P. Ulmaschneider, E. R. Prist, and R. Rosner (Springer, Berlin, 1991), p. 615.
5. E. Antonucci, D. Alexander, J. L. Culhane, *et al.*, in *The Many Faces of the Sun: A Summary of the Results from NASA's Solar Maximum Mission*, Ed. by K. T. Strong, J. L. R. Saba, B. M. Haisch, and J. T. Schmelz (Springer, Berlin, 1998), p. 331.
6. P. L. Bornmann, in *The Many Faces of the Sun: A Summary of the Results from NASA's Solar Maximum Mission*, Ed. by K. T. Strong, J. L. Saba, B. M. Haisch, and J. T. Schmelz (Springer, Berlin, 1998), p. 301.
7. J. V. Hollweg, Astrophys. J. **277**, 392 (1984).
8. J. V. Hollweg, in *Mechanisms of Chromospheric and Coronal Heating*, Ed. by P. Ulmschneider, E. R. Prist, and R. Rosner (Springer, Berlin, 1991), p. 423.
9. J. Ionson, Astrophys. J. **276**, 357 (1984).
10. J. M. Davila, Astrophys. J. **317**, 514 (1987).
11. L. Ofman, J. M. Davila, and R. S. Steinolfson, Astrophys. J. **421**, 360 (1994).
12. D. B. Melrose and G. A. Dulk, Astrophys. J. **282**, 308 (1984).
13. M. J. Aschwanden, J. S. Newmark, J.-P. Delabourdiniére, *et al.*, Astrophys. J. **515**, 842 (1999).
14. M. J. Aschwanden, D. Alexander, N. Hurlburt, *et al.*, Astrophys. J. **531**, 1129 (2000).
15. F. Reale and G. Pares, Astrophys. J. **528**, L45 (2000).
16. A. Schlüter and L. Biermann, Z. Natürforsch. **5A**, 237 (1950).
17. T. G. Cowling, *Magnetohydrodynamics* (Adam Hilger, London, 1976), p. 117.
18. K. Shibasaki, Astrophys. J. **557**, 326 (2001).
19. N. M. Bakhareva, V. V. Zaitsev, and M. L. Khodachenko, Sol. Phys. **139**, 299 (1992).
20. M. Suzuki and J. I. Sakai, Astrophys. J. **465**, 393 (1996).
21. M. Suzuki and J. I. Sakai, Astrophys. J. **487**, 921 (1997).
22. M. Ryutova, R. Shine, A. Title, and J. I. Sakai, Astrophys. J. **492**, 402 (1997).
23. K. Furusawa and J. I. Sakai, Astrophys. J. **540**, 1156 (2000).
24. E. R. Priest, *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982; Mir, Moscow, 1985), p. 84.
25. J. C. Raymond, D. P. Cox, and B. W. Smith, Astrophys. J. **204**, 290 (1976).
26. R. Rosner, W. H. Tucker, and G. S. Vaiana, Astrophys. J. **220**, 643 (1978).
27. G. Peres, R. Rosner, S. Serio, and G. S. Vaiana, Astrophys. J. **252**, 791 (1982).
28. V. V. Zaitsev and M. L. Khodachenko, Radiophys. Quantum Electron. **40**, 114 (1997).
29. T. Gold and F. Hoyle, Mon. Not. R. Astron. Soc. **120**, 89 (1960).
30. M. J. Aschwanden, R. W. Nightingale, and P. Alexander, Astrophys. J. **541**, 1059 (2000).
31. J. E. Vernazza, E. H. Avrett, and R. Loeser, Astrophys. J., Suppl. Ser. **45**, 635 (2000).
32. V. E. Golant, A. P. Zhilinskii, and I. E. Sakharov, *Fundamentals of Plasma Physics* (Atomizdat, Moscow, 1975; Wiley, New York, 1980).
33. A. Burgess, Astrophys. J. **139**, 776 (1964).
34. H. Zirin, *The Solar Atmosphere* (Mir, Moscow, 1969; Cambridge Univ. Press, 1988), p. 62.
35. J. D. Huba, *NRL Plasma Formularly* (NRL, Washington, 1994), p. 54.
36. M. Landini and M. Monsignori Fossi, Sol. Phys. **20**, 322 (1971).
37. N. K. Jain and U. Naran, Sol. Phys. **50**, 361 (1976).
38. J. M. Fontela, E. N. Avrett, and R. Loeser, Astrophys. J. **355**, 700 (1990).
39. C. Litwin and R. Rosner, Astrophys. J. **412**, 375 (1993).
40. D. Lenz, E. E. De Luka, L. Golub, *et al.*, Astrophys. J. **517**, L155 (1999).
41. J. Heyvaertz, Astron. Astrophys. **37**, 65 (1974).

Translated by A. Getling

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 1018–1027. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1137–1147.

Energy Conservation in the Restricted Elliptical Three-Body Problem

L. G. Luk'yanov

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

Received June 5, 2005; in final form, July 6, 2005

Abstract—The energy conservation law for the elliptical three-body problem is derived using an invariant relation corresponding to the Jacobi integral in the circular problem. The minimum-energy surfaces are constructed, which transform in the case of zero eccentricity into zero-velocity surfaces. Some astronomical applications of the results are considered. In particular, it is shown that Roche-lobe overflow displays pulsational behavior in the elliptical three-body problem. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The equations of motion in the restricted, circular three-body problem admit the existence of the Jacobi integral, which can be used to find the domains of possible motions. The boundary of these domains, which is defined by the condition that the velocity of a low-mass body in the rotating coordinate system be equal to zero, is usually named the *zero-velocity surface*. Hill [1] was the first to obtain and analyze zero-velocity surfaces in the restricted circular three-body problem when constructing a theory for the motion of the Moon. Later, such surfaces became widely used in many other problems of celestial mechanics. These surfaces are also commonly named *Hill surfaces*, after the person who was the first to discover and study them. These are not the *surfaces of zero kinetic energy* known for the general problem of three and more bodies [2], since the kinetic energy of a small body in the restricted three-body problem contains first- and zeroth-order terms with respect to the velocity components, in addition to second-order terms.

Attempts have been made to construct approximate zero-velocity surfaces in the restricted, elliptical three-body problem [3–5]. Ovenden and Roy [3] showed that, in the case of small eccentricity, the zero-velocity surfaces for the circular problem can be used over a time interval of approximately several revolutions of the primaries. In his fundamental analysis of the restricted three-body problem, Szebehely [5] pointed out the possibility of constructing approximate zero-velocity surfaces for the planar elliptical three-body problem over a negligibly short time interval near the passage of the main bodies through the pericenters of their orbits. The construction of the domains of possible motion in the restricted elliptical three-body problem is made difficult by the lack of a Jacobi integral, although an integral invariant relation is known for this case. In the current paper, we use this relation to derive the energy-conservation law for the elliptical problem, which contains one unknown periodic function. We use this law to construct pulsating domains of possible motions and the bounding surfaces of these domains. Such surfaces have not been considered before and have no special names. These are surfaces where the left-hand side of the invariant relation, which has the meaning of some sort of energy of the low-mass body, takes its minimum value (among other things, the velocity of the small body is equal to zero at such a surface). We will name these surfaces as *minimum-energy surfaces*. If the eccentricity of the orbit of the main bodies is zero, these surfaces become the zero-velocity surfaces of the circular problem. The results obtained open new opportunities for astronomical applications, some of which we consider in this work.

2. THE INVARIANT RELATION

The motions of a small body M in the restricted elliptical three-body problem can conveniently be analyzed using differential equations in the following rotating and pulsating coordinate system xyz [6–9]:

$$x'' - 2y' = \frac{\partial \Omega}{\partial x}, \quad y'' + 2x' = \frac{\partial \Omega}{\partial y}, \qquad (1)$$
$$z'' = \frac{\partial \Omega}{\partial z},$$

where

$$\Omega = \rho[(x^2 + y^2 - ez^2\cos v)/2 + U],$$
$$U = (1-\mu)/r_1 + \mu/r_2, \quad \rho = 1/(1 + e\cos v),$$

$$r_1 = \sqrt{(x+\mu)^2 + y^2 + z^2}, \qquad (2)$$
$$r_2 = \sqrt{(x+\mu-1)^2 + y^2 + z^2};$$

1063-7729/05/4912-1018$26.00

μ is the mass of the smaller of the two main bodies ($\mu \le 1/2$), and e is the eccentricity of their orbits. A prime denotes differentiation with respect to the true anomaly v of the primaries, which is chosen as an independent variable. The units are chosen to make the sum of the masses of the primaries, the focal parameter of their relative orbit, and the gravitational constant all equal to unity. If $e = 0$, Eqs. (1) determine the motion of the small body in the restricted circular problem, which admits the existence of the Jacobi integral. An attempt to find a similar integral in the elliptical problem does not yield a positive result since the function Ω depends explicitly on v. Indeed, if we multiply (1) by x', y', and z', respectively, sum the resulting equations, and integrate the sum over the true anomaly v of the main bodies from v_0 to v, we obtain

$$\frac{V^2}{2} - \frac{V_{v_0}^2}{2} = \int_{v_0}^{v} \left(\frac{\partial \Omega}{\partial x} x' + \frac{\partial \Omega}{\partial y} y' + \frac{\partial \Omega}{\partial z} z' \right) dv, \quad (3)$$

where $V^2 = x'^2 + y'^2 + z'^2$, $V_{v_0}^2 = \left(V^2\right)_{v=v_0}$.

The integrand is not a total differential if $e \neq 0$. Therefore, (3) is not a first integral, and must be viewed as an integral invariant relation. We now add and subtract from the integrand the partial derivative $\partial\Omega/\partial v$, select the total differential of the function Ω, and integrate the result to derive another form of the invariant relation from (3):

$$\frac{V^2}{2} - \Omega = \frac{V_0^2}{2} - \Omega_0 - \int_{v_0}^{v} \frac{\partial \Omega}{\partial v} dv. \quad (4)$$

We then compute explicitly the partial derivative

$$\frac{\partial \Omega}{\partial v} = \frac{e \sin v}{(1 + e \cos v)^2} \left[\frac{1}{2} \left(x^2 + y^2 + z^2 \right) + U \right] \quad (5)$$

and substitute this into (4) to obtain

$$\frac{V^2}{2} - \Omega = \frac{V_0^2}{2} - \Omega_0 - \int_{v_0}^{v} \frac{e \sin v}{(1 + e \cos v)^2} W dv, \quad (6)$$

where $W = (x^2 + y^2 + z^2)/2 + U$. If in addition we make the change of integration variable from the true anomaly v to the quantity $\rho = 1/(1 + e \cos v)$, which is equal to the dimensionless distance between the main bodies, we can write the invariant relation in the form

$$\frac{V^2}{2} - \Omega = \frac{V_0^2}{2} - \Omega_0 - \int_{\rho_0}^{\rho} W d\rho. \quad (7)$$

The function W takes on only positive values, so that the inequalities

$$\int_{\rho_0}^{\rho} W d\rho > 0 \quad \text{at } \rho > \rho_0, \quad (8)$$

$$\int_{\rho_0}^{\rho} W d\rho < 0 \quad \text{at } \rho < \rho_0$$

will always be satisfied.

3. THE ENERGY CONSERVATION LAW

To compute the integral in the right-hand part of the invariant relation (6), we must know the explicit dependences of the coordinates x, y, z on the true anomaly v. If such dependences are defined by a certain partial solution of (1), relation (6) can be reduced to a quadrature, which is valid only along the trajectory corresponding to the partial solution in question. Another partial solution yields a different quadrature with a different functional dependence. If we assume that the general solution of (1) is known in the form of explicit functions of time and six arbitrary constants, this solution can be substituted into the quadrature to obtain

$$\int_{v_0}^{v} \frac{\partial \Omega}{\partial v} dv = u(v) - u(v_0), \quad (9)$$

where, strictly speaking, $u(v)$ is an unknown function, because determining this function requires knowledge of the general solution of the initial equations. We now use (9) to write the invariant relation in the form

$$\frac{V^2}{2} - \Omega + u(v) = h, \quad (10)$$

where $h = V_0^2/2 - \Omega_0 + u(v_0)$.

The first two terms on the left-hand side of this equation have a form that is similar to the left-hand side of the Jacobi integral in the circular problem. They are sometimes referred to as the energy of the body M, although this is not so from the viewpoint of theoretical mechanics. This energy has no generally accepted name; Roy [10] calls it the relative energy. However, the sum of these terms would better be called the *Jacobi energy*, after the person who first discovered and studied it. The third term $u(v)$ determines the *additional energy* of the body M. This is a potential energy that the body M receives or gives to the primaries as they move in their elliptical orbits. Even if the body is at rest, its energy $u(v)$ varies periodically because of the motion of the primaries. The

right-hand side of (10) contains the same quantities, but computed at the initial time, $v = v_0$.

Equation (10) is the energy-conservation law for body M in the restricted elliptical three-body problem. This law can be formulated as follows. The total energy of body M, which consists of the Jacobi energy $V^2/2 - \Omega$ and the additional energy $u(v)$, is a constant that depends on both the initial Jacobi energy and the chosen trajectory (i.e., on six arbitrary independent variables). Along each trajectory, its own energy-conservation law is satisfied. The quantity h can be viewed as an energy constant which has a particular value on each particular trajectory. When $e = 0$, relation (10) transforms into the Jacobi integral. We now set $v_0 = \pi$ and then $v_0 = 0$ in (10), to rewrite the energy-conservation law in the form

$$\frac{V_a^2}{2} - \Omega_a + u(\pi) = \frac{V^2}{2} - \Omega + u(v) = \frac{V_p^2}{2} - \Omega_p + u(0), \quad (11)$$

where the subscripts p and a refer to the pericenter and apocenter of the orbit of the main bodies, respectively. Equation (10) cannot be considered a first integral of the equations of motion, since the function $u(v)$ is unknown and the coordinates in the integral (9) are substituted with their (unknown) relations, which are functions of v.

The function $u(v)$ is unknown, but we can determine some of its properties. Because $\rho(v + 2\pi) = \rho(v)$ and $W > 0$, we can easily establish from the equation

$$\int_{v_0}^{v} \partial\Omega/\partial v dv = \int_{\rho_0}^{\rho} W d\rho = u(v) - u(v_0)$$

that $u(v)$ is a periodic function with period 2π. Indeed, the equation

$$u(v + 2\pi) - u(v_0) = \int_{v_0}^{v+2\pi} \partial\Omega/\partial v dv = \int_{\rho_0}^{\rho(v+2\pi)} W d\rho = \int_{\rho_0}^{\rho(v)} W d\rho = \int_{v_0}^{v} \partial\Omega/\partial v dv = u(v) - u(v_0)$$

implies that $u(v + 2\pi) = u(v)$. In view of (8), we note that the additional energy $u(v)$ monotonically increases as the primaries move from the pericenter to the apocenter, and the body M acquires an additional (potential) energy from the primaries. When the main bodies move in the opposite direction, the additional energy decreases, and the body M transfers energy to the primaries. The following inequalities are always satisfied when $\pi \geq v \geq 0$:

$$u(\pi) \geq u(v) \geq u(0). \quad (12)$$

The function $u(v)$ reaches its maximum and minimum values at the apocenter and pericenter of the orbit of the primaries, respectively. The difference $u(\pi) - u(0)$ is a small quantity if the eccentricity is small, and equals zero when $e = 0$. In this case, the energy conservation law (11) is close to the Jacobi integral, and coincides with it when $e = 0$. However, $u(\pi) - u(0)$ increases with increasing eccentricity and, as we will see below [see (19)], it tends to infinity as $e \to 1$. The behavior of $u(v)$ agrees with the symmetry properties of the first integrals of the restricted three-body problem [11]. It is obvious from (11) and (12) that in the restricted elliptical three-body problem, the Jacobi energy is a periodic function with period 2π. It reaches its maximum and minimum at the pericenter and apocenter of the orbits of the main bodies, respectively, and the following inequalities are always satisfied:

$$\frac{V_a^2}{2} - \Omega_a \leq \frac{V^2}{2} - \Omega \leq \frac{V_p^2}{2} - \Omega_p. \quad (13)$$

A certain (minimum) part of the additional energy $u(v)$ can be determined explicitly. To this end, we must determine the minimum of the function W.

4. FINDING THE MINIMUM OF W

In the coordinate system considered, the function W coincides with the potential of the Lagrange's problem of three fixed centers [12]: the problem of two fixed centers producing a Newtonian attractive force and the third fixed center located at the center of mass, which produces a repulsive (or attractive) force that obeys Hooke's law. The function W must have a minimum, since W is continuous in any closed domain and is bounded from below: $W > 0$. The necessary conditions for an extremum of function W can be written in the form

$$\frac{\partial W}{\partial x} = x - \frac{1-\mu}{r_1^3}(x + \mu) - \frac{\mu}{r_2^3}(x + \mu - 1) = 0, \quad (14)$$

$$\frac{\partial W}{\partial y} = y\left(1 - \frac{1-\mu}{r_1^3} - \frac{\mu}{r_2^3}\right) = 0,$$

$$\frac{\partial W}{\partial z} = z\left(1 - \frac{1-\mu}{r_1^3} - \frac{\mu}{r_2^3}\right) = 0.$$

Equations (14) coincide to within constant factors in some of the terms with the equations for the coordinates of the libration points for the motion of a body in the field of two fixed attracting galaxies located

in the repulsive field of the cosmic vacuum [13]. Luk'yanov [13] showed that an equation of the form (14) admits the existence of three collinear libration points and a libration ring—a circle whose distance from each of the two primaries is equal to the separation between them. The function W takes on its minimum value $W_{\min}$ on this circle, which is defined by the equations $x = (1-2\mu)/2$, $y^2 + z^2 = 3/4$:

$$W_{\min} = [1 - \mu(1-\mu)]/2. \tag{15}$$

We can now introduce the functions $\widetilde{W} = W - W_{\min} = (x^2+y^2+z^2)/2 + U - [1-\mu(1-\mu)]/2$ and $\tilde{u}(v) = u(v) - W_{\min}/(1+e\cos v)$, to write the energy-conservation law (11) in the form

$$\begin{aligned} &\frac{V_a^2}{2} - \Omega_a + \frac{W_{\min}}{1-e} + \tilde{u}(\pi) \\ &= \frac{V^2}{2} - \Omega + \frac{W_{\min}}{1+e\cos v} + \tilde{u}(v) \\ &= \frac{V_p^2}{2} - \Omega_p + \frac{W_{\min}}{1+e} + \tilde{u}(0) = h. \end{aligned} \tag{16}$$

We then use the inequalities

$$\widetilde{W} \geq 0, \qquad \int\limits_{1/(1-e)}^{\rho} \widetilde{W} d\rho \leq 0, \tag{17}$$

$$\int\limits_{1/(1+e)}^{\rho} \widetilde{W} d\rho \geq 0, \qquad \tilde{u}(\pi) \geq \tilde{u}(v) \geq \tilde{u}(0)$$

to obtain

$$\begin{aligned} &\frac{V_a^2}{2} - \Omega_a + \frac{W_{\min}}{1-e} \leq \frac{V^2}{2} - \Omega \\ &+ \frac{W_{\min}}{1+e\cos v} \leq \frac{V_p^2}{2} - \Omega_p + \frac{W_{\min}}{1+e}. \end{aligned} \tag{18}$$

The inequalities (18) generalize (13), since the former include additional terms containing $W_{\min}$. Inequalities (17) and (18) imply the following inequality for the range of variation of the Jacobi energy:

$$\begin{aligned} &\frac{V_p^2}{2} - \Omega_p - \left(\frac{V_a^2}{2} - \Omega_a\right) \\ &= u(\pi) - u(0) \geq \frac{e}{1-e^2}\left[1 - \mu(1-\mu)\right]. \end{aligned} \tag{19}$$

It is obvious from this inequality that this range tends to infinity as $e \to 1$.

5. DOMAINS OF POSSIBLE MOTION

If we use the inequality $V^2 \geq 0$, which is valid for all real motions, (10) yields the domain of possible motions in the form $u - \Omega \leq h$. If we further use (12), we obtain the domain of possible motions in the form

$$\Omega \geq \Omega_p - \frac{V_p^2}{2}. \tag{20}$$

The boundary of this domain is the *surface of zero velocity and zero additional energy*

$$x^2 + y^2 - ez^2\cos v + 2U = C_p(1 + e\cos v), \tag{21}$$

where $C_p = 2\Omega_p - V_p^2$ is the Jacobi energy when $v = 0$. The use of the inequality $V^2 \geq 0$ jointly with inequalities (17) yields the domain of possible motions in the form

$$\Omega - \frac{W_{\min}}{1+e\cos v} \geq \Omega_p - \frac{V_p^2}{2} - \frac{W_{\min}}{1+e}. \tag{22}$$

We refer to the boundary of this domain as the *surface of minimum energy*. Its equation can be written in the form

$$\begin{aligned} &x^2 + y^2 - ez^2\cos v + 2U - 2W_{\min} \\ &= C_p'(1 + e\cos v), \end{aligned} \tag{23}$$

where $C_p' = C_p - 2W_{\min}/(1+e)$. The body M cannot go beyond these boundaries for given values of μ, e, C_p' and v. When $W_{\min} = 0$, the minimum-energy surfaces transform into the surfaces of zero velocity and zero additional energy, and, when $e = 0$, they transform into the zero-velocity surfaces of the restricted circular problem. By moving all terms containing $W_{\min}$ into the right-hand side of (23), we obtain the nonnegative term $2eW_{\min}(1-\cos v)/(1+e)$. This means that the domain of possible motions bounded by the surface (23) is smaller than surface (21). These surfaces coincide only when $v = 0$. In other words, (23) defines the domain of possible motions more rigorously than (21). We therefore specify the boundary of the domain of possible motions below using the surface of minimum energy (23). As is evident from (23), the surface of minimum energy varies periodically with v (i.e., with changes of time), and is always located between the two surfaces corresponding to the true anomaly values of 0 and π:

$$x^2 + y^2 - ez^2 + 2U - 2W_{\min} = C_p'(1+e), \tag{24}$$

$$x^2 + y^2 + ez^2 + 2U - 2W_{\min} = C_p'(1-e). \tag{25}$$

Thus, the surfaces (24) and (25) for fixed C_p', μ, and e define the boundaries inside which all the surfaces (23) are located for all v.

In his monograph, Szebehely [5] uses the invariant relation (6) for the planar elliptical problem to derive the equation of zero velocity curves in the form

$$x^2 + y^2 + 2U = C_p(1 + e\cos v).$$

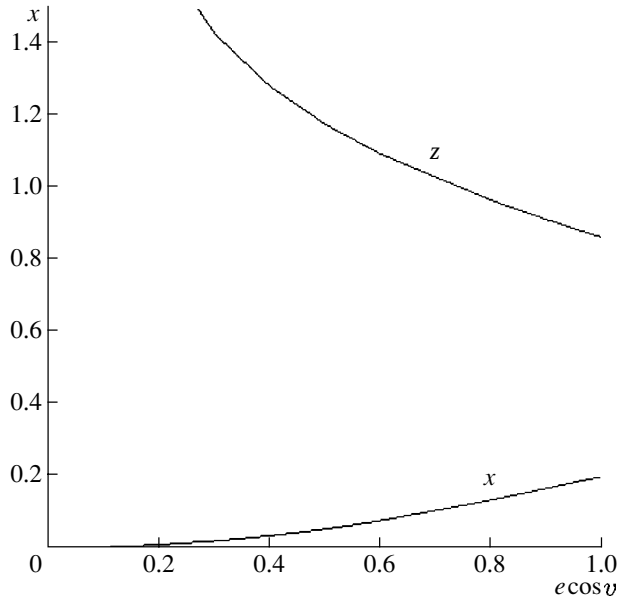

Fig. 1. Coordinates of the singular points $\tilde{L}_6(x, 0, z)$ and $\tilde{L}_7(x, 0, -z)$ as a function of $e\cos v$ for $\mu = 0.3$.

Szebehely [5] calls these curves "approximate pulsating zero-velocity curves," which are valid in the infinitesimal interval $[0, v]$. However, these are actually pulsating curves of zero velocity and zero additional energy which can be obtained from (21) when $z = 0$. Szebehely [5] did not introduce the function W and did not consider inequalities (13), and so had to assume that the interval $[0, v]$ was infinitesimal to provide a basis for the requirement

$$\int_{v_0}^{v} \partial\Omega/\partial v dv = 0.$$

In view of inequalities (13), we can assert that Szebehely's [5] result is valid in finite (not just infinitesimal) intervals $[0, v]$ and for any v, and not just for the planar, but also for the three-dimensional problem.

6. SINGULAR POINTS OF THE MINIMUM-ENERGY SURFACES

The algebraic equations for the singular points of the family of curves (23) are the same as those for the libration points in the restricted elliptical three-body problem [14]:

$$x - \frac{1-\mu}{r_1^3}(x+\mu) - \frac{\mu}{r_2^3}(x+\mu-1) = 0, \quad (26)$$

$$y\left(1 - \frac{1-\mu}{r_1^3} - \frac{\mu}{r_2^3}\right) = 0,$$

$$-z\left(e\cos v + \frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3}\right) = 0.$$

The restricted circular and elliptical problems have three collinear libration points L_1, L_2, L_3 and two triangular libration points L_4, L_5. The circular problem has two additional infinitely distant libration points $L_{\pm\infty}$ [15]. All these libration points are also singular points of the family of curves (23). There are no other libration points, but there are other singular points. The total number of libration points is not equal to the number of singular points, although both are determined from the same equations. This is due to the fact that the surfaces of minimum energy are constructed separately for each value of the true anomaly v; i.e., the quantity v in (26) is considered to be constant when the coordinates of the singular points are determined. However, the true anomaly is considered to be variable when we determine the partial stationary solutions of the equations of motion (i.e., the libration points). Therefore, the singular points include not only libration points, but also the solutions of (26) for $z \neq 0$ and $v = \text{const}$. To find these, we substitute the relation $(1-\mu)/r_1^3 + \mu/r_2^3 = -e\cos v$ derived from the third equation of (26) into the second equation, which can then be written in the form $y(1 + e\cos v) = 0$, which implies $y = 0$. This means that the points to be found are located in the xz plane. We obtain the following set of two equations for these points:

$$x(1 + e\cos v) - \mu(1-\mu)\left(\frac{1}{r_1^3} - \frac{1}{r_2^3}\right) = 0, \quad (27)$$

$$e\cos v + \frac{1-\mu}{r_1^3} + \frac{\mu}{r_2^3} = 0.$$

It is obvious from the second equation of this set that there are no real solutions when $\cos v > 0$; solutions can exist only when $\cos v \leq 0$. Because z enters (27) only in squared form, these solutions must exist in pairs that are symmetric about the x axis. We refer to such solutions as *coplanar singular points* and denote them $\tilde{L}_6 = (x_6, 0, z_6)$ and $\tilde{L}_7 = (x_7, 0, z_7)$, where $x_7 = x_6, z_7 = -z_6$. When $v = \pi/2$, the coplanar singular points coincide with the infinite libration points $L_{\pm\infty} = (0, 0, \pm\infty)$ known for the circular problem. Equations (27) also have an obvious triangular solution $r_1 = r_2 = 1$ when $\cos v = -1$ and $e = 1$. This is one of the points of the libration ring in the collinear problem. In other cases, the coordinates of coplanar singular points can be determined only numerically. Figure 1 shows the coordinates of the singular point $\tilde{L}_6$ for the mass parameter $\mu = 0.3$ as a function of the product $e\cos v$. When $e \to 0$, the coplanar singular points tend to the infinitely distant libration points.

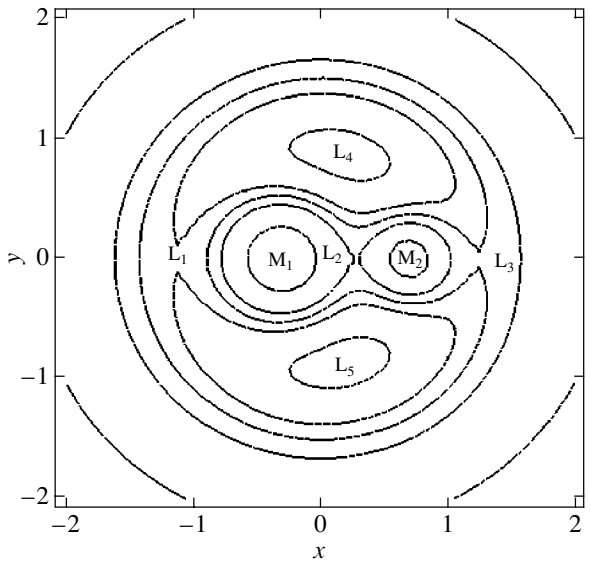

Fig. 2. Intersections of the surfaces of minimum energy by the xy plane for $v = 0$, $\mu = 0.3$, $e = 0.2$. Shown are the sections of the surfaces with energy levels $C'_{\rm p} = 4.342$, 2.622, 2.302, 2.082, and 1.742.

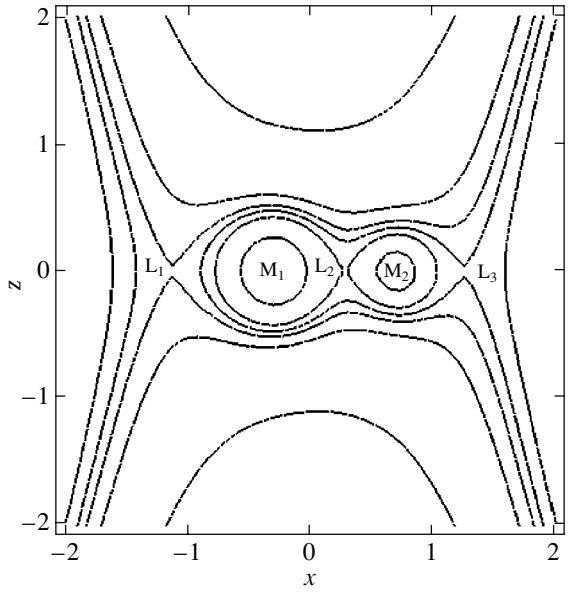

Fig. 3. Intersections of the surfaces of minimum energy by the xz plane for the same parameter values as in Fig. 2. Also shown is the surface with the energy level $C'_{\rm p} = 0.542$.

7. CONSTRUCTION OF THE SURFACES OF MINIMUM ENERGY

We now turn to an analysis of the surfaces of minimum energy. We note in this connection that, at very large distances from the coordinate origin, (25) can be written in the form

$$x^2 + y^2 + ez^2 + \varepsilon = C'_{\rm p}\,(1 - e)\,, \qquad (28)$$

where ε is a small quantity and $C'_{\rm p}$ is a large quantity. It is obvious that, at large distances, the surface considered is close to a prolate ellipsoid of revolution. At large distances from the coordinate origin, (24) transforms to the form

$$x^2 + y^2 - ez^2 + \varepsilon = C'_{\rm p}\,(1 + e) \qquad (29)$$

and defines a surface that is close to a one-sheet or two-sheet hyperboloid of revolution when $C'_{\rm p} > 0$ and $C'_{\rm p} < 0$, respectively. If we move the term $-2W_{\min}$ to the right-hand side of (23), the remaining left-hand side for $z = 0$ will exactly coincide with the analogous expression for the circular problem. Therefore, the values of $C_1, C_2, C_3, C_4 = C_5$ calculated at the libration points L_1, L_2, L_3, L_4, L_5 fully coincide with the analogous values in the circular problem. We then determine the corresponding $C'_{\rm p}$ values at the singular points, including $\tilde{L}_6$, $\tilde{L}_7$, using the formula

$$C'_{\rm p} = (C_i - 2W_{\min})/(1 + e\cos v) \quad (i = 1, 2, \ldots, 7). \qquad (30)$$

If $C'_{\rm p}$, e, μ, and v are related by (30), the curves of minimum energy for $i = 1, \ldots, 5$ in the $z = 0$ plane have the same appearance as in the circular problem. The surfaces of minimum energy (23) for fixed $C'_{\rm p}$, μ, e and varying true anomaly v are pulsating surfaces contained between the two surfaces (24) and (25).

Figures 2–5 show the intersections of surfaces (24) and (25) by the xy and xz planes for $\mu = 0.3$ and $e = 0.2$. Let us see how the minimum-energy surfaces change with monotonically decreasing $C'_{\rm p}$, starting from infinitely large values. At very large $C'_{\rm p}$, the surfaces of minimum energy pulsate either between the ellipsoid (27) and the one-sheet hyperboloid (28), or between two surfaces that are close to spheres surrounding the body M_1, $2(1-\mu)/r_1 + \varepsilon = C'_{\rm p}(1+e)$, $2(1-\mu)/r_1 + \varepsilon = C'_{\rm p}(1-e)$, or the body M_2, $2\mu/r_2 + \varepsilon = C'_{\rm p}(1+e)$, $2\mu/r_2 + \varepsilon = C'_{\rm p}(1-e)$. The sizes of these surfaces are such that the ellipsoid is always located inside the hyperboloid, and, on the contrary, the spheres corresponding to $v = 0$ are located inside the spheres for $v = \pi$. Other $v \in (0, \pi)$ yield other intermediate surfaces.

Depending on the initial position of the body M, the domains of possible motion are located either inside the pulsating spheres indicated above (satellite-type motion) or outside a surface pulsating between an ellipsoid and hyperboloid. The sizes of the spheres increase and those of the ellipsoid and hyperboloid decrease with decreasing $C'_{\rm p}$. The shapes of the domains are distorted but the qualitative picture remains the same up to the intersection of the surfaces at

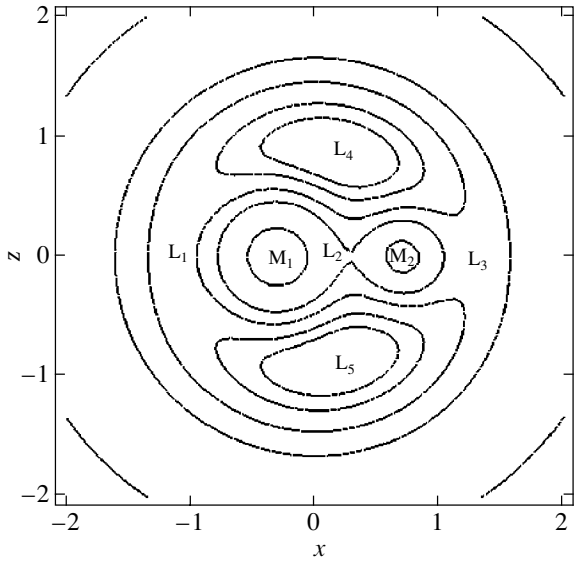

Fig. 4. Intersections of the minimum-energy surfaces by the xy plane for $v = \pi$, $\mu = 0.3$, $e = 0.2$. Shown are the sections of the surfaces with energy levels $C'_{\rm p} = 7.342$, 3.942, 3.342, 2.942, and 2.742.

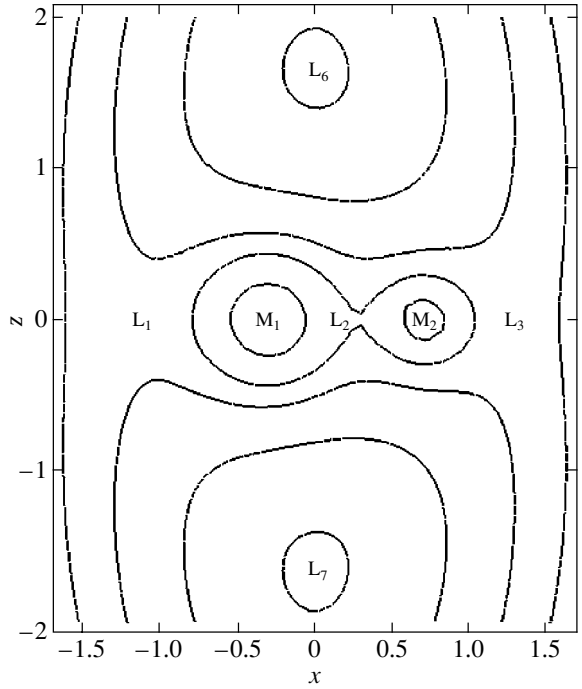

Fig. 5. Intersections of the minimum-energy surfaces by the xz plane for the same parameter values as in Fig. 4. Shown are the sections of the surfaces with energy levels $C'_{\rm p} = 7.342$, 3.942, 2.942, 1.942, and 1.242.

the point L_2 with $C'_{\rm p} = (C_2 - 2W_{\rm min})/(1-e)$. Further decrease of $C'_{\rm p}$ results in the appearance of a "dumbbell"-type domain enclosing both main bodies. Such domains first appear for true-anomaly values near the apocenter of the orbit of the main bodies. With further decrease of $C'_{\rm p}$, these domains encompass increasingly large values of the true anomaly. The behavior of the motions with further decrease in $C'_{\rm p}$ depends strongly on the eccentricity of the orbits of the primaries.

If the eccentricity is small, the variations of the minimum-energy surfaces qualitatively reproduce the well-known variations of the zero-velocity surfaces in the circular problem. However, the minimum-energy surfaces for $v = \pi$ and $v = 0$ may be located in the domains of the same or different (adjacent) topological types. For example, when $C'_{\rm p}(1-e) < C_2 - 2W_{\rm min} < C'_{\rm p}(1+e)$, the minimum-energy surfaces have the form of pulsating surfaces, which are located in the domain of satellite-type motions and in a dumbbell-type domain for true anomaly values in the vicinity of the pericenter and apocenter, respectively. With further decrease in $C'_{\rm p}$, when $C_3 - 2W_{\rm min} < C'_{\rm p}(1-e) < C'_{\rm p}(1+e) < C_2 - 2W_{\rm min}$, all the pulsating surfaces are located inside the dumbbell domain. When $C'_{\rm p}(1-e) < C_3 - 2W_{\rm min} < C'_{\rm p}(1+e)$, the motions of the low-mass body become unbounded in the vicinity of the primaries moving near their apocenter, but the motions remain bounded near the pericenter. If $C'_{\rm p}(1-e) < C_1 - 2W_{\rm min} < C'_{\rm p}(1+e)$, the shape of the intersection of the pulsating surface with the xy plane varies from horseshoelike (near the pericenter) to tadpolelike (near the apocenter). On the other hand, if $C'_{\rm p}(1-e) < C_{4,5} - 2W_{\rm min} < C'_{\rm p}(1+e)$, the surfaces considered vary from a tadpole shape to "curtains" that do not cross the xy plane. At negative $\cos v$, the infinite "curtains" give way to some ovals, which with decreasing $C'_{\rm p}$ shrink toward the singular points $L_{6,7}$. When $C'_{\rm p}(1-e) < C_{6,7} - 2W_{\rm min} < C'_{\rm p}(1+e)$, the motions in the vicinity of the apocenter are completely unbounded. When $C'_{\rm p}(1-e) < C_{6,7} - 2W_{\rm min}$ and $C'_{\rm p}(1+e) > C_{6,7} - 2W_{\rm min}$, all the motions of the small body become completely unbounded for values of the true anomaly of the primaries in the interval $v \in [\pi/2, \pi]$. When $v \le \pi/2$, there remain inaccessible domains bounded by "curtains" similar to those in the circular problem. The pulsating minimum-energy surfaces at $v \le \pi/2$ persist at negative $C'_{\rm p}$, to $C'_{\rm p} \to -\infty$, when all the surfaces shrink to an infinitely distant two-sheet hyperboloid.

The behavior of the minimum-energy surfaces may differ from the pattern described above in the case of sufficiently large eccentricities. With variations in the true anomaly, the pulsating surfaces may encompass not only one libration point, as considered

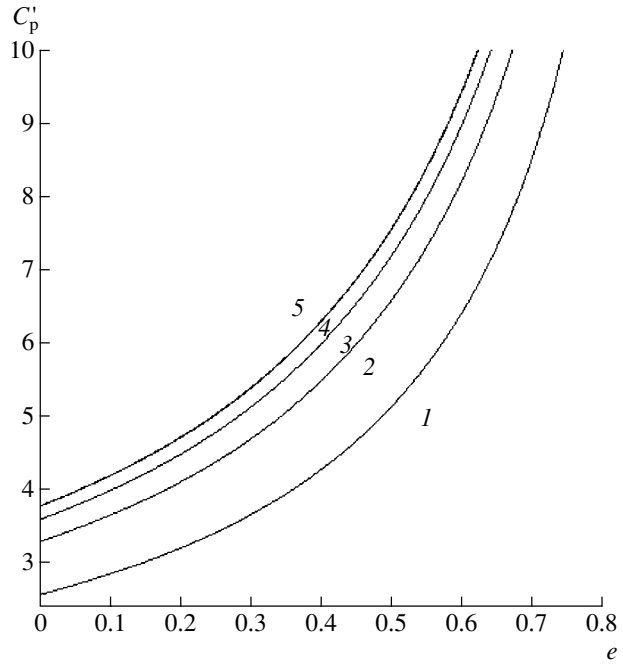

Fig. 6. Domains of satellite-type motions stable in Hill's sense for $\mu = 0$ (curve *1*), 0.1 (curve *2*), 0.2 (curve *3*), 0.3 (curve *4*), and 0.5 (curve *5*). The stability domain is located above the corresponding curve.

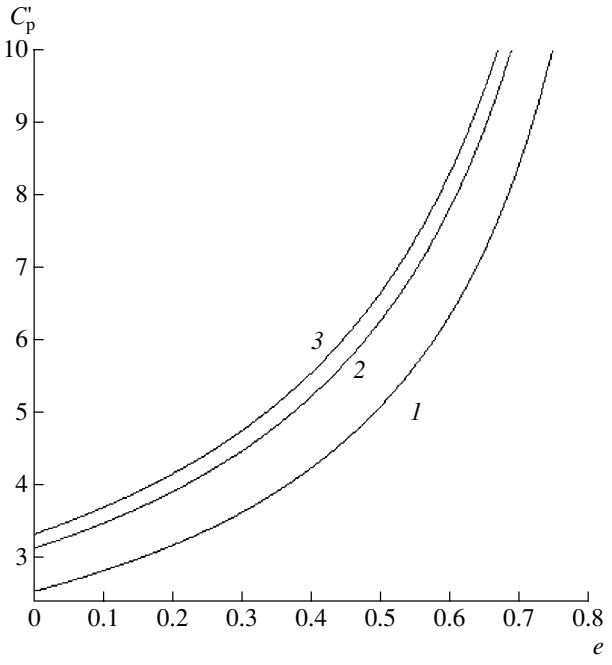

Fig. 7. Domains of the Hill stability for motions inside the "dumbbell" at $\mu = 0$ (curve *1*), 0.1 (curve *2*), and 0.5 (curve *3*).

above, but two or more. Our construction of the minimum-energy surfaces performed above shows that these surfaces resemble somewhat the zero-velocity surfaces in the restricted circular problem. At the same time, however, there are fundamental differences due to the pulsation of the minimum-energy surfaces, the existence of coplanar singular points, the existence of surfaces with arbitrarily large positive energies, etc.

8. THE HILL STABILITY

The concept of stability in Hill's sense remains applicable to the elliptical problem. The Hill stability criterion for satellite-type motions in the vicinity of one of the primaries can be derived from (30) in the form

$$C'_p > (C_2 - 2W_{min})/(1 - e). \qquad (31)$$

In this case, $C'_p(1 + e\cos v) > C_2 - 2W_{min}$ for any v. On the other hand, if $C'_p(1 + e) > C_2 - 2W_{min} > C'_p(1 - e)$, the satellite-type motions are unstable in Hill's sense, although motions that are stable in Hill's sense persist at v values in the vicinity of the pericenter. When $C'_p(1 + e) < C_2 - 2W_{min}$, satellite-type motions are unconditionally unstable.

Figure 6 shows the stability domains in the eC'_p plane for satellite-type motions for various μ. It is evident that the Hill stability depends rather weakly on the mass parameter μ, whereas the dependence on the eccentricity is very strong. The stability domains decrease in size with increasing e. As $e \to 1$, all satellite-type motions become unstable in Hill's sense. The Hill stability of the motions inside the dumbbell encompassing both primaries can be analyzed in a similar way. The Hill stability criterion of such motions is given by the inequality

$$C'_p > (C_3 - 2W_{min})/(1 - e). \qquad (32)$$

Motions become unbounded if this condition is not satisfied. Figure 7 shows the stability domains for this case. Hill's [1] analysis of the Sun–Earth–Moon restricted circular problem established that the motion of the Moon relative to the Earth is stable because the inequality $C'_p = 3.25440 > C_2 = 3.0009264$ is satisfied. The above results make it possible to verify whether the Hill stability condition is satisfied when allowance is made for the ellipticity of the Earth's orbit. To this end, we must check whether inequality (31) is satisfied, where $C_2 = 3.0009264$ is the Jacobi constant at the libration point L_2 (this value coincides with the analogous value for the circular problem). Taking the eccentricity of the Earth's orbit to be $e = 0.01675104$ and the ratio of the sum of the masses of the Earth and the Moon to the mass of the Sun to be 1 : 328906, we obtain $C'_p(1 - e) + 2W_{min} = 3.23284 > 3.0009264$. Hence the orbit of the Moon remains stable in Hill's sense

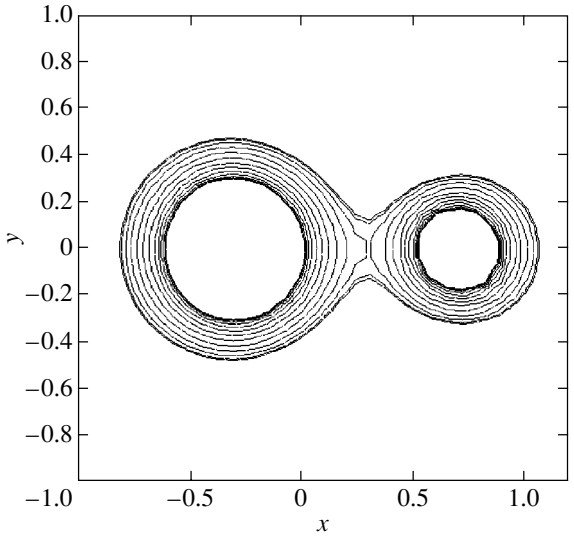

Fig. 8. Intersections of the pulsating minimum-energy surfaces by the xy plane for $\mu = 0.3$, $e = 0.2$, and $C_\mathrm{p}' = 3.9$.

in the framework of the elliptical problem. The motion of the Moon would be unstable in Hill's sense if the eccentricity of the Earth's orbit were greater by a factor of five.

9. THE TISSERAND CRITERION

We can apply the results obtained here to use the Tisserand criterion [16] to identify minor bodies in the Sun–Jupiter system with allowance for the ellipticity of Jupiter's orbit. Denoting the true anomaly of Jupiter, the function Ω, and the velocity V of the minor body during its first observation as v_1, Ω_1, and V_1, respectively, and the corresponding quantities at the time of the second observation as v_2, Ω_2, and V_2, we can write the inequality

$$\frac{V_1^2}{2} - \Omega_1 + \frac{W_\mathrm{min}}{1 + e\cos v_1} - \frac{V_2^2}{2} + \Omega_2 - \frac{W_\mathrm{min}}{1 + e\cos v_2} \begin{cases} > 0, \\ < 0, \end{cases} \tag{33}$$

where the upper and lower inequality signs are valid if $\cos v_1 > \cos v_2$ and $\cos v_1 < \cos v_2$, respectively.

We now introduce the standard notation a_1, p_1, i_1 and a_2, p_2, i_2 for the osculating elements of the orbit of the minor body during the first and second observations, respectively, to rewrite inequalities (33) in the form

$$\frac{1}{a_1} - 2\sqrt{p_1}\cos i_1 - \frac{2W_\mathrm{min}}{1 + e\cos v_1} - \frac{1}{a_2} + 2\sqrt{p_2}\cos i_2 + \frac{2W_\mathrm{min}}{1 + e\cos v_2} \begin{cases} > 0, \\ < 0. \end{cases} \tag{34}$$

Here, the inequality signs are to be chosen in the same way as in (33). It is obvious that the Tisserand criterion for the elliptical restricted three-body problem has the form of an inequality instead of an equality. The sign of the inequality depends on the distance of Jupiter from the coordinate origin at the times of the observations. Only if $\cos v_1 = \cos v_2$, i.e., if the Jupiter is at the same distance from the coordinate origin, do the inequalities (33) and (34) become equality. If we set the eccentricity of Jupiter's orbit equal to zero, $e = 0$, the inequalities (34) are transformed into the equality of the classical Tisserand criterion. In view of (18), it is safe to say that $1/a - 2\sqrt{p}\cos i - 2W_\mathrm{min}/(1 + e\cos v)$ is always located inside the interval bounded by the maximum and minimum values of the Jacobi energy and the known part of the additional energy. In this case, the numerical values of these energies remain unknown. Of course, one can perform numerical integration to evaluate

$$\int_{v_0}^{v} (\partial\Omega/\partial v)\, dv$$

and derive from the energy-conservation law the Tisserand criterion in the form of an equality linking arbitrary "old" and "new" observations. However, in this case, the use of the Tisserand criterion loses any meaning, since the problem of identifying the minor body is solved completely via numerical integration.

10. THE ROCHE MODEL

Roche [17] models are used extensively in analyses of the evolution of close binary systems. The equilibrium figures and sizes of the stars are represented by spherical and pear-shaped equipotential surfaces. Mass overflow is explained by considering dumbbell- and pear-shaped cavities. Mass flows through the throat of the dumbbell-shaped Roche lobe from the giant star to the other, possibly relativistic, star. The Roche lobe is constructed using the zero-velocity surface of the restricted circular three-body problem, i.e., assuming that the stars move in circular orbits. Strictly speaking, the model for circular orbits is not applicable to elliptical orbits, which is the situation for most stars (some stars move in orbits with eccentricities of 0.8 and higher); however, this model has been used due to the lack of a more suitable celestial-mechanics model.

Our results open the possibility of using a new model—that for the restricted elliptical three-body problem. Mass overflow during the expansion of the giant star begins when the energy level characterizing the dumbbell-shaped minimum-energy surface is reached. Unlike the circular-orbit model, the mass

overflow can be either intermittent or continuous (at small eccentricities), but with a periodically varying intensity. Observations of many close binaries exhibit well-defined pulsating behavior of the mass overflow. The minimum-energy surfaces in the elliptical problem provide a theoretical explanation for this phenomenon.

Figure 8 illustrates the typical pattern for the configuration and variation of minimum-energy surfaces in the case of pulsating mass overflow. This figure shows intersections of the dumbbell-shaped minimum-energy surfaces by the xy plane for true anomaly values from periastron to apoastron and back to periastron in steps of 15°. Our results can be used to determine the time of mass overflow. We find from the condition that the energy level be that corresponding to the libration point L_2, $C'_p(1+e\cos v) = C_2 - 2W_{min}$, that $\cos v = (S-1)/e$, where $S = (C_2 - 2W_{min})/C'_p$. The relative overflow time τ (the overflow time during one rotation of the stars about their axes divided by the orbital period) can then be determined from the formula

$$\tau = 1 - (E - e\sin E)/\pi, \qquad (35)$$

where $E = 2\arctan(\sqrt{(1-e)/(1+e)} \times \sqrt{(1-\cos v)/(1+\cos v)})$ is the eccentric anomaly of the stars. If the giant continues to expand despite the mass flow onto its companion, some mass may find its way through the funnel near the libration points L_3, L_1, and even L_4 and L_5. In this case, the relative overflow time can be calculated using the same formulas setting $S = (C_i - 2W_{min})/C'_p$ $(i = 3, 1, 4, 5)$.

Our results indicate that pulsating mass overflow occurs when the stars are located near apastron, since the Jacobi energy is greater at periastron than at the apastron. Sometimes, on the contrary, the Jacobi energy at apastron is considered to be higher than at periastron. The authors who use this hypothesis base it on the increase in the separation of the stars as they move toward apastron and the resulting increase of the potential part of the Jacobi energy. In addition, the Jacobi integral is assumed to preserve its "osculating" form in the elliptical problem. However, this reasoning is in contradiction with the energy-conservation law (10), since an increase of the energy at periastron results in a transition to a new energy level that exceeds the initial energy level. The energy-conservation law was derived in pulsating coordinates, in which the separation between the stars always remains constant and equal to unity in the adopted system of units. The variation of the energy of the small body due to the motion of the stars results exclusively from the variation of the independent variable v. The additional energy $u(v)$ increases and the Jacobi energy decreases due to the increase of the factor $\rho = 1/(1+e\cos v)$ in Ω [see (2)] as stars move from periastron to apastron. The total energy in this case is conserved in accordance with the energy-conservation law (10). Therefore, as the stars move toward apastron, the Jacobi energy must decrease, in accordance with (13).

11. CONCLUSIONS

The well-known integral invariant relation in the restricted elliptical three-body problem had neither theoretical nor practical applications. Our analysis shows that this relation can be applied to a variety of astronomical problems hitherto analyzed only approximately using the restricted circular problem or formations without any theoretical basis. It is also possible to use the models of the restricted parabolic and hyperbolic three-body problems, for which all the main results of this paper remain valid except for the periodicity of the function $u(v)$.

REFERENCES

1. G. W. Hill, Am. J. Math. **1**, 284 (1905).
2. L. G. Luk'yanov and G. I. Shirmin, Pis'ma Astron. Zh. **27**, 948 (2001) [Astron. Lett. **27**, 816 (2001)].
3. M. W. Ovenden and A. E. Roy, Mon. Not. R. Astron. Soc. **123**, 1 (1961).
4. Z. Kopal and R. A. Lyttleton, Icarus **1**, 455 (1963).
5. V. Szebehely, *Theory of Orbits* (Acad. Press, New York and London, 1967).
6. W. Scheibner, Crelle J. Reine Angew. Math. **65**, 291 (1866).
7. K. Petr and V. Nechvile, Casopis Pestovani Mat. Fys. (Praha) **47**, 268 (1918).
8. V. Nechvil, Compt. Rend. **182**, 310 (1926).
9. N. F. Rein, Tr. Gos. Astron. Inst. im. P.K. Shternberga **14** (1), 85 (1940).
10. A. Roy, *Orbital Motion* (Adam Hilger, Bristol, 1978; Mir, Moscow, 1981).
11. L. G. Luk'yanov, Pis'ma Astron. Zh. **4**, 326 (1978) [Sov. Astron. Lett. **4**, 177 (1978)].
12. J. L. Lagrange, *Mécanique Analytique* (Gauthier-Villars, Paris, 1888; Gostekhizdat, Moscow, 1950), Vol. 2.
13. L. G. Luk'yanov, in *Proceedings of the All-Russia Astronomical Conference VAK-2004, "Horizons of the Universe," Moscow, Russia, 2004*, p. 212.
14. G. N. Duboshin, *Celestial Mechanics. Analytical and Qualitative Methods* (Nauka, Moscow, 1964) [in Russian].
15. L. G. Luk'yanov, Astron. Zh. **65**, 422 (1988) [Sov. Astron. **32**, 215 (1988)].
16. F. Tisserand, *Traite de Mechanique Celeste* (Gauthier-Villars, Paris, 1896).
17. E. Roche, Mem. Acad. Sci. Montpellier **1**, 243 (1849).

Translated by A. Dambis

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 1028–1032. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1148–1152.

The Russian Virtual Observatory Project

O. B. Dluzhnevskaya and O. Yu. Malkov

Institute of Astronomy, ul. Pyatnitskaya 48, Moscow, 119017 Russia

Received May 10, 2005; in final form, May 18, 2005

Abstract—We describe the Russian Virtual Observatory (RVO), a prestigious international project sponsored by the Russian Academy of Sciences (RAS). In 2001, the RAS Scientific Council on Astronomy included this project in a list of the most important international projects of the RAS. Its main goal to create and develop the RVO, intended to provide Russian astronomers with direct and effective access to worldwide astronomical data resources. The RVO is one component of the International Virtual Observatory (IVO), a system in which vast astronomical archives and databases around the world, together with analysis tools and computational services, are linked together into an integrated facility. The IVO unites all important national and international projects to create virtual observatories, coordinated by the International Virtual Observatory Alliance. The RVO is one of the organizers and an important participant of the IVO Alliance. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Currently, astronomy faces an avalanche of newly acquired data due to important achievements in observing, detecting, and computing technologies. These datasets cover the sky in a variety of wavebands: from gamma-rays and X-rays, ultraviolet, optical and infrared, to radio. In a few years, it will be much easier to "dial-up" a digitized part of the sky than wait for access to a telescope; even now, the accumulated observations are sufficient to solve most current problems in astrophysics. With the advent of inexpensive data-storage technologies and the availability of high-speed networks, the concept of multiterabyte online databases that are seamlessly interoperational is no longer far-fetched. More and more catalogs are becoming interlinked, query engines are becoming more and more sophisticated, and research results from online databases are becoming as rich as those from new data acquired with telescopes. New survey telescopes now being planned will image the entire sky every few days, yielding data volumes measured in petabytes. These technological developments will fundamentally change the character of astronomical research, inevitably having dramatic effects on the sociology of astronomy itself.

The problems described above became most acute for the world astronomical community, and were realized towards the end of the previous century. In 2000, the General Assembly of the International Astronomical Union (IAU) resolved to create the International Virtual Observatory (IVO), and a special working group was formed within IAU Commission 5, "Documentation and Astronomical Data," aimed at creating virtual observatories, facilitating their activity, and linking them into the IVO.

2. THE INTERNATIONAL VIRTUAL OBSERVATORY

Over the past several years, the concept of virtual observatories aimed at satisfying the current requirements of astronomers for the management, analysis, and distribution of data have become widely popular. A virtual observatory is a system in which vast astronomical archives and databases around the world are linked together into an integrated facility, together with analysis tools and computational services. Modern astronomy stands at the edge of a new frontier for discovery, enabled by modern information technology, as well as international political and technical collaboration.

In 2001–2002, three major international projects aimed at developing and realizing modern approaches to astronomical-data use and creating virtual observatories received funding and entered their implementation phase. These are the Astrogrid (the United Kingdom), Astrophysical Virtual Observatory (Europe), and National Virtual Observatory (USA) projects. The total investment of funds in these projects is about 20 million US dollars over the first three to five years. The scope of this effort is not limited to national boundaries, and extends over the range of space and ground facilities utilized by the international astronomical community. Each project seeks to provide astronomers and experts and researches in related fields with powerful tools for the fully modern evaluation of observational data. The

1063-7729/05/4912-1028$26.00

number of virtual-observatory projects continues to increase as more and more national astronomical communities begin to face the problems described above.

For the IVO to become a reality, it is necessary to define its mission and the steps needed to realize it as a coordinated international effort. Each of the existing national projects has scientific and technological problems of its own. While this diversity can be beneficial to the success of the IVO, there are also common elements to be agreed upon if the IVO is to be designed effectively. Most of these common elements have to do with standards for data and interfaces. Other common (or shared) elements are software packages and source-code libraries. Other examples of areas in which international cooperation is desirable are security problems, funding, other political issues. The problems of interoperability standards for the data sources used should be solved in sufficiently early stages on the road toward the IVO. Demonstrations of new virtual-observatory capabilities with various levels of complexity are essential to gain credibility in the broader astronomical community. Possibilities for the open exchange of information and experience between different virtual-observatory projects should be created. It is therefore necessary to create a publicly visible roadmap for the IVO. Some of the milestones on this map have already been achieved, while others will be met in the very near future. Thus, the virtual-observatory community has expressed the need for a mechanism for the facilitation of international cooperation in order to achieve common goals.

The International Virtual Observatory Alliance provides just such a mechanism. The idea to join efforts from the national projects was discussed at the XXIV General Assembly of the IAU (Manchester, the United Kingdom, August 2000), and the first international dialog on interoperability problems was held at a conference in Strasbourg (France, January 2002). The IVO Alliance was formed in June 2002, at a virtual-observatory conference in Garching, Germany. This alliance consists of representatives from all funded virtual-observatory projects, who meet on a regular basis and hold discussions in order to coordinate solutions of their common problems and to reach consensus when developing common bases, without which the IVO could not function. Currently (October 2005), 16 projects participate in the IVO Alliance.

The main functions of the virtual observatories currently being developed throughout the world are:

- creating a registry of existing and planned key national resources and databases; creating an environment linking archives of observations;
- providing a connecting link between the IVO Alliance and the national astronomical community;
- contributing to the development of standard data-access and data-analysis interfaces;
- producing astronomical data;
- creating virtual-observatory prototypes that demonstrate search for, evaluation of, and access to distributed sources of astronomical data of various kinds.

3. THE RUSSIAN VIRTUAL OBSERVATORY

In December 2001, the Russian Academy of Sciences (RAS) Scientific Council on Astronomy (based at the Department of Physical Sciences of the RAS) approved an initiative aimed at creating the Russian Virtual Observatory (RVO), whose ultimate objective is to provide Russian astronomers with access to vast data and metadata sources created in the course of various observational projects.

Efforts aimed at creating the RVO represent a very important part of the scientific research plan of Section 13 of the RAS Scientific Council on Astronomy, "Databases and Information Services."

The RVO is a national member of the IVO Alliance, currently uniting 16 projects aimed at creating virtual observatories, and a representative of the RVO sits on the Alliance's Executive Committee.

Creating the RVO is very topical, first and foremost because the disintegration of the Soviet Union deprived the Russian astronomical community of almost all instrumental facilities and observatories that had been in the southern part of the USSR. An optimal way out of the situation is the integration of the Russian astronomical community into the worldwide astronomical-data information grid, most importantly, to western archives of observations. As its top achievement, creating the RVO will provide Russian scientists with online access to all significant worldwide astronomical-data resources.

Virtual observatories are an effective democratization instrument in astronomy: small scientific and educational astronomical institutions can obtain essentially the same scientific research possibilities based on modern observing material as can the world's leading astronomical institutions.

The contribution of the RVO to the worldwide astronomical resources can also be quite significant. About 30 astronomical institutes and structures operate in Russia, many possessing large data archives and cooperating with the RAS Institute of Astronomy (INASAN) in order to create the RVO. Russia has

the largest extent in longitude of any single country in the world (11 time zones), and many Russian observatories are located on the opposite side of the globe from most other large observatories. Thus, one advantage of Russian astronomical databases is the possibility to compile, in combination with data from other countries' observatories, long series of observations.

The main objectives of the RVO are in agreement with the general functions of national virtual observatories:

- to provide the Russian astronomical community with a convenient and effective mechanism for accessing foreign data sources;
- to unite Russian astronomical information resources as an important component for integration in the IVO;
- to provide access to observing facilities in the absence of required data in the united archive;
- to develop Russian electronic educational facilities in astronomy.

To provide the Russian astronomical community with a convenient and effective means of access to foreign data sources, the INASAN Center of Astronomical Data (CAD) mirrors and maintains several well-known foreign databases (ADS, VizieR, INES). It is planned to mirror other popular astronomical information resources in addition to the CAD, among them the most important stellar database, SIMBAD.

It is also planned to mirror several popular Russian astronomical information resources at the CAD, when access to them at their original sites encounters difficulties of any kind.

The CAD also provides access to offline astronomical resources: the center's archive possesses a continuously updated collection of CD-ROMs, published mainly by leading western observatories and containing astronomical catalogs, databases, image archives, and other resources.

The CAD developed the IRinA information system (**I**nternet **R**esources **IN A**stronomy), a collection of more than 200 references to worldwide astronomical internet resources, which is the RVO's information hub. The system uses a multilevel scheme of content classification. We plan to update and expand this information system in future. We continue the traditional services of the CAD (providing users with copies of catalogs, creating and scientifically analyzing samples from catalogs and databases, providing users with software and algorithms for search, visualization, cross-identification, and classification of objects from catalogs, surveys, and databases, as well as expert examination of astronomical data).

The problem of integrating and providing the foreign community a comprehensive description of the format and contents of Russian astronomical information resources based on IVO standards is also being addressed in the CAD information system. In particular, we completed the creation of a structured list of astronomical internet resources in Russia (as well as several other countries of the former USSR), some resources provided offline, and facilities currently being developed. The list, which contains more than 50 original resources, is highly valued by our foreign colleagues, and will be continuously updated.

One of the CAD's obligations to the international community is processing and verifying tables from papers published in Russian journals on astronomy and star catalogs prepared in Russian astronomical institutions (including those prepared by CAD staff members) and sending them to the Strasbourg Astronomical Data Center. The CAD compiles English-language computer-readable versions of published Russian catalogs and consults authors on catalog compilation.

We compiled a complete bilingual list of Russian astronomical institutions, with their addresses (including electronic addresses) and information on the telescopes and other observing facilities with which they are equipped. The list also includes several astronomical institutions in other countries of the former USSR. We are going to update this list in order to facilitate access to information on these institutions for local as well as foreign users (in particular, to prepare a version of the list in the format of a map of Russia).

Over the next few years, we expect to be able to digitize photographic plates accumulated in astronomical observatories of Russia (and of some countries of the Commonwealth of Independent States) in the framework of the RVO project. Currently, this project is being implemented jointly by the INASAN and the Sternberg Astronomical Institute. The CAD (INASAN) is the project's initiator and coordinator. Similar projects are underway in observatories of Bulgaria, Hungary, and Germany. The resulting scans enter a database developed at the Astronomical Institute of the Bulgarian Academy of Sciences by Dr. M. Tsvetkov. Work on the preservation of information from plate archives is carried out under the auspices of the IAU.

The results of activity on the RVO project are reflected in the regularly updated RVO web site, http://www.inasan.rssi.ru/rus/rvo.

When this program is completed, Russian astronomers will be able to visit the RVO site and obtain

online all information on objects of interest currently available in the world.

4. SCIENTIFIC RESEARCH USING VIRTUAL OBSERVATORIES

Making it possible to quickly compile and study high-volume datasets, virtual observatories considerably simplify the solution of a wide range of scientific problems (often providing the only means for such solutions). Virtual observatories enable astronomers to put into practice studies that would require, under different conditions, expenditure of time and resources too high for them to be possible. Of most value is the possibility of combining large datasets over the entire electromagnetic spectrum, over wide spatial and temporal ranges. The result minimizes the selection effects that necessarily influence any particular observation or survey, making it possible to reveal new information contained in the data but impossible to discover when studying a single dataset on its own. Virtual-observatory-oriented research includes studies of the large-scale structure of the Universe; Galactic structure, the population of active galactic nuclei in the Universe; variability of various astronomical objects over wide ranges of time scales, wavelengths, and radiation fluxes; and other, previously poorly studied, regions of the space of observational parameters. The virtual observatory also makes it possible to search for rare, unusual, or completely new types of astrophysical objects and phenomena. For the first time, astronomers are able to compare the results of extensive numerical modeling to equally extensive and complex observational datasets. The variety of virtual-observatory studies extends from large, key projects to support of information services and preparation of starting data samples for new studies, focused at problems of interest for ground- as well as space-based observatories. Thus, in addition to addressing specific scientific problems, virtual observatories are expected to conduct statistical astronomical studies with very large datasets and to determine planning strategies for future surveys and space missions.

Currently, a number of virtual-observatory projects use scientific prototypes, or demonstrations, to show the user community the advantages of sharing archives, catalogs, and computing facilities. For example, the demonstration of the Astrophysical Virtual Observatory (Europe) provided astronomers with a powerful interface for exploring the GOODS (Great Observatories Origins Deep Survey) deep sky survey, analyzing multiwavelength images and obtaining spectral energy distributions for the GOODS catalog objects, and measuring uncataloged objects dynamically. The selected scientific scenarios were galaxy formation in the early Universe and the identification of supernova candidates. The National Virtual Observatory (USA) demonstrated several possibilities at once: a transient event follow-up service, a brown-dwarf candidate search, and a galaxy morphology analysis. These demonstrations utilized new standard interfaces and protocols for accessing catalogs and images, and the National Virtual Observatory galaxy-morphology demonstration employed grid-based computing for carrying out parallel computations. The brown-dwarf search was primarily intended to validate previous results obtained using various methods, but yielded discoveries of new objects of this type in the course of the demonstration (the study covered 1/200 of the sky). The Astrogrid and Canadian Virtual Observatory also demonstrated their possibilities (deep field surveys, high redshift quasars, searches for faint galaxies, etc.).

One of the most recent successes of the European Astrophysical Virtual Observatory is the discovery of 30 type-II quasars that are very red X-ray sources, by combining data from the VLT, HST, and Chandra telescopes. One of the latest scientific applications developed by the National Virtual Observatory (USA) is an engine for comparing the results of globular-cluster evolution modeling and observations.

Similarly, an important part of the RVO project is developing other scientific applications, including determination of fundamental stellar parameters from data on binary components, preparation of a three-dimensional map of the Galaxy's interstellar extinction from photometric surveys, etc.

Many member projects of the IVO Alliance have active scientific teams including representatives of various astronomical communities (optical astronomy, radio astronomy, high-energy astrophysics, space- and ground-based astronomy). The common activity focus of the scientific teams is to form a clear understanding of the scientific requirements for a working virtual observatory. These requirements include a combination of new technologies and algorithms with new virtual-observatory standards of data presentation. One of the main results of the joint work of the scientific teams is a better understanding of the IVO's role and position in modern astronomy. Within several years, the member projects of the IVO Alliance will create a new astronomical infrastructure controlled by the community of researchers and incorporating continuously emerging new technologies. As soon as this fundamental infrastructure comes into operation, the international astronomical community will be able to compose new research programs and publish the data and results they obtain on the basis of much more advanced and scientifically desirable methods than those in use today.

5. CONCLUSION

The Russian Virtual Observatory is a prestigious international project sponsored by the RAS. In 2001, the RAS Scientific Council on Astronomy included the RVO in a list of the most important international RAS projects.

The creation and development of the RVO has the main objective to provide Russian astronomers with direct and effective access to the worldwide astronomical data resources. Most astronomical institutions of Russia will participate in creating the RVO.

The RVO is a component of the IVO, a system in which the vast astronomical archives and databases around the world, together with analysis tools and computational services, are linked together into an integrated facility. The IVO is one of the most important projects developed by the IAU Commission 5, "Documentation and Astronomical Data." The IVO combines all the significant national and international projects aimed at creating virtual observatories, coordinated by the IVO Alliance. The RVO is one of the founders and an important participant of the Alliance.

The realization of the project will open new wide scientific research prospects for Russian astronomers, and the virtual observatory will make the solutions of many scientific problems much simpler, and sometimes provide the only possibility for obtaining these solutions.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 96-07-89184 and 02-07-90-391).

Translated by N. Samus'

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 947–957. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1059–1070.

A High-Accuracy Method for the Removal of Point Sources from Maps of the Cosmic Microwave Background

A. T. Bajkova

Main (Pulkovo) Astronomical Observatory, St. Petersburg, Russia

Received December 28, 2004; in final form, May 18, 2005

Abstract—A new method for removing point radio sources and other non-Gaussian noise is proposed as a means of improving the accuracy of estimates of the angular power spectrum of the cosmic microwave background (CMB). The main idea of the method is to reconstruct fluctuations of the CMB in places contaminated by such emission, while traditional methods simply exclude these regions from consideration, leading to the appearance of "holes" in the resulting maps. The fundamental possibility of reconstructing the CMB signal in such holes follows from the analytical properties of a function with a finite spatial spectrum (the Silk damping frequency). A two-dimensional median filter is used to localize the point radio sources. Results of simulations of the method for maps of modest size are presented. The efficiency of applying the method to reconstruct the CMB from data with limited resolution and contaminated by appreciable pixel noise is investigated. The fundamental possibility of applying the method to reconstruct the CMB distribution in the region of the Galaxy is also demonstrated. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In connection with numerous experiments on measuring the anisotropy of the cosmic microwave background (CMB) with the aim of deriving high-accuracy estimates of the main cosmological parameters of the Universe [1], it is currently of interest to find new approaches to data reduction that could increase the accuracy with which various background contaminating signals of Galactic and extragalactic origin can be removed from radio observations of the CMB. This problem has been widely discussed in the literature [2–4].

We will restrict our consideration to the removal of pointlike radio sources and other non-Gaussian noise, which most strongly distorts the angular power spectrum of the CMB at high spatial harmonics (multipoles). This problem has been considered by a number of authors [5–9]. The currently available methods differ primarily in the means used to localize the contaminating signal, with the aim of subsequently removing the corresponding sections from the measured maps.

The following three strategies for the removal of pointlike sources are all natural: (1) reconstructing the point sources, then subtracting them from the measured CMB maps; (2) excluding contaminated pixels from the maps; and (3) reconstructing the true values of the CMB in contaminated pixels.

Experience shows that the first strategy leads to larger errors in the angular power spectrum of the CMB than the other two methods, since it is not possible to completely accurately reconstruct the brightness distribution of the point sources when the spatial spectra of the individual distinguished signals overlap [10]. For this reason, maps obtained with this method are characterized by the presence of appreciable non-Gaussian residuals. The second approach, which requires only the determination of the locations of contaminated pixels and their elimination before estimating the CMB power spectrum, is substantially more effective. Precisely this method is currently the most widely used.

However, this second method for the removal of contaminating point sources is likewise not free of drawbacks. First, the total effective area of the map that is used to estimate the angular power spectrum is reduced, leading to appreciable errors for more distant sections of the CMB. Second, the simple elimination of contaminated pixels does not fully restore the Gaussian statistics of the CMB, since it leaves "holes" in the maps that are used. In addition, the sharp boundaries of the cutout sections lead to so-called Gibbs phenomena in the behavior of the spectrum, which are especially strongly manifest at high spatial harmonics, and can appreciably complicate studies of the secondary anisotropy of the CMB.

Note as well that increasing the resolving power of a system used to measure the CMB anisotropy will lead to the detection of a larger number of weak sources [11], so that an increasingly large number of pixels in the CMB map are contaminated

1063-7729/05/4912-0947$26.00

by these signals. A special problem is also presented by the strong contamination due to Galactic emission. Simple removal of the zone of Galactic emission, which comprises an appreciable fraction of the overall area of the celestial sky, unavoidably leads to appreciable errors when estimating the angular power spectrum compared to the use of completely uncontaminated data.

Obviously, if the third strategy is realizable in practice, which depends primarily on the noise characteristics of the apparatus, it will be free of the drawbacks of the first two methods. Therefore, we propose and develop here a new method for the removal of high-multipole non-Gaussian noise, including both the localization of such noise and reconstruction of the true values of the CMB anisotropy in the contaminated regions. For the specific conditions considered in this paper, this could lead to an increase in the accuracy with which the angular spectrum of the CMB can be constructed, compared to the simple elimination of contaminated map pixels.

Thus, our aim is to present and develop a new method for eliminating the contamination of a CMB signal by point sources and other sources of high-multipole noise. We estimate the internal accuracy of the method and its stability to the input noise, and also investigate its application to data obtained with finite resolution for various levels of the instrumental pixel noise, thereby evaluating possibilities for applying the method to real systems.

The following sections of the paper present a statistical description of the CMB anisotropy; describe our model for the observed map, the method used to localize the point sources, and the method used to reconstruct the CMB values in noise-contaminated regions; and present the results of numerical simulations carried out for maps with modest angular dimensions but having the same statistical properties as the CMB distribution over the entire celestial sphere.

2. MODEL FOR AN OBSERVED MAP OF THE CMB

The distribution of the CMB temperature over the celestial sphere can be presented by the following expansion in spherical harmonics $Y_l^m(\theta,\phi)$ [12]:

$$\frac{\Delta T}{T} = \sum_{l,m} a_{lm} Y_l^m(\theta,\phi),$$

where T and ΔT are the mean temperature and temperature fluctuations of the CMB and the a_{lm} are the coefficients of the expansion.

The angular power spectrum of the fluctuations C_l is determined as the mean square of the coefficients a_{lm}:

$$C_l = \langle |a_{lm}|^2 \rangle,$$

where l is the multipole number.

If the fluctuations in the early Universe satisfy Gaussian statistics, as is expected in most cosmological theories, each coefficient a_{lm} should be statistically independent. Thus, the power spectrum C_l provides a full statistical description of the CMB anisotropy, which is a fundamental characteristic of the Universe that can be obtained directly from observations via a spherical-harmonic analysis.

For Gaussian fields, the expansion coefficients a_{lm}, $l \neq 0$, likewise represent Gaussian fields with random phases, zero means, and the dispersions

$$\langle a_{lm} a_{l'm'} \rangle = \delta_{ll'} \delta_{mm'} C_l, \quad l \neq 0.$$

Here, we consider sections of the sky that are modest in size. In this case, it is expedient to operate with the Gaussian fields in a flat, two-dimensional space, and to replace the spherical-harmonic analysis with a Fourier analysis, which appreciably simplifies the testing of proposed methods. The CMB fluctuations $\Delta T = T - \langle T \rangle$ can then be generated via the simple calculation of the Fourier series [12]

$$\frac{\Delta T(\theta_x,\theta_y)}{T} = \sum_{n_u=0}^{N_u-1} \sum_{n_v=0}^{N_v-1} D(n_u,n_v) \times \exp\left[i \frac{2\pi}{L} (n_u \theta_x + n_v \theta_y) \right], \quad (1)$$

where L is the linear size of the region considered in radians, (θ_x,θ_y) are Cartesian coordinates on the sky (in the spatial region), and (n_u,n_v) is the number of the Fourier component $D(n_u,n_v)$ in the region with spatial frequencies u and v.

The amplitudes of the Fourier components $D(n_u,n_v)$ obey a Gaussian distribution with zero mean and the dispersion

$$\langle |D(n_u,n_v)|^2 \rangle = C_l, \quad l = \frac{2\pi}{L}\sqrt{n_u^2 + n_v^2}, \quad (2)$$

while the phases are uniformly distributed in the interval $(0, 2\pi)$. Here, C_l is the angular power spectrum of the CMB temperature when it is expanded in spherical harmonics.

Relation (2) describes the circular symmetry of the power spectrum, i.e., its independence of the azimuthal number m, which we will use below as additional *a priori* information when reconstructing the CMB fluctuations in noise-contaminated regions.

Another, more important, characteristic of the power spectrum is its finite spatial extent, which follows from the existence of the so-called Silk damping frequency (l_D) [13], above which the power-spectrum fluctuations fall off sharply and the contribution of the CMB to the total observed signal becomes negligibly small with increasing frequency. The finiteness of the spectrum enables us to apply the theory of analytical functions [14] to describe the CMB fluctuations, which implies the possibility of reconstructing functions over an entire region based on knowledge for part of the region or at some set of points [15]. This principle lies at the basis of the proposed algorithm for reconstructing the CMB fluctuations in regions contaminated by noise.

When modeling the background of point sources (PS), we assume that they are randomly distributed over the sky in accordance with a Poisson law. The instrumental pixel noise is "white" Gaussian noise with zero mean. The measured CMB map can be represented by the model

$$\mathrm{CMB}_{\mathrm{meas}} = (\mathrm{CMB} + \mathrm{PS}) * \mathrm{BEAM} + \mathrm{N}, \qquad (3)$$

where N represents noise, BEAM the antenna beam, and $*$ a linear convolution.

The required data processing consists of solving an equation of the form of the convolution (3) for the CMB signal. The sequence of operations we have used during the reconstruction of the CMB is the following.

(1) The noise is filtered using a Wiener filter or a modification of such a filter that does not distort the form of the angular power spectrum [16]; this yields an estimate of the signal (CMB + PS) $*$ BEAM.

(2) This quantity is deconvolved from the antenna beam using a regularized inversion filter [17, 18], in order to derive an estimate of the signal CMB + PS.

(3) Positions contaminated by point sources are localized using a median filter (see Section 4).

(4) The CMB signal is reconstructed at the contaminated positions using the method proposed in the following section.

3. RECONSTRUCTION OF THE CMB FLUCTUATIONS AT CONTAMINATED POSITIONS

The proposed method for reconstructing the fluctuations in holes in a CMB map is a modification of the algorithm of Fienup [19], intended for the reconstruction of an image of an object with a finite carrier from the amplitude of its Fourier spectra (the phase problem). In our modified version, we determine the limitation on the spectral region using information about the finiteness and circular symmetry (2) of the amplitude spectrum, rather than the amplitude spectrum itself, and determine the limitation on the spatial region using the known map values.

The algorithm is iterative, and consists of the following sequence of operations.

(1) An initial approximation for the map is made. We recommend the use of the initial map with zero brightness in specified locations (holes) as a first approximation.

(2) The Fourier transform of the initial approximation is calculated, bringing about a transformation to the spatial-frequency domain.

(3) A condition for the spatial limitation of the fluctuation spectra, which comes about because the values of the Fourier components derived in the first step with numbers $l > l_D$ have been set equal to zero, is imposed. An additional constraint on the region of spatial frequencies that appreciably speeds up the convergence of the algorithm is the circular symmetry of the power spectrum of the fluctuations. To satisfy this last condition, the form of the Fourier components is modified so that the power spectrum has a form that is consistent with relation (2). For each value of l, the squared amplitudes of the spectra measurements are averaged in radius over a length l, after which these measurements are replaced by values with the derived squared amplitudes, retaining the phases.

(4) The inverse Fourier transform of the spectrum obtained in the previous step is taken, bringing about a transformation to the domain of the CMB map.

(5) The constraints on the spatial region of the map are imposed. The brightness values outside the holes in the map are replaced by the known values, while the values inside holes are not changed.

(6) The Fourier transform of the map obtained in step 5 is taken.

(7) Return to step 3 until the image obtained in step 5 ceases to change in accordance with a chosen convergence criterion.

Simulations show (see Section 6) that this modified Fienup algorithm leads rather rapidly to the correct solution. For this to be the case, the number of unknown values of the CMB fluctuations should be approximately half the number of knowns, if the discretization frequency of the map is twice its upper spatial frequency. If the discretization frequency is increased, a larger number of map values can be reconstructed, since the insufficient information in the spatial domain can be compensated by information in the spatial-frequency domain.

Obviously, a high computational speed for the method can be reached via the application of fast-Fourier-transform algorithms.

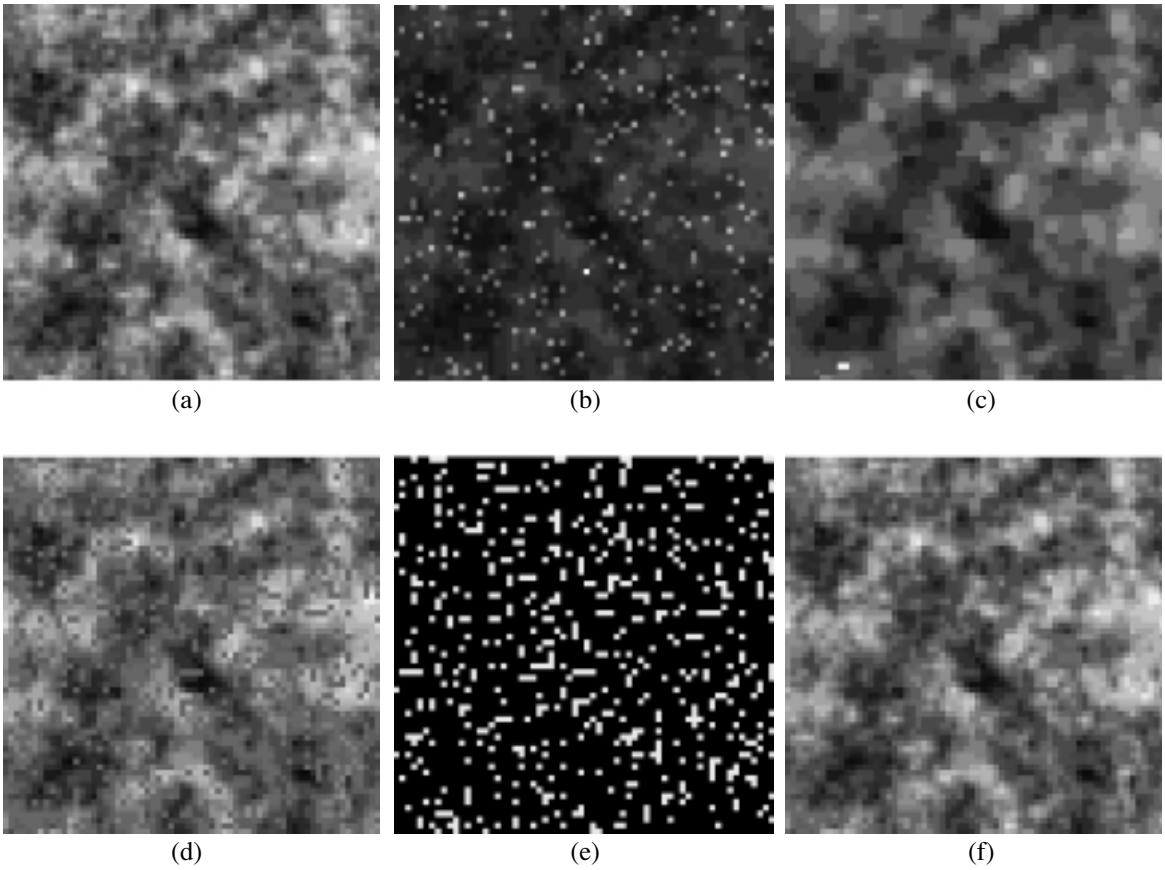

Fig. 1. Maps for experiment 1.

4. METHOD FOR LOCATING CONTAMINATED SECTIONS

The proposed method for localizing the point sources is based on applying a two-dimensional median filtration, which is a row-by-row column algorithm consisting of one-dimensional n-point median filtrations [20, 21]. The number n is taken to be odd. If $n = 2k + 1$, one-dimensional median filtration consists of ascribing to the current value for the sequence the mean of the series that is obtained when the $(2k + 1)$-point sequence is placed in increasing order, with the first k values located to the left and the last k values to the right of the given value. As a result of applying this operation to a one-dimensional series, all impulsive noise is eliminated. In our case, such noise corresponds to point sources.

To obtain the series of removed sources itself, we subtract the output map from the input map. Obviously, since the nonlinear transform is applied to the useful signal together with the noise, the resulting difference map obtained from the subtraction contains additional noise, whose magnitude is appreciably lower than the desired impulsive noise when n is chosen correctly. Sections of the resulting difference map that are contaminated by point sources can be identified by applying a cutoff to the levels of this map. The lower the cutoff level, the more pixels are subject to distortion. In our case, the cutoff level was usually chosen to be fairly low (10% of the peak value of the CMB fluctuations), in order to avoid missing genuine places that are contaminated by weak radio sources. On the other hand, the number of identified points must not be too large to ensure convergence of the subsequent reconstruction of the CMB-fluctuation values at these points.

In contrast to numerous methods that have been proposed in the literature (wavelet analysis [7, 8], clean and maximum-entropy deconvolution [10], optimal linear filtration [9], etc.), the proposed method for localizing the point sources is distinguished by its simplicity and the speed with which it operates, which is important in the reduction of large datasets obtained over the entire celestial sphere.

Obviously, this method will operate more reliably the brighter the point sources and the fewer the number of point sources in the map being analyzed. We have investigated the method for the cases of both

Table 1. Dispersion σ (in units of 10^{-5}) of the maps in experiment 1

CMB	PS	CMB + PS	CMB with holes	Residual map	CMB_{recon}	Residual map	CMB_{recon} (SNR ≈ 10)	Residual map
4.10	4.75	6.23	3.97	0.99	4.10	0	4.11	0.31

relatively bright (resolved) and relatively weak (unresolved) radio sources, whose number grows with the resolution of the instrument used [11].

5. FILTRATION OF THE INSTRUMENTAL NOISE AND DECONVOLUTION

We used a modification of a Wiener filter that preserves the form of the angular power spectrum [16], called a power filter, for the filtration of the instrumental white noise. A power filter can be applied in two ways. In the first, the dispersion of the instrumental (pixel) white noise is taken to be known *a priori*, whereas in the second, it is not. We present a preliminary estimate of this noise based on applying a median filter to the input signal. The dispersion of the instrumental noise was determined by subtracting the output signal of the median filter from the measured map, CMB_{meas}. We present here results obtained using the second approach.

Reconstruction of the map taking into account the antenna beam was carried out using a regularized inversion filter containing a regularizing parameter that depends on the level of the residuals after filtration of the noise [17, 18].

6. SIMULATION RESULTS

In this section, we present the results of applying the proposed method to model CMB measurements. The anisotropy of the CMB was simulated using a function for the angular power spectrum that corresponds to a standard ΛCDM cosmological model for the Universe with $\Omega_b h^2 = 0.02$, $\Omega_\Lambda = 0.65$, $\Omega_m = 0.3$, $h = 0.65$, and $n = 1$. We numerically generated a map of the CMB $7.5^\circ \times 7.5^\circ$ in size in accordance with formulas (1) and (2); this map, therefore, possesses all the properties of the CMB fluctuations distributed over the entire celestial sphere. The upper frequency of the angular power spectrum of the CMB corresponds to the multipole $l = 1536$.

6.1. Effect on the CMB Reconstruction in Point-Source Contaminated Locations in Systems with High Angular Resolution

The goal of this experiment (experiment 1) is to illustrate the proposed method and to estimate the effect of applying the method in systems with high angular resolution that detect a large number of relatively weak radio sources. We tested the method assuming that the preliminary noise filtration and deconvolution were already carried out (see the end of Section 2). This was done with the aim of estimating the internal accuracy of specifically the method for reconstructing the CMB in holes in the initial CMB map.

We took the map to be 64×64 pixels in size, which corresponds to discretization at the Nyquist frequency for the specified maximum multipole of the CMB spectrum, $l_{max} = 1536$. The results of this numerical experiment are shown in Fig. 1, which presents (a) the initial (model) map of the CMB fluctuations; (b) the map of the CMB contaminated by radio sources (CMB + PS; 234 weak sources with amplitudes exceeding the CMB-fluctuation peak by a factor of three, distributed randomly over the map according to a Poisson distribution); (c) the map obtained by processing the map in (b) with a row–column, three-point median filter; (d) the map obtained by replacing the CMB values at the detected contaminated points with zero brightness; (e) identified sections of contamination (which, in general, differ from the intrinsic distribution of point sources, see Section 3) obtained by applying an amplitude cutoff of 10% of the peak CMB fluctuation to the difference of maps (b) and (c); and (f) the map of

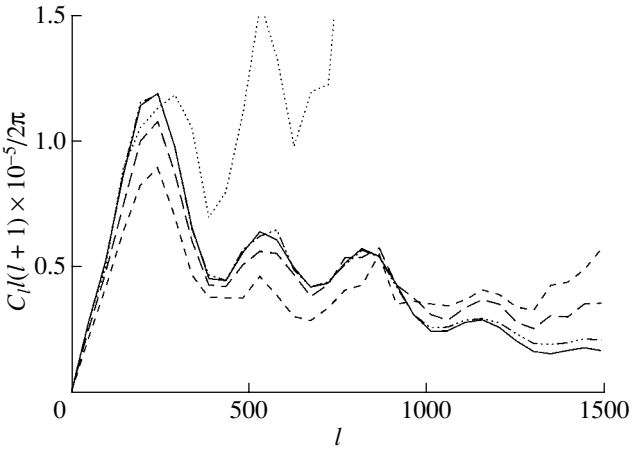

Fig. 2. Angular power spectra for the maps for experiment 1. See text for details.

Table 2. Maps and their dispersions for experiment group 2

Experiment	Map	Dispersion σ, 10^{-5}	Dispersion of map with holes σ, 10^{-5}
2	CMB	4.10	3.90
	PS $*$ BEAM	1.99	0
	CMB $*$ BEAM	4.51	4.29
	(CMB + PS) $*$ BEAM	4.93	4.29
2.1	Pixel noise	0	–
	CMB_{meas}	4.93	–
	CMB_{recon}	4.09	3.89
	Residual map	0.39	1.28
2.2	Pixel noise	0.65	–
	CMB_{meas}	4.97	–
	CMB_{recon}	4.05	3.86
	Residual map	0.74	1.38
2.3	Pixel noise	2.61	–
	CMB_{meas}	5.58	–
	CMB_{recon}	4.27	4.10
	Residual map	1.19	1.59
2.4	Pixel noise	5.23	–
	CMB_{meas}	7.19	–
	CMB_{recon}	4.14	3.93
	Residual map	1.47	1.84

the CMB reconstructed using the modified Fienup algorithm, which nearly precisely coincides with the input model CMB map in (a).

Figure 2 presents the angular power spectra of the maps; the solid curve refers to the original and reconstructed CMB maps, the dotted curve to the contaminated map (b), the short-dashed curve to the map obtained via median filtration (c), and the long-dashed curve to the map with zero brightness in contaminated locations (d). The dispersions of the signals represent a quantitative characteristic of the experiment, and are presented in Table 1.

We can estimate the effect of applying our approach compared to the traditional approach that does not reconstruct the CMB fluctuations in holes by comparing the angular power spectra in Fig. 2 (as well as the dispersions of the corresponding residual maps). We can see that simple removal of 234 contaminated measurements from the CMB map leads to distortion of the true angular power spectrum, especially at high harmonics: this distortion is from 30% to 100% when $l > 1000$. Applying the proposed method to remove the point sources leads to virtually no such distortion. Simulation of the case of a background of unresolved point sources with amplitudes a factor of three smaller than the maximum CMB fluctuations likewise yielded very high accuracy for the reconstruction in the absence of pixel noise.

To investigate the stability of the method with regard to the input instrumental noise, we added Gaussian white noise with a signal-to-noise ratio of SNR ≈ 10 to the map in panel (b), taking the signal level to be that of the CMB fluctuations. The dash–dotted curve in Fig. 2 depicts the angular power spectrum of the reconstructed map in this case. We can see that small variations in the input data gave rise to only small changes in the solution; i.e., the proposed algorithm displays a fairly high stability. Analysis of these results indicates that simple zeroing of the CMB values, even in precisely determined contaminated locations, yields appreciable errors in the reconstruction of the angular power spectrum, especially on small angular scales. If the contami-

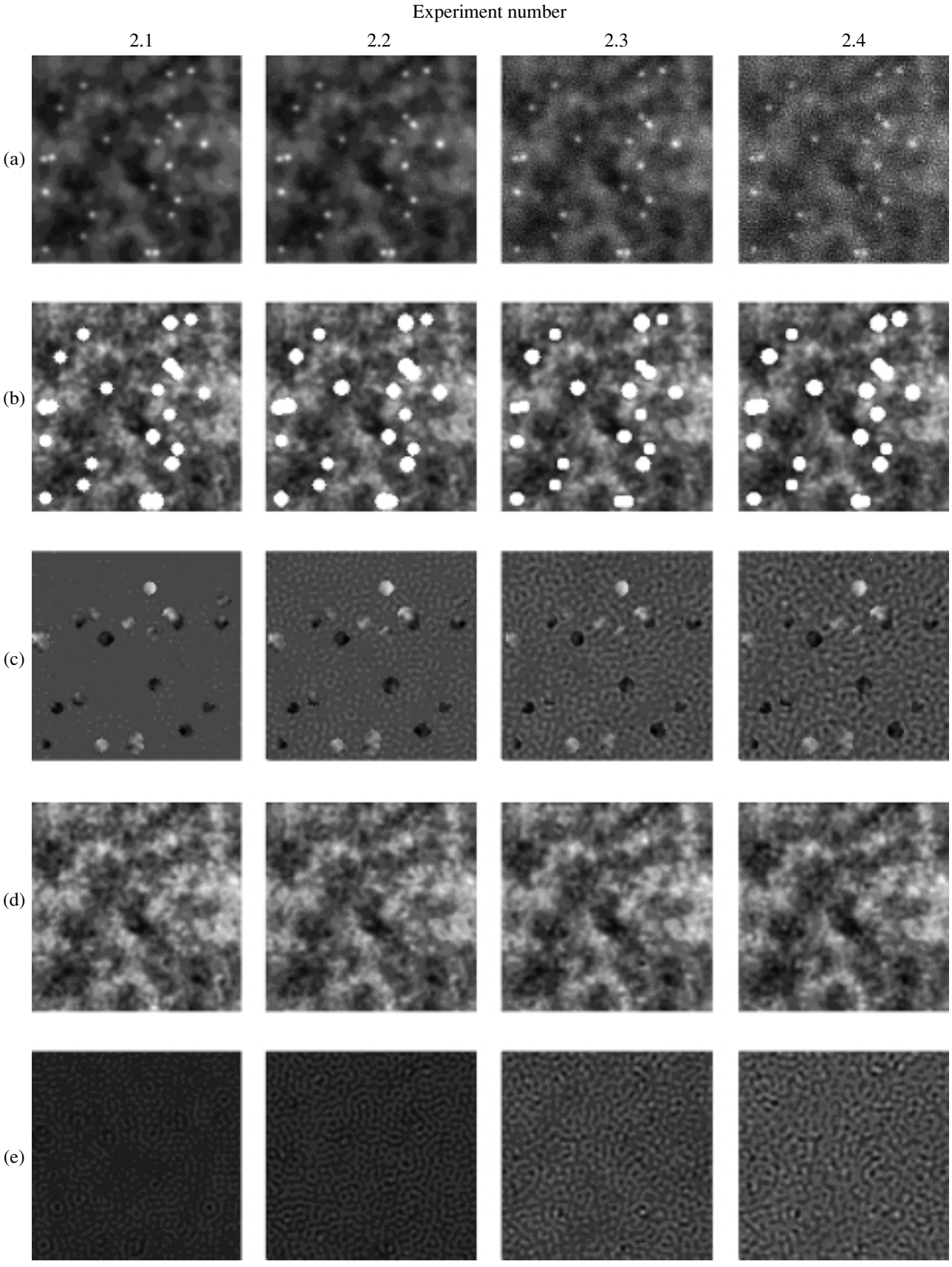

Fig. 3. Maps for experiment group 2.

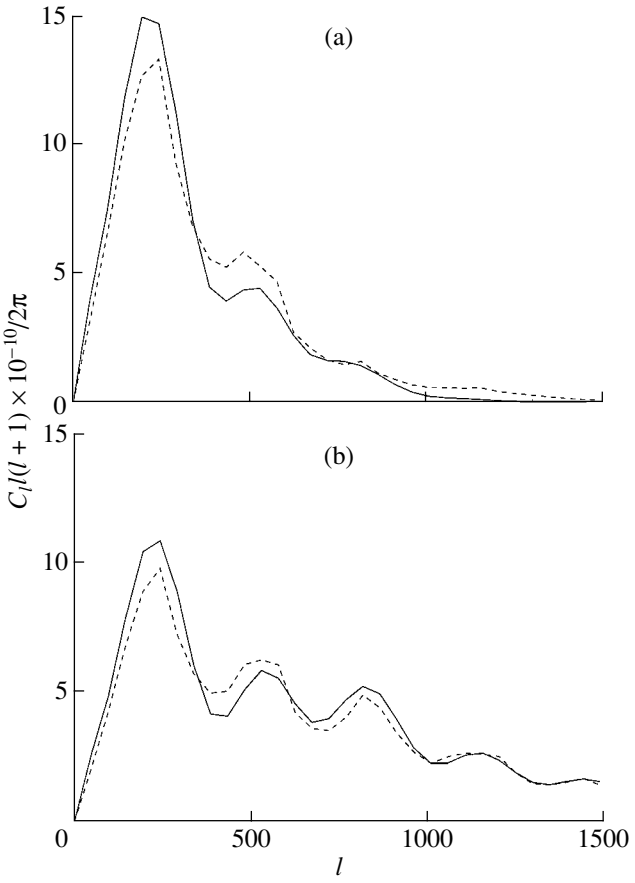

Fig. 4. Illustrating the effect of excluding sections contaminated by point sources. See text for details.

nated sections occupy a large area (Fig. 1e), the error in the desired angular power spectrum can be still larger, possibly leading to substantial errors in derived estimates of cosmological parameters.

Thus, in this experiment, without allowance for the real resolution of a system and the input instrumental noise, it was possible to reach very high accuracy in the CMB-map reconstruction. As we indicated above, this testifies to a high internal accuracy of the proposed method for reconstructing the CMB in holes in a map. Note that convergence of the algorithm to the required solution can also be achieved without placing constraints on the circular symmetry of the power-spectrum fluctuations. However, use of this information as an additional constraint appreciably speeds up the process of convergence. Experience shows that, when the map is discretized with a frequency that is twice the bandwidth of the spectrum signal, high accuracy in the reconstruction is achieved only if the number of unknowns is no more than one third of the total number of pixels in the map. As is shown above, the method displays a fairly high degree of stability to the noise in the input data; i.e., small variations in these data correspond to only small variations in the reconstruction. When the input noise level is high, a preliminary filtration should be applied using a power filter (see Section 5), as was done in the experiments discussed in the following section.

6.2. Effect on the CMB Reconstruction in Contaminated Locations in the Model $CMB_{meas} = (CMB + PS) * BEAM + N$

Let us make the problem more complex and consider a model for the observed CMB map that corresponds to expression (3), which includes the effect of the limited resolution of the system and the input pixel noise. In the following set of experiments, the map size was 128×128 pixels, the pixel size was $3.51'$, and the full width at half maximum of the antenna beam $10.53'$. For an antenna with a diameter of 1 m, this resolution corresponds to an observing frequency of 98 GHz. These system parameters are close to those for one of the channels of the PLANCK mission [11].

We carried out four experiments based on these parameters (experiment group 2), with various levels of white pixel noise, beginning with zero noise. The dispersions of the input, noise, reconstructed, and residual maps can be used as a measure of the accuracy of the reconstruction of the CMB anisotropy, and are given in Table 2. The results are shown in Fig. 3, where the numbers (columns) indicate the experiment number and the letters (rows) indicate the physical meaning of the map: (a) measured CMB map satisfying (3); (b) reconstructed CMB map with holes in contaminated pixels, including the effect of the antenna beam; (c) residual map, equal to the difference of the given CMB map and the reconstructed map shown in row "b"; (d) reconstructed CMB map obtained using our method to interpolate functions in the holes; (e) residual map, equal to the difference of the given CMB map and the reconstructed map shown in row "d." Obviously, each column of maps presented in the figure corresponds to a particular value for the pixel-noise level.

The results presented in Fig. 3 were obtained using pixel-noise filtration based on applying a median filter to derive a preliminary estimate of the noise dispersion (see Section 5). The effect of the antenna beam in the reconstructed map was included using a regularized inversion filter.

Analysis of the results presented in Fig. 3 shows that the main advantage of the proposed method for removing point sources is that the residual maps shown in row "e" are free from non-Gaussian features, and resemble residual Gaussian noise, whose amplitude depends on the level of the input Gaussian

noise (see the dispersion in Table 2). As we can see in map row "c," the traditional method, which simply excludes contaminated pixels from the CMB map, does not fully eliminate the contribution of the point sources, which are manifest as small non-Gaussian features determined by the magnitude of the CMB fluctuations in the contaminated locations. In addition, the sharp edges of the cutout sections lead to high-frequency noise that extends far beyond the frequency edge, which corresponds in our case to $l_{\max} = 1536$.

We illustrate the effect of applying our method in pure form in Fig. 4, which shows (a) the angular power spectra of the convolved map CMB $*$ BEAM without (solid curve) and with holes, whose size is determined by the size of the antenna beam (dashed curve), and (b) the same spectra for the input CMB map. We can see from the curves that the presence of holes in the maps leads to errors of about 13% in the angular power spectrum of the CMB for multipoles $l = 200-300$ and errors of about 30% for multipoles $l = 450-550$. The relative errors are even higher for higher l values.

Note that, in spite of the fact that the radio sources considered are point sources (i.e., they are essentially δ functions), we have taken the contaminated locations to include a region surrounding the source coordinates, determined by the characteristics of the inverse filter, which does not provide an ideal reconstruction of a δ function. The presence of even a small amount of input noise requires regularization of the algorithm, which leads to a solution with finite resolution. To ensure reliable exclusion of all contaminated pixels, we have assumed that the region distorted by each point source occupies an area equal to the area of the base of the antenna beam, which exceeds somewhat the total area of the excluded CMB measurements. However, this is not a serious concern for our method, since we reconstruct the intrinsic CMB values in the resulting holes.

Obviously, the maximum effect from reconstructing the CMB signal in contaminated locations is achieved in the absence of pixel noise. The quantitative effect is lowered as the pixel-noise level is increased, although the qualitative effect of a complete elimination of non-Gaussian features from the CMB maps is maintained even in the presence of fairly high noise levels.

Figures 5a and 5b present the angular power spectra of the signals corresponding to experiments 2.1 and 2.4, respectively. The results obtained in experiments 2.2 and 2.3 occupy intermediate positions and are not shown. The solid curve shows the angular power spectrum of the initial CMB map, the dotted curve the measured CMB map, which satisfies (3), the dashed curve the angular spectrum of the reconstructed CMB map with the pixels distorted by point sources zeroed, and the dash–dotted curve the CMB map with reconstructed values for the fluctuations at the distorted locations.

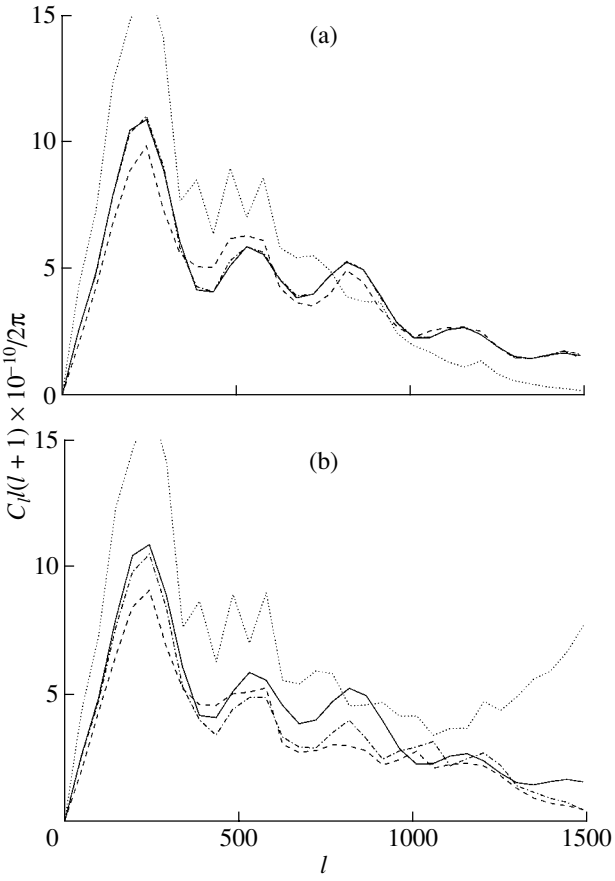

Fig. 5. Angular power spectrum for the maps in experiment group 2. See text for details.

We can see from Fig. 5a that it is possible to reconstruct the CMB map nearly perfectly in the absence of pixel noise. The error in the reconstruction is determined only by the error in the inversion filtration. The effect of applying our method in this case is close to that shown in Fig. 4.

Figure 5b shows that a high level of pixel noise leads to a substantial loss in the accuracy of the reconstructed CMB angular power spectrum, for both a simple exclusion of contaminated sections and the reconstruction of the CMB signal in holes. However, the accuracy is nonetheless higher in the latter case—quite appreciably for multipoles $l = 24-500$. The dispersion of the residual-noise map for the specified interval of the input pixel noise is 25–27% lower due to the full removal of point sources from the CMB maps (Table 2).

Although the simulations show that the effect of applying our method decreases with growth in the

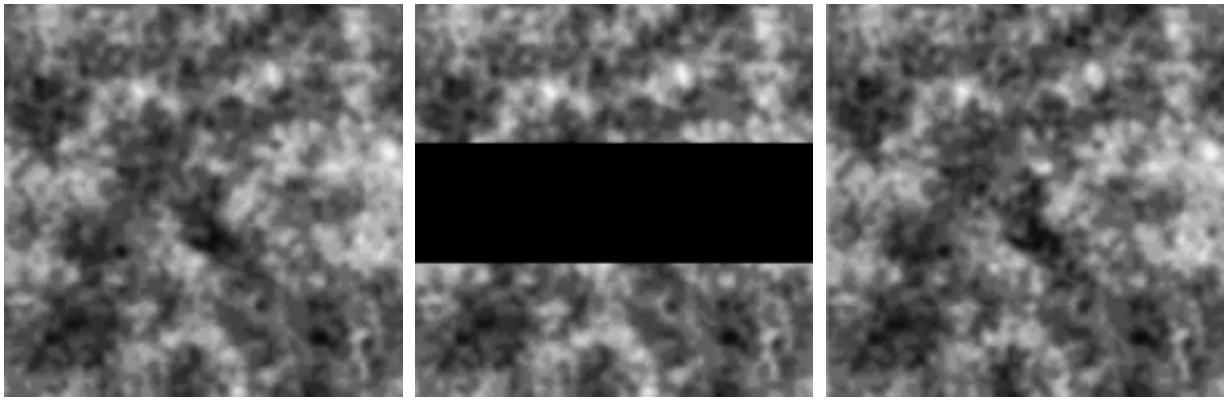

Fig. 6. Maps for experiment 3.

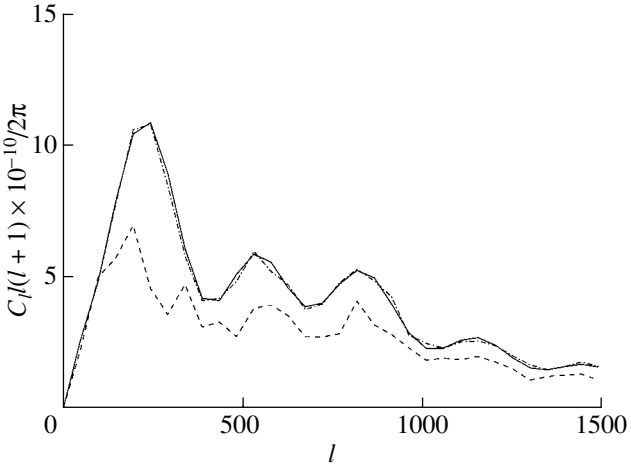

Fig. 7. Angular power spectra for the maps for experiment 3. See text for details.

pixel-noise level in a natural way, it nonetheless remains fairly high for a system with noise characteristics similar to those planned for the PLANCK mission [11].

The method for reconstructing the CMB signal proposed here can also be applied in connection with constructing catalogs of point sources. For this, it is sufficient to subtract the reconstructed CMB map from the estimated CMB + PS map.

6.3. Reconstruction of the CMB Anisotropy over a Wide Region of the Map

From the point of view of enhancing the accuracy of an estimated angular power spectrum based on CMB measurements over the entire celestial sphere, it is of interest to consider the reconstruction of the CMB signal in the zone of the Galaxy where powerful non-Gaussian noise is observed.

In the traditional strategy, this region of the sky is simply not taken into account, which clearly leads to a loss of accuracy in the derived CMB angular power spectrum compared to that which would be obtained using a full set of undistorted data. Let us consider the reconstruction of the CMB signal in the zone of the Galaxy using the same method as in the previous experiment. Let the noise occupy the middle part of the map in the form of a band that is elongated in the horizontal direction. We cut out the contaminated region from the map and reconstruct the absent components of the CMB in this region using our method (experiment 3).

Our simulations show that it is possible to obtain a nearly perfect reconstruction of the CMB, right up to the case when the width of the contaminated band comprises a third of the linear size of the map. Further increase in the width of the band leads to a growth in the errors of the reconstruction. Simulation results for band widths comprising about 30% of the linear size of the map are presented in Fig. 6, which shows the (a) initial CMB map, (b) CMB map with the Galactic band cut out, and (c) reconstructed CMB map. The angular power spectra are presented in Fig. 7, where the solid curve corresponds to the initial CMB map, the dashed curve to the map with the band cut out, and the dash–dotted curve to the reconstructed map.

Analysis of the maps presented in Fig. 7 shows that simple elimination of sections of the CMB occupying a substantial area leads to appreciable errors in the derived angular power spectrum, which can reach 50% (for example, for multipoles $l = 200–300$). Applying our method led to a virtually perfect reconstruction.

Thus, this experiment provides hope that reconstructing the CMB in the zone of the Galaxy based

on real measurements over the entire celestial sphere can substantially improve the accuracy with which the real angular power spectrum of the CMB can be derived.

7. CONCLUSION

We have investigated the fundamental possibility of improving the accuracy of estimates of the angular power spectrum of the CMB by reconstructing the CMB anisotropy in regions contaminated by point radio sources and other high-multipole noise. A series of examples have been used to demonstrate the effect of applying this approach, compared to the currently standard strategy of simply excluding contaminated sections of CMB maps.

We have shown that, in the absence of instrumental noise, the proposed method ensures the removal of point sources and other non-Gaussian multipole noise with very high accuracy, and is fairly stable to variations in the noise level in the input data.

The main advantage of the method is that it is able to fully remove non-Gaussian features from CMB maps, while the simple exclusion of contaminated locations does not achieve this goal. The full removal of residual point sources from the maps makes it possible to appreciably lower the dispersion of the residual noise, even in the presence of a relatively high level of input pixel noise. Our simulations show that the effect of applying our method decreases in a natural way as the input noise level grows, but nonetheless remains fairly high for noise characteristics corresponding to those planned for the PLANCK mission.

We have also shown the fundamental possibility of enhancing the accuracy of estimates of the angular power spectrum of the CMB by reconstructing the CMB anisotropy in the zone of the Galaxy where the strongest level of background contamination is observed.

ACKNOWLEDGMENTS

This work was partially supported by the Basic Research Program of the Presidium of the Russian Academy of Sciences "Nonstationary Phenomena in Astronomy."

REFERENCES

1. E. Gawiser and J. Silk, Phys. Rep. **333**, 245 (2000).
2. M. Tegmark and G. Efstathiou, Mon. Not. R. Astron. Soc. **281**, 1297 (1996).
3. M. Tegmark, A. de Oliveira-Costa, and A. Hamilton, Phys. Rev. D **68**, 123523 (2003).
4. V. Stolyarov, M. P. Hobson, M. A. J. Ashdown, *et al.*, Mon. Not. R. Astron. Soc. **336**, 97 (2002).
5. M. Tegmark and A. de Oliveira-Costa, Astrophys. J. **500**, L83 (1998).
6. M. P. Hobson, R. B. Barreiro, L. Toffolatti, *et al.*, Mon. Not. R. Astron. Soc. **306**, 232 (1999).
7. J. L. Sanz, R. B. Barreiro, L. Cayon, *et al.*, Astron. Astrophys. **140**, 99 (1999).
8. P. Vielva, E. Martinez-Gonzalez, L. Cayon, *et al.*, Mon. Not. R. Astron. Soc. **326**, 181 (2001).
9. R. Vio, L. Tenorio, and W. Wamsteker, Astron. Astrophys. **391**, 789 (2002).
10. A. T. Bajkova, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **45**, 909 (2002); Radiophys. Quantum Electron. **45**, 835 (2002).
11. M. Bersanelli *et al.*, *COBRAS/SAMBA, ESA Report D/SCI* (1996).
12. J. R. Bond and G. Efstathiou, Mon. Not. R. Astron. Soc. **281**, 655 (1987).
13. J. Silk, Nature **215**, 1155 (1967).
14. Ya. I. Khurgin and V. P. Yakovlev, *Finite Functions in Physics and Technology* (Nauka, Moscow, 1971) [in Russian].
15. G. I. Vasilenko and A. M. Taratorin, *Image Reconstruction* (Radio i Svyaz', Moscow, 1986) [in Russian].
16. K. M. Gorski, astro-ph/9701191 (1997).
17. A. N. Tikhonov and V. Ya. Arsenin, *Solutions of Ill-Posed Problems* (Halsted, New York, 1977; Nauka, Moscow, 1986).
18. A. V. Goncharskii, A. M. Cherepachshuk, and A. G. Yagola, *Numerical Method for Solving Inverse Problems in Astrophysics* (Nauka, Moscow, 1978) [in Russian].
19. J. R. Fienup, Opt. Lett. **3**, 27 (1978).
20. B. Yustusson, in *Two-Dimensional Digital Signal Processing II. Transforms and Median Filters,* Ed. by T. S. Huang (Springer-Verlag, New York, 1984; Radio i Svyaz', Moscow, 1984), p. 156.
21. S. G. Tyan, in *Two-Dimensional Digital Signal Processing II. Transforms and Median Filters*, Ed. by T. S. Huang (Springer-Verlag, New York, 1984; Radio i Svyaz', Moscow, 1984), p. 191.

Translated by D. Gabuzda

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 958–966. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1071–1080.

A Spectroscopic and Photometric Study of NGC 2685, the Prototype Polar-Ring Galaxy

V. A. Hagen-Thorn[1], L. V. Shalyapina[1], G. M. Karataeva[1], V. A. Yakovleva[1], A. V. Moiseev[2], and A. N. Burenkov[2]

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

[2] *Special Astrophysical Observatory, Russian Academy of Sciences, Nizhniĭ Arkhyz, Karachaĭ-Cherkessian Republic, 369167 Russia*

Received February 8, 2005; in final form, February 18, 2005

Abstract—We present and analyze spectroscopic and photometric observations for NGC 2685, the prototype polar-ring galaxy. The spectroscopic data were acquired using the 6 m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences with the UAGS spectrograph and a scanning Fabry–Perot interferometer, while the photometric analysis was based on images from the Hubble Space Telescope archive. We demonstrate that the subsystem usually called the "inner polar ring" is a highly inhomogeneous gas and dust disk rotating approximately in the polar plane relative to the galaxy's main body. When the self-absorption in the disk is taken into account, a comparison of its color indices with those from model computations of the color evolution of galaxies results in an age for the disk of about 1.4×10^9 years, much lower than the previously accepted estimate. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Many studies have been devoted to NGC 2685 (Fig. 1), the prototype polar-ring galaxy (PRG). Observations of this galaxy, called by Vorontsov-Velyaminov [1] in 1972 "the most peculiar of all galaxies," were commenced even before PRGs were identified as a special type of galaxies. Its morphology was discussed by Burbidge and Burbidge [2] and Sandage [3] in the late 1950s and early 1960s based on photographs obtained with the 5 m telescope.

The galaxy's main body is $100'' \times 25''$ in size, and is probably a disk viewed almost edge-on, which is surrounded with luminous matter reaching out to $\approx 50''$ and having a sharp outer boundary in the form of a ring in a plane approximately perpendicular to the main plane of the main body ("the polar ring"). All this structure is immersed in a faint outer envelope, whose shape and alignment follow the main body's isophotes but are a factor of 2.5–3 larger. The outer envelope and polar ring are clearly visible in the λ 21 cm line of neutral hydrogen [4] (see also [5]).

Several studies in the 1960s–1980s dealt with the kinematic characteristics of all three subsystems. Optical observations showed that the main body (stellar component) rotates about its minor axis at a rate typical of S0 galaxies [6], while the gas component rotates about the main body's major axis in a plane perpendicular to the rotational plane of the stellar component [6–8]. The presence of two mutually orthogonal kinematic systems is the definitive feature when attributing galaxies to the PRG class. The observations of Shane [4] at λ 21 cm confirmed the existence of two mutually orthogonal kinematic systems, one related to the polar ring (Shane was also

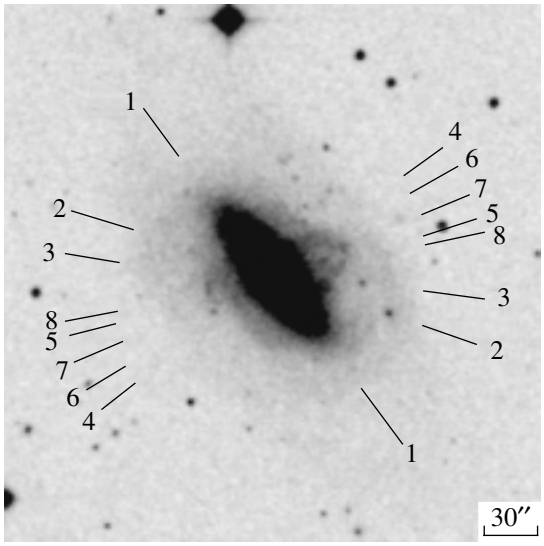

Fig. 1. NGC 2685 in the *R* band, from the DSS (N is upward and E to the left). The positions of the slit of the UAGS spectrograph are marked.

1063-7729/05/4912-0958$26.00

Table 1. Log of the spectroscopic observations

Instrument, date	PA	Exposure time, s	Field of view	Scale, ″ per pixel	Spectroscopic resolution, Å	Seeing	Spectral range, Å
UAGS	38° (1)	2 × 1200	2 × 140″	0.4	3.6	1.5″	6040–7280
Oct. 2–4, 1999	72° (2)	3 × 1200	2 × 140	0.4	3.6	1.5	6040–7280
	84° (3)	2 × 1800	2 × 140	0.4	3.6	1.5	6040–7280
	128° (4)	2 × 1800	2 × 140	0.4	3.6	1.5	6040–7280
	106° (5)	2 × 1500	2 × 140	0.4	3.6	1.5	6040–7280
	121° (6)	2 × 1800	2 × 140	0.4	3.6	1.5	6040–7280
	113° (7)	2 × 1800	2 × 140	0.4	3.6	1.5	6040–7280
	102° (8)	2 × 1800	2 × 140	0.4	3.6	1.5	6040–7280
IFP		32 × 180	5′ × 5′	0.71	2.5	1.8	Hα
Dec. 22, 2003							

the first to note that this was probably not a spiral coiling around the main body, as had been believed earlier, but instead a system of projected coplanar rings), and the other system associated with the faint outer envelope (the main body of NGC 2685 does not radiate line emission and, like most S0 galaxies, is almost free of gas).

Recent observations of the central (8″ × 12″) part of NGC 2685 with the multipupil spectrograph of the 6 m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences (SAO) [9] enabled the reconstruction of the velocity fields of the stellar and gaseous components, confirming rotation characteristic of an oblique disk of the stellar component about the galaxy's minor axis, and the presence of a gaseous disk in the galaxy's central region, which rotates in an orthogonal plane. These observations did not include the polar-ring region, where no detailed velocity field of the ionized gas is available. To address this shortcoming, we obtained spectroscopic observations of NGC 2685 with the long-slit UAGS spectrograph [10] and a scanning Fabry–Perot interferometer [11].

The formation and evolution of polar rings has been frequently discussed in the literature (see, for example, [12]). An important aspect of these questions is the age of the polar rings. The age of the polar ring in NGC 2685 cannot be considered to be firmly known. Until recently, age estimates were based on either surface colorimetry and comparison with models of galactic color evolution [13] or determinations of the chemical composition of the emitting gas in the HII regions [14]. Peletier and Christodoulou [13] and Eskridge and Pogge [14] both suggest that the ring is a long-lived feature, because its colors do not differ strongly from those of the main body, and the metallicity in the HII regions is nearly solar.

However, the recent paper of Karataeva *et al.* [15], which used archive images from the Hubble Space Telescope (HST) to resolve the polar ring into stars, revealed blue supergiants in the ring, testifying to active star formation not long ago. Analysis of the color–absolute magnitude diagram for the resolved stars did not confirm the presence of a high metallicity in the ring. Unfortunately, the brightness level of red giants could not be reached, due to insufficient exposure time, so that the problem of the existence of old stars in the ring remained unresolved.

Peletier and Christodoulou [13] themselves point out that the colometric data they used were obtained on a nonphotometric night and they are not quite in agreement with the earlier surface photometry of [16, 17]; therefore, we also attempted surface photometry of the galaxy using the same B, V, I_c images from the HST archive that had been used earlier to resolve the ring into stars.

NGC 2685 is one of the closest PRGs. Its Galactocentric velocity from our analysis (see below) is 957 km/s, corresponding for $H_0 = 65$ km s^{-1}Mpc^{-1} to a distance of 14.7 Mpc and a scale of 71 pc per arcsecond.

2. OBSERVATIONS AND REDUCTION

2.1. Spectroscopic Data

Our spectroscopic observations were acquired with the 6 m telescope of the SAO with the UAGS spectrograph and a scanning Fabry–Perot interferometer (IFP). Table 1 presents a log for these observations.

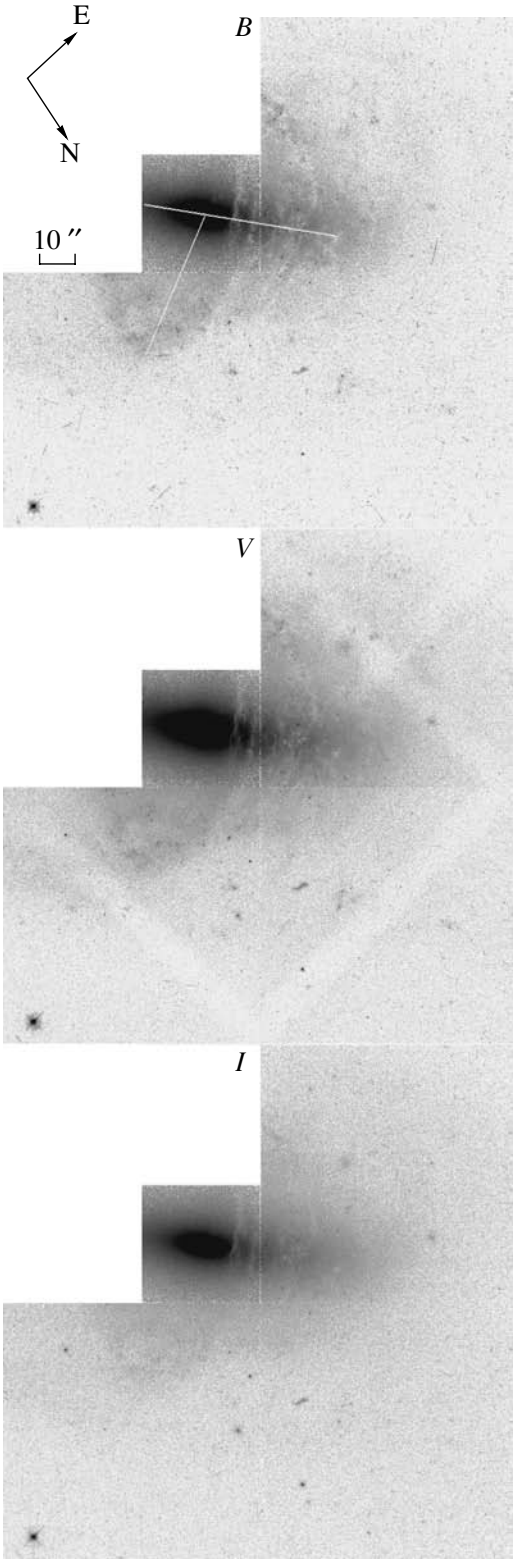

Fig. 2. Images of NGC 2685 in the B, V, and I_c color bands from the Hubble Space Telescope archive. The sections used to study the color variations and associated brightness distributions are indicated.

When using the UAGS, the $2'' \times 140''$ spectrograph slit was placed in eight position angles (Fig. 1) to provide the best coverage of the polar-ring region. The scale at the slit was $0.4''$ per pixel and the resolution 3.6 Å; the seeing was about $1.5''$. The light detector was a Photometrics CCD (1024×1024). The recorded spectral range, λ6040–7280 Å, included Hα and [NII] and [SII] lines, which were used to derive the radial-velocity curve (the [SII] lines were used to reduce the spectrum at position angle 38°, where the Hα and [NII] emission lines were distorted by the Hα absorption). The spectrogram for a slit position angle of 38° (along the major axis of the galaxy's main body) also displays absorption lines that can be used to plot the radial-velocity curve for the stellar component. The data reduction was done using the ESO/MIDAS software package. We applied a cross-correlation technique to derive the radial-velocity curve for the stellar component, using data at λ6090–6560 Å and a spectrum of the galaxy's nucleus as a template. The uncertainties of the radial velocities derived from emission lines were 15 km/s, while the uncertainties for those derived from absorption lines were about 10 km/s for the central region and up to 30 km/s for the peripheral regions.

Our observations with the scanning IFP behind the SCORPIO focal reducer were made near Hα. The preliminary conversion to monochromatic light was achieved using a filter with a central wavelength of 6589 Å and a full width at half maximum (FWHM) of 13 Å. The detector was an EEV 42-40 CCD. The readout was done in a 4×4-pixel binning mode, so that images 512×512-pixels in size were obtained in each individual channel. These data were used to derive the velocity field and brightness distribution for the Hα line and continuum. The reduction was performed using a software package developed at the SAO by Moiseev [18]. The radial-velocity uncertainties did not exceed 10 km/s and the spatial resolution after smoothing during the data reduction was $2.5''$.

2.2. Photometric Data

We used images from the HST archive for our photometric study. These were taken in January 1999 during program No. 6633 using the WFPC2 camera with the F450W (exposure time 2300 s), F555W (1000 s), and F814W (730 s) filters. The images obtained after the standard reduction and transformations to the BVI_c system are shown in Fig. 2. Unfortunately, the available data do not cover the entire galaxy and the exposure times were not sufficient to detect its faint outer envelope.

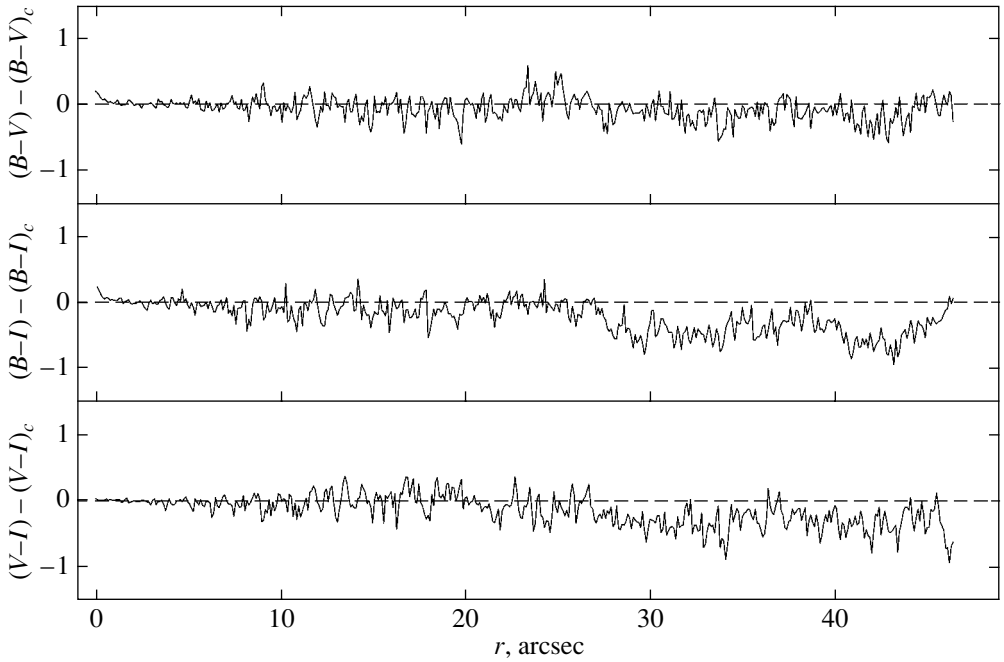

Fig. 3. Color variations along the section at position angle 293°.

3. OBSERVATIONAL RESULTS AND ANALYSIS

3.1. Photometric Data

Inspection of Fig. 2 reveals that, although the galaxy's structure is generally similar in all the color bands, there exist considerable differences between the images in different bands in the polar-ring regions, outside the galaxy's main body, as well as in the region where the ring is projected against the main body. Numerous details outside the main body stand out much better in the B band, and the general background seems somewhat attenuated compared to the V and I backgrounds. Inhomogeneities in the brightness distribution within the main body decrease progressively from B to I.

To determine the age of the rings, we are primarily interested in color variations between the galaxy's main body and the ring; thus, we present all subsequent data in differential form, relative to the brightness and colors in the central region of the galaxy. This also enables us to neglect possible uncertainties in standardization of the images.

Figure 3 shows the distributions of the $B-V$, $B-I$, and $V-I$ color indices along a section through the galaxy's center at position angle 293° (indicated in Fig. 2). To reduce the fluctuations, we first averaged the data in 10-pixel bins (1 pixel corresponds to 0.1″). The color indices remain nearly unchanged to distances of ≈20″. The major contribution to the light observed in this region comes from the main body (note that a similar result was obtained in [16]; see Table 2 there). Beyond 20″, the colors become bluer (by approximately 0.2^m in $B-V$, 0.3^m in $V-I$, and 0.5^m in $B-I$). Here, the contribution from the galaxy's main body is negligible, and it is obvious that the ring is considerably bluer than the main body.

The brightness distributions in the B, V, and I bands along the major axis of the galaxy's main body are shown in Fig. 4 (the distribution to the NE from the center is shown by the solid curve, and that to the SW by the circles). The profiles for both parts of the galaxy coincide virtually perfectly to a distance of $r = 5.5''$ (except for a small feature at $r = 1''$, which is probably real, since it is visible in all the three distributions). Beginning at $r = 5.5''$, the NE profile displays regions of lower brightness compared to the SW profile. Individual features are repeated in all three distributions, with the difference decreasing with increasing wavelength. The amounts of these dimmings for several features in the profile and the dimming ratios for different bands are collected in Table 2. The mean ratio of the amounts of dimming in B and V is 1.32, while this mean ratio for I and V is 0.80. The former of these values is in excellent agreement with the ratio expected if the dimming is due to absorption by dust; the latter is somewhat higher than the value given by the standard interstellar-extinction curve, but not so much as to force us to reject the obvious conclusion that the dimming is due to dust in the ring projected against the galaxy's main body.

The nonuniform dimming can also be followed along the NE profile beyond 18″, where no data are

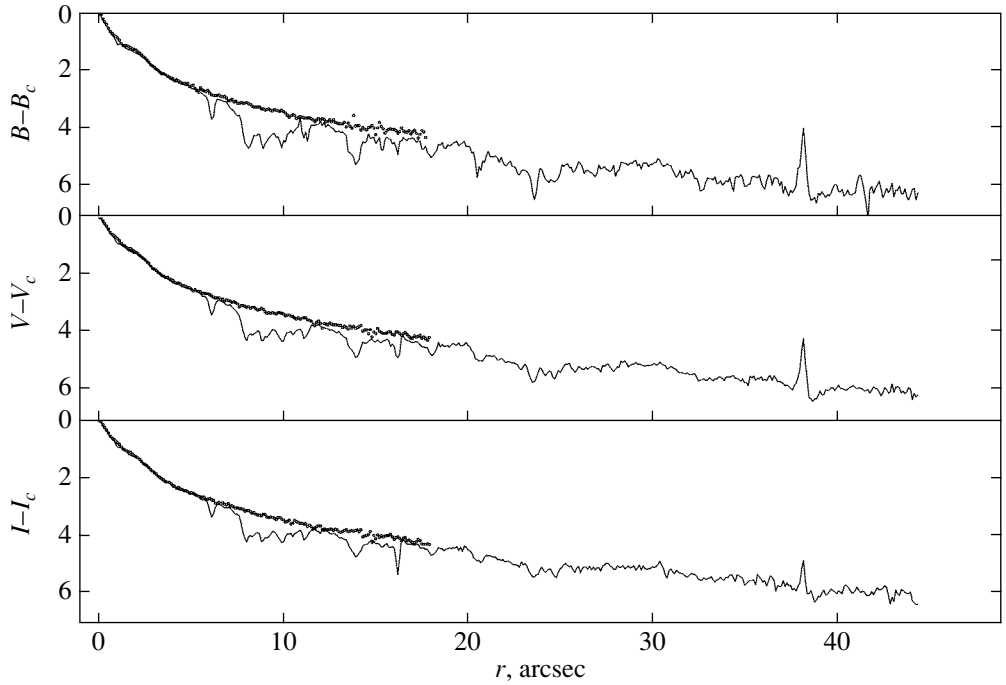

Fig. 4. Brightness distributions along the galaxy's major axis.

available for the SW profile. It seems that the dimming is due to a very inhomogeneous dusty disk projected onto the galaxy's main body, with the absorption A_V reaching 1.2^m in some regions. Adopting $\overline{A_V} = 0.7^m$ as the mean absorption, we find that the disk's mean optical depth (assuming it to be an "absorbing screen") is $\tau_V = 0.64$. Obviously, reddening in the disk must also be present outside the main body. If the dust and stars in the disk are mixed, then, according to Disney *et al.* [19], the mean absorption in the disk for this optical depth is $A_V = 0.32^m$. For the mean ratios of the absorptions in different bands we have derived, this leads to $A_B = 0.42^m$, $A_I = 0.26^m$, and the color excesses $E_{B-V} = 0.10^m$, $E_{B-I} = 0.16^m$, and $E_{V-I} = 0.06^m$. Thus, with the internal absorption in the disk taken into account, the color differences between the main body and the disk reach 0.30^m in $B-V$ and 0.66^m in $B-I$.

Table 2. Absorption in the disk

r	A_B	A_V	A_I	A_B/A_V	A_I/A_V
1″	0.25	0.17	0.15	1.47	0.88
6″	0.85	0.68	0.59	1.25	0.87
8″	1.55	1.20	1.10	1.29	0.92
9″	1.40	1.15	0.92	1.22	0.80
10″	1.30	0.97	0.80	1.34	0.82
14″	1.50	1.03	0.93	1.46	0.90
18″	0.63	0.52	0.28	1.21	0.54

The mean color index of the central region of NGC 2685, corrected for the reddening in our Galaxy, is $\overline{B-V} = 0.87^m$ [13, 16]. We can then adopt $\overline{B-V} = 0.57^m$ for the disk. The metallicity in the rings was shown to be 0.008 in [15]. For this metallicity, the age of the disk is 1.4×10^9 years, according to model computations of galactic color evolution [20]. It is claimed in [9] that the age of the stellar population in the galaxy's central parts is at least 10^{10} years.

3.2. Spectroscopic Data

The results of our observations with the Fabry–Perot interferometer are shown in Fig. 5. Figure 5a presents the Hα brightness distribution (gray scale) with superposed continuum isophotes. It is mainly the outer part of the disk (the ring) that emits in Hα, and the brightness distribution in this region is extremely nonuniform. We can also see individual HII regions belonging to the outer envelope. Figure 5b presents the derived velocity field, which is consistent with the picture of a gas disk rotating in the polar (with respect of the galaxy's main body) plane.

Figure 6 presents the radial-velocity curves based on our observations with the UAGS. The velocities found from the [NII] and Hα lines are plotted as triangles and pluses, respectively. The velocities measured with the IFP are also plotted (crosses). All the

data are in good mutual agreement. In the upper left panel, the filled squares show the radial-velocity curve derived from absorption lines, and the open squares the curve derived from [SII] lines. All the velocities are heliocentric. The mean value for the galaxy's center, based on all the spectra, is 873 ± 3 km/s, in excellent agreement with data in the literature (e.g., Shane [4] gives $v_{\rm hel} = 870$ km/s from measurements at $\lambda 21$ cm). After reduction to the Galactic center, we obtain $v_{\rm sys} = 957$ km/s, as is given in the Introduction.

An inspection of all the curves presented in Fig. 6 indicates that the rotation in the inner region (approximately to 10″) is consistent with that for a rigid body. The variations of the velocity gradient with position angle confirm this conclusion. Table 3 presents these variations according to our data, with the gradients derived from [6, 8] given for comparison. Our data agree well with those published earlier. In rigid-body rotation, the gradient is proportional to the cosine of the position angle. This is precisely the variation behavior that follows from Table 3. The best-fit sine curve representing the data for the gaseous component gives a largest gradient of 11.5 km s^{-1} arcsec^{-1} at position angle 126°. This angle is in good agreement with that given in [9], although the largest gradient we have found is much lower than the gradient in Table 2 of [9] (obviously, this table erroneously contains the velocities at distances given in Fig. 3 in the same paper instead of the velocity gradients; after correction, we obtain satisfactory agreement). The gradient we find for the gaseous component is close to that for the stellar component in the direction of the galaxy's major axis found by both ourselves and Schechter and Gunn [6].

An analysis of the velocity field using the "tilted ring" technique [21] yielded a model for the disk's circular rotation in good agreement with the observations for a position angle of the dynamical axis $\mathrm{PA_d} = 115°$ and an inclination $i_{\rm d} = 50°$. The residual velocities do not exceed 20 km/s. For the position angle and inclination of the main body of the galaxy (assumed to be disk-shaped), $\mathrm{PA_g} = 38°$ and $i_{\rm g} = 70°$ [17], the angle between the galaxy's main plane and the plane of the disk will be 87°, so that the disk is actually a polar disk.

Figure 7 presents the galaxy's rotation curve derived combining our data for the central region (triangles) with the neutral-hydrogen data for the peripheral regions [4] (squares).

The beginning of this curve cannot be represented with the standard "bulge + disk" model, but is satisfactorily represented with a "disk + disk" model. At the periphery, it is necessary to add a halo component. A three-component model with

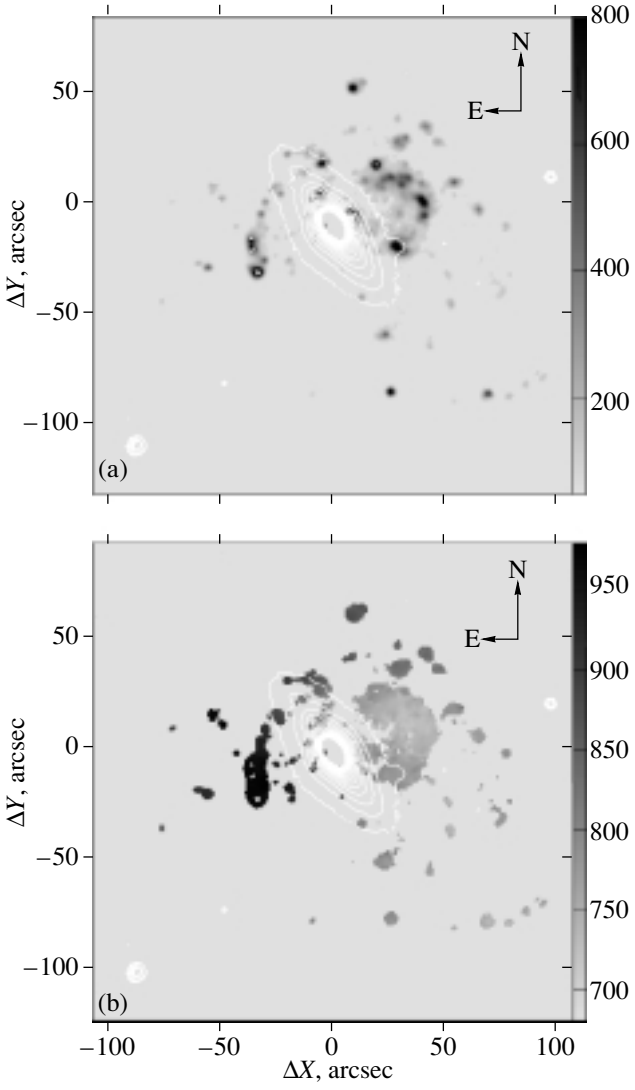

Fig. 5. Results from the Fabry–Perot interferometer: (a) the Hα and continuum (isophotes) brightness distributions; (b) the velocity field plotted over the continuum isophotes.

$\{\sigma_1 = 13 \times 10^3\, M_\odot\ \mathrm{pc}^{-2},\ h_1 = 206\ \mathrm{pc};\ \sigma_2 = 0.9 \times 10^3\, M_\odot\ \mathrm{pc}^{-2},\ h_2 = 2.7\ \mathrm{kpc};\ \rho = 0.008\, M_\odot\ \mathrm{pc}^{-3},\ r_c = 7.1\ \mathrm{kpc}\}$ successfully represents the rotation curve (the solid curve in Fig. 7).

4. DISCUSSION AND CONCLUDING REMARKS

Our results testify to the following structure for the inner region of NGC 2685, including the main body and the region usually called the inner polar ring (our data on the outer envelope are sparse, but it appears from Fig. 5b that the velocities of several HII regions belonging to the envelope are consistent with the suggestion of Shane [4] that this envelope is

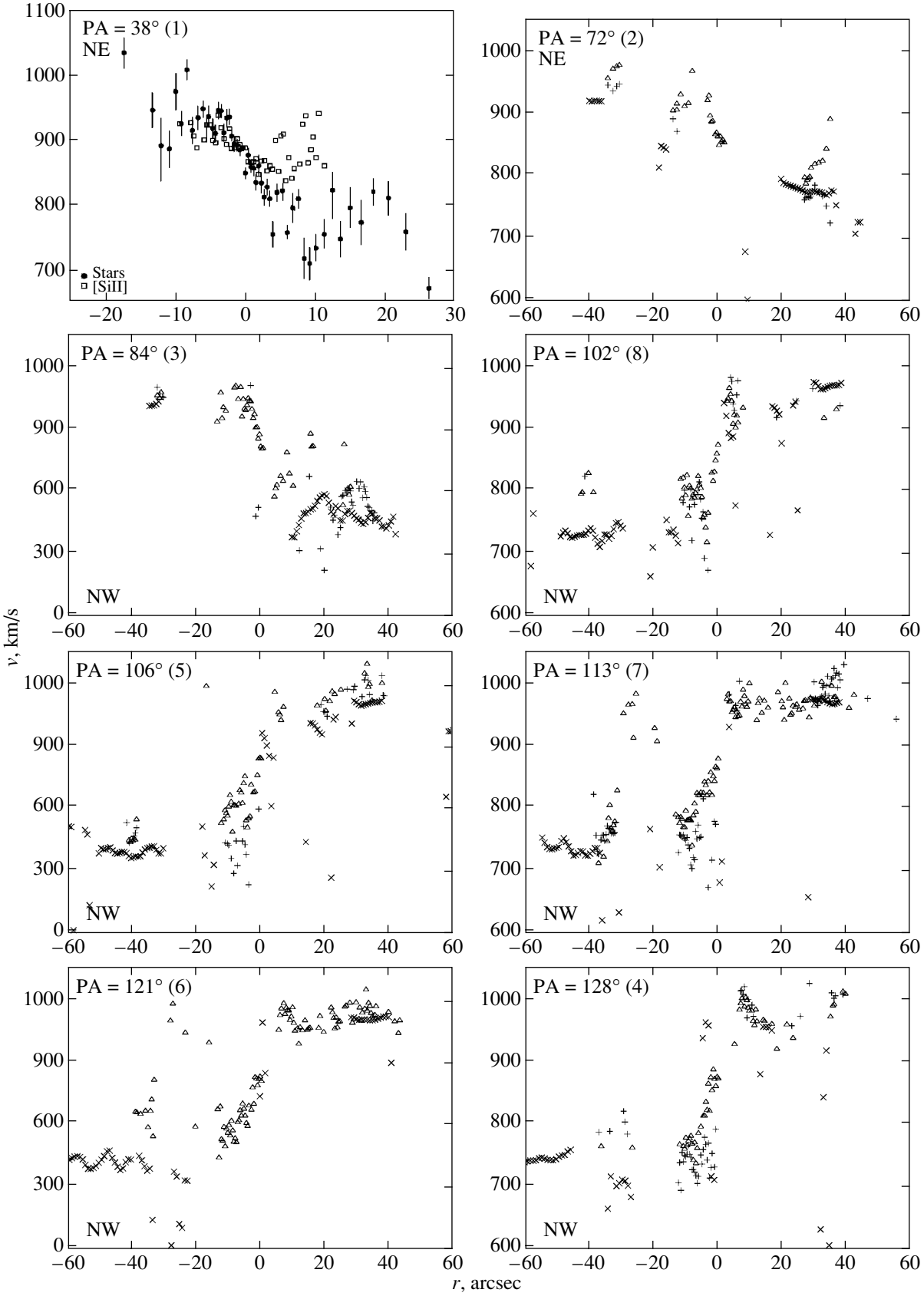

Fig. 6. Radial-velocity curves (UAGS: triangles are for [NII], plus signs are for Hα; IFP: crosses). In the upper left frame, the filled squares correspond to absorption (stellar) lines, and the open squares to [SII] lines.

a ring rotating in the same direction as the galaxy's main body—an equatorial ring). The galaxy's disk-shaped main body is crossed at nearly a right angle by a highly nonuniform gas and dust disk. The system is aligned with respect to the line of sight, so that, in the NE part, the disk is projected against the main body, while the disk is behind the main body in the SW part. The disk may be somewhat curved, as is indicated by the difference between the position angles of the dynamical axis derived for the central region (130° using data from [9] and 126° from the UAGS data) and peripheral regions (115° from the IFP data).

The disk's central part (to distances of ≈1 kpc) is purely gaseous, and it emits in forbidden lines. Beginning at a distance of 1 kpc, the presence of dust is revealed in the disk (according to Fig. 4, the first strong dust lane is observed 8″ from the center, corresponding to a disk inclination of $i = 50°$ to 13″, or ≈1 kpc in the plane of the disk), and Hα emission appears, being strongest at the edges of the disk. Here, the radial velocities derived from the Hα and [NII] lines coincide, suggesting that we are dealing with the same gaseous formation. The differences in emission at the center and peripheral parts could be due to difference in the gas-excitation conditions. The same pattern is observed for the galaxy IC 1689, for which an inner polar disk with the same properties has been discovered [22]. As for IC 1689, the absence of Hα emission in the central region is probably due to an absence of ionizing radiation (collisional excitation occurs there). According to [15], there are many young hot stars capable of exciting Hα emission (radiative excitation) in the peripheral regions.

Though we have not been concerned here with the formation of the polar disk, we can express several considerations. It seems to us that the head-on collision hypothesis [23] cannot explain the overall structure of the NGC 2685 system. Such a collision should have disturbed the galaxy's outer envelope, which is not the case [5]. On the other hand, if the disk was formed by accretion, a donor galaxy should be present in the polar-disk plane, which is also not observed. Thus, we believe that the formation of the disk was accretion-related, but that the donor galaxy was consumed by the primary galaxy after the polar disk had formed. A similar case was considered in [12], and Fig. 7 of [12] is very similar to what should be observed in the case of NGC 2685 if the line of sight were coincident with the normal to the disk.

It is difficult to present final conclusions concerning the disk's age (and stability), since it is not known at what stage of the disk formation the star formation commenced. Our data indicate only that 1.4×10^9 years have elapsed since the onset of the star formation. This estimate is considerably lower than the age suggested in [13]. With the circular velocity we have derived, 180 km/s at a distance of 50″, one orbit takes 0.12×10^9 years, and the disk has revolved some ten times since the onset of the star formation. If the star formation commences at an early stage in the formation of the disk, the disk's age is low, and it can be considered to be still forming rather than a long-lived structure.

In conclusion, we note that any age estimates based on colorimetric data (including our own estimate) have some uncertainty, since they are based on models of galactic color evolution that incorporate several assumptions that may not be satisfied. Direct age estimates will become possible when the disk is resolved into stars in deep HST images, which should reveal old stars, if any are present. The absence of old stars will testify that the disk is young.

Table 3. Position-angle relation of the observed velocity gradient in the central region

PA, degrees	dv/dr, km s^{-1} arcsec^{-1}
This paper	
Gas	
38	5.0
72	6.4
84	8.6
102	7.8
106	9.5
113	10.5
121	10.0
128	15.5
Stars	
38	14.8
Ulrich [8]	
Gas	
127	12.5
130	11.5
135	13.4
Schechter and Gunn [6]	
Gas	
128	12.0
Stars	
38	10.9

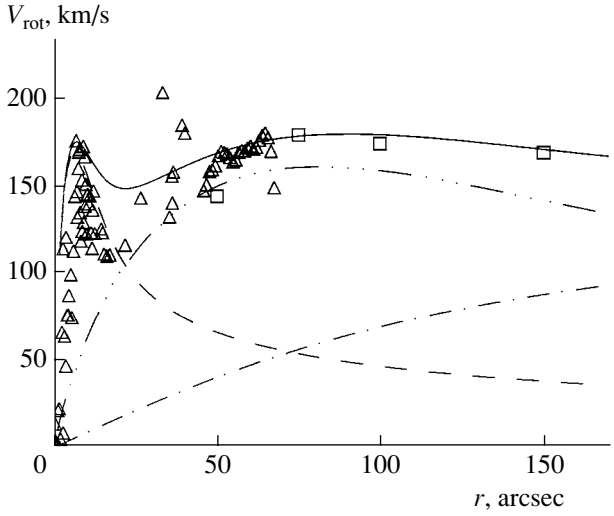

Fig. 7. The galaxy's rotation curve. The triangles show the data derived in the current paper, the squares the data from [4], and the solid curve the three-component model (the dashed curve represents the minor disk, the three-dots-dash-dot curve is the major disk, and the dash-dot curve the halo).

ACKNOWLEDGMENTS

The authors are grateful to the Program Committee for Large Telescopes for the allocation of observing time on the 6 m telescope. This study was supported by the Russian Foundation for Basic Research (project no. 02-02-16033), grant no. E02-11.0-5 from the Ministry of Education of the Russian Federation (currently the Ministry of Education and Science of the Russian Federation), and grant no. 40.022.1.1.1001 of the Federal Science and Technology Program "Astronomy."

REFERENCES

1. B. A. Vorontsov-Vel'yaminov, *Extragalactic Astronomy* (Nauka, Moscow, 1972) [in Russian].
2. E. M. Burbidge and G. R. Burbidge, Astrophys. J. **130**, 20 (1959).
3. A. R. Sandage, *The Hubble Atlas of Galaxies* (Washington, 1961).
4. W. W. Shane, Astron. Astrophys. **82**, 314 (1980).
5. E. Schinnerer and N. Scoville, Astrophys. J. Lett. **577**, L103 (2002).
6. P. L. Schechter and J. E. Gunn, Astron. J. **83**, 1360 (1978).
7. M.-H. Demoulin, Comp. Rend. Acad. Sci. **260**, 3287 (1965).
8. M.-H. Ulrich, Publ. Astron. Soc. Pac. **87**, 965 (1975).
9. O. K. Sil'chenko, Astron. Astrophys. **330**, 412 (1998).
10. V. L. Afanas'ev, A. N. Burenkov, V. V. Vlasyuk, and S. V. Drabek, Otchet Spets. Astrofiz. Obs. Ross. Akad. Nauk, No. 234 (1995).
11. V. L. Afanas'ev and A. V. Moiseev, Pis'ma Astron. Zh. **31**, 214 (2005) [Astron. Lett. **31**, 194 (2005)].
12. F. Bournaud and F. Combes, Astron. Astrophys. **401**, 817 (2003).
13. R. F. Peletier and D. M. Christodoulou, Astron. J. **105**, 1378 (1993).
14. P. B. Eskridge and R. W. Pogge, Astrophys. J. **486**, 259 (1997).
15. G. M. Karataeva, I. O. Drozdovsky, V. A. Hagen-Thorn, *et al.*, Astron. J. **127**, 789 (2004).
16. V. A. Hagen-Thorn, I. I. Popov, and V. A. Yakovleva, Astrofizika **19**, 599 (1983) [Astrophys. **19**, 325 (1983)].
17. V. V. Makarov, V. P. Reshetnikov, and V. A. Yakovleva, Astrofizika **30**, 15 (1989) [Astrophys. **30**, 6 (1989)].
18. A. V. Moiseev, Bull. Spec. Astrophys. Observ. **54**, 74 (2002).
19. M. Disney, J. Davies, and S. Phillips, Mon. Not. R. Astron. Soc. **239**, 939 (1989).
20. O. M. Kurth, U. Fritze-v. Alvensleben, and K. J. Fricke, Astron. Astrophys. **138**, 19 (1999).
21. A. V. Moiseev and V. V. Mustsevoi, Pis'ma Astron. Zh. **26**, 190 (2000) [Astron. Lett. **26**, 565 (2000)].
22. V. A. Hagen-Thorn and V. P. Reshetnikov, Astron. Astrophys. **319**, 430 (1997).
23. K. Bekki, Astrophys. J. **499**, 635 (1998).

Translated by N. Samus'

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 967–972. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1081–1086.

The Physical Parameters of Several Extragalactic Radio Sources Displaying Rapid Variability

S. A. Tyul'bashev

Pushchino Radio Astronomy Observatory, Astro Space Center, Lebedev Institute of Physics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142290 Russia

Received November 5, 2004; in final form, July 6, 2005

Abstract—We present the results of a study of the sources PKS 0405−385, B0917+624, PKS 1257−326, and J1819+3845, which display variability on time scales from 1−7 h (at 5 GHz). Estimates of the physical parameters (magnetic-field intensity, density of relativistic particles, energies of the magnetic field and relativistic particles) are given for B0917+624 and J1819+3845. It is shown that these sources are not in a state of energy equipartition. A number of indirect arguments indicate that the shortest flares in sources with long-time-scale variability have the same parameters as flares in the studied sources. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Variability of the radio emission of extragalactic sources on time scales of the order of one day was first detected more than 30 years ago (see the review [1] and references therein). Two main hypotheses have been proposed to explain this variability. In the first, the variability is due to intrinsic variations of the source, while, in the second, it is due to interstellar scintillations [1]. If the variability is related to the source itself, the estimated brightness temperatures exceed the Compton limit by many orders of magnitude. Attempts to explain these temperatures as an effect of relativistic brightening of the source result in unrealistically high apparent velocities (up to 100 times the speed of light and higher). Explanations of the variability invoking interstellar scintillation seem more convincing. This hypothesis is favored by the detection of seasonal variations of the characteristic time scale for the observed flux-density fluctuations for a number of sources (see, e.g., [2]).

As is noted in [3], if the angular size of a source is less than or about equal to the first Fresnel zone for the interstellar medium, it will necessarily scintillate. If the source also displays intrinsic variability, the scintillations will be superimposed on the intrinsic variability. Thus, in a study of sources with fast variability, we must distinguish cases of pure scintillations and of scintillations + intrinsic variability. A method for checking these cases using the asymmetry index was proposed in [3]. The sources PKS 0405−385, B0917+624, PKS 1257−326, and J1819+3845 were studied, and the variability observed in these sources could be explained by interstellar scintillations. It was also shown in [3] that the scintillating component comprises 50−100% of the total flux density in the studied sources, and that the intrinsic angular sizes of the sources at 5 GHz are 10−40 microarcseconds (μas). In spite of the fact that several dozen sources displaying rapid (shorter than a day) variability are known, only four sources were considered in [3]. This is because, to estimate the angular sizes, one must have data on the fluxes of the rapidly variable sources obtained in long-term multifrequency observations for which the initial observations are coeval. Unfortunately, there are few such series of observations.

The angular sizes of the sources studied in [3] turned out to be quite small, though not as small as expected if the variability were intrinsic.

VLBI observations show that radio sources contain compact features that can be identified with hot spots, jets, cores, or circumnuclear regions. The estimated characteristic linear dimensions of these features are from hundreds of parsecs to approximately one parsec. The intrinsic variability of the sources on time scales of hundreds of days or years demonstrates that the sources contain features that are essentially unresolved in VLBI observations and whose linear dimensions could be less than a parsec. The linear scales of the sources displaying fast variability can be even smaller. Therefore, the question of the physical characteristics of the compact features observed in sources displaying rapid (<1 d) and slow (>1 year) variability is of considerable interest.

Studies of the physical parameters in compact radio sources have yielded some interesting results, including information about the ratio of the energies of the magnetic field and relativistic particles in

1063-7729/05/4912-0967$26.00

Table 1. Estimates of the brightness temperatures and δ

Name	z	θ, μas	T_b, K	δ
PKS 0405−385	1.285	30	9.3×10^{13}	4.5
B0917+624	1.446	10	1.8×10^{15}	12.2
PKS 1257−326	1.256	30	3.4×10^{12}	1.5
J1819+3845	0.54	40	1.7×10^{13}	2.7

sources of various classes. In particular, a study of compact steep-spectrum sources [4–6] showed that, for a mean size of the compact sources of 300 pc, these sources had with almost equal probability any possible ratio of the energies of the magnetic field and relativistic particles: $E_H/E_e \gg 1$, $E_H/E_e \sim 1$, $E_H/E_e \ll 1$ (since the accuracy of the method used to estimate the physical parameters is low, $E_H/E_e \sim 1$ implies equality to unity within an order of magnitude). Studies of faint sources with a gigahertz-peaked spectrum indicated that $E_H/E_e \ll 1$ in 70% of the studied sources [7]. In this case, the mean size of the compact sources was 3 pc, and the sources themselves were probably in the circumnuclear region. A study of the physical conditions in the nuclei of giant radio galaxies has suggested a possible relation between the morphological type of a galaxy and the ratio of E_H and E_e (in radio galaxies with narrow jets, the energy of the magnetic field exceeds the energy of the relativistic particles by many orders of magnitude, while, in radio galaxies without jets, the opposite is observed [8]). The sizes of the studied nuclei in these radio galaxies were about 1 pc.

Sources displaying rapid variability have linear dimensions of a fraction of a parsec and are probably regions of the radio source that are adjacent to the nucleus. In our view, a study of their physical parameters is of special interest.

This paper deals with two problems. The first is to estimate the physical parameters of sources displaying fast variability studied in [3]. There are only four such sources, but, unfortunately, no other estimates of angular sizes obtained using the method proposed in [3] are currently available. Three of these four sources have the fastest variability known. The second problem is to study the physical sizes of sources displaying fast variability and long-time-scale variability.

2. ESTIMATED PHYSICAL PARAMETERS

We estimated the physical parameters using the technique published in [9], which is based on the following formula of Slysh [10]:

$$\theta = 4.3 \times 10^3 S^{1/2} \nu^{-5/4} H_\perp^{1/4} (1+z)^{1/4}, \quad (1)$$

where θ is the angular size of the source in radians, S the flux at the spectral peak in Jansky, ν the frequency in hertz at which the spectral peak is observed, $H_\perp$ the transverse component of the magnetic field in Gauss, and z the redshift of the source.

In the original article, the angular sizes of the sources were estimated based on some reasonable assumptions about the magnetic-field intensity. The formula could be applied if synchrotron self-absorption was observed in the spectrum of the source. However, if the angular size of the source is known, formula (1) can be used to estimate the perpendicular component of the magnetic field. The estimated angular sizes can be obtained from VLBI observations. Assuming that the magnetic-field intensity obtained from (1) refers to the entire source, we can also estimate the density of relativistic particles (see [9] for details). The main advantage of this technique is that these estimates do not use the assumption that the energies of the magnetic field and relativistic particles in the source are equal (energy equipartition).

To use the technique for the estimation of the physical parameters, we must know the redshift and angular size of the source. It is also necessary that the spectrum have a cutoff due to synchrotron self-absorption. Note that, in the case of superluminal outflows, the observed component will have an additional factor, δ, in formula (1), which takes into account the relativistic motion of the source of radiation [11]:

$$\delta = \gamma^{-1}(1 - \beta\cos\phi)^{-1}, \quad \gamma = (1-\beta^2)^{-1/2},$$
$$\beta = v/c,$$

where β is the velocity of the component in units of the speed of light, and ϕ is the angle between the line of sight and the direction of the motion. According to [12], δ can be estimated from the formula

$$\delta = (T_b/10^{12})^{1/3}, \quad (2)$$

where T_b is the brightness temperature of the source in Kelvins, which, in turn, can be estimated from the formula given in the review [13]:

$$T_b = 1.22 \times 10^{12} S \nu^{-2} \theta^{-2} (1+z). \quad (3)$$

Here, S is the flux density in Jansky, ν the frequency in gigahertz at which the brightness temperature T_b is estimated, and θ the angular size of the source at this frequency in milliarcseconds.

Our estimates of the brightness temperatures and δ are listed in Table 1, which gives (i) the source name, (ii) the redshift, (iii) the angular size of the

Table 2. Estimates of physical parameters of sources displaying fast variability

Source name	z	θ, μas	r, light days	$H_\perp$, G	$E_{H_\perp}$, erg	n_e, cm^{-3}	E_e, erg	E_H/E_e
B0917+624	1.446	10	33	10^{-4}	2×10^{42}	6×10^{7}	5×10^{56}	3×10^{-15}
J1819+3845	0.54	40	138	3×10^{2}	10^{54}	0.2	10^{47}	10^{7}

source, (iv) the estimated brightness temperature, and (v) the estimated value of δ.

The values of δ given in Table 1 show that, when estimating the physical parameters, it is necessary to use the formula of Slysh [10] with the correction of Marscher [11], which takes into account superluminal motions.

As is noted above, a cutoff must be observed in the spectra of the studied sources in order to apply the technique [9]. All four sources have signs of a cutoff in their spectra. Let us consider the spectra of these sources in more detail.

PKS 0405−385. The redshift of this quasar is $z = 1.285$ [14]. Light curves at 8.6, 4.8, 2.4, and 1.4 GHz are available [15]. It was shown in [3] that the compactness of the source [i.e., the ratio of the flux of the compact (scintillating) component to the flux of the extended (not scintillating) component] is $R = 1$ at 8.6 GHz; i.e., there is no extended component within the accuracy of the estimates. The source compactness is 0.5 at 4.8 GHz. The observed flux density decreases beginning from 8.6 GHz ($S_{8.6\,\mathrm{GHz}} = 1.8$ Jy, $S_{4.8\,\mathrm{GHz}} = 1.5$ Jy, $S_{2.4\,\mathrm{GHz}} = 1.1$ Jy, $S_{1.4\,\mathrm{GHz}} = 0.8$ Jy). The spectral index $\alpha\,(S \sim \nu^{-\alpha})$ near the spectrum cutoff is $\alpha = 0.37$, which increases to $\alpha \simeq 1.6$ when the known compactness of the source is taken into account. Thus, the presence of a spectral cutoff is beyond doubt. However, we cannot estimate the physical parameters because we do not know the peak frequency of the spectrum.

B0917+624. The redshift of this quasar is $z = 1.446$ [16]. Light curves are available at 15, 8.4, 5.0, 2.7, and 1.4 GHz [17]. The observed flux densities (during a flare in May 1989) were $S_{15\,\mathrm{GHz}} = 1.52$ Jy, $S_{8.4\,\mathrm{GHz}} = 1.57$ Jy, $S_{5\,\mathrm{GHz}} = 1.53$ Jy, $S_{2.7\,\mathrm{GHz}} = 1.36$ Jy, and $S_{1.4\,\mathrm{GHz}} = 1.14$ Jy. Thus, a weakly marked maximum is visible in the spectrum at 8.4 GHz. According to [17], the accuracy of the estimated flux densities was better than 2%, so that the spectral cutoff is significant. The spectral index near the cutoff is $\alpha \simeq 0.2$. We estimated the physical parameters of the source, assuming that the cutoff is due to synchrotron self-absorption (Table 2).

Table 2 presents (i) the source name, (ii) the redshift, (iii) the source angular size, and (iv) the linear dimension in light days. The linear size of the source was estimated using the formulas given in the reference book of Lang [18]:

$$D_L = \frac{c}{H_0 q_0^2} z q_0 + (q_0 - 1)\left(-1 + \sqrt{2q_0 z + 1}\right), \tag{4}$$

$$r = D_L \theta (1+z)^{-2}, \tag{5}$$

where D_L is the photometric distance, c the speed of light in km/s, $H_0 = 75$ km/s Mpc the Hubble constant, $q_0 = 1/2$ the cosmological deceleration parameter, and r the estimated linear size (diameter) of the source if the angular size of the source is θ. The remaining columns of Table 2 list the estimated physical parameters: the magnetic-field intensity, magnetic-field energy, and the density and energy of the relativistic electrons.

PKS 1257−326. This quasar's redshift is $z = 1.256$ [19]. Unfortunately, light curves of this source are available only at 8.6 and 4.8 GHz. The observed flux decreases with decreasing frequency [19]. Since no maximum is observed in the spectrum, we cannot estimate the physical parameters.

J1819+3845. This quasar's redshift is $z = 0.54$ [20]. Light curves are available at 8.4, 4.9, 2.7, and 1.4 GHz. According to [3], the source compactness at 4.9 and 8.4 GHz is unity, indicating an absence of a halo at these frequencies. The integrated spectrum is also presented in [20]; its shape suggests the presence of a cutoff beginning from 43 GHz. The spectral index near the cutoff is $\alpha = 0.9$. Assuming that the cutoff is due to synchrotron self-absorption, we can estimate the physical parameters of the source.

The accuracy of the estimated physical parameters depends in the first instance on the accuracy of the estimated linear dimensions of the source, which we suppose can differ from the values listed in Table 2 by a factor of two. Therefore, the accuracy of our estimates of the magnetic-field energy is two to three orders of magnitude of the value listed in Table 2 ($E_{H_\perp} \sim H_\perp^2$, $H_\perp \sim \theta^4$). Even with this large error in the estimated physical parameters, we can see that $E_H/E_e \ll 1$ in B0917+624, while $E_H/E_e \gg 1$ in J1819+3845; i.e., these sources are not in energy equipartition. Recall that an absence

Table 3. Estimates of brightness temperatures for sources from [24]

Name	z	ΔS, Jy	τ, days	T_b, K
B0048−097	0.475	1	115	5×10^{13}
B0235+164	0.940	3	183	1.5×10^{14}
OJ287	0.306	5	183	5×10^{13}

of energy equipartition was shown for the majority of compact steep-spectrum sources, faint sources with gigahertz-peaked spectra, and giant radio galaxies [6–8, 21, 22].

3. DURATION OF FLARES IN SOURCES WITH LONG- AND SHORT-TIME-SCALE VARIABILITY

Let us consider the characteristic sizes of the regions of flares in sources with long- and short-time-scale variability. The mean sizes of the four sources with rapid variability studied by us are of the order of 100 light days at 4.8 GHz (for PKS 0405−385 and PKS 1257−326, which are not included in Table 2, the linear sizes are 100 and 103 light days, respectively). At first glance, these sizes differ by an order of magnitude from those for the sources with long-time-scale variability. However, special studies of the durations of flares at frequencies from 37 GHz to 408 MHz yield a different picture.

In particular, a study of 294 individual flares in 40 active galactic nuclei (AGN) at 37 and 22 GHz was carried out in [23]. A statistical study showed that the mean size of the flare regions was 95 light days at 37 GHz and 153 light days at 22 GHz. For each flare, the sizes were corrected for the redshift and Doppler brightening before the analysis. Prior to this correction, the mean sizes of the flare regions at both frequencies were a factor of three larger. The study [23] dealt specifically with the properties of average flares; i.e., some flare regions were smaller than average, while some were larger. Note also that these sizes are, in fact, only upper limits for the linear dimensions, since they were obtained based on the variability time scale by estimating the size of the flare region as $r = ct$ (where c is the speed of light and t is the variability time scale). A comparison of the linear dimensions of the sources from Table 2 with those for the flare regions of the sources from [23] shows that they approximately coincide.

We can obtain information on the characteristic sizes of the flare regions in 51 AGN (the monitoring of some AGN lasted longer than 25 years) at 14.5, 8.5, and 4.8 GHz from [24]. The observed characteristic durations of flares vary from 115 to 3650 days. Since no correction for the redshift or Doppler beaming was applied, the intrinsic sizes of the flare regions could be considerably smaller (recall that the apparent and intrinsic sizes at 37 and 22 GHz differ by a factor of three).

The study [25] of the low-frequency (408 MHz) variability of 11 AGN showed that the characteristic durations of flares vary from 73 to 1280 days.

Note that the three papers noted above are concerned with studies of the durations of flares of a large number of AGN in a very broad frequency band, and that the lower boundary of the flare durations essentially coincides with the mean size of the flare sources we have studied.

4. DISCUSSION

(1) We can see from [23–25] that the durations of flares (and, consequently, the linear sizes of the flare regions) at frequencies from 37 GHz to 408 MHz can vary over at least an order of magnitude. However, we can assert with a high degree of confidence that the characteristic sizes of regions corresponding to flares of extremely short duration are from several tens to hundreds of light days. Interstellar scintillations should be observed for flares with such durations, since a linear size of 100 light days for a source with $z = 1$ corresponds to an angular size of $\theta = 15$ μas [see (4) and (5)], and, according to [3], such sizes correspond approximately to the first Fresnel zone for the interstellar medium.

Thus, the real sizes of the regions producing short-time-scale variability (due to interstellar scintillations) on time scales from one to several hours (for the four sources we have studied) coincide with the real sizes of the regions of some flares in sources with a long-time-scale (intrinsic) variability.

(2) As we can see from Table 1, the brightness temperatures of the sources we have studied are 10^{12}–10^{14} K. The estimated brightness temperatures of 11 AGN observed at 408 MHz [25] were from 10^{12} to 10^{15} K. The brightness temperatures of flares observed at 37 and 22 GHz [23] were estimated in [26]; of 40 observed AGN, brightness temperatures from 10^{12} to 10^{14} K were observed in 24 sources. Unfortunately, we were not able to find data on the brightness temperatures of the sources observed in [24]. Therefore, we have taken three sources with the fastest observed variability and estimated their brightness temperatures using (3); the results are listed in Table 3, which gives (i) the source name, (ii) the redshift, (iii) the flux fluctuation of a "mean" flare at 4.8 GHz, taken from the UMRAO database (since the peak flux densities in individual flares can

differ widely, we took an average value), (iv) the flare duration in days, taken from the structure-function data of [24], and (v) the brightness temperatures of the sources. The sources with the fastest variability observed at 4.85 GHz have brightness temperatures of 10^{13}–10^{14} K, which again coincide with those of the sources we have studied. Thus, the brightness temperatures of sources displaying long-time-scale variability at frequencies from 37 GHz to 408 MHz coincide with those for the sources with short-time-scale variability studied here.

(3) We have estimated the physical parameters for two of the four radio sources studied. The estimated magnetic-field intensities and densities of relativistic electrons differ strongly for the two sources. However, a comparison of these estimates with those obtained using the same technique for giant radio galaxies, compact steep-spectrum sources, and faint sources with gigahertz-peaked spectra [6–8, 21, 22] demonstrates that the estimates obtained earlier do not contradict (within the errors) those obtained in the current study. Thus, a comparison of the physical parameters of supercompact sources and the sources studied earlier does not provide any reason to consider these sources as belonging to a special class.

Summarizing items (1), (2), and (3), we can see that the angular sizes of some sources with long-time-scale variability coincide with those for sources with short-time-scale variability; their brightness temperatures also coincide, and the estimated physical parameters do not differ fundamentally from those for compact sources studied earlier. This suggests that sources displaying short-time-scale variability belong to the same class as sources displaying long-time-scale variability.

The arguments presented in items (1), (2), and (3) do not represent direct proofs that sources with variability on time scales shorter than a day represent the same class as sources with variability on time scales of about one year. However, it seems to us that these indirect arguments clearly testify that sources with fast and slow variability have the same nature.

A good test that can yield direct indications about the nature of sources displaying short-time-scale variability is long-term multifrequency monitoring of sources with appreciable variability on time scales shorter than a year. Such sources should also exhibit variability on time scales shorter than a day. In this case, the estimated upper limits for the linear dimensions based on the durations of long-time-scale flares should be less than or equal to the linear dimensions estimated using scintillation methods.

5. CONCLUSION

(1) We have estimated the physical parameters of two sources that display rapid variability on time scales from one to several hours. A comparison of the energies of the magnetic field and relativistic particles demonstrates an absence of energy equipartition.

(2) We present indirect arguments suggesting that sources with variability time scales shorter than half a day and of the order of a hundred days are not clearly distinguished, and belong to the same class of sources as those displaying long-time-scale variability.

ACKNOWLEDGMENTS

The author is grateful to T.V. Smirnova and V.I. Shishov for useful discussions and comments, as well as to G.E. Tyul'basheva for help with the manuscript. This work was supported by the Russian Foundation for Basic Research (project nos. 03-02-509 and 03-02-16522) and the Basic Research Program of the Division of Physical Sciences of the Russian Academy of Sciences "Extended Objects in the Universe." This research has made use of the UMRAO database.

REFERENCES

1. S. J. Wagner and A. Witzel, Annu. Rev. Astron. Astrophys. **33**, 163 (1995).
2. D. L. Jauncey and J.-P. Maquart, Astron. Astrophys. **370**, L9 (2001).
3. V. I. Shishov, T. V. Smirnova, and S. A. Tyul'bashev, Astron. Zh. **82**, 281 (2005) [Astron. Rep. **49**, 250 (2005)].
4. V. S. Artyukh, S. A. Tyul'bashev, and P. A. Chernikov, Astron. Zh. **76**, 3 (1999) [Astron. Rep. **43**, 1 (1999)].
5. S. A. Tyul'bashev and P. A. Chernikov, Astron. Zh. **77**, 331 (2000) [Astron. Rep. **44**, 286 (2000)].
6. S. A. Tyul'bashev and P. A. Chernikov, Astron. Astrophys. **373**, 381 (2001).
7. S. A. Tyul'bashev, Astron. Zh. **78**, 498 (2001) [Astron. Rep. **45**, 428 (2001)].
8. V. S. Artyukh, M. A. Ogannisyan, and S. A. Tyul'bashev, Pis'ma Astron. Zh. **20**, 258 (1994) [Astron. Lett. **20**, 211 (1994)].
9. V. S. Artyukh, Proc. Lebedev Phys. Inst. **189**, 289 (1988); http://lnfm1.sai.msu.ru/~patch/css/.
10. V. I. Slish, Nature **199**, 682 (1963).
11. A. P. Marscher, Astrophys. J. **264**, 296 (1983).
12. R. D. Blandford and A. Königl, Astrophys. J. **232**, 34 (1979).
13. D. R. Altschuler, Fundam. Cosm. Phys. **14**, 37 (1989).
14. B. A. Peterson and J. G. Bolton, Astrophys. Lett. **10**, 105 (1972).
15. L. Kedziora-Chudczer, D. L. Jauncey, M. H. Wieringa, *et al.*, Astrophys. J. **490**, L9 (1997).

16. M. Stickel and H. Kühr, Astron. Astrophys., Suppl. Ser. **101**, 521 (1993).
17. A. Quirrenbach, A. Kraus, A. Witzel, *et al.*, Astron. Astrophys., Suppl. Ser. **141**, 221 (2000).
18. K. R. Lang, *Astrophysical Formulae: A Compendium for the Physicist and Astrophysicist* (Springer-Verlag, Berlin, 1974; Mir, Moscow, 1978).
19. E. Perlman, P. Padovani, P. Giommi, *et al.*, Astron. J. **115**, 1253 (1998).
20. J. Dennett-Thorpe and A. G. de Bruyn, Astrophys. J. **529**, L65 (2000).
21. V. S. Artyukh, M. A. Ogannisyan, and S. A. Tyul'bashev, Pis'ma Astron. Zh. **14**, 706 (1988) [Sov. Astron. Lett. **14**, 301 (1988)].
22. V. S. Artyukh, S. A. Tyul'bashev, and P. A. Chernikov, Astron. Zh. (2005) (in press).
23. E. Valtaoja and H. Teräsranta, Astron. Astrophys. **289**, 35 (1994).
24. P. A. Hughes, H. D. Aller, and M. F. Aller, Astrophys. J. **396**, 469 (1992).
25. M. Bondi, L. Padrielli, R. Fanti, *et al.*, Astron. Astrophys. **308**, 415 (1996).
26. H. Teräsranta and E. Valtaoja, Astron. Astrophys. **283**, 51 (1994).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 973–983. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1087–1098.

A Solution to the Problem of Phaseless Mapping for a High-Orbit Space–Ground Radio Interferometer

A. T. Bajkova

Main (Pulkovo) Astronomical Observatory, St. Petersburg, Russia

Received January 20, 2005; in final form, May 18, 2005

Abstract—We consider the problem of mapping with ultra-high angular resolution using a space–ground radio interferometer with a space antenna in a high orbit, whose apogee height exceeds the radius of the Earth by a factor of ten. In this case, a multielement interferometer essentially degenerates into a two-element interferometer. The degeneracy of the close-phase relations prevents the use of standard methods for hybrid mapping and self-calibration for the correct reconstruction of images. We propose a new phaseless mapping method based on methods for the reconstruction of images in the complete absence of phase information, using only the amplitudes of the spatial-coherence function of the source. In connection with this problem, we propose a new method for the reliable solution of the phase problem, based on optimizing information-carrying nonlinear functionals, in particular, the Shannon entropy. Results of simulations of mapping radio sources with various structures with ultra-high angular resolution in the framework of the RADIOASTRON mission are presented. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The problem of phaseless mapping in Very Long Baseline Interferometry (VLBI) is considered in detail in [1], together with the uniqueness of the solutions and methods for solving for the phases. The current study is a logical continuation of this work, since it is concerned with the phaseless mapping of radio sources with ultra-high angular resolution using a ground–space interferometer with a high-apogee orbiting antenna, and the development of more trustworthy methods for the reconstruction of images based only on the amplitudes of their Fourier spectra (the spatial-coherence function).

The problem of phaseless mapping in VLBI arises when the measured phases are subject to large errors introduced by the medium through which the radio wave propagates, and it is not possible to use the closure phases and adaptive-calibration methods (hybrid mapping and self-calibration [2]) traditionally applied in VLBI to correctly reconstruct images.

This is the case for a two-element interferometer [1], and also for a ground–space radio interferometer with an antenna in a high orbit whose apogee exceeds the Earth's radius by an order of magnitude or more. In the latter case, the multielement interferometer essentially degenerates into two elements, independent of the number of ground stations, leading to degeneracy in the phases summed around triangles. As a result, mapping using the standard methods becomes meaningless.

This problem is relevant for the future Russian space mission RADIOASTRON [3] (planned to have an antenna in a high-apogee orbit reaching heights of 350 000 km), intended to map extragalactic radio sources with ultra-high angular resolution reaching hundredths of a milliarcsecond (mas), with the goal of revealing the workings of the central engines of distant quasars and galaxies.

As we indicated above, an effective strategy for phaseless VLBI mapping was presented in [1]. This method is based on (1) making a preliminary reconstruction of the amplitude of the visibility function (spectrum) over the entire UV plane by reconstructing an intermediate image with zero spectral phase using the data measured on a limited set of points, and (2) reconstructing the desired image based on the amplitude of the source spectrum obtained in the first stage using methods designed to solve the phase problem.

The goal of our current study is to develop a method for the realization of the second step of this algorithm that is more reliable than the method of Fienup that has been applied earlier.

The goal of this work is topical. In spite of the fundamental existence (apart from degenerate cases, defined on a set of measure zero) of solutions of the phase problem for multidimensional ($\geq$2), spatially restricted signals described by nonnegative real functions [4] that are unique with accuracy to within a class of equivalent functions (a linear shift or rotation of the image by 180°), no trustworthy, practical algorithm has been developed to obtain these solutions.

For example, the algorithms of Fienup [5, 6], which are the most efficient and most widely applied in practice, do not possess the property of compression [7], so that they cannot guarantee convergence to the correct solution in all cases. A successful application of Fienup algorithms in the case of comparatively simple source structures is demonstrated in [1].

Here, we attempt to fill this gap by proposing a more fundamental method for solving the phase problem, based on using the methods of nonlinear optimization to search for global extrema.

The following sections discuss VLBI mapping with an antenna in a high orbit, describe the proposed method for solving the phase problem, and present tests of the method and results of simulations of the RADIOASTRON mission, aimed at mapping radio sources with ultra-high angular resolution.

2. VLBI MAPPING WITH AN ANTENNA IN A HIGH ORBIT

Placing at least one VLBI station beyond the Earth, and thereby increasing the maximum baseline of the resulting interferometer, makes it possible to appreciably enhance the resolving power of the instrument.

It is planned in the near future to realize the space project RADIOASTRON [3] of the Russian Academy of Sciences (the Astro Space Center, Lebedev Physical Institute), which aims to construct a ground–space radio interferometer for the mapping of radio sources with ultra-high angular resolution, of the order of hundredths of a mas. Such resolutions are provided by observations at 1.35 cm on the maximum baseline of 350 000 km, which is achieved when the orbiting antenna is at apogee [8].

The desire to obtain ultra-high angular resolution by increasing the length of the baseline joining the ground stations and the one orbiting station leads to certain mathematical problems in the mapping, associated with the degeneracy of the multielement ground–space interferometer into a two-element interferometer in terms of the effective filling of the UV plane, independent of the number of ground stations [1]. One consequence of this is the degeneracy of the relations for the closure phases that are traditionally used in VLBI to correctly reconstruct the spectral phases via adaptive-calibration methods.

We can write the equations for the closure phases [2]:

$$\widetilde{C}_{ijk} = \widetilde{\phi}_{ij} + \widetilde{\phi}_{jk} - \widetilde{\phi}_{ik} = \phi_{ij} + \phi_{jk} - \phi_{ik} + \text{noise term} = C_{ijk} + \text{noise term}, \quad (1)$$

where $\widetilde{\phi}_{ij} = \phi_{ij} + \theta_i - \theta_j + \text{noise term}$, ϕ_{ij} is the spectral phase of the source on the baseline (ij), θ_i and θ_j are the phases of the complex gains of antennas i and j, which include both instrumental and atmospheric components that cancel out when the phases are summed around a triangle, and "noise term" is the random, residual component of the phase noise, which is usually small. Here, a tilda denotes measured quantities.

The degeneracy of the closure-phase equations is a consequence of the geometrical degeneracy of the triangles, whose apices correspond to the ground antennas i and j and the space antenna k, which is very distant from the Earth. The right-hand side of relation (1) essentially vanishes, independent of the real spectral phase of the source: $\widetilde{C}_{ijk} \approx$ noise term ≈ 0. It is obvious that, in this situation, applying the closure equations will always yield symmetrical structures, independent of the real source structure.

The problem of poor UV-plane coverage, which leads to large sidelobes in the synthesized antenna beam, can be partially solved by applying the technique of multifrequency synthesis [9, 10]. However, this is not sufficient to correctly reconstruct the phases. Therefore, we propose to use phaseless mapping based on the presented methods to reconstruct the structure of the source. We can correctly reconstruct the spatial orientation of the source using a solution obtained via adaptive-calibration methods based on observations with the ground-based (low-frequency) part of the VLBI array.

3. METHOD FOR SOLVING THE PHASE PROBLEM

Let us formulate the problem of reconstructing two-dimensional images in discrete form. Let the discretization of the map be carried out in accordance with the theorem of Kotel'nikov–Shannon, and the dimensions of the map be $N \times N$. The spectrum of the source is the N-point discrete Fourier transform of the two-dimensional distribution x_{ml} over the source radiation with a finite carrier:

$$X_{nk} = \frac{1}{N}\sum_{m=0}^{N-1}\sum_{l=0}^{N-1} x_{ml} \times \exp\left(\frac{-i\,2\pi(nm+kl)}{N}\right) = A_{nk} + iB_{nk} = M_{nk}\exp(i\Phi_{nk}), \quad (2)$$

where A_{nk} is the real part, B_{nk} the imaginary part, M_{nk} the magnitude, and Φ_{nk} the phase of the spectrum X_{nk}, with

$$A_{nk} = M_{nk}\cos\Phi_{nk}, \quad B_{nk} = M_{nk}\sin\Phi_{nk}. \quad (3)$$

We formulate this problem as follows. We wish to use known values of the amplitude of a spectrum (the

spatial-coherence function in VLBI) M_{nk} on some set of points in the spatial-frequency domain (the UV plane) to reconstruct the image x_{ml}, which is equivalent to reconstructing the spectral phase Φ_{nk}, since the distribution x_{ml} is the inverse Fourier transform of the total spectrum X_{nk} (2), taking into account both the amplitudes and phases.

Let the measurements of the spatial-coherence function taking into account (2) and (3) satisfy the relations

$$\sum_m \sum_l x_{ml} a_{ml}^{nk} + \eta_{nk}^r = A_{nk} = M_{nk} \cos \Phi_{nk}, \quad (4)$$

$$\sum_m \sum_l x_{ml} b_{ml}^{nk} + \eta_{nk}^i = B_{nk} = M_{nk} \sin \Phi_{nk}, \quad (5)$$

separately for the real and imaginary parts of the spectrum, where $a_{ml}^{nk} = \cos(2\pi(mn + lk)/N)/N$, $b_{ml}^{nk} = \sin(2\pi(mn + lk)/N)/N$, and η_{nk}^r and η_{nk}^i are the measurement errors for the real and imaginary parts of the spectrum, respectively, which obey a Gaussian distribution with zero mean and dispersion σ_{nk}^2.

Let us represent the relations for the spectral phase as follows:

$$\cos \Phi_{nk} = 2 \cos^2(\Phi_{nk}/2) - 1, \quad (6)$$

$$\sin \Phi_{nk} = [\sin(\Phi_{nk}/2) + \cos(\Phi_{nk}/2)]^2 - 1 \quad (7)$$

and introduce the new nonnegative variables

$$t_{nk} = 2 \cos^2(\Phi_{nk}/2), \quad (8)$$

$$s_{nk} = [\sin(\Phi_{nk}/2) + \cos(\Phi_{nk}/2)]^2. \quad (9)$$

The reconstruction of the image can then be represented as the solution of the following optimization problem with the linear constraints (4), (5), into which we substitute the variables t_{nk} and s_{nk} in accordance with (6)–(9):

$$Q(x_{ml}, t_{nk}, s_{nk}) + \sum_n \sum_k \frac{(\eta_{nk}^{r\,2} + \eta_{nk}^{i\,2})}{2\sigma_{nk}^2} = \min, \quad (10)$$

$$\sum_m \sum_l x_{ml} a_{ml}^{nk} - M_{nk} t_{nk} + \eta_{nk}^r = -M_{nk}, \quad (11)$$

$$\sum_m \sum_l x_{ml} b_{ml}^{nk} - M_{nk} s_{nk} + \eta_{nk}^i = -M_{nk}, \quad (12)$$

$$x_{ml}, t_{nk}, s_{nk} \geq 0. \quad (13)$$

Here, Q is a nonlinear functional determining the chosen criteria for the quality of the reconstruction. The second term in the optimization functional (10) is an estimate of the disagreement between the solution and the measurements according to an χ^2 criterion.

In addition, it is easy to show that the variables t_{nk} and s_{nk} satisfy the nonlinear constraint

$$\cos^2 \Phi_{nk} + \sin^2 \Phi_{nk} = (t_{nk} - 1)^2 + (s_{nk} - 1)^2 = 1, \quad (14)$$

which is key for the correct reconstruction of the spectral phase.

Including the nonlinear constraints (14) in the Lagrange functional in the standard way appreciably complicates the reconstruction algorithm. We therefore propose the following scheme for the solution of problems (10–(14). The problem with the linear constraints (10)–(13) is first solved in the standard way via direct optimization using Lagrange multipliers [7]. In the course of the numerical iterative search for the extremum of the corresponding dual functional (for example, using a coordinate-descent method [11]), the nonlinear constraints (14) are placed on the variables t_{nk} and s_{nk}, similar to a projection onto a convex set [7]. In this case, t_{nk} and s_{nk} cannot vary independently of each other in their determination of the spectral phase. This combined algorithm incorporates the advantages of both methods for the optimization of nonlinear functionals with linear constraints, characterized by the presence of a global extremum, and iterative methods, which are distinguished by the simplicity of their allowance for various constraints on the solution directly in the computational algorithm.

When Q is a functional of the Shannon entropy [12], namely,

$$Q(x_{ml}, t_{nk}, s_{nk}) = \sum_m \sum_l x_{ml} \ln x_{ml} + \sum_n \sum_k (t_{nk} \ln t_{nk} + s_{nk} \ln s_{nk}), \quad (15)$$

we have the maximum-entropy method (MEM). The basis for the legitimacy of the functional (15), which is the total entropy of the image and the new variables t_{nk} and s_{nk}, can be obtained based on a ray model for the formation of the image [7], maximizing the joint probability for the formation of the image and the field of the statistically independent variables $\{t_{nk}, s_{nk}\}$.

An important constraint appearing in the system of equations (11) and specified by the necessary normalization of the image is the constraint on the total flux of the source M_o (the zeroth harmonic of the spectrum):

$$\sum_m \sum_l x_{ml} = M_o.$$

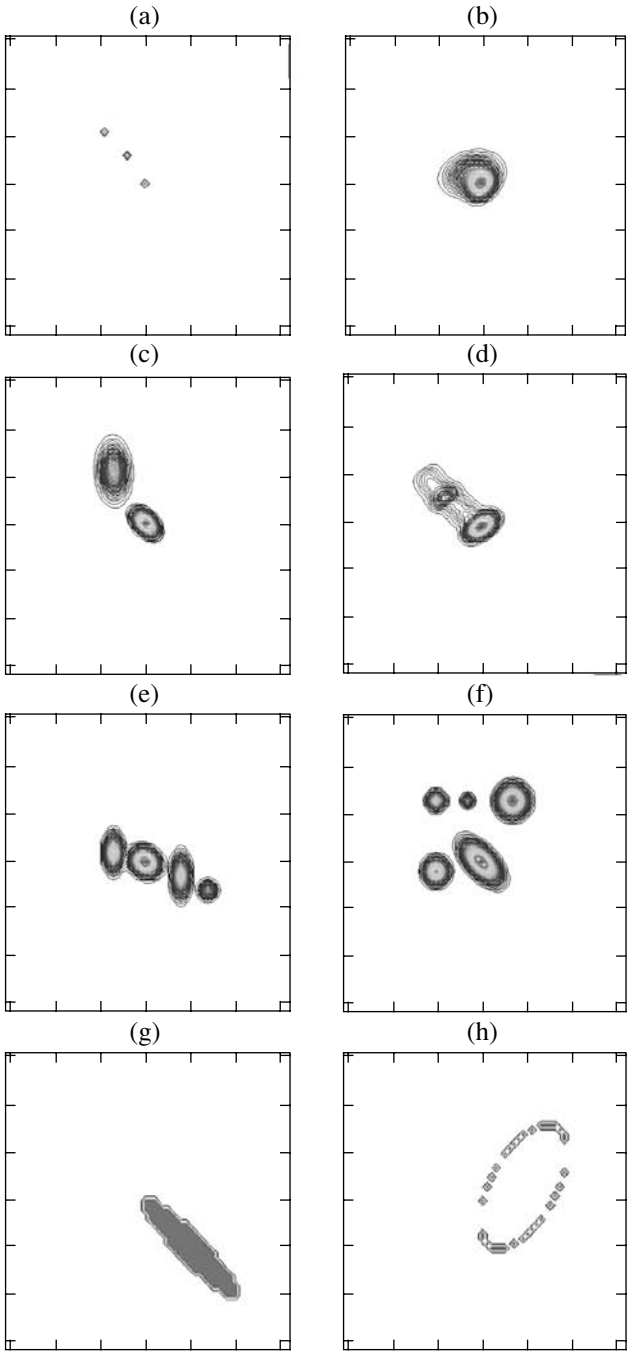

Fig. 1. Specified images of model radio sources.

Another well-studied effective method is the minimum measure of Rényi [13]. In this case, the functional Q has the appearance

$$Q(x_{ml}, t_{nk}, s_{nk}) = \sum_m \sum_l x_{ml}^{\alpha} + \sum_n \sum_k (t_{nk}^{\alpha} + s_{nk}^{\alpha}),$$

where $\alpha \neq 1$.

In practice, in VLBI mapping, the Shannon entropy has become the most widely accepted among the various nonlinear functionals specifying various measures of the quality of reconstructed images. Accordingly, we consider here a developed phase-reconstruction method based on the MEM.

Applying the standard Lagrange-multiplier method to problems (10)–(13) and using (15) yields the following absolute optimization problem:

$$L = \sum_m \sum_l x_{ml} \ln x_{ml} + \sum_n \sum_k (t_{nk} \ln t_{nk} + s_{nk} \ln s_{nk}) + \sum_n \sum_k \frac{({\eta_{nk}^r}^2 + {\eta_{nk}^i}^2)}{2\sigma_{nk}^2} + \sum_n \sum_k \alpha_{nk} \left(\sum_m \sum_l x_{ml} a_{ml}^{nk} - M_{nk} t_{nk} + \eta_{nk}^r + M_{nk} \right) + \sum_n \sum_k \beta_{nk} \left(\sum_m \sum_l x_{ml} b_{ml}^{nk} - M_{nk} s_{nk} + \eta_{nk}^i + M_{nk} \right) = \min, \quad (16)$$

where α_{nk}, β_{nk} are the Lagrange multipliers, or the dual variables.

The necessary condition for the existence of an extremum of the functional L is that the gradient $\dfrac{\partial L}{\partial y_{op}} = 0$, and that the Hess matrix composed of the elements $\dfrac{\partial^2 L}{\partial y_{op} \partial y_{qr}}$ be positive semidefinite at the point where this gradient is zero (here, the letter y denotes the generalized variables $x_{ml}, t_{nk}, s_{nk}, \eta_{nk}^r, \eta_{nk}^i$, and op, qr the two-dimensional indices for these variables).

A sufficient condition for the existence of a local extremum is that the Hess matrix be positive definite. If the Hess matrix is positive definite everywhere, the functional will be convex, and the local extremum will be a global extremum.

We can obtain a solution for the desired distribution (image) x_{ml} and the variables t_{nk}, s_{nk} determining the spectral phase from the necessary condition for the existence of an extremum of the functional L:

$$x_{ml} = \exp\left(\sum_n \sum_k \alpha_{nk} a_{ml}^{nk} + \beta_{nk} b_{ml}^{nk} - 1 \right), \quad (17)$$

$$t_{nk} = \exp(\alpha_{nk} M_{nk} - 1), \quad s_{nk} = \exp(\beta_{nk} M_{nk} - 1), \quad (18)$$

$$\eta_{nk}^r = -\alpha_{nk} \sigma_{nk}^2, \quad \eta_{nk}^i = -\beta_{nk} \sigma_{nk}^2.$$

As we can see from (17) and (18), the solutions for the variables x_{ml}, t_{nk}, s_{nk} are always positive; i.e., the condition (13) is satisfied automatically, which is an

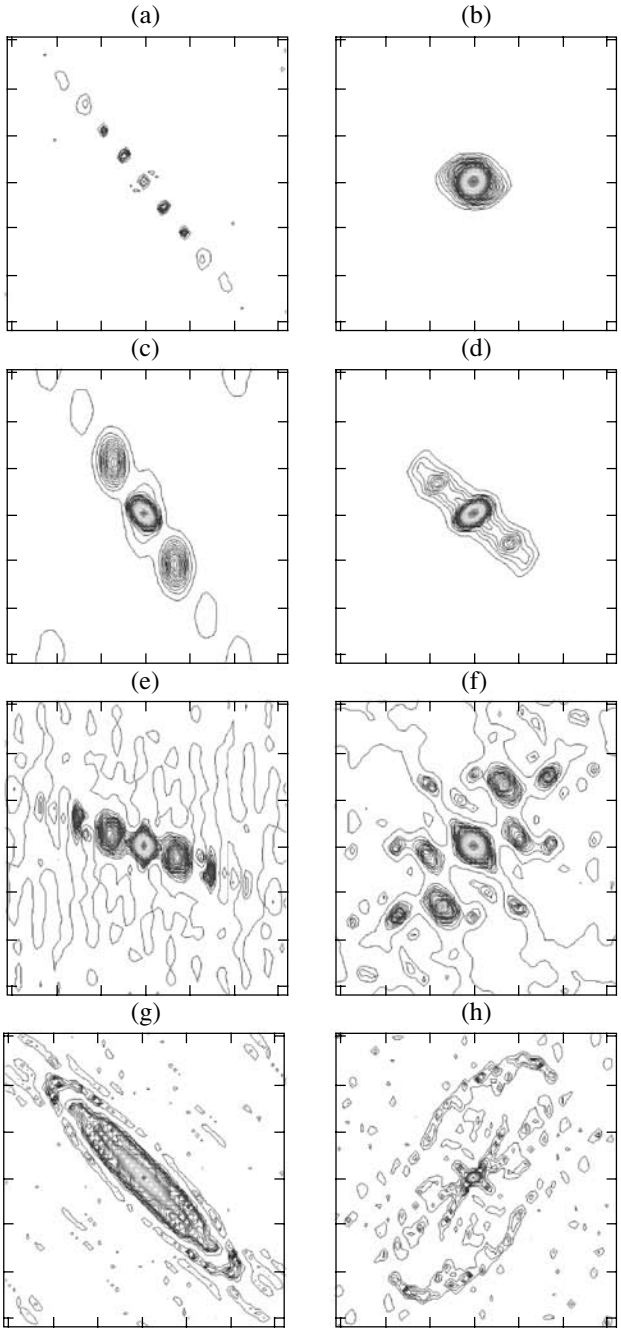

Fig. 2. Images of model radio sources reconstructed using standard mapping methods in the case of degenerate phase triangles (zero spectral phase).

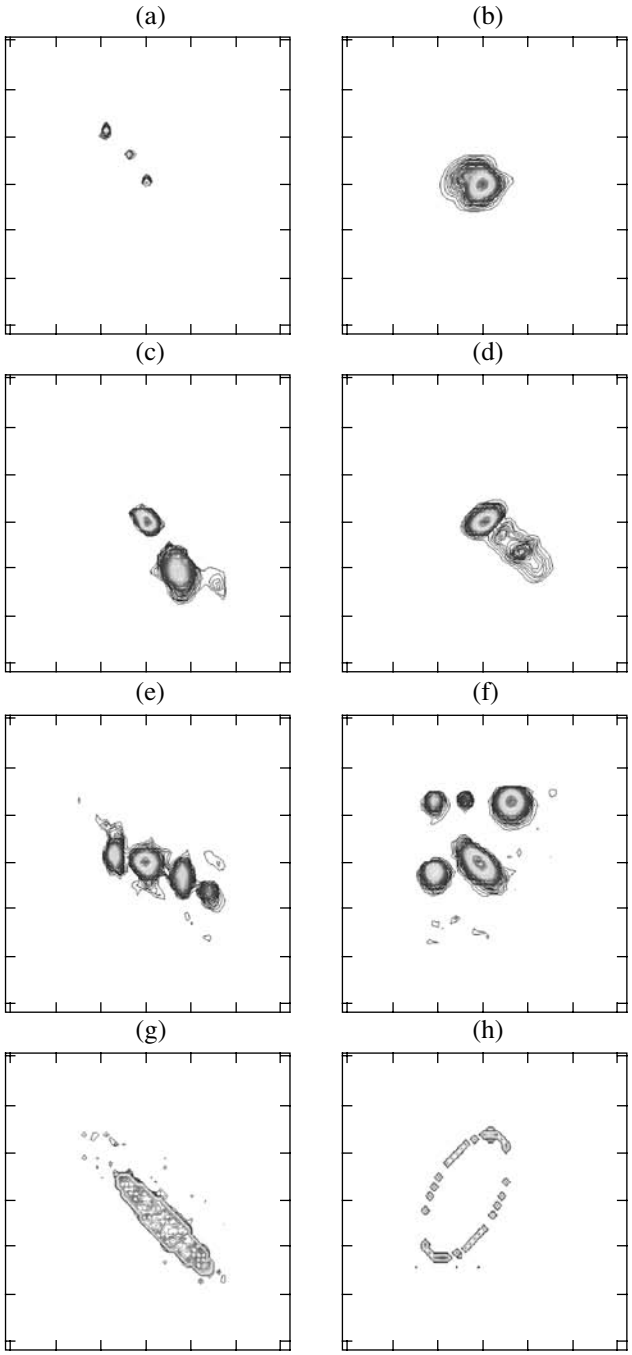

Fig. 3. Images of the model radio sources constructed using the proposed phase-reconstruction method.

internal property of the entropy functional. It is not difficult to show that the Hess matrix is diagonal and has positive elements, which means that it is positive definite everywhere, so that the Lagrange functional (16) is convex and the solution is global, i.e., unique.

However, the uniqueness of the solution in this case refers only to the uniqueness of the reconstruction of the shape of the source. Images obtained as a result of linear shifts or 180° rotations of the solution (17) also satisfy (10)–(14), as follows from the properties of the trigonometric functions determining the desired phase in terms of the variables t_{nk} and s_{nk} [see (8), (9)], and so are likewise solutions of the functional (16). All these solutions comprise a class of equivalent functions that differ by linear shifts or rotations by 180°. Thus, the proposed method for solving

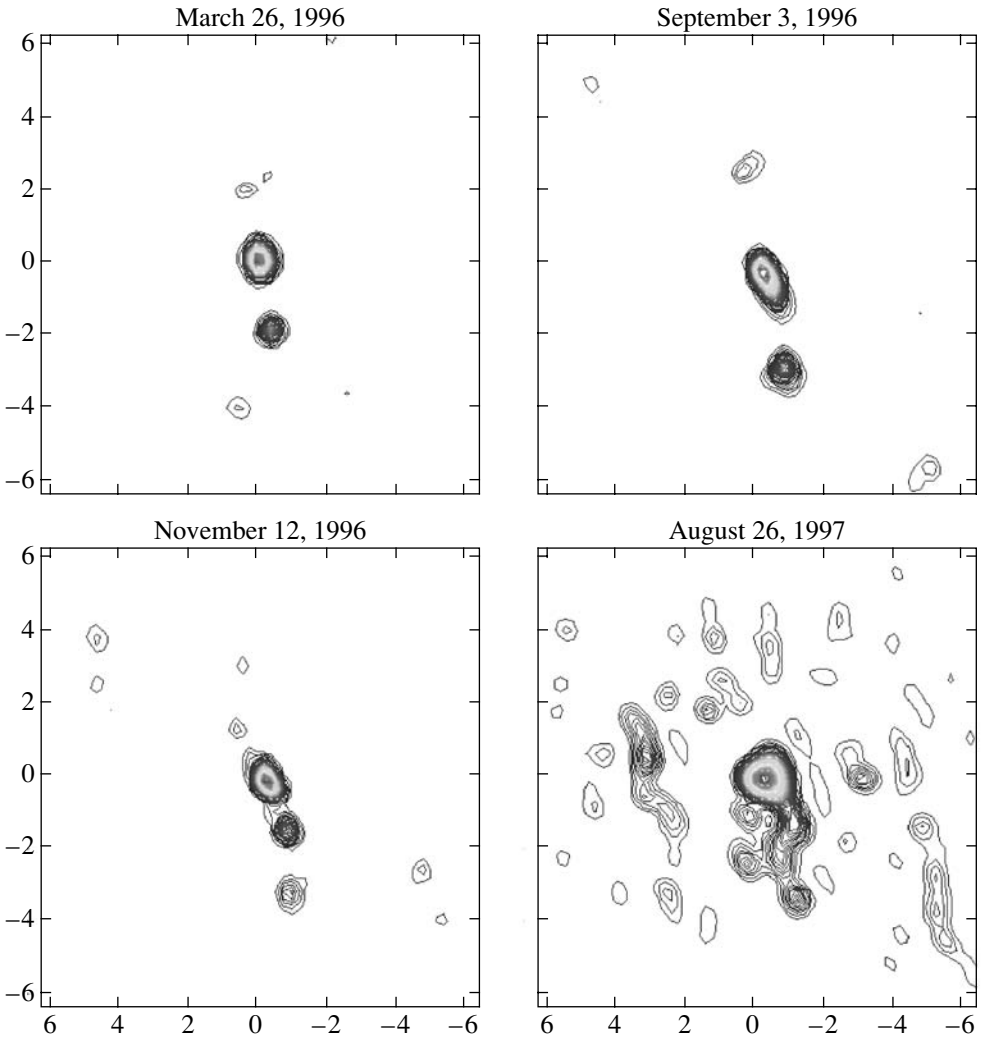

Fig. 4. Results of mapping the radio source 2200+420 using data from a global VLBI array. The scales along the axes are in mas.

the phase problem yields a unique solution with accuracy to within a class of equivalent functions. We used the coordinate-descent method studied in detail in [11] for various functionals to numerically realize the absolute optimization (16).

In addition, note that the proposed mapping method based on the MEM possesses higher stabilizing properties with respect to the noise than, for example, the CLEAN algorithm that is traditionally used in adaptive-calibration methods in VLBI, or the method of Fienup used in [1]. The high stability of the proposed method is due to both the properties of the nonlinear entropy functional and the possibility of including the real signal-to-noise ratio in the χ^2 criterion [see (10)] [14].

4. TESTING THE METHOD

We tested the proposed method using the eight model radio sources shown in Fig. 1. The models and the radio sources themselves will be referred to in accordance with the notation in the figure, using the letters "a–h."

These models reflect the various characters of possible brightness distributions—from a collection of unresolved point sources and various numbers of Gaussian components with various relative positions, to uniform distributions within specified boundaries and rings (see the following section for more detail on these source structures). Note that the lowest contour and the constant step between contours in all images is 1% of the peak of the map.

Figure 2 shows the images reconstructed in the case of zero spectral phase. As was noted in Section 2, this is characteristic of situations in which the phase triangles are completely degenerate. These images were produced using standard mapping methods. We can see that the source structures have taken on a symmetrical form. Mapping methods that enable

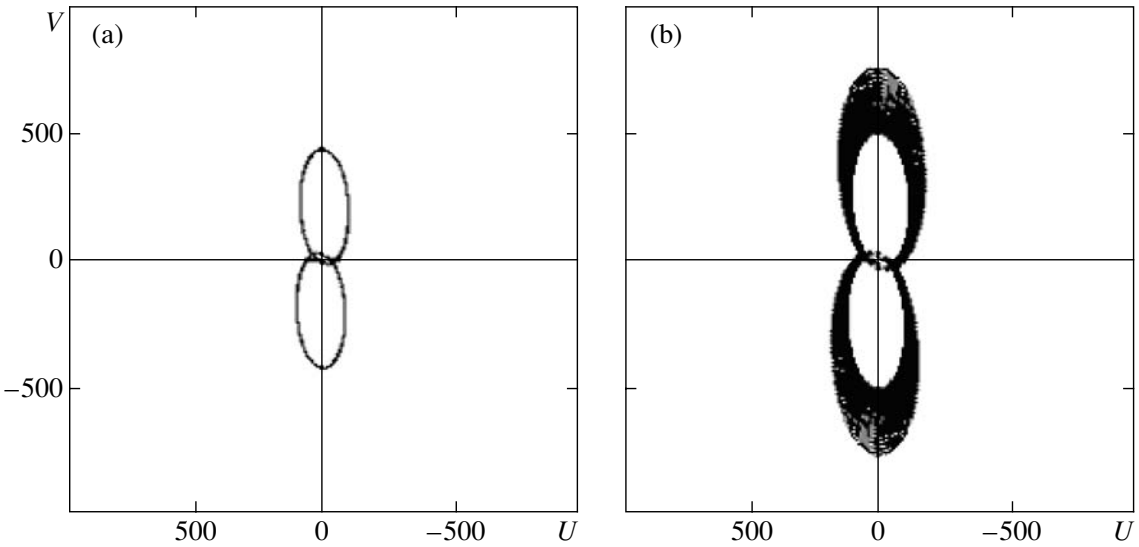

Fig. 5. Simulations of a ground–space interferometer for (a) single-frequency synthesis and (b) multifrequency synthesis. The scales along the axes are in units of 10^8 wavelengths.

reconstruction of the spectral phase are required if we wish to derive the intrinsic structure of the source. Images obtained using our proposed method to reconstruct the phases are presented in Fig. 3.

We took the full set of measurements of the amplitude of the spectrum as the input data. A comparison of the model (Fig. 1) and reconstructed (Fig. 3) maps demonstrates the fairly high internal accuracy of the method, which is from 1% to 4% for these examples. We can see from Fig. 3 that some sources are reconstructed with accuracy to within a linear shift ("g" and "h") or within a rotation by 180° ("c" and "d").

To test the application of the proposed method to real VLBI data, we carried out phaseless mapping of the radio source 2200+420 using data obtained during a global astrometric/geodetic observing program at 8.2 GHz (wavelength 3.5 cm) in 1996–1997. The UV-plane coverages corresponding to these data are presented in [1]. Figure 4 shows images reconstructed with angular resolutions of the order of 0.5–0.7 mas. These maps qualitatively and quantitatively agree with the results obtained in [1] using alternative mapping methods—both phaseless methods based on the algorithm of Fienup and standard methods based on self-calibration. Real geodetic VLBI data are not intended for astrophysical mapping, and therefore do not have the highest quality in terms of calibration and signal-to-noise ratio. Nevertheless, our analysis of maps made from such data shows that the proposed method is able to construct high-quality images with resolutions determined by the geometry of the interferometer.

5. SIMULATIONS OF THE RADIOASTRON MISSION

We will now present the results of simulating mapping radio sources using a ground–space VLBI system with the following parameters.

The ground stations used (the choice is not of fundamental importance here) were the Svetloe, Zelenchuk, and Badary stations of the QUASAR network. The wavelength of the simulated observations was 1.35 cm. In the case of multifrequency synthesis, we used a frequency band with a width equal to 30% of the frequency corresponding to the chosen wavelength. The orbit of the RADIOASTRON spacecraft [8] is inclined to the equatorial plane by 51.5° and has an angle from the Vernal Equinox to the line of nodes of −45°, an angle from the line of nodes to perigee of 30°, and perigee and apogee heights of 20 000 and 350 000 km, respectively. The period of the spacecraft in its orbit around the Earth is 9.5 days.

Figures 5a and 5b depict the coverage of the UV plane obtained for the cases of single-frequency and multifrequency syntheses over a time equal to one period of the spacecraft around the Earth. The spatial frequencies are plotted along the U and V axes in units of 10^8 wavelengths.

Note that, since the bandwidth for the multifrequency synthesis images was taken to be no more than 30% of the central frequency, the effect of the frequency dependence on the images can be neglected. According to the estimates of [15], the influence on the synthesized image due to this effect is no more than 1%, even when the source has a large spectral

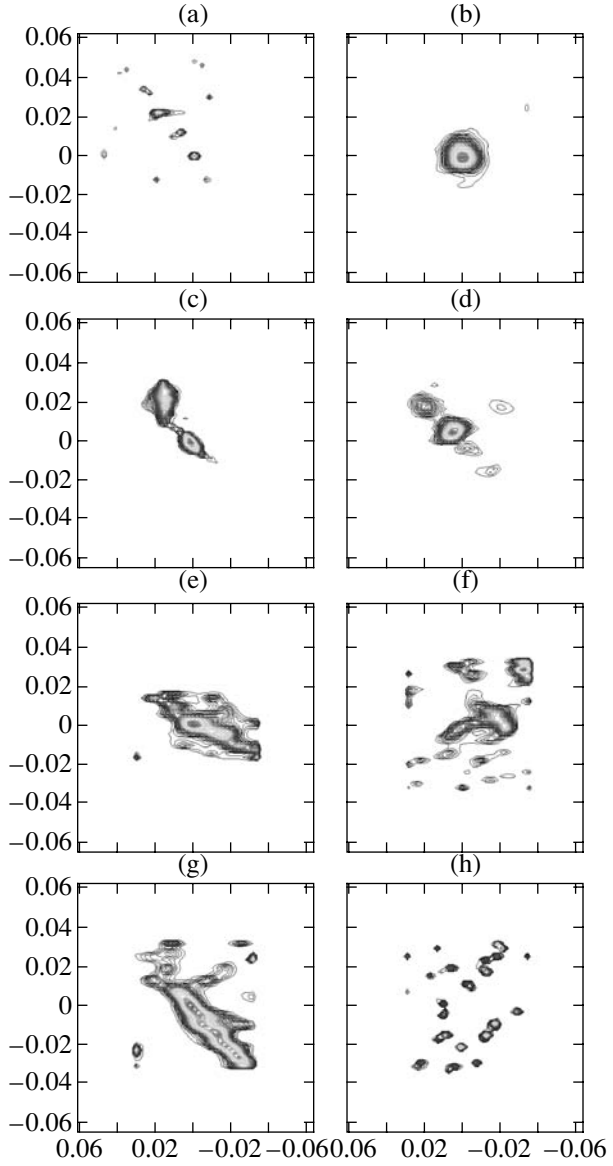

Fig. 6. Images of model radio sources constructed from model single-frequency data for a ground–space interferometer with the parameters of RADIOASTRON. The scales along the axes are in mas.

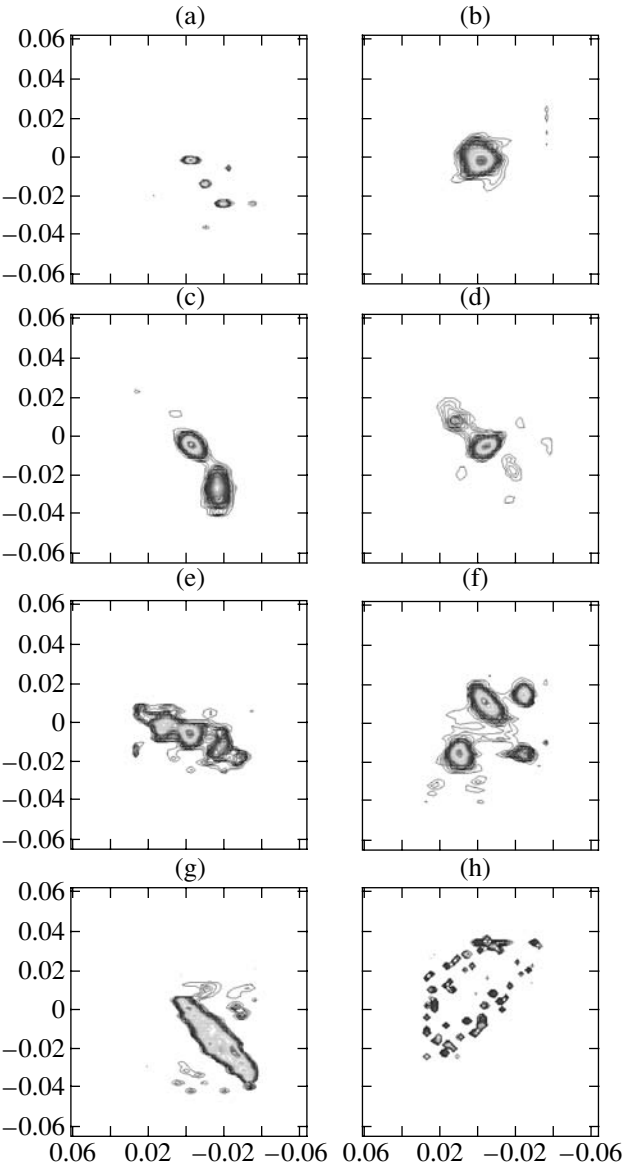

Fig. 7. Same as Fig. 6 but for multifrequency model data.

index. Correcting for the frequency can lower this influence to 0.1%. Our aim here is only to demonstrate the effect of the multifrequency synthesis regime on the coverage of the UV plane, compared to the case of single-frequency synthesis. If a wider band is used for the frequency synthesis, this requires application of frequency corrections, in accordance with the well known algorithms described in [15].

These parameters for the mapping system correspond to a maximum angular resolution of $\lambda/D = 0.007$ mas, where λ is the wavelength and D is the maximum baseline of the ground–space interferometer. The resolution in the east–west direction is roughly half that in the north–south direction. The following simulations will show that, thanks to the application of nonlinear methods, it is possible to reconstruction the images with a resolution higher than that indicated above.

As the model radio sources depicted in Fig. 1, we used simplified Gaussian models derived from maps on mas scales constructed from VLBA [16] data at 8.2 GHz, but translated to the angular scales considered here (tens of microarcsecond).

Thus, the model-source structures presented in Fig. 1 represent (a) a collection of point features corresponding to the three brightest components in the source 1223+395 [17], (b) a core and short curved jet adjacent to the core (a model for the quasar

Characteristics of the measurements and reconstructed images

Map	Peak input noise relative to the measured amplitude of the visibility function	Signal-to-noise ratio in the input data	Peak relative brightness of the map	Entropy of the map	Peak false feature, %
Fig. 1d	–	–	1.00	−28.65	–
Fig. 6d	0.1	11.67	0.93	−31.74	6
Fig. 7d	0.1	10.04	1.06	−33.07	3

0215+015), (c) an extended core and a comparatively bright one-sided, extended jet (a model for the quasar 1607+268), (d) a core and weak, one-sided, extended jet (a model for the quasar 1022+194), (e) a core and two-sided, multicomponent, bright jet (a model for the quasar 0238−084), (f) a collection of several bright and weak Gaussian components in a region about 50 microarcsecond in diameter (a model for the unidentified radio source 0259+121), (g) a uniform brightness distribution within specified boundaries, and (h) a thin ring. These two last source structures ("g" and "h") are the most exotic and demanding in terms of obtaining acceptable images by extrapolating the spectral data to the high-frequency domain, and are considered here in order to demonstrate the limits of the proposed reconstruction method.

Figures 6 and 7 present the images reconstructed based on the amplitude data for both the single-frequency and multifrequency syntheses. Note that we added to the data uniform noise in the range ±10% of each measurement of the visibility function, which yielded an input signal-to-noise ratio for the algorithm of about ten, in agreement with the sensitivity parameters for the RADIOASTRON mission presented in [18].

We give here a qualitative analysis of the reconstruction results. We will not take into consideration rotations of the images by 180°. We will refer to single-frequency and multifrequency syntheses as the first and second cases, respectively.

Source "a." The quality of the reconstruction was slightly worse in the first than in the second case. We can see several false point features (artefacts), but their brightnesses are appreciably lower than the brightnesses of the correctly reconstructed features (<4% of the peak value). More than the other distortions, the false features that are located along a line passing through the center of the source manifest the incomplete reconstruction of the spectral phase. In the second case, these artefacts are not present. In the first case, the shape of the source and the amplitude ratios of the features are somewhat distorted, while it was possible to reconstruct the correct amplitude ratios in the second case. The brightnesses of the small number of point artefacts in the second case is no more than 1–2%. The point features were reconstructed with finite resolution, but with a resolution that is higher than that of the array used. The resolution in the north–south direction is about twice that in the east–west direction. In both cases, the coordinates of the source components were accurately reconstructed. This simulation is useful because it clearly demonstrates the resolving capability of the reconstruction method.

Source "b." In both cases, it was possible to reconstruct the core well, but the adjacent jet much more poorly; in the second case, however, the jet is manifest more distinctly. In the first case, the features are elongated in the upward direction (in amplitude), making the weak, extended feature virtually invisible. This can be explained by the redistribution of the high-frequency components of the spectrum as a result of the extrapolation of the data belonging to a limited number of tracks in the UV plane [1]. The larger the region occupied by the data, the more exact the values of the extrapolated spectrum. The level of false extended features in the maps does not exceed 2%.

Source "c." As in the previous example, the Gaussian components are elongated in amplitude in the first case, likewise due to the redistribution of the spectrum at high spatial frequencies. The reconstruction of the shapes of the components and their brightness ratios is more accurate in the second case. The level of false features does not exceed 1%.

Source "d." A fundamental difference of this source structure from that considered in the previous example is that the brightness of the jet is appreciably lower than that of the core. In this case, the reconstruction of the image is more complex, since the spectral phases are expressed less strongly. This was visible in the results of the simulations, which are presented in the table (as an example of the most typical type of radio source). Compared to the previous simulation, the incomplete reconstruction of the phases led to a growth in the brightnesses of the false components, which could be taken to be a counterjet. In the second case, however, the quality of the image is much better; the shapes and amplitude ratios of the components are reconstructed

more accurately, and the level of artefacts is reduced by a factor of two.

Source "e." In the first case, the direction of the jet is reconstructed well, but individual components in the jet are expressed appreciably worse (due to the lower resolution in the east–west direction) than in the model, although they are nonetheless resolved. The maximum level of extended false features is about 10%. In the second case, both the shapes (although there is some elongation) and amplitude ratios of the components are reconstructed well. The artefacts occupy fewer regions, and their level does not exceed 5%.

Source "f." In these simulations, we have approximately the same relative qualities of the maps obtained for the single-frequency and multifrequency synthesis data as in the previous example. In the first case, there is some smearing near adjacent components due to the lower resolution in the east–west direction. All the components except for the weakest were reconstructed fairly well in the second case, although we can also see some elongation of their shapes. The weakest pointlike component was not able to be reconstructed in either case, reflecting the comparatively low dynamical range of the resulting maps.

Source "g." In both cases, we have a fairly good reconstruction of the boundary of the source, although there are three fairly bright artefacts (about 8% of the peak brightness) adjacent to the outer boundary in the first case. The resulting brightness distribution inside the boundary is not uniform. The brightness of artefacts is appreciably lower in the second case; they do not press up against the boundaries of the source and they have a pointlike character. The brightness distribution inside the source is substantially more uniform.

Source "h." The shape of the ring is reconstructed more distinctly and with a larger number of features in the second case. In both cases, there are pointlike artefacts adjacent to the ring, but the shape of the ring is clearly traced.

Thus, our analysis of the results of each simulation demonstrates that we were able to obtain an acceptable reconstruction, even in the case of poor UV-plane coverage (the single-frequency synthesis), at least for those sources with comparatively simple core + jet structures (such as sources "c," "d"). The multifrequency synthesis provided appreciably more accurate reconstructions than the single-frequency synthesis, with this difference being more pronounced as the source structure became more complex (see the models for sources "e"–"h"). In all eight of the simulations, the multifrequency synthesis enabled more accurate reconstruction of the shapes and amplitude ratios of the components, provided higher resolution, and lowered the level of artefacts.

Analysis of the results (Figs. 6 and 7) shows that the reconstructed maps have a resolution exceeding that of the instrument, especially in the case of multifrequency synthesis (the maps of sources "e" and "f"), although a ratio of 2 : 1 is preserved for the resolution in the north–south and east–west directions. The achievement of higher resolution is due to the fact that the image-reconstruction method is based on the substantially nonlinear maximum-entropy method, which has the property of super-resolution; i.e., the ability to trace the spectrum outside the aperture of the instrument. This frequency extrapolation is more accurate in the case of multifrequency synthesis, due to the more complete coverage of the UV plane.

If we convolve the solutions obtained (Figs. 6 and 7) with a "clean" beam corresponding to half the width of the main lobe of the synthesized "dirty" beam for the system, in order to use the most reliably reconstructed part of the spectrum and to avoid the effect of overdetermination of the spectrum, on average, we obtain maps with angular resolution of the order of 0.01 mas.

Thus, when measuring the visibility function with a signal-to-noise ratio of about ten using a radio interferometer with the geometry specified for the RADIOASTRON mission, we can achieve a resolution of no worse than 0.010 mas at 1.35 cm. At lower signal-to-noise ratios, an appreciable role in the optimized functional (10) begins to be played by the regularizing term χ^2, which can lead to some loss of resolution. Estimates show that this loss will constitute a factor of about 1.5 for a signal-to-noise ratio of about three.

It follows that, when used together with multifrequency synthesis, the proposed method for phaseless mapping can be successfully applied to obtain maps of good quality on angular scales of the order of tens of mas for data obtained on the RADIOASTRON ground–space interferometer, when standard methods of adaptive calibration cannot be used due to the degeneracy of the closure-phase equations.

6. CONCLUSION

The task of phaseless VLBI mapping is especially topical now, in connection with future radio interferometers that may include a space antenna in a high orbit whose apogee exceeds the radius of the Earth by a factor of several dozen or more. In this case, the closure-phase relations become degenerate, making application of the standard methods of adaptive calibration incorrect. Our results on simulations for the RADIOASTRON ground–space radio interferometer, designed for imaging with ultra-high

angular resolution, demonstrate the ability to achieve acceptable image reconstructions when our proposed phaseless mapping methods are used together with multifrequency synthesis techniques.

ACKNOWLEDGMENTS

The author is grateful to N.S. Kardashev and A.V. Stepanov for support of this work. This work was partially supported by the Basic Research Program of the Presidium of the Russian Academy of Sciences "Nonstationary Phenomena in Astronomy."

REFERENCES

1. A. T. Baikova, Pis'ma Astron. Zh. **30**, 253 (2004) [Astron. Lett. **30**, 218 (2004)].
2. T. J. Cornwell and E. B. Fomalont, in *Synthesis Imaging in Radio Astronomy II. A Collection of Lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School,* Ed. by G. B. Taylor, C. L. Carilli, and R. A. Perley (Astron. Soc. Pac., San Francisco, 1999); Astron. Soc. Pac. Conf. Ser. **180**, 187 (1999).
3. N. S. Kardashev, Exp. Astron. **7**, 329 (1997).
4. Yu. M. Bruck and L. G. Sodin, Opt. Commun. **30**, 304 (1979).
5. J. R. Fienup, Opt. Lett. **3**, 27 (1978).
6. J. R. Fienup, Appl. Opt. **21**, 2758 (1982).
7. G. I. Vasilenko and A. M. Taratorin, *Image Restoration* (Radio i Svyaz', Moscow, 1986) [in Russian].
8. *RADIOASTRON (Space VLBI Project)*, http://www.asc.rssi.ru/radioastron/description/orbit_eng.htm.
9. *RADIOASTRON (Space VLBI Project)*, http://www.asc.rssi.ru/radioastron/description/mfs_eng.htm.
10. A. T. Bajkova, Preprint No. 24, IPA AN SSSR (Inst. Appl. Astron., USSR Acad. Sci., Moscow, 1990).
11. A. T. Bajkova, Preprint No. 58, IPA AN (Inst. Appl. Astron., Russian Acad. Sci., Moscow, 1993).
12. B. R. Frieden, J. Opt. Soc. Am. **62**, 511 (1972).
13. B. R. Frieden and A. T. Bajkova, Appl. Opt. **34**, 4086 (1995).
14. T. J. Cornwell, R. Braun, and D. S. Briggs, in *Synthesis Imaging in Radio Astronomy II. A Collection of Lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School,* Ed. by G. B. Taylor, C. L. Carilli, and R. A. Perley (Astron. Soc. Pac., San Francisco, 1999); Astron. Soc. Pac. Conf. Ser. **180**, 151 (1999).
15. R. J. Sault and J. E. Conway, in *Synthesis Imaging in Radio Astronomy II. A Collection of Lectures from the Sixth NRAO/NMIMT Synthesis Imaging Summer School,* Ed. by G. B. Taylor, C. L. Carilli, and R. A. Perley (Astron. Soc. Pac., San Francisco, 1999); Astron. Soc. Pac. Conf. Ser. **180**, 419 (1999).
16. C. Ma and M. Feissel, IERS Technical Note No. 23 (1997).
17. D. R. Henstock, I. W. Browne, P. N. Wilkinson, *et al.*, Astrophys. J., Suppl. Ser. **100**, 1 (1995).
18. *RADIOASTRON (Space VLBI Project)*, http://www.asc.rssi.ru/radioastron/description/observations_eng.htm.

Translated by D. Gabuzda

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 984–992. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1099–1108.

Atmospheric Chemical Abundances of the Components of the Quadruple System ADS 11061. 40 Draconis

Yu. Yu. Balega[1], V. V. Leushin[1], and M. K. Kuznetsov[2]

[1] *Special Astrophysical Observatory, Russian Academy of Sciences, Nizhniĭ Arkhyz, Karachaĭ-Cherkessian Republic, 357147 Russia*

[2] *Rostov State University, ul. Bol'shaya Sadovaya 105, Rostov-on-Don, 334006 Russia*

Received November 22, 2004; in final form, February 17, 2005

Abstract—As part of our study of the components of the hierarchic quadruple system ADS 11061, we acquired spectroscopic observations of the binary 40 Dra. Echelle spectra showing the separation of the components' lines were obtained in the spectral range 3700–9200 Å. Effective temperatures and surface gravities were derived for the components from BV photometry and the hydrogen-line profiles. The components of the 40 Dra system have parameters close to $T^a_{\rm eff} = 6420$ K, $\log g_a = 4.17$, $T^b_{\rm eff} = 6300$ K, and $\log g_b = 4.20$. We find the microturbulence velocity in the component atmospheres to be $V_t = 2.6$ km/s. The abundances of iron, carbon, nitrogen, and oxygen in the atmospheres of both components are estimated to be $\log N(\mathrm{Fe})^a = 7.50$, $\log N(\mathrm{Fe})^b = 7.46$, $\log N(\mathrm{C})^a = 8.39$, $\log N(\mathrm{C})^b = 8.45$, $\log N(\mathrm{N})^a = 8.12$, $\log(\mathrm{N})^b = 8.15$, $\log N(\mathrm{O})^a = 8.77$, $\log N(\mathrm{O})^b = 8.74$. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The spectroscopic binary 40 Dra in the hierarchic quadruple system ADS 11061 has a period of about 10 days and an orbital eccentricity of $e = 0.3$ [1]. The distance between its components varies during one period between 26.25×10^6 km at periastron and 48.75×10^6 km at apoastron (the orbital semimajor axis is 0.125 AU [2]). Since the diameters of the components are nearly equal (about 1.1×10^6 km, comparable to the component separation), the effects of the companion's gravitational attraction on each star's internal structure and evolutionary status should be significant. All the members of the ADS 11061 system have similar spectral types, F5–F7, and fundamental parameters, but form pairs with strongly different orbital parameters. This could be due to evolutionary changes of the system as a whole, as well as to specific features of each of the components of the ADS 11061 system.

We analyzed the atmospheres of the components of the binary 41 Dra in our earlier paper [3]; here, we analyze the atmospheres of the components of 40 Dra. Using high-resolution echelle spectra, we have computed model atmospheres for each component, and derived effective temperatures, surface gravities, microturbulence velocities, and the abundances of iron, carbon, nitrogen, and oxygen in the stellar atmospheres.

2. OBSERVATIONS

Table 1 presents the dates when the spectra of the binary 40 Dra were taken. The spectra vz4935, vz4936, and vz4937 were obtained at the Coudé focus of the 1 m Zeiss telescope of the Special Astrophysical Observatory (SAO) of the Russian Academy of Sciences using the CEGS echelle spectrometer [4]. Each spectrum has 41 orders, each containing 1242 pixels. The spectra covers the range $\lambda\lambda$7120–9160 Å; and the resolution was 40 000 and the signal-to-noise ratio 100. The rest of the spectra were taken with the Zeiss-2000 telescope of the International Center for Astronomical, Medical and Ecological Research (ICAMER) at Peak Terskol [5].

Figure 1 shows fragments of the studied spectra near the Hα line. The lines of metals are well separated at all phases, as is demonstrated by the example of the FeI λ6546.239 line. In most of the spectra, the doubling of the hydrogen line is revealed by the asymmetry of its profile and, at phases 0.822 and 0.920, by the splitting of the line's central region. The lines of atmospheric water (e.g., H_2O λ6552.629) are clearly visible in the SAO spectra, but barely visible in the spectra taken at Peak Terskol (the SAO and Peak Terskol telescopes are at altitudes of 2100 m and 3100 m, respectively). We reduced the spectra using Galazutdinov's code DECH20T [6]. This code was also used to calculate radial velocities and reduce them to the Sun. The radial velocities measured with respect to the instrument ($V^a_{\rm r}$ and $V^b_{\rm r}$)

1063-7729/05/4912-0984$26.00

Table 1. Log of spectra and radial velocities of the components of 40 Dra

Spectrum	Date	JD	Phase	V_r^a, km/s	V_r^b, km/s	V_a, km/s
t05413	21.01.2001	2451930.503	0.342	29.6 ± 1.5	-24.1 ± 1.8	-3.5
t05511	26.01.2001	2451935.578	0.822	-31.4 ± 1.2	40.3 ± 1.4	-4.1
t05614	27.01.2001	2451936.610	0.920	-36.3 ± 1.2	58.4 ± 1.6	-4.2
vz4935	8.06.2001	2452068.541	0.451	24.9 ± 1.5	-1.5 ± 1.8	-1.6
vz4936	8.06.2001	2452068.551	0.452	23.7 ± 1.2	-0.2 ± 1.5	-1.6
vz4937	8.06.2001	2452068.559	0.453	23.7 ± 1.3	1.1 ± 1.8	-1.6
t11404	19.11.2002	2452598.062	0.749	-24.3 ± 1.5	20.4 ± 1.5	3.7

and the values reduced to the Sun (V_a) are given in Table 1. We used the elements $T = \text{JD } 2452053.260$ and $P = 10.52785$ days to calculate the phases.

3. PARAMETERS OF THE ORBIT AND COMPONENTS OF 40 DRA

The first orbital elements for 40 Dra were published by Boothroyd [1] in 1920; the system was later studied in detail by Tokovinin [2] and other authors [7, 8]. The data from these papers are collected in Table 2. These observations indicate the presence of apsidal motion in the system, with this motion's rate in the interval between 1918 (JD 2421764.648) and 1990 (JD 2448000.008), which covers 2492 cycles, the same as in the interval between 1990 and 2001 (385 cycles), namely 0.0044° per one orbit. The apsidal rotation testifies that the spherical shape of the system's components is distorted by their gravitational interaction. This should lead to brightness variations of 40 Dra, as are suspected in the catalog [9].

4. EFFECTIVE TEMPERATURES OF THE COMPONENTS OF 40 DRA

The spectroscopic binary 40 Dra is a close binary system that cannot be resolved by means of speckle interferometry, and the parameters of the components can be analyzed only using high-resolution spectroscopy. Our echelle spectra, which cover the spectral region between 3500 and 9200 Å, enable reliable determination of the component brightness ratio as a function of the wavelength. To estimate this ratio, we measured the depths of each component's resolved, unblended lines. In the case of the 40 Dra system, the ratio of these depths is equal to the component brightness ratio, since the two stars have similar spectral types. Figure 2 is the wavelength dependence of the corresponding ratios, $E^b/E^a = R^{b,\mathrm{obs}}/R^{a,\mathrm{obs}}$.

The linear regression relation yielded by our observations is

$$E^a_\lambda/E^b_\lambda = 1.820 - 0.376 \times 10^{-5}\lambda[\text{Å}].$$

Thus, the brightness difference in the B band

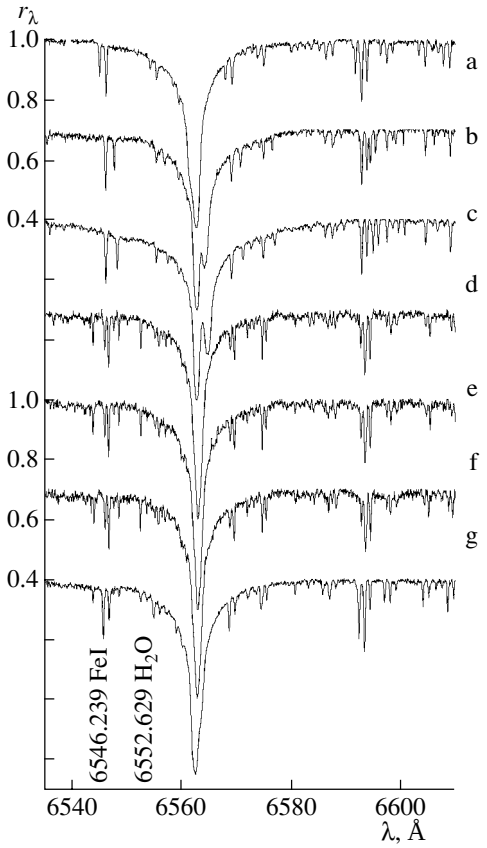

Fig. 1. Fragments of the observed spectra of 40 Dra near Hα for phases (a) 0.342, (b) 0.822, (c) 0.920, (d) 0.451, (e) 0.452, (f) 0.453, and (g) 0.749. All the spectra clearly show the splitting of metal lines into components (for example, for the FeI λ6546.239 line). Atmospheric-water lines (H_2O λ6552.629) are apparent in the SAO spectra (d, e, f) and are virtually invisible in the Peak Terskol spectra (a, b, c, g).

Table 2. Parameters of the orbit and components of 40 Dra

Parameter	Boothroyd [1]	Tokovinin [2]	Al-Wardat [7]	Tokovinin *et al.* [8]
P, days	10.5217	10.52785 ± 0.00007		10.52785
T, JD	2421764.648 ± 0.15	2448000.008 ± 0.011		2452053.260 ± 0.040
e	0.310 ± 0.015	0.374 ± 0.003		0.380 ± 0.008
a, AU	0.156	0.125		0.125
ω, deg	256.8	246.2 ± 0.4		244.35 ± 1.33
$\imath$, deg		35		35
π, arcseconds		0.028–0.025	0.023 ± 0.002	0.0224 ± 0.0005
Δm		0.79	0.51 ± 0.03	0.56 ± 0.03
q	0.897	0.912		0.910
$(M^a + M^b)\sin^3 i, M_\odot$	0.87	0.480		0.480
$M^a/M_\odot$			1.315	1.32
$M^b/M_\odot$			1.229	1.20

(λ4400 Å) is $\Delta m^{B,\mathrm{obs}} = 0.528^m$, and in the V band (λ5500 Å) is $\Delta m^{V,\mathrm{obs}} = 0.496^m$.

The system's mean combined color index is $B - V = 0.510 \pm 0.001$ [10–12].

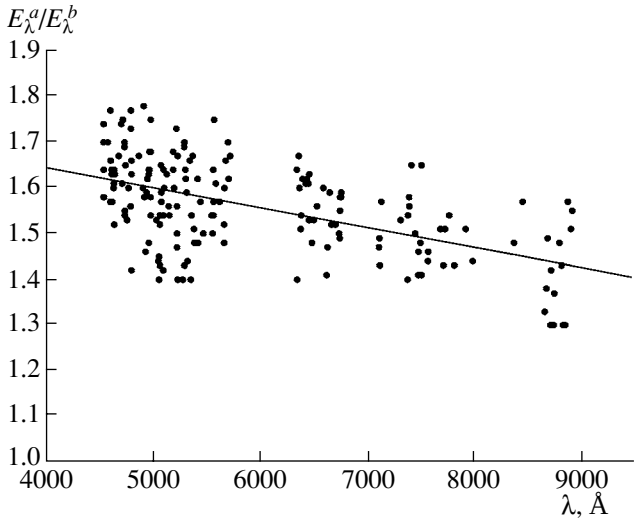

Fig. 2. Wavelength dependence of the component brightness ratio, $E^a/E^b = R^{a,\mathrm{obs}}/R^{b,\mathrm{obs}}$, derived from measurements of line depths for each of the components of 40 Dra in spectra with well separated spectral lines of metals. The straight line corresponds to the linear regression fit $E^a_\lambda/E^b_\lambda = 1.820 - 4.376 \times 10^{-5}\lambda[\text{Å}]$.

Thus, for component a:

$$(B-V)^a = B - V + 2.5\log\frac{1+10^{-0.4\Delta m^{B,\mathrm{obs}}}}{1+10^{-0.4\Delta m^{V,\mathrm{obs}}}} = 0.498,$$

and for component b:

$$(B-V)^b = (B-V)^a + (\Delta m^{B,\mathrm{obs}} - \Delta m^{V,\mathrm{obs}}) = 0.530.$$

The effective temperatures, estimated for 40 Dra using computations from Kurucz [13] for models with a normal chemical composition and an atmospheric turbulence velocity of 2 km/s are:

$$T^a_{\mathrm{eff}} = 6420\ \mathrm{K}, \quad T^b_{\mathrm{eff}} = 6300\ \mathrm{K}.$$

We can derive characteristics for each of the components separately using the mean value $V = 6.043$ [10–12] for the combined brightness of 40 Dra, the parallax from [8] ($d = 44.6$ pc), the bolometric correction for F7 stars (B.C. $= -0.02$), and the components' brightness difference:

Component a

$$V^a = V + 2.5\log(1 + 10^{-0.4\Delta m^{V,\mathrm{obs}}}) = 6.576^m,$$
$$M^a_{\mathrm{bol}} = 5 + V^a - 5\log d + \mathrm{B.C.} = 3.309^m,$$
$$\log(L/L_\odot)^a = (4.72 - M^a_{\mathrm{bol}})/2.5 = 0.564,$$
$$\log(R^a/R_\odot) = (\log(L/L_\odot)^a - 4\log(T/T_\odot))/2 = 0.191;$$

Component b

$$V^b = V^a + \Delta m^{V,\mathrm{obs}} = 7.072^m,$$

$$M^b_{\mathrm{bol}} = 3.825^m,$$

$$\log(L/L_\odot)^b = 0.366,$$

$$\log(R^b/R_\odot) = 0.108.$$

With the derived radii of the components and their estimated masses from [8], we can also determine the components' surface gravities:

$$\log g^a = 4.17, \qquad \log g^b = 4.29.$$

5. HYDROGEN LINE PROFILES

In order to test the correctness of the atmospheric parameters determined for the components of 40 Dra, we computed a combined synthetic spectrum for the system near the Hα line. We summed the derived theoretical profiles using Tsimbal's SintVa code [14], with the radial-velocity shifts taken into account. The radial velocities of the components in our spectra are given in Table 1. The theoretical profiles were found for several sets of effective temperatures and surface gravities for the component model atmospheres. The microturbulence and rotational velocities were derived from our analysis of the metallic lines. The rotational and microturbulence velocities for the two components are similar, $V \sin i = 6.5$ km/s and $V^a_t = V^b_t = 2.60$ km/s. Figure 3 compares the calculated theoretical profiles with the observed profile for phase 0.920. The theoretical synthetic spectra were obtained for three sets of models:

(1) component a: $T_{\mathrm{eff}} = 6500$ K and $\log g = 4.00$, component b: $T_{\mathrm{eff}} = 6500$ K and $\log g = 4.50$;

(2) component a: $T_{\mathrm{eff}} = 6420$ K and $\log g = 4.17$, component b: $T_{\mathrm{eff}} = 6300$ K and $\log g = 4.29$;

(3) component a: $T_{\mathrm{eff}} = 6250$ K and $\log g = 4.00$, component b: $T_{\mathrm{eff}} = 6250$ K and $\log g = 4.50$.

When summing the spectra, we assumed, in agreement with our derived brightness difference, that the component luminosity ratio was $L^b/L^a = 0.64$. The agreement between the theoretical and observed profiles is generally good, and is consistent with our estimates of the components' effective temperatures and surface gravities. Unfortunately, the sensitivity of the line profiles to the parameters is weak in the temperature range of our stars, and corresponds to uncertainty in the temperature of ± 250 K and in $\log g$ of ± 0.25 for our spectra. In addition, note that the existing deviations between the observed and theoretical profiles are within several hundredths in r_λ, and could be due to the fact that our spherically symmetric models are not adequate descriptions of the real stars, whose shapes are distorted by their companion. Nevertheless, the chosen parameters describe the observations satisfactorily, within the uncertainties in our spectra.

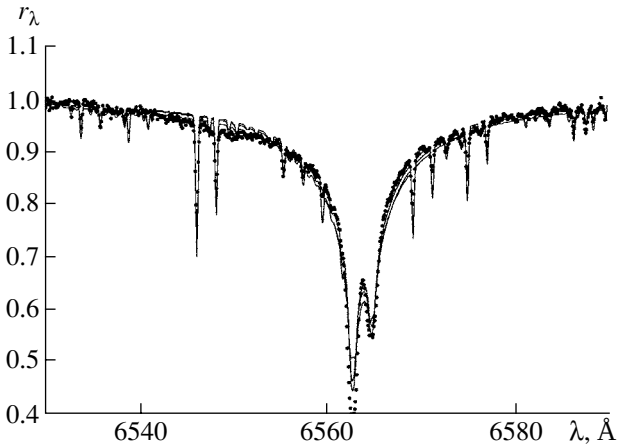

Fig. 3. Comparison between the observed spectrum of the 40 Dra system at phase 0.920 (dots) and the theoretical synthetic spectra (solid curves). The top spectrum was calculated for component model atmospheres with the parameters $T^a_{\mathrm{eff}} = 6500$ K, $\log g^a = 4.00$ and $T^b_{\mathrm{eff}} = 6500$ K, $\log g^b = 4.50$; the middle spectrum for $T^a_{\mathrm{eff}} = 6420$ K, $\log g^a = 4.17$ and $T^b_{\mathrm{eff}} = 6300$ K, $\log g^b = 4.29$; and the bottom spectrum for $T^a_{\mathrm{eff}} = 6250$ K, $\log g^a = 4.00$ and $T^b_{\mathrm{eff}} = 6250$ K, $\log g^b = 4.50$.

6. IRON ABUNDANCES AND MICROTURBULENT VELOCITIES IN THE ATMOSPHERES OF THE COMPONENTS OF 40 Dra

To determine the iron abundance, we measured the equivalent widths of 76 FeI lines and 16 FeII lines with $\lambda > 5260$ Å, where the line blending is comparatively weak. The lines of components a and b can be separated fairly easily. The equivalent widths in the spectra of each of the components are determined by the component brightness ratio, $l = E^b_\lambda/E^a_\lambda$, and their observed values are described by the relations $W^a_\lambda = W^a_{\lambda,\mathrm{obs}}(1+l) = 1.64 W^a_{\lambda,\mathrm{obs}}$ for component a and $W^b_\lambda = W^b_{\lambda,\mathrm{obs}}(1+l)/l = 2.56 W^b_{\lambda,\mathrm{obs}}$ for component b. The equivalent widths of the FeI lines, along with the atomic parameters and the values of $\log N(\mathrm{Fe})$ derived for $V_t = 2.6$ km/s for each, are collected in Table 3.

Our computations made use of the KONTUR code [15] and were carried out for models with $T_{\mathrm{eff}} = 6420$ K and $\log g = 4.17$ for component a and $T_{\mathrm{eff}} = 6300$ K and $\log g = 4.29$ for component b, for various microturbulence velocities. The resulting

Table 3. Line parameters and FeI equivalent widths and iron abundances from each line in the atmospheres of components *a* and *b* of 40 Dra

λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{Fe})^a$	W_λ^b, mÅ	$\log N(\mathrm{Fe})^b$	λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{Fe})^a$	W_λ^b, mÅ	$\log N(\mathrm{Fe})^b$
5263.31	3.27	−0.88	114.5	7.27	173.3	7.90	5266.56	3.00	−0.39	190.4	7.57	–	–
5269.54	0.86	−1.62	363.7	7.98	365.4	7.80	5273.16	3.29	−0.99	93.8	7.09	–	–
5273.37	2.48	−2.16	100.8	7.63	–	–	5281.79	3.04	−0.83	135.4	7.34	133.4	7.19
5283.62	3.24	−0.43	175.9	7.61	154.3	7.25	5285.13	4.44	−1.64	42.1	7.89	50.2	7.97
5288.53	3.70	−1.51	49.0	7.25	60.4	7.37	5321.11	4.34	−0.95	61.5	7.47	–	–
5322.04	2.28	−2.80	49.2	7.31	–	–	5324.18	3.21	−0.10	182.8	7.38	201.1	7.43
5328.04	0.92	−1.47	287.0	7.63	239.2	7.15	5328.53	1.56	−1.85	168.8	7.56	156.6	7.24
5339.93	3.27	−0.65	129.6	7.26	136.7	7.25	5341.02	1.61	−1.95	154.3	7.50	–	–
5349.74	4.39	−1.30	44.1	7.61	30.3	7.23	5353.37	4.10	−0.84	97.7	7.67	85.4	7.41
5379.57	3.70	−1.51	78.0	7.73	70.4	7.54	5383.377	4.31	0.64	156.6	7.13	178.3	7.25
5389.48	4.42	−0.41	89.2	7.38	82.6	7.22	5391.46	4.15	−0.82	84.2	7.50	117.6	7.88
5393.17	3.24	−0.72	122.4	7.19	126.0	7.14	5394.68	4.19	−1.62	30.8	7.54	–	–
5397.13	0.92	−1.99	220.5	7.73	202.4	7.42	5397.19	4.45	−2.34	5.5	7.56	–	–
5398.28	4.45	−0.67	95.9	7.76	–	–	5400.50	4.37	−0.16	131.4	7.66	151.7	7.78
5403.82	4.08	−1.03	70.2	7.44	72.2	7.40	5404.11	4.31	0.22	200.3	7.92	163.2	7.54
5405.35	4.39	−1.39	48.2	7.75	–	–	5405.77	0.99	−1.84	187.7	7.30	221.6	7.52
6355.04	2.84	−2.35	59.1	7.51	51.0	7.24	6358.69	4.14	−1.66	55.1	7.87	61.7	7.91
6364.37	4.79	−1.43	37.5	7.87	–	–	6380.75	4.19	−1.34	54.5	7.61	48.2	7.45
6393.61	2.43	−1.43	127.0	7.18	145.8	7.34	6400.01	3.60	−0.29	176.0	7.66	157.3	7.38
6408.03	3.69	−1.02	96.7	7.43	89.2	7.25	6411.66	3.65	−0.60	108.9	7.17	123.8	7.29
6419.96	4.73	−0.24	77.0	7.27	110.3	7.65	6421.36	2.28	−2.03	108.6	7.36	124.5	7.48
6430.86	2.18	−2.01	98.4	7.12	100.9	7.05	6469.19	4.84	−0.77	43.7	7.42	68.2	7.73
6475.63	2.56	−2.94	57.4	7.80	49.0	7.61	6481.88	2.28	−2.98	47.7	7.47	51.8	7.41
6494.99	2.40	−1.27	175.1	7.63	–	–	6495.74	4.84	−0.94	24.0	7.11	–	–
6496.47	4.80	−0.57	50.2	7.24	–	–	6498.95	0.96	−4.70	21.2	7.54	19.1	7.29
6533.94	4.56	−1.46	21.4	7.46	–	–	6546.24	2.76	−1.54	88.5	7.03	98.8	7.07
6569.22	4.73	−0.42	88.5	7.62	109.9	7.82	6575.04	2.59	−2.71	53.0	7.55	55.3	7.49
6593.88	2.43	−2.42	79.8	7.46	–	–	6597.57	4.80	−1.07	50.5	7.78	–	–
6609.12	2.56	−2.69	49.4	7.43	–	–	6625.04	1.01	−5.35	10.5	7.72	–	–
6627.56	4.55	−1.68	21.6	7.62	26.3	7.67	6633.43	4.84	−1.49	24.2	7.70	16.3	7.53
6633.43	4.56	−0.80	71.4	7.62	–	–	6639.72	4.61	−1.72	32.2	7.88	–	–
6653.91	4.15	−2.52	9.4	7.64	–	–	6663.25	4.56	−1.13	45.5	7.58	36.5	7.32
6663.44	2.42	−2.48	52.2	7.08	–	–	6678.00	2.69	−1.42	118.9	7.26	–	–
6705.11	4.61	−1.50	31.4	7.70	–	–	6713.75	4.80	−1.60	–	–	13.3	7.52
6715.39	4.61	−1.64	31.9	7.81	18.6	7.54	6726.67	4.10	−0.83	68.9	7.18	69.5	7.12
6737.98	4.56	−1.75	25.3	7.74	28.1	7.75	6739.52	1.56	−4.79	4.4	7.38	–	–
6750.16	2.42	−2.62	66.8	7.48	–	–	6752.72	4.64	−1.20	27.0	7.33	26.8	7.24
6786.86	4.19	−2.07	17.9	7.60	16.3	7.54	6806.86	2.75	−3.21	37.0	7.85	28.1	7.64
6810.27	4.61	−0.97	49.4	7.52	48.7	7.43							

Table 4. Coefficients in the relation between relative iron abundance and line equivalent width for various turbulence velocities, the corresponding mean values of $\log N(\mathrm{Fe})$, and the rms errors of the mean, $\Delta(\log N)$, for the components of 40 Dra

V_t, km/s	Component a				Component b			
	k	$\log N(\mathrm{Fe})_0$	$\log N(\mathrm{Fe})$	$\pm\Delta(\log N)$	k	$\log N(\mathrm{Fe})_0$	$\log N(\mathrm{Fe})$	$\pm\Delta(\log N)$
2.30	+0.00038	7.537	7.569	0.047	+0.00116	7.487	7.509	0.064
2.50	+0.00011	7.536	7.546	0.044	+0.00011	7.466	7.478	0.055
2.60	+0.00009	7.503	7.505	0.044	−0.00032	7.456	7.453	0.033
2.70	−0.00002	7.503	7.504	0.045	−0.00049	7.448	7.431	0.035

Table 5. Line parameters and FeII equivalent widths and iron abundances from each line in the atmospheres of 40 Dra a and b

λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{Fe})^a$	W_λ^b, mÅ	$\log N(\mathrm{Fe})^b$	λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{Fe})^a$	W_λ^b, mÅ	$\log N(\mathrm{Fe})^b$
4583.83	2.79	−1.87	228.7	7.65	248.1	7.71	4620.52	2.83	−3.29	75.0	7.15	103.3	7.60
5234.63	3.22	−2.27	147.2	7.52	120.4	7.16	5264.79	3.25	−3.23	89.8	7.64	–	–
5316.62	3.15	−1.85	216.3	7.73	204.5	7.63	5325.55	3.22	−3.32	62.5	7.28	49.2	7.11
5337.73	3.23	−3.89	57.1	7.81	–	–	5525.13	3.27	−3.94	30.6	7.49	–	–
5529.93	6.73	−1.88	14.3	7.74	–	–	5534.85	3.24	−2.73	91.6	7.15	127.4	7.69
5567.84	6.73	−1.89	12.0	7.68	11.0	7.70	6369.46	2.89	−4.16	40.0	7.52	31.6	7.37
6383.72	5.55	−2.27	26.2	7.61	14.5	7.35	6416.93	3.89	−2.85	60.2	7.33	52.5	7.27
6456.39	3.90	−2.10	102.8	7.22	110.7	7.38	6516.08	2.89	−3.32	90.8	7.35	128.9	7.92

$\log N(\mathrm{Fe})$ estimates for various V_t were used to find linear-regression relations,

$$\log N(\mathrm{Fe}) = \log N(\mathrm{Fe})_0 + kW_\lambda,$$

which enabled selection of the optimal V_t value for the atmosphere of the star under investigation. The corresponding results are given in Table 4, which shows the components' atmospheric microturbulence velocities to be approximately the same, close to 2.6 km/s.

Figure 4 is an example of the relation between $\log N(\mathrm{Fe})$ and the line equivalent widths for $V_t = 2.6$ km/s; the straight line is the linear fit, with virtually no variations of $\log N(\mathrm{Fe})$.

The lines of neutral iron, FeI, give for the abundances in the atmospheres of the components

component a $\log N(\mathrm{Fe}) = 7.505 \pm 0.044$,
component b $\log N(\mathrm{Fe}) = 7.456 \pm 0.033$.

Table 5 presents the equivalent widths of the lines of ionized iron, FeII, and the corresponding $\log N(\mathrm{Fe})$ values.

The mean iron abundances derived from the FeII lines for each of the components are

component a $\log N(\mathrm{Fe}) = 7.492 \pm 0.035$,
component b $\log N(\mathrm{Fe}) = 7.503 \pm 0.078$.

The rms uncertainties of the mean values describe the internal accuracy of the results, and show that these values coincide within errors. The abundance of iron in the atmospheres of the components of 40 Dra nearly coincides with the solar value ($\log N(\mathrm{Fe})_\odot = 7.50$)[16].

7. ATMOSPHERIC ABUNDANCES OF CARBON, NITROGEN, AND OXYGEN FOR THE SYSTEM'S COMPONENTS

We determined the atmospheric abundances of carbon, nitrogen, and oxygen for the components of the spectroscopic binary 40 Dra from lines measured mainly in the red part of the spectrum. The techniques

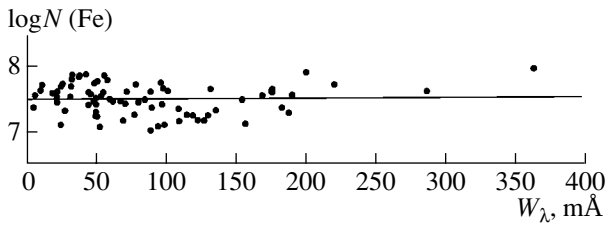

Fig. 4. Relation between $\log N(\mathrm{Fe})$ and line equivalent width for $V_t = 2.6$ km/s. The straight line is the linear regression fit with the coefficients in Table 4.

Table 6. Line parameters and CI equivalent widths and carbon abundances from each line in the atmospheres of 40 Dra *a* and *b*

λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{C})^a$	W_λ^b, mÅ	$\log N(\mathrm{C})^b$	λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{C})^a$	W_λ^b, mÅ	$\log N(\mathrm{C})^b$
4762.31	7.48	−2.46	24.7	8.53	–	–	4762.53	7.48	−2.34	30.2	8.53	–	–
4792.66	7.95	−2.79	8.7	8.60	6.1	8.56	4817.37	7.48	−3.04	8.1	8.53	7.1	8.53
4932.05	7.68	−1.88	35.5	8.28	34.2	8.37	5039.05	7.95	−1.79	29.6	8.30	–	–
5039.10	7.95	−2.29	14.8	8.50	–	–	5052.17	7.68	−1.65	57.7	8.48	41.1	8.24
5380.34	7.68	−1.84	35.9	8.28	–	–	5515.55	8.85	−2.34	2.0	8.59	–	–
5793.12	7.95	−2.06	18.9	8.42	14.8	8.36	5800.60	7.95	−2.34	10.9	8.44	–	–
6586.27	9.00	−1.89	8.2	8.61	10.2	8.76	6587.61	8.54	−1.60	18.6	8.47	35.2	8.84
6588.64	9.17	−2.20	1.9	8.62	–	–	6595.24	8.85	−2.41	1.2	8.50	–	–
6602.41	8.85	−1.38	12.0	8.12	10.1	8.18	6605.78	8.85	−2.31	1.5	8.44	–	–
7116.99	8.65	−0.91	45.6	8.40	35.5	8.31	7119.66	8.64	−1.15	27.0	8.22	27.0	8.41
9061.43	7.48	−0.35	157.4	8.02	160.3	8.15	9062.49	7.48	−0.46	155.4	8.11	152.4	8.19
9078.29	7.48	−0.58	150.8	8.19	200.8	8.70	9088.52	7.48	−0.43	164.9	8.17	178.9	8.42
9094.83	7.49	0.15	257.6	8.36	220.8	8.14							

Table 7. Line parameters and NI equivalent widths and nitrogen abundances from each line in the atmospheres of 40 Dra *a* and *b*

λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{N})^a$	W_λ^b, mÅ	$\log N(\mathrm{N})^b$	λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{N})^a$	W_λ^b, mÅ	$\log N(\mathrm{N})^b$
7468.31	10.34	−0.19	21.4	8.11	18.4	8.14	8210.71	10.33	−0.71	9.9	8.07	9.2	8.13
8216.33	10.34	0.13	25.7	7.90	16.6	7.71	8223.13	10.33	−0.27	16.4	7.98	15.1	8.03
8242.39	10.34	−0.26	22.3	8.12	20.6	8.18	8680.28	10.34	−0.46	17.6	8.13	16.4	8.21
8683.40	10.33	0.09	47.2	8.30	38.2	8.27	8686.15	10.33	−0.30	27.0	8.24	28.4	8.42
8703.25	10.33	−0.32	24.6	8.20	19.1	8.18	8711.71	10.33	−0.23	28.3	8.21	23.4	8.21
8718.83	10.34	−0.33	16.5	8.02	19.0	8.19							

used are described in our paper on the analysis of the components of 41 Dra [3].

Tables 6, 7, and 8 present the parameters of the CI, NI, and OI lines measured in the spectrum of 40 Dra. They also contain the equivalent widths W_λ, in mÅ and the abundances of carbon ($\log N(\mathrm{C})$), nitrogen ($\log N(\mathrm{N})$), and oxygen ($\log N(\mathrm{O})$) for each of the components. The values of W_λ were determined from measurements of separated lines, in the same manner as for the iron lines.

The model atmospheres and the techniques used here for the abundance determinations were the same as in the analysis of the iron lines. The values of $\log gf$ were mainly taken from the VALD list [17].

The mean carbon abundances derived by comparing the observed and calculated equivalent widths for the atmospheres of components *a* and *b* of 40 Dra are

component *a* $\log N(\mathrm{C}) = 8.388 \pm 0.036$,
component *b* $\log N(\mathrm{C}) = 8.448 \pm 0.031$.

The mean nitrogen abundances in the atmospheres of components *a* and *b* of 40 Dra are

component *a* $\log N(\mathrm{N}) = 8.116 \pm 0.054$,
component *b* $\log N(\mathrm{N}) = 8.152 \pm 0.077$.

The atmospheric oxygen abundances for the components of 40 Dra were derived from lines of four triplets in the red, with $\lambda > 6100$ Å. These lines are fairly strong and nearly unblended, so that the abundance of oxygen can be determined with certainty. The values for the atmospheres of each component are

component *a* $\log N(\mathrm{O}) = 8.768 \pm 0.031$,
component *b* $\log N(\mathrm{O}) = 8.735 \pm 0.028$.

8. DISCUSSION OF THE RESULTS

To analyze the uncertainties of our results and identify characteristic features of the components' atmospheres in ADS 11061, we compared the derived

Table 8. Line parameters and OI equivalent widths and oxygen abundances from each line in the atmospheres of 40 Dra *a* and *b*

λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{O})^a$	W_λ^b, mÅ	$\log N(\mathrm{O})^b$	λ, Å	ε_i, eV	$\log gf$	W_λ^a, mÅ	$\log N(\mathrm{O})^a$	W_λ^b, mÅ	$\log N(\mathrm{O})^b$
6155.96	10.74	−1.36	2.1	8.84	–	–	6155.97	10.74	−1.01	3.9	8.79	–	–
6155.98	10.74	−1.12	3.2	8.81	–	–	6158.18	10.74	−1.00	3.0	8.53	–	–
6158.19	10.74	−0.41	14.8	8.82	–	–	7771.94	9.15	0.37	143.4	8.92	119.9	8.80
7774.16	9.15	0.22	111.7	8.71	108.1	8.81	7775.39	9.15	0.00	105.6	8.85	102.3	8.73
8446.25	9.52	−0.46	50.0	8.73	41.5	8.72	8446.36	9.52	0.24	100.0	8.75	82.9	8.66
8446.76	9.52	0.01	80.0	8.70	66.3	8.65							

Table 9. Abundances of the light elements and iron, $\log N(\mathrm{Elem})$, in the atmospheres of the components of ADS 11061 and comparison stars

Element	Sun	Procyon A	40 Dra		41 Dra	
			component			
			a	*b*	*a*	*b*
C	8.69 ± 0.05	8.73 ± 0.05	8.39 ± 0.04	8.45 ± 0.03	8.52 ± 0.04	8.58 ± 0.03
N	8.08 ± 0.03	8.18 ± 0.06	8.12 ± 0.05	8.15 ± 0.08	8.05 ± 0.07	7.99 ± 0.03
O	8.81 ± 0.04	8.86 ± 0.14	8.77 ± 0.03	8.74 ± 0.03	8.73 ± 0.07	8.76 ± 0.06
Fe	7.50 ± 0.05	7.43 ± 0.04	7.50 ± 0.05	7.46 ± 0.05	7.55 ± 0.04	7.60 ± 0.04

elemental abundances to the atmospheric chemical compositions of the Sun and Procyon A. For this purpose, we applied our technique to determine the abundances of iron, carbon, nitrogen, and oxygen in the solar atmosphere, adopting a model atmosphere with $T_{\mathrm{eff}} = 5770$ K, $\log g = 4.44$, and $V_t = 0.85$ km/s. The equivalent widths of the FeI, FeII, CI, NI, and OII lines were taken from [18]. The results are given in Table 9, and coincide with the generally accepted values for the chemical composition of the Sun and meteorites [16].

We carried out a similar abundance analysis for the same elements for the atmosphere of Procyon A using our measurements of line equivalent widths for echelle spectra taken with the Zeiss-2000 telescope (ICAMER, Peak Terskol) by F.A. Musaev [5]. The list of spectral lines is the same as for the components of ADS 11061. The model atmosphere for Procyon A was selected by interpolating the models of Kurucz [13], using parameters ($T_{\mathrm{eff}} = 6530$ K, $\log g = 3.96$, $V_t = 2.40$ km/s) taken from [19]. The results are collected in Table 9, along with our determinations for the components of ADS 11061.

The data for each of the elements coincide within the errors for all the stars; nevertheless, several trends can be found that testify to the presence of different evolutionary changes in the atmospheric chemical compositions for the studied stars. A slight deficiency of carbon and a slight excess of nitrogen in the atmospheres of the components of 40 Dra, compared to the Sun, provides evidence for weak mixing between the core and the atmosphere in these stars. During the core-hydrogen-burning stage in the equilibrium CN cycle, such mixing increases the atmospheric nitrogen abundance and decreases the carbon abundance, since virtually all the carbon in the core becomes nitrogen as the equilibrium CN cycle is approached.

At the same time, the absence of such chemical composition differences for the atmospheres of the components of 41 Dra shows that there is no mixing in these stars. Such differences among the components of ADS 11061 are most likely due to the substantial difference in the periods of 40 Dra and 41 Dra. In the former case, the period is close to 10 days, the component separation is close to 0.1 AU, and the interaction between the components is constant, whereas, for 41 Dra, the period is 1246.7 days, the component separation is mostly in excess of 3 AU, and only during the periastron passage does this distance become less than 0.1 AU, and then for a very short time only. It is probably the more constant and stronger interaction in the case of 40 Dra than in the case of 41 Dra that leads to the mixing. The "shock" gravitational stresses during periastron passages, when the components approach each other at

distances of several stellar radii, cannot give rise to considerable mixing in 41 Dra.

On average, the iron abundance in the atmospheres of the components of ADS 11061 is solar. Thus, the components of the ADS 11061 system reveal no indications of the gravitational separation of elements, in contrast to Procyon A [19]. Since the parameters of all the components of ADS 11061 and Procyon A are similar and their ages are approximately the same (about 2.5 Gyr), difference are due to the different orbital parameters of the spectroscopic binaries. In ADS 11061, the equal-mass components are close to each other, whereas in the Procyon system, the white dwarf is fairly far from the primary, and its action on the primary cannot lead to mixing.

The effects discussed above are at the detection limit, and the question of the uncertainties in the results is crucial. The coincidence of the elemental abundances we derived for the Sun and Procyon A with the data in [19, 20] testifies to the trustworthiness of the data in Table 9, and thereby the correctness of our conclusions. Note that, in our case, the principal source of errors are uncertainties in the parameters of the model atmospheres for the studied stars, but the accurate description of the observations confirms the correctness of the parameters used. Applying three-dimensional models [21] changes the elemental abundances from those obtained for the one-dimensional models we have used. These corrections are comparatively small for iron—$\log N(\mathrm{Fe})_\odot = 7.45 \pm 0.05$, $\log N(\mathrm{Fe})_{\mathrm{Proc}} = 7.36 \pm 0.03$—but are more significant for C, N, and O, reaching $-0.2 \ldots -0.3$ dex. However, the introduction of such corrections does not change the differences between the stars that led us to the above conclusions.

9. CONCLUSION

We have used spectroscopic observations of the binary 40 Dra, a member of the hierarchic multiple system ADS 11061, with the 1 m telescope of the SAO and the 2 m telescope of the ICAMER to analyze the atmospheric parameters separately for each component of 40 Dra. Within the uncertainties, all the observations indicate that the effective temperatures of the system's components are 6420 and 6300 K, respectively, for the primary and secondary.

The iron abundance in the components' atmospheres is essentially solar, as are the observed ratios of the abundances of carbon, nitrogen, and oxygen. Nevertheless, slight differences in the elemental abundances of the components of 40 Dra and 41 Dra suggest the presence of slight mixing in the atmospheres of the components of 40 Dra and no mixing for the components of 41 Dra. The mixing in 40 Dra is stimulated by the continuous gravitational interaction between the components; in 41 Dra, such interaction cannot lead to any observed effects due to its short duration (only during periastron passages).

ACKNOWLEDGMENTS

The authors wish to thank F.A. Musaev, who took the spectra of 40 Dra. This study was supported by the Russian Foundation for Basic Research (project no. 04-02-17563).

REFERENCES

1. S. L. Boothroyd, Publ. Dom. Astrophys. Obs. Victoria BC **1**, 246 (1920).
2. A. A. Tokovinin, Pis'ma Astron. Zh. **21**, 286 (1995) [Astron. Lett. **21**, 250 (1995)].
3. Yu. Yu. Balega, V. V. Leushin, and G. Vaigel't, Astron. Zh. **82**, 1 (2005) [Astron. Rep. **49**, 217 (2005)].
4. F. A. Musaev, Pis'ma Astron. Zh. **19**, 776 (1993) [Astron. Lett. **19**, 315 (1993)].
5. F. A. Musaev, G. A. Galazutdinov, A. V. Sergeev, *et al.*, Kinematika Fiz. Nebesnykh Tel **15**, 282 (1999).
6. G. A. Galazutdinov, ftp.sao.ru (2002).
7. M. A. Al-Wardat, Bull. Spec. Astrophys. Obs. **53**, 51 (2002).
8. A. Tokovinin, Yu. Yu. Balega, E. A. Pluzhnik, *et al.*, Astron. Astrophys. **409**, 245 (2003).
9. A. H. Batten, J. M. Fletcher, and D. G. MacCarthy, *Eighth Catalogue of the Orbital Elements of Spectroscopic Binary Systems,* Publ. Dom. Astrophys. Obs. **17** (1989).
10. H. L. Johnson and W. W. Morgan, Astrophys. J. **117**, 313 (1953).
11. V. G. Kornilov, I. M. Volkov, A. I. Zakharov, *et al.*, *UBVR Catalog of Bright Stars of the Northern Hemisphere*, Tr. Gos. Astron. Inst. im. P.K. Shternberga **63** (1991).
12. *The Hypparcos and Tycho Catalogues,* ESA SP-1200 (ESA, 1997).
13. R. Kurucz, http://kurucz.harvard.edu.
14. V. V. Tsimbal, *Magnetic Stars* (SAO RAN, Nizhnii Arkhyz, 2004) (in press) [in Russian].
15. V. V. Leushin and G. P. Topil'skaya, Astrofizika **22**, 121 (1985) [Astrophys. **21**, 74 (1985)].
16. N. Grevesse and A. J. Sauval, Astron. Astrophys. **347**, 348 (1999).
17. F. Kupka, N. E. Piskunov, T. A. Ryabchikova, *et al.*, Astron. Astrophys., Suppl. Ser. **138**, 119 (1999).
18. Ch. E. Moore, M. G. J. Minnaert, and J. Houtgast, *The Solar Spectrum 2935 Å to 8770 Å*, NBS Monograph 61 (Washington, 1966).
19. P. Kervella, F. Thevenin, P. Morel, *et al.*, astro-ph/0309148 (2003).
20. C. A. Prieto, M. Asplund, R. J. G. Lopez, and D. L. Lambert, Astrophys. J. **576**, 544 (2002).
21. M. Asplund, astro-ph/0302409 (2003).

Translated by N. Samus'

Astronomy Reports, Vol. 49, No. 12, 2005, pp. 993–1000. Translated from Astronomicheskiĭ Zhurnal, Vol. 82, No. 12, 2005, pp. 1109–1118.

Neutron Stars in Close Binaries with Elliptical Component Orbits

A. V. Tutukov

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

Received June 14, 2005; in final form, July 6, 2005

Abstract—We consider the evolution of close binaries in which the initial secondary component is a nondegenerate helium star with mass $M_{\rm He} = 0.4–60\,M_\odot$, while the initially more massive primary has evolved into a black hole, neutron star, or degenerate dwarf. The neutron star is assumed to originate as a result of the evolution of a helium star with a mass of $2.5\,M_\odot \le M_{\rm He} \le 10\,M_\odot$ after the explosion of a type Ib,c supernova. If the axial rotation of the helium star before the explosion is rigid-body and synchronized with the orbital rotation, for $P_{\rm orb} \le 0.16$ day, the rotational energy of the young neutron star will exceed the energy of an ordinary supernova. If the magnetic field of the neutron star is sufficiently strong, the necessary conditions for a magnetic–rotational supernova are provided. The initial rotational period of a young neutron star originating in a system with an orbital period shorter than ∼50 days is shorter than ∼4 s, which, according to observations, is required for the appearance of a radio pulsar. A helium star whose mass exceeds $\sim 10\,M_\odot$ in a close binary with an orbital period shorter than one day and with the axial rotation of the helium presupernova synchronous with the orbital rotation evolves into a Kerr black hole, whose formation is likely to be accompanied by a gamma-ray burst with a duration longer than two seconds. In particular, we consider close binaries in which the second supernova results in the formation of a neutron star that remains in the binary. The theoretical distribution of orbital periods and eccentricities for such systems is consistent with that observed for radio pulsars in the Galactic disk in binaries with compact components and orbital eccentricities exceeding ∼0.09, providing an explanation for the observed correlation between the orbital eccentricities and orbital periods for these systems.

1. INTRODUCTION

Helium nondegenerate stars are an intermediate product of the evolution of stars in close binaries with initial masses of their main-sequence components exceeding $\sim 2.8\,M_\odot$. The evolution of these stars is related to important astrophysical objects and phenomena: neutron stars, black holes, type Ia and Ib,c supernovae, gamma-ray bursts, and others. The final product of the evolution of a helium star is determined by its initial mass. We will assume here that the final products of helium stars with masses of $0.35–2.0\,M_\odot$, $2–2.5\,M_\odot$, and $2.5–10\,M_\odot$, respectively, are carbon–oxygen degenerate dwarfs (CO), oxygen–neon dwarfs (ONe), and neutron stars (NS), while helium stars with higher initial masses than this end their evolution as black holes (BH) [1, 2].

Here, we continue our studies of the conditions for the formation of helium stars in close binaries and the role of these objects in the formation of degenerate CO and ONe dwarfs, neutron stars, and black holes [1, 2]. The observations indicate that the formation of at least some of these stars is determined by the fact that they are in close binaries [1–4]. This requires that we analyze the evolution of these binaries using scenarios based on the current theory of the evolution of close binaries with various masses [1, 2]. We aim here to analyze the basic parameters of the products of the evolution of nondegenerate helium stars in close binaries with compact components, and to reveal possible connections between these stars and the products of their evolution (CO and ONe dwarfs, neutron stars, and black holes), as well as events in the evolution of massive close binaries (supernovae, the formation of single radio pulsars and pulsars in binaries, gamma-ray bursts). Section 2 discusses the products of the evolution of helium stars and phenomena that may be related to their formation. Section 3 is concerned with the formation of radio pulsars observed in close binaries with elliptical orbits ($e > 0.09$) in the Galactic disk. Finally, we formulate our basic conclusions: a number of parameters of radio pulsars, such as the rate of their formation in the Galaxy, the distribution of orbital periods and orbit eccentricities for observed radio pulsars in close binaries, and the rate of formation of gamma-ray bursts, must be analyzed in the framework of modern concepts about the evolution of close binaries, based on evolution scenarios for these stars.

1063-7729/05/4912-0993$26.00

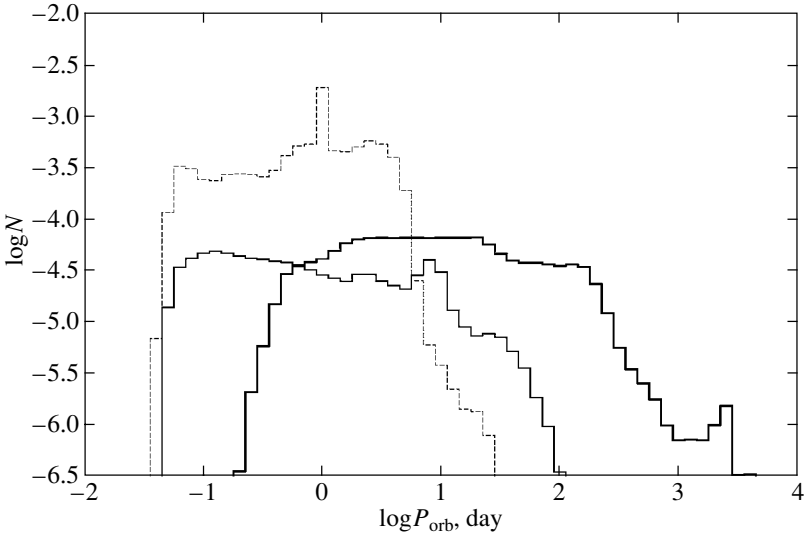

Fig. 1. The rate of formation, N, of close binaries with nondegenerate helium secondaries that produce black holes (thick solid curve), neutron stars (thin solid curve), degenerate dwarfs (dotted curve), and with the compact final remnants of their primaries as a function of the orbital period, $P_{\rm orb}$, of the system before the supernova explosion of the initial secondary (in the first two cases) or before the formation of a degenerate dwarf (in the last case).

2. ORBITAL-PERIOD DISTRIBUTION FOR NONDEGENERATE HELIUM STARS WITH COMPACT SECONDARIES

In this section, we will assume that the evolution of the initially more massive component of a close binary has resulted in the formation of a compact remnant: a degenerate CO or ONe dwarf, neutron star, or black hole. In addition, we consider only systems that remain gravitationally bound, even when the evolution of the primary results in the supernova explosion. Let us suppose that the secondary component evolves into a nondegenerate helium star with a mass exceeding $\sim 0.35\,M_\odot$, and consider the further evolution of such binaries. Figure 1 presents the distribution of orbital periods for their rate of formation. All such systems are divided into three families, according to the final evolutionary product for the helium stars. Most of these stars possess small masses ($<2.0\,M_\odot$), and so produce degenerate CO or ONe dwarfs. These systems obviously tend to display short orbital periods of from 1 h to 5 days (Fig. 1).

Further evolution of the systems considered with components whose nuclear evolution is complete depends on the orbital period and the nature of the components [1]. For extremely close systems, the radiation of gravitation waves becomes an important evolutionary factor which causes the components to approach each other and finally merge. The result of this process depends on the nature of the helium star's companion. In systems with orbital periods shorter than ~ 0.1 day, the coalescence of nondegenerate helium stars with their companions precedes the depletion of helium in the core of the helium star if the latter has a mass of $\sim 0.5\,M_\odot$. According to Fig. 1, the total rate of such events in the Galaxy is ~ 0.001 per year. As a result, either a helium subdwarf (if the primary is a degenerate helium dwarf), R CrB-type star (if the primary is a degenerate CO or ONe dwarf), or Thorne–Zhytkov object (if the primary is a neutron star or black hole [2]) can originate.

In systems with orbital periods (~ 0.1 day $< P_{\rm orb} < \sim 0.3$ day), the helium stars have time to become CO or ONe dwarfs even before the coalescence. Their merging with degenerate He dwarfs—the products of the evolution of the primaries—can result in the formation of R CrB-type stars [2]. The coalescence of CO and ONe dwarfs will probably lead to either a type-Ia supernova (for a super-Chandrasekhar total mass [5–7]) or the formation of an ONe dwarf (for a lower mass). Our model yields a SNIa rate for our Galaxy of ~ 0.0023 per year, which is close to generally accepted observational estimates. However, in spite of the importance of mergers of degenerate dwarfs for current theories of stellar evolution, the Galaxy, and even the Universe (whose expansion is measured with observations of these supernovae), models for mergers of degenerate dwarfs remain poorly developed. Thus far, 18 close-binary degenerate dwarfs with orbital periods shorter than 30 days have been detected; in accordance with Fig. 1, the logarithms of their orbital periods are almost uniformly distributed within the interval 0.1–10 days [8]. Eight of these systems display orbital periods that are sufficiently short for their components to merge over the Hubble time. It is particularly important that, for one of these systems, the components are both able to merge over the Hubble time and have a total mass exceeding the Chandrasekhar

limit. Thus, observations of binary degenerate dwarfs confirm current model scenarios both qualitatively and quantitatively, but with only poor statistics. In the absence of outer effects, degenerate CO and ONe dwarfs forming in systems with $P_{\rm orb} > 0.3$ day (Fig. 1) will remain in gravitationally bound binaries for a time longer than the Hubble time.

Let us now consider the orbital-period distribution (Fig. 1) for the more massive ($2 < M_{\rm He}/M_\odot < 10$) helium stars that produce neutron stars as a result of their evolution. The observations indicate that main-sequence stars in close binaries with orbital periods shorter than ~10 days display synchronous axial and orbital rotation [9]. We suppose that the same can occur for helium components in the closest binaries. During the formation of a neutron star as a result of a supernova at the end of the nuclear evolution of a helium star, the radius of the collapsing core of the supernova decreases by almost three orders of magnitude, making it possible to relate the orbital period of these systems at the time of the explosion $P_{\rm orb}$ with the rotational period of the young pulsar: $P_{\rm p} = 10^{-6} P_{\rm orb}$ [10]. For $P_{\rm orb} < 0.16$ day, simple estimates yield a rotational period for a young neutron star shorter than 0.014 s. This means that the rotational energy of such a pulsar will exceed the characteristic energy of a supernova (~10^{50} erg) [11]. In the presence of a sufficient magnetic field of the young pulsar, its rotational energy can be transferred to the envelope, resulting in its supernovalike ejection with a high velocity [12]. Such phenomena were named magnetic-rotational explosions; according to Fig. 1, their rate in our Galaxy is $\sim 2 \times 10^{-4}$ per year. For a total rate of Galactic supernovae of ~0.03 per year [6], the model rate of magnetic-rotational explosions will be only about 1% of this value.

The fate of a young neutron star is determined by its initial rotational rate. Assuming that the above relation between $P_{\rm orb}$ and $P_{\rm p}$ is valid for such systems, a young neutron star forming in a system with an orbital period shorter than ~50 days will rotate with a period shorter than ~4 s. According to the observations, only such neutron stars are manifest as radio pulsars [13], given a sufficient magnetic field. The model scenario yields a formation rate for such pulsars in our Galaxy of ~0.001 per year (Fig. 1). Taking into account the remaining substantial uncertainties in both theoretical and empirical estimates of the rate of formation of radio pulsars in our Galaxy (0.01–0.001 per year) [14–16], this value is well within the present "empirical" limits.

Neutron stars originating in wider systems or evolving from single massive stars probably rotate too slowly to become radio pulsars, even in the presence of a strong magnetic field. As a result, the total supernova rate (SNII, SNIb,c), and, hence, the estimated birthrate of neutron stars (0.01–0.03 per year) [2] may substantially exceed the estimated rate of formation of radio pulsars. In other words, the birthrate of neutron stars can exceed the formation rate of radio pulsars by a factor of 10 to 30. To draw final conclusions about observational estimates of the rate of formation of radio pulsars, we must know what fraction of radio pulsars remain undetected due to the narrow beam pattern of their radiation. This may help to explain, at least partly, the reason for the substantial deviation between the above observational estimates. It remains possible that neutron stars originating in wide binaries, or as a result of the evolution of single massive stars, display such long rotational periods (several hours, days) that they are not manifest as ordinary radio pulsars. Since most young neutron stars evolve from wide massive binaries or single massive stars, they may belong to this set of objects. A recently detected transient source of intense radio radiation with a period for its outbursts of about 76 min [17] may be such a slowly rotating neutron star. This scenario deserves further detailed study. Another way of forming this type of neutron star is represented by X-ray pulsars accompanying Be stars, whose axial rotation has been slowed to 1–15 min by the stellar wind of the Be star. The supernova of the Be star at the end of its evolution could transform the slowly rotating neutron star into a potential transient super-long-period radio pulsar. This is another possible scenario of getting neutron stars with super-long periods of axial rotation.

The distribution of orbital periods for helium stars producing neutron stars in close binaries (Fig. 1) yields a theoretical estimate for the interval of initial rotational periods for most neutron stars of 0.1–1 s. Recent observational estimates for the characteristic periods of young radio pulsars are in the interval 0.1–0.5 s [17], consistent with our model. In addition, according to the orbital-period distribution for the precursors of neutron stars (Fig. 1), young neutron stars with periods longer than ~10 s cannot originate in close binaries in which the axial rotation of the helium stars is synchronous with their orbital motion, since the orbital periods of the precursors are shorter than ~10 days. Indeed, the maximum observed rotational periods of magnetars—young neutron stars with strong magnetic fields—are no longer than 10 s. These facts provide supplementary evidence supporting the idea that a substantial fraction of radio pulsars form only in close binaries.

It is also important that the disruption of systems with orbital periods of 0.05–10 days (Fig. 1) results in high (100–1000 km/s) spatial velocities for the corresponding young neutron stars [15, 18]. Thus, the

formation of radio pulsars in close binaries may give rise not only to their observed rapid initial rotations, but also to the high average velocities of these objects. Note also that the minimum rotational periods of young pulsars formed in extremely close binaries ($\sim$0.01 s) yield rotational velocities for neutron stars almost an order of magnitude higher than those required for their rotational instability. As we will see below, this distinguishes neutron stars from black holes, which are able to reach very high rotational velocities and produce Kerr black holes in the closest binary systems.

Let us now consider the most massive ($M \geq 10\,M_\odot$) helium stars [19], known as Wolf–Rayet stars. Precisely these stars produce black holes. The initial masses of such stars in close binaries in the main sequence stage are currently estimated to be $\sim 25\,M_\odot$ [19]. Based on the above assumption that the rotation of the helium star is synchronous with the orbital motion, we can derive a constraint on the orbital period of a system containing a Wolf–Rayet star, such that a Kerr black hole (i.e., a black hole with (super)critical rotation) can be formed. According to one model, the formation of such a black hole may be accompanied by a gamma-ray burst [19–21]. If we take the density in the core of a massive star collapsing into a black hole to be 10^4 g/cm^3 [22] and the density of a black hole with a mass of $5\,M_\odot$ to be $\sim 2 \times 10^{16}$ g/cm^3, then, if angular momentum is conserved as the core collapses into a black hole, we conclude that a Kerr black hole can be formed if the orbital period of the system before the collapse is shorter than roughly one day [21]. In this case, the helium star should have a black-hole or neutron-star companion. The estimated limiting orbital period remains fairly uncertain due to uncertainties in the used parameters. It yields a rate of formation of Kerr black holes in the Galaxy of $\sim 10^{-4}$ per year (Fig. 1). To bring this value closer to the estimated observed rate of gamma-ray bursts in the Galaxy, which is comparable with our estimate (10^{-6}–10^{-7} per year), we should take into account the probable collimation of the gamma-ray radiation within a narrow cone with an opening angle of only a few degrees [19, 23, 24]. This makes our estimate for the rate of formation of Kerr black holes in the Galaxy the most likely explanation for long (>2 s) gamma-ray bursts [21].

Let us consider several other possible ways of forming Kerr black holes during the evolution of massive stars, either single or in wide binaries. Stars with masses exceeding $\sim 50\,M_\odot$ bypass the red supergiant stage; after the hydrogen in their cores is depleted, they turn into Wolf–Rayet stars. The initial central densities of these very massive stars on the main sequence are ~ 2 g/cm^3, while their final densities at the stage of the formation of a black hole are $\sim 2 \times 10^{16}$ g/cm^3. If we assume that the cores of these stars can evolve into black holes conserving their angular momentum, then, by the formation of the black hole, the angular velocity of the rotation of the core will increase by a factor of $\sim 5 \times 10^{10}$. Therefore, assuming evolution with conservation of the local angular momentum, all stars with initial masses $>50\,M_\odot$ and initial rotational periods on the main sequence shorter than $\sim$30 days can end their evolution as Kerr black holes. For such massive stars, these rotational periods correspond to very moderate equatorial velocities $\sim$20 km/s. Most massive stars rotate substantially more rapidly ($\sim$300 km/s) [25]; according to the star-formation function for our Galaxy adopted here [1, 10], this yields a rate of formation of Kerr black holes of $\sim$0.001 per year. This exceeds appreciably the observed rate of gamma-ray bursts, even taking into account the high degree of collimation of their radiation. Consequently, the assumption that the local angular momentum is conserved during the evolution of at least the most massive single stars ($>50\,M_\odot$) is probably not correct, and, by the time of its collapse into a black hole, the core of such a star manages to transfer the bulk of its initial angular momentum to the envelope, possibly via the internal magnetic field of the star [20]. Thus, the formation of a Kerr black hole and a gamma-ray burst can probably be excluded from this scenario.

The coalescence of close binary helium subdwarfs due to the radiation of gravitation waves may provide another way to explain the rapid rotation of some helium stars. However, simple estimates for our Galaxy show that this rare ($<10^{-4}$ per year) possibility exists only for systems for which the total mass of the nondegenerate helium components is $<3\,M_\odot$. Consequently, the product of such a merger will not be sufficiently massive to produce a black hole; recall that the mass of the helium star must exceed ($\sim 10\,M_\odot$) in order for a black hole to be formed [19]. On the other hand, this scenario can explain the formation of young, single millisecond radio pulsars with large initial magnetic fields. This can occur in the case of insufficient deceleration of the rotation of the core of a helium star with an initial mass of ~ 2–$3\,M_\odot$ that collapses to a neutron star due to the action of the stellar wind and the expansion of the stellar envelope during its evolution. Such young, single millisecond pulsars with large magnetic fields are formed due to the rapid rotation of their helium precursors, rather than the usual accretion of matter from a low-mass companion. They should be studied in more detail. The strong magnetic field should result in very short lifetimes for such millisecond pulsars, which could complicate their detection, even if their rate of formation is comparatively high.

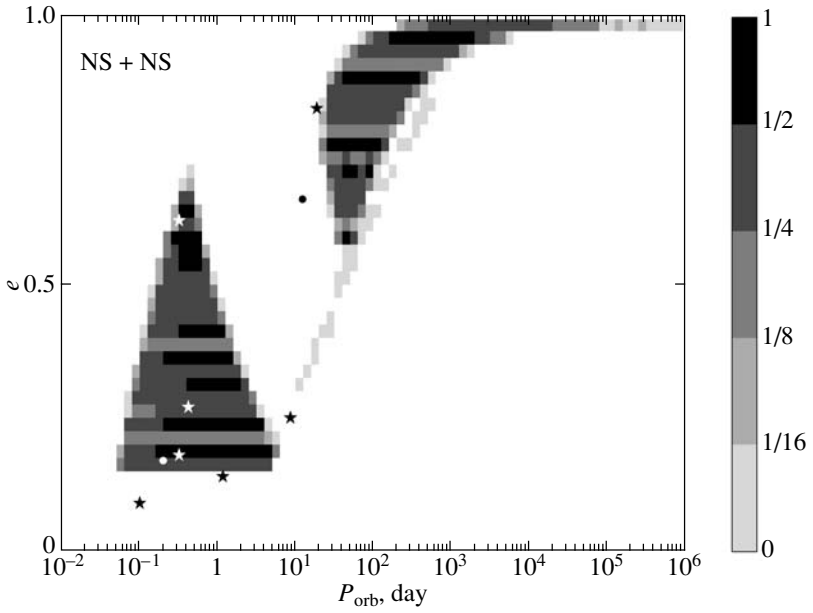

Fig. 2. Model distribution in the orbital eccentricity e–orbital period P_{orb} plane for close binary final systems consisting of two neutron stars, NS + NS, at the time of formation of the second NS. The theoretical distribution for young neutron stars with carbon–oxygen degenerate dwarfs as companions (CO + NS) essentially coincides with that for (NS + NS) systems. The density reflects the relative number of the systems. The points denote binary radio pulsars with $e > 0.09$ observed in the Galactic disk with degenerate-dwarf companions, while the asterisks denote neutron stars [27–37].

Finally, the merging of close binary neutron stars with orbital periods shorter than ∼8 h in our Galaxy and our model rate of $\sim 10^{-4}$ per year present another possibility for the formation of Kerr black holes. Such phenomena could explain short (<2 s) gamma-ray bursts [22–25]. In summary, we conclude that the core collapse of a Wolf–Rayet star in a massive close binary containing a black hole with an orbital period shorter than approximately one day is the most realistic scenario for long (>2 s) gamma-ray bursts. In this case, the gamma-ray burst should mark the beginning of a SNIb,c explosion, as has indeed been observed in several cases [26].

3. DISTRIBUTION OF ORBITAL PERIODS AND ORBITAL ECCENTRICITIES FOR YOUNG NEUTRON STARS IN BINARIES WITH BLACK HOLES, NEUTRON STARS, AND DEGENERATE DWARFS

The second supernova in a system usually results in the disruption of the binary, due to either mass loss or the "kick" (≤100 km/s) experienced by the young neutron star. However, some neutron stars (observational selection effects current prevent clarification of their relative number) remain components of close binaries with appreciable orbital eccentricities after the second supernova. In the absence of a kick, for zero orbital eccentricity, the eccentricity e before the supernova provides a criterion for the mass lost in the explosion: $e = dM/M_R$, where dM is the mass lost and M_R is the mass of the remaining gravitationally bound binary.

The young neutron star originating from the initial secondary can be a black hole, neutron star, or degenerate ONe or CO dwarf [2]. The number of systems with a black hole and an orbital period of 1–200 days is small, as is the number of systems with a degenerate dwarf. In the model, the most common secondary for a radio pulsar should be a neutron star. Figure 2 presents the distribution of the initial orbital periods for these systems, obtained using our scenario code [2] (for the parameters of the common envelope $\alpha_{CE} = 2$ and $\gamma = 1.5$). The distribution for systems with degenerate carbon–oxygen and oxygen–neon dwarf remnants of the initial primary is essentially the same as that in Fig. 2. Independent of the nature of the secondary, this makes it possible to compare the positions of the observed binary radio pulsars in the disk of our Galaxy with the theory. It is obvious from Fig. 2 that the model systems are clearly divided into two groups with different orbital periods and eccentricities. Analyzing the evolution of the precursors for these systems indicates that final systems with orbital periods shorter than 10 days are produced as a result of two common envelope stages. This follows from the large initial mass ratio of the components. The evolution of systems with final orbital periods exceeding 10 days during the filling of the Roche lobe of the primary is conservative since the initial masses of the components are comparable, which increases

the orbital period of the system prior to the filling of the secondary's Roche lobe. Only after this does the common envelope form causing the components to become closer. The conservativeness of the mass exchange is ensured by the closeness of the initial masses of the main-sequence components. This can explain the existence of two groups of binary radio pulsars in the Galactic disk in Fig. 2.

Currently, the parameters of nine neutron stars located outside globular clusters in close binaries with orbital eccentricities exceeding ~0.09 and with compact secondaries are known [27–37]. In seven systems (J1537+1155, J1915+1606, J0737−3039, J1829+2456, J1756−2251, J1518+4904, J1811−1756), the secondaries are likely neutron stars, while, in two (J2305+4707, J1141−6545), they are degenerate dwarfs. The high eccentricity indicates that the pulsar was produced during the supernova of the secondary, and the fact that these systems do not belong to any globular cluster suggests that they were formed in the course of their natural evolution as isolated massive close binaries, which are the object of our present study. The orbital periods for these systems are from 0.1 to 20 days, and the eccentricities from 0.1 to 0.8. The number of "reliable" observed systems in the Galactic disk remains small, but their distribution is close to the model presented in Fig. 2, which provides evidence for the existence of two groups of binary radio pulsars in the $e-P_{\rm orb}$ plane (Fig. 2). Recall that we assumed no kick during the formation of the neutron star when constructing the theoretical distribution. Simple estimates show that a roughly 300 km/s kick would disrupt even observed systems with orbital periods of ~8 h, while a 100 km/s kick would disrupt systems with periods of ~20 days. Naturally, such a kick would make it impossible for the two groups of binary radio pulsars in Fig. 2 to exist. Consequently, at least for these neutron stars, the kick obtained during their formation was below the above limits.

It is obvious from a comparison of the data in Fig. 2 that the observed positions of the systems in the Galactic disk is generally consistent with the model scenario for the adopted parameters of the common envelope. Systems with $P_{\rm orb} < 10$ days are close to the theoretical predictions for systems that have undergone two common-envelope stages, while those with $P_{\rm orb} > 10$ days have likely undergone only one such stage. The absence of observed systems with orbital periods exceeding 100 days probably follows from their small rate of formation (Fig. 1). The theoretical distribution of the systems in Fig. 2 can explain the observed "correlation" between orbital eccentricity and orbital period: this is simply due to the existence of two groups of binary radio pulsars. In order for a system to undergo two common-envelope stages, the mass of the second presupernova must not exceed $\sim 4\,M_\odot$. This can explain the small values for both the final orbital period and the orbital eccentricity. The conservativeness of the first mass exchange between the components results, on the one hand, in relatively large final orbital periods and, on the other hand, in large masses for the presupernovae (up to $\sim 5.6\,M_\odot$), leading to the disruption of the systems during the explosions of their secondaries.

4. CONCLUSION

We have continued our study of the evolution of close binaries with nondegenerate helium components [1]. According to current evolutionary concepts, the fate of helium stars is basically determined by their masses. Helium stars with masses smaller than $\sim 2\,M_\odot$ produce degenerate CO and ONe dwarfs; in the closest of these systems, the components may merge to produce type-SNIa supernovae, given their extra-Chandrasekhar total masses. Helium stars with masses of $2-10\,M_\odot$ produce neutron stars as a result of SNIb,c explosions. Assuming the axial rotation and orbital motion of a helium presupernova are synchronous, we can estimate the limiting orbital periods for the formation of different types of neutron stars. In the presence of a sufficiently strong magnetic field, the formation of a neutron star in a close binary with an orbital period shorter than 0.16 day is accompanied by a magnetic-rotational supernova explosion [11, 12]. Neutron stars originating in close binaries with orbital periods shorter than ~50 days can become radio pulsars with initial rotational periods shorter than ~4 s.

Helium stars with masses exceeding $\sim 10\,M_\odot$ are able to produce black holes [19]. The assumption of synchronicity of the axial rotation and orbital motion yields the conditions for the formation of a Kerr black hole, which is likely to be accompanied by a long (>2 s) gamma-ray burst [19]. A Kerr black hole can form if the orbital period of a black hole + helium star system is shorter than roughly one day at the time of the supernova explosion. For a galaxy with a mass of the order of our Galaxy, the expected rate of gamma-ray bursts is close to the observed rate, if we take into account the collimation of the gamma-ray radiation within only a few degrees [19, 38]. Short (<2 s) gamma-ray bursts may be due to the coalescence of close binary neutron stars, whose rate was previously estimated as $\sim 10^{-4}$ per year for a galaxy with the mass of our Galaxy [39, 40]. In this case, taking into account the collimation of the gamma-ray radiation will also likely be able to reconcile the merging rate for close binary neutron stars and the observed rate of short gamma-ray bursts.

It is obvious from the orbital-period distribution for young neutron stars formed from the secondaries that their most common companions in the resulting binaries with orbital periods of 0.1–1 days are neutron stars and carbon–oxygen dwarfs (Figs. 1, 2). The most probable companions in the case of orbital periods of 1–200 days are black holes (Fig. 1). According to the model [2], the number of ONe-dwarf companions to young neutron stars does not exceed several percent of the number of CO-dwarf companions.

The distribution of nine binaries with radio pulsars observed in the Galactic disk with elliptical orbits ($e > 0.07$) in the e–$P_{\rm orb}$ plane is consistent with the positions expected for the two groups predicted by the model, with orbital periods shorter and longer than ~10 days. In addition, this distribution provides evidence for the dependence $e \sim P_{\rm orb}^{1/3}$ between the eccentricity and orbital period of the system (which, however, displays a large dispersion—see Fig. 2). The total observed distribution for 20 known radio pulsars found in binaries with appreciable (>0.09) orbital eccentricities is in fairly good consistency with the prediction of the model scenario (Fig. 2). Within the currently available statistics and the remaining uncertainties, the observations support the division of the studied systems into two distinct model groups: highly eccentric ($e > 0.6$) systems with long (>10 days) orbital periods, and short-period systems ($P_{\rm orb} \sim 0.1$–10 days) with small orbital eccentricities ($e < 0.6$). This can fully explain the observed correlation between the periods and orbital eccentricities of the studied binaries. The remaining 11 binary radio pulsars [27–37], which, as a rule, belong to the spherical component of the Galaxy, display the same division (12 + 8), which is clear and coincides with the model (Fig. 2), along with the same eccentricity–period correlation. It is important that, while theoretical estimates for the characteristic orbital periods of these groups depend on the model parameters adopted for their common envelopes, the existence of the groups themselves is due to the different evolution of close binaries in the conservative and nonconservative (common envelope) stages. Thus, the difference between binaries that have undergone common-envelope stages or conservative mass exchange after the depletion of hydrogen in the core of the primary can be traced to the formation of the final binaries with compact components. This problem deserves further study.

The problem of the existence of binary neutron stars like J0737–3039 with $e = 0.09$ remains unsolved. A similar system has been found in a globular cluster (B0021-72H, $e = 0.07$ [28], $P = 2.36$ day). In the absence of a kick due to an "instantaneous" supernova explosion, such a small orbital eccentricity suggests a very small mass lost in the explosion—0.2–0.25 $M_\odot$. In this case, the resulting mass of the presupernova should be 1.6–1.7 $M_\odot$, which certainly exceeds the Chandrasekhar mass limit for degenerate configurations, which makes it possible for a supernova to occur. Theoretically, however, it remains unclear if such a small (1.6–1.7 $M_\odot$) mass for a helium SNIb,c presupernovae can exist, in the absence of an appreciable stellar wind from such a small helium subdwarf. Observational restrictions imposed on the secular variations of the orbital period rule out the possibility that these radio pulsars have massive (~10 $M_\odot$) black-hole companions. Of course, a kick of 30–50 km/s could solve this problem [41]; however, the physics of this kick, factors specifying its value, and its universality remain unclear. Another explanation for such small eccentricities can be obtained if we assume that the lower mass limit for neutron-stars progenitors can be lowered to 7–8 $M_\odot$. Again, there is no basis for this assumption at present.

No systems consisting of a black hole and radio pulsar [42] have been observed. Essentially, these are expected to display orbital periods of 0.4–200 days (Fig. 1) and eccentricities of 0.05–0.65. The relatively small eccentricities of these systems are due to the insignificant mass loss in the second explosion in the close binary.

For WD + NS and NS + NS systems, the interval of observed orbital periods is close to that estimated using the "scenario machine," confirming the consistency of current evolutionary concepts for massive close binaries. However, despite the reasonable correspondence between the model and observed distributions of radio pulsars in binaries (Fig. 2), our understanding of the evolution of close binaries may be insufficient to accurately predict the positions of radio pulsars in close binaries in the e–$P_{\rm orb}$ plane. For example, one weak point of our model scenario is the assumption that the common-envelope parameter is the same for all systems. In fact, this parameter can vary by a factor of a few [43] due to structural variations of the donor envelopes when these stars fill their Roche lobes. If we take this into account, the orbital-period distribution for the final systems (Figs. 1 and 2) may become different, resulting, in particular, in a broadening of the orbital-period distributions for the two groups of binary radio pulsars in Fig. 2.

The adopted "flat" initial distributions for the logarithm of the semimajor axis and the component-mass ratio for the close binaries [44] may provide another source of uncertainty. According to recent empirical estimates, these distributions can vary within a factor of two to three [45]. It is obvious that our model cannot be used directly for binaries in globular clusters, since the high stellar density there could substantially

alter the parameters of binaries with orbital periods exceeding ~30 days [46]. Future studies should improve the currently poor statistical data for binaries with neutron stars; last, but not least, the possible role of various observational selection effects must be taken into consideration.

ACKNOWLEDGMENTS

The author thanks G.S. Bisnovatyĭ-Kogan, R. Manchester, and B. Paczynski for the discussions of magnetic-rotational supernovae, empirical estimates of the birthrates of single and binary radio pulsars, and models of gamma-ray bursts. He also thanks L.R. Yungel'son for assistance in constructing the figures. This work was supported by the Program of Support for Leading Scientific Schools of the Russian Federation (grant no. NSh-162.2003.2), the Russian Foundation for Basic Research (project no. 03-02-16254), and the Basic Research Programs "Nonstationary Phenomena in Astronomy" (established by the Presidium of the Russian Academy of Sciences) and "Extended Objects in the Universe" (established by the Department of Physical Sciences of the Russian Academy of Sciences).

REFERENCES

1. L. R. Yungel'son and A. V. Tutukov, Astron. Zh. (2005) (in press).
2. A. V. Tutukov and L. R. Yungel'son, Astron. Zh. **79**, 738 (2002) [Astron. Rep. **46**, 667 (2002)].
3. E. Green and B. Chaboyer, Bull. Am. Astron. Soc. **30**, 1404 (1998).
4. Z. Han, P. Podsiadlowski, P. Maxted, and T. Marsh, Mon. Not. R. Astron. Soc. **341**, 669 (2003).
5. A. V. Tutukov and L. R. Yungel'son, Nauchn. Inform. Astrosovet Akad. Nauk SSSR **49**, 57 (1983).
6. I. Iben and A. Tutukov, Astrophys. J., Suppl. Ser. **54**, 335 (1984).
7. R. Webbink, Astrophys. J. **279**, 252 (1984).
8. R. Napiwotzki, C. Karl, and G. Nelemans, Rev. Mex. Astron. Astrofis. **20**, 113 (2004).
9. F. Pont, F. Bouchy, and C. Melo, astro-ph/0501615.
10. I. Iben, A. Tutukov, and L. Yungelson, Astrophys. J., Suppl. Ser. **100**, 233 (1995).
11. G. S. Bisnovatyĭ-Kogan and A. V. Tutukov, Astron. Zh. **81**, 797 (2004) [Astron. Rep. **48**, 724 (2004)].
12. G. S. Bisnovatyĭ-Kogan, Astron. Zh. **47**, 813 (1970) [Sov. Astron. **14**, 652 (1970)].
13. R. Manchester, A. Lyne, J. Taylor, *et al.*, Mon. Not. R. Astron. Soc. **185**, 409 (1979).
14. G. Fan, K. Cheng, and R. Manchester, Astrophys. J. **557**, 297 (2001).
15. Z. Arzoumanian, D. Chernov, and J. Cordes, Astrophys. J. **568**, 289 (2002).
16. N. Vranesevich, R. Manchester, D. Lorimer, *et al.*, Astrophys. J. Lett. **617**, L139 (2004).
17. S. Hyman, T. Lazio, N. Kassim, *et al.*, astro-ph/0503052.
18. A. V. Tutukov, N. N. Chugaĭ, and L. R. Yungel'son, Pis'ma Astron. Zh. **10**, 586 (1984) [Sov. Astron. Lett. **10**, 244 (1984)].
19. A. V. Tutukov and A. M. Cherepashchuk, Astron. Zh. **611**, 452 (2004) [Astron. Rep. **48**, 39 (2004)].
20. A. Heger, S. Woosley, and H. Spruit, astro-ph/0409422.
21. B. Paczynski, Astrophys. J. **494**, L45 (1998).
22. M. El Eid, B. Meyer, and L. The, Astrophys. J. **611**, 452 (2004).
23. A. V. Tutukov, Astron. Zh. **80**, 692 (2003) [Astron. Rep. **47**, 637 (2003)].
24. C. Firmani, V. Avila-Reese, G. Ghisellini, *et al.*, Astrophys. J. **611**, 1033 (2004).
25. H. Abt, H. Levato, and M. Grosso, Astrophys. J. **573**, 359 (2002).
26. T. Matheson, in *The Fate of the Most Massive Stars*, Ed. by R. Humphreys and K. Stanek (Astron. Soc. Pac., San Francisco, 2005); Astron. Soc. Pac. Conf. Ser. **332**, 416 (2005).
27. M. Burgay, N. D'Amico, A. Possenti, *et al.*, Nature **426**, 531 (2003).
28. J. Taylor, R. Manchester, and A. Lyne, Astrophys. J., Suppl. Ser. **88**, 529 (1993).
29. W. Johnstone, http://www.johnstonearchive.net/relativity/binpulstable.html (2005).
30. A. Lyne, F. Camilo, and R. Manchester, Mon. Not. R. Astron. Soc. **312**, 698 (2000).
31. H. Chaurasia and M. Bailes, astro-ph/0504021.
32. V. Kaspi, J. Lackey, J. Mattox, *et al.*, Astrophys. J. **528**, 445 (2000).
33. D. Champion, D. Lorimer, M. McLaughlin, *et al.*, Mon. Not. R. Astron. Soc. **350**, L61 (2004).
34. F. Camilo and F. Rasio, astro-ph/0501226.
35. A. Faulkner, M. Kramer, A. Lyne, *et al.*, Astrophys. J. **618**, L119 (2005).
36. R. Johnston, Astrophys. J. **575**, L119 (2002).
37. H. K. Chaurasia and M. Bailes, astro-ph/0504021.
38. S. Mao and B. Paczynski, Astrophys. J. **388**, L45 (1992).
39. A. Tutukov and L. Yungelson, Mon. Not. R. Astron. Soc. **260**, 675 (1993).
40. R. Voss and T. Tauris, Mon. Not. R. Astron. Soc. **342**, 1169 (2003).
41. J. Dewi and O. Pols, Mon. Not. R. Astron. Soc. **344**, 629 (2003).
42. V. Lipunov, A. Bogomazov, and M. Abubekerov, astro-ph/0503341.
43. T. Tauris and J. Dewi, Astron. Astrophys. **369**, 170 (2001).
44. E. I. Popova, A. V. Tutukov, and L. R. Yungel'son, Pis'ma Astron. Zh. **8**, 297 (1982) [Sov. Astron. Lett. **8**, 160 (1982)].
45. D. Hubber and A. Whitworth, astro-ph/0503412.
46. N. Ivanova, K. Belzynski, J. Fregeau, *et al.*, Mon. Not. R. Astron. Soc. **358**, 572 (2005).

Translated by K. Maslennikov