The Minimum Stellar Mass in Early Galaxies

E. O. Vasil'ev[1, 2] and Yu. A. Shchekinov[1]

[1]Rostov State University, Rostov-on-Don, Russia
[2]Physics Research Institute, Rostov State University, Rostov-on-Don, Russia
Received July 3, 2004; in final form, February 17, 2005

Abstract—The conditions for the fragmentation of the baryonic component during mergers of dark matter halos in the early Universe are studied. We assume that the baryonic component undergoes a shock compression. The characteristic masses of protostellar molecular clouds and the minimum masses of protostars originating in these clouds decrease with increasing halo mass. This may indicate that the initial stellar mass function in more massive galaxies was shifted towards lower masses during the initial stages of their formation. This would result in an increase in the number of stars per unit mass of the halo, i.e., in an increase in the efficiency of star formation. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In the hierarchical theory of the origin of structures in the Universe, gravitationally bound objects—dark matter halos, in which stars can form—first appear at redshifts $z \sim 20$ and have masses $M \sim 10^7 \, M_\odot$, including the dark and baryonic components [1−3]. These halos form due to multiple collisions and the merging of similar objects of smaller mass. The dynamics of the early stages of these processes are determined by the gravitational instability of the dark component, which consists of cool (nonrelativistic) particles. Density perturbations of virtually any wavelength can develop in the dark component, since the minimum critical scale of the perturbations—the Jeans mass, which depends on the velocity dispersion in a collisionless gas—is negligible compared with the mass of the halo (see, for example, [4]). The spatial scale of the perturbations can be constrained from below by a truncation of the perturbation spectrum. The amplitude of the density perturbations on small scales is known to decrease with wavelength according to a power law ($\propto \lambda^{-3}$) [5, 6]. Thus, the formation of structure includes relatively developed short-wavelength motions modulated by slowly growing long-wavelength flows [7].

Under these conditions, a nearly one-dimensional motion develops, in which perturbations are compressed mainly along a single direction, giving rise to a disklike configuration [8, 9]. Baryons confined by the gravitational field of the dark matter and involved in this one-dimensional flow undergo shock compression and are heated to the temperature $T_f \sim m_p v_c^2/2k$, where v_c is the velocity of the inward gas

flow; as a result, a dense layer or disk is formed. The compressed baryonic layer is efficiently cooled by radiative processes with neutral hydrogen atoms or molecules. The formation of molecular hydrogen is substantially accelerated behind shock fronts, so that the cooling of the gas can basically be governed by the H_2 molecules. When the temperature decreases down to 200 K, deuterium efficiently transforms into HD, which results in an even higher thermal-energy loss rate, providing favorable conditions for the fragmentation of the compressed gas [10−17]. Thus, during the formation of galaxies, encountering flows result in the formation of evolving, dense baryonic condensations in which stellar objects can subsequently form. This concept was recently discussed in [16, 17]. During the formation of massive protogalaxies, dense baryonic layers are maintained over longer times, so that the gas is cooled substantially and higher densities are reached. Therefore, one may expect that, in the formation of more massive galaxies, baryonic fragments of lower mass will be gravitationally unstable, and less massive stars will be formed, which increases the number of stars per unit of mass of the halo—in other words, the star-formation rate. Here, we study this possibility.

Section 2 discusses the model adopted to describe the thermal evolution of baryons behind shock fronts. Section 3 presents the results, and Section 4 contains our conclusions and final discussion. Our calculations were based on the ΛCDM model for the Universe: $(\Omega_0, \Omega_\Lambda, \Omega_m, \Omega_b, h) = (1.0, 0.71, 0.29, 0.047, 0.72)$, with the abundance of deuterium assumed to be $n(\text{D})/n = 2.6 \times 10^{-5}$ [18].

2. HALO VIRIALIZATION
AND THE THERMO-CHEMICAL
EVOLUTION OF GAS

In the linear stage of its growth, the evolution of a dark-matter density perturbation $\delta_{dm}(t)$ can be described by the equation (see, for example, [6])

$$\ddot{\delta}_{dm} + 2H\dot{\delta}_{dm} = \frac{3}{2}H^2(\Omega_b\delta_b + \Omega_{dm}\delta_{dm}), \quad (1)$$

where H is the Hubble constant, δ_b is the perturbation of the baryon density, and $\Omega_{dm} = \Omega_m - \Omega_b$. As we can see from this equation, there is no limiting minimum scale in dark matter. In reality, the Jeans mass in collisionless dark matter depends on the velocity dispersion; however, this value turns out to be negligibly small compared with the mass of the halo. As was already noted, if a density perturbation in dark matter is essentially super-Jeans, then its compression will proceed more rapidly along one direction [8, 9]. The dissipative baryonic component follows the dark matter potential and forms a disklike configuration in the symmetry plane. The growth of the perturbation is essentially a collision between flows of dark matter and baryons and results in the formation of a shock front in the baryonic component. Subsequently, the collisionless dark matter becomes virialized via violent relaxation [19], and the baryonic component undergoes shock compression in the colliding flows, forming a dense, radiatively cooled layer. Starting from some time, the increased density of the gas may exceed the density of the dark matter, and the dynamics of the gas will be determined by its own parameters. Therefore, we can neglect the influence of the dark matter and consider only the baryonic component. This is valid at least within the characteristic time for a collision between regions, $t_d \sim D/v$.

When gas flows collide, a thin dense gas layer and two diverging shock fronts around the contact region are initially formed; behind the shock fronts, the gas is heated to a temperature of $T_f \sim m_p v_c{}^2/2k$, where m_p is the proton mass and v_c the collisional velocity. The compression and formation of the dense layer lasts for approximately a time $t_d = D/v_c$. In the transverse direction, the gas is not restrained by dynamical pressure and can expand freely on a time scale $t_\perp \sim D/c_s$, where c_s is the sound speed. Thus, the ratio of the characteristic time scales is $t_d/t_\perp \simeq 1/\mathcal{M}$, where $\mathcal{M}$ is the Mach number, and we have for a collision between baryonic flows with velocities $v_c > c_s$ the condition $t_d < t_\perp$. Therefore, for supersonic collisions, we can neglect the transverse motion in the layer when $t < t_d$ and solve the one-dimensional problem. It was shown via two-dimensional modeling that, indeed, only an insignificant amount of the mass is lost to transverse outflow during a collision of gas clouds [20].

The velocity of the gas motions during the formation of a dark halo with mass M is

$$v_c = \sqrt{3}\sigma, \quad (2)$$

where

$$\sigma(M) = \sqrt{GM^{2/3}(3\pi^3\Omega_m\rho_0)^{1/3}(1+z)} \quad (3)$$

is the one-dimensional velocity dispersion and ρ_0 is the present mean total density of the Universe. It is clear that the collisional velocity and gas temperature behind the shock fronts will be higher when more massive halos are formed.

The thermal evolution of baryons behind a shock front can be described by a system of ordinary differential equations written for a single Lagrangian element of the fluid:

$$\dot{x} = \beta xn(1 - x - 2f) - k_1 nx^2, \quad (4)$$

$$\dot{f} = k_m n(1 - x - 2f)x, \quad (5)$$

$$\dot{g} = k_{D1} fxnd_c - nx(k_{D1}f + k_{D2})g, \quad (6)$$

$$\dot{T} = \frac{2}{3}\frac{\dot{n}}{n}T + \Sigma\Lambda_i, \quad (7)$$

where $x = n(e)/n$, $f = n(H_2)/n$, $g = n(HD)/n$ are the relative concentrations of electrons, H_2, and HD, $d_c = n(D)/n$ is the cosmological abundance of deuterium, k_i are the rates of reactions [3, 21], β is the rate of collisional ionization of hydrogen [14], Λ_i is the rate of cooling and heating due to Compton interactions with photons of the cosmic microwave background radiation (CMB), emission in atomic and molecular hydrogen lines [22], and also HD molecules [21, 23, 24]. Initially, the gas density behind the shock front is $n = 4n_0$, where n_0 is the density before the shock front. Further, we will assume that each element of the gas behind the shock front evolves isobarically and that the density is described by the expression

$$n = \frac{p}{\mu kT}, \quad (8)$$

where $\mu = \rho/nm_p$.

In the ΛCDM model, the spectrum of perturbations in the dark matter is rather steep ($n = -3$), which implies that halos with masses of $10^4 - 10^8 \, M_\odot$ condense out over times shorter than the corresponding Hubble time [25]. Some time after condensing, the halos become virialized. In addition, since they are involved in large-scale motions, they merge with each other, forming halos of higher mass. Both virialization and merging take comparable times. For this reason, two limiting possibilities can be suggested for the initial conditions for the halo formation. In the first, halos of smaller mass (subhalos) collide and soon reach the
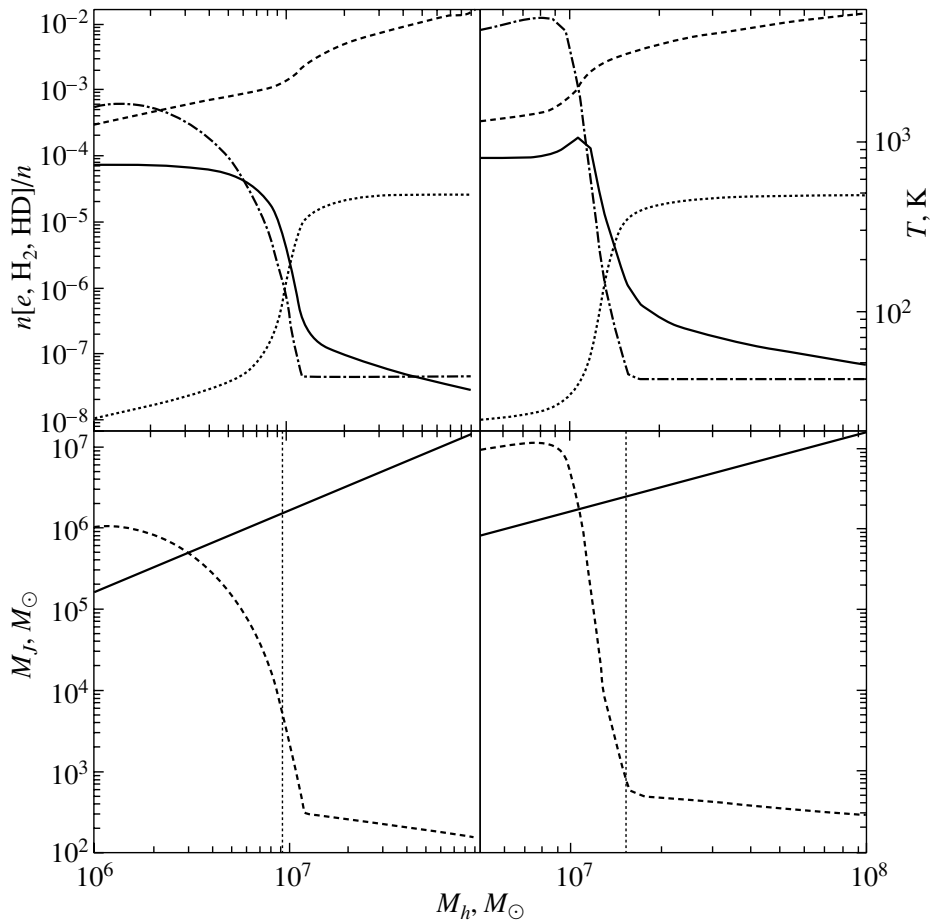
Fig. 1. Top: concentrations of electrons (solid curve), H_2 (dashed curve), and HD (dotted curve), and the temperatures (dot−dashed curve) reached in the gas during the virialization of a halo with mass M at $z = 15$. Bottom: total baryonic mass in the halo (solid curve) and the Jeans mass in the layer behind the shock front. The left and right graphs correspond to the first and second models for the initial conditions (see the text for more details).

virial state at larger redshifts. In the second, virialization and merging of subhalos occur at the same time at a given redshift. In both cases, the process can be presented as a one-dimensional compression. The two cases differ in the initial characteristics of the matter in the flow or, more precisely, in the density and the velocity of the collision. In the first model, a halo is formed due to collisions of subhalos, and the parameters of the matter correspond to those in the objects that have been virialized by the beginning of the formation of the larger halo (z_{ta}), i.e., by the time it separates from the cosmological expansion. The matter density can be taken to be $\rho \simeq 18\pi^2\rho_0(z_{ta})$ [25], and the relative concentrations of e, H_2, and HD to their values inside the subhalos. In the second model, subhalos are virialized and collide at the same redshift, and the initial density of the matter is $\rho \simeq 18\pi^2\rho_0(z_{vir})$, where $\rho_0(z)$ is the background density of matter at the redshift z; the relative concentrations of electrons, H_2, and HD are assumed to be equal to their values inside the halo. In both models, a colli-

sion results in the formation of a shock front, behind which gas is rapidly heated to the temperature $T_f = m_p v_c^2/3k$, which is taken as the initial temperature of an element of gas behind the front. Since the initial conditions in the models considered differ only in their densities, the results will be qualitatively similar. Therefore, we will describe only the first model, but present graphs for both. Calculations for halos with the masses corresponding to 3σ perturbations were carried out only for the second model.

3. RESULTS

3.1. Thermal and Chemical Evolution of Baryons

Let us investigate the chemical and thermal evolution of the gas behind the shock front using the example of halos formed at redshifts near $z = 15$. Figure 1 presents the concentrations of electrons, H_2, and HD and the temperatures reached in the gas behind the shock front at the end of one compression
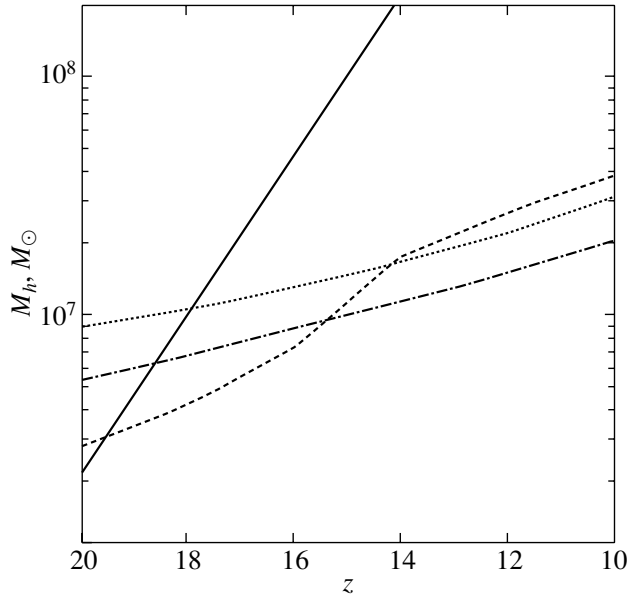
Fig. 2. The minimum halo mass for which the gas layer behind the shock front becomes unstable according to the criterion (9) during the formation of the halo for two halo-formation models (dot−dashed and dotted curves). The solid line indicates the mass of 3σ perturbations, and the dashed line the minimum mass for which baryons can be cooled and form gravitationally bound objects [3].

Fig. 3. Dependence of the Jeans mass behind the shock front on the time when the halo forms. The dashed and dotted lines correspond to the minimum halo masses needed in order for the layer behind the shock front to be unstable according to criterion (9) for the first and second models, respectively. The solid line corresponds to the mass of the 3σ perturbations.

time during the formation of a halo with mass M. It is clearly seen that, starting from some halo mass, the relative concentration of H_2 molecules exceeds 5×10^{-4} and the temperature of the gas decreases substantially ($T \sim 500$ K). This corresponds to the minimum H_2 concentration needed in order for the gas to cool rapidly, in one comoving Hubble time [3]. In more massive halos, the H_2 concentration continues to grow and can reach 10^{-2} for halos with $M_h \gtrsim 4 \times 10^7 M_\odot$. Radiative losses in H_2 lines cool the gas to $T \lesssim 200$ K, and all the deuterium rapidly becomes bound in HD molecules due to the effects of chemical fractionation. Due to strong cooling in HD lines, the temperature drops to several tens of Kelvin, which is close to the temperature of the CMB radiation at this redshift ($2.73\,(1+z)$ K). HD molecules provide efficient heat exchange between the CMB and baryons, due to the fact that the baryons absorb background photons and subsequently transfer the excitation energy to the gas through collisions [24, 26, 27]. Under these conditions, the Jeans mass behind the shock front (dashed curves in lower panels of Fig. 1) decreases and, starting from some value, becomes substantially lower than the baryonic mass of the forming halo ($\Omega_b M_h/\Omega_m$, solid curve). For ex-
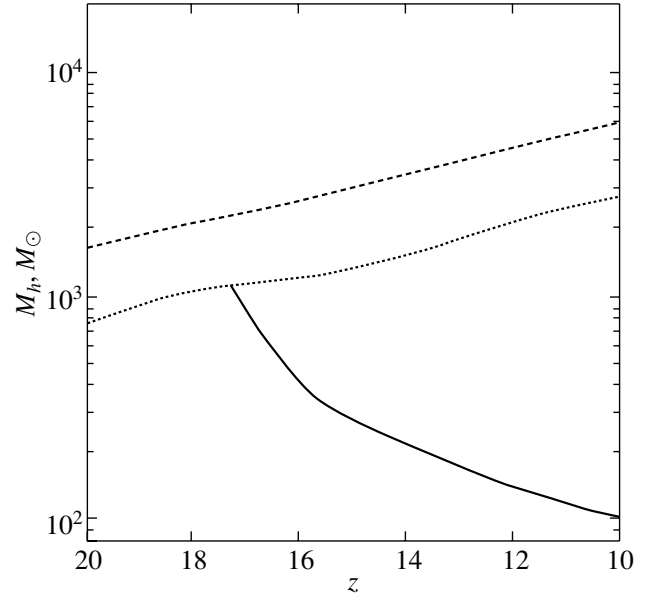
ample, for a halo with $M = 10^7 M_\odot$, the Jeans mass is $M_J \simeq 10^3 M_\odot$.

Let us now consider the instability of the cool compressed layer and the possible formation of dense baryonic condensations in it. We will assume that the gas in the layer is gravitationally unstable and can fragment, provided the critical perturbation length λ_m is shorter than the initial size of the cloud D and that the corresponding time t_m is shorter than the compression time [28, 29]:

$$\lambda_m/D \leq 1, \quad t_m/t_d \leq 1. \tag{9}$$

The vertical line in Fig. 1 indicates the minimum halo mass for which the criterion (9) is fulfilled during its formation. Thus, the gas behind the shock front originating during the formation of a halo with mass $M \geq 10^7 M_\odot$ at a redshift $z \simeq 15$ is unstable by the end of the compression phase; accordingly, the first baryonic objects with masses roughly equal to the Jeans mass $M_J \leq 10^3 M_\odot$ can form inside the halo. Since the temperature of the gas in the unstable fragments is lower than 200 K, the concentration of HD also increases substantially, thereby determining the subsequent thermal evolution of the gas behind the shock fronts forming during the formation of the first halos.

Figure 2 shows the relation between the redshift and the minimum halo mass for which the layer behind the shock front becomes unstable according to criterion (9) during its formation. For comparison, the mass of 3σ perturbations is displayed, as well as the minimum mass for which baryons can cool and form gravitationally bound objects (the latter value was derived in [3]). A gravitationally unstable gas layer is formed during the virialization behind the shock front in a halo with mass $M_{\min} \simeq 5 \times 10^6 \, M_\odot$ at $z = 20$. This derived minimum halo mass is comparable to that obtained in [3]; however, the temperature of the gas is substantially lower in our model.

Baryonic objects with masses approximately equal to the Jeans mass can condense inside this unstable layer. Figure 3 displays the dependence of the Jeans mass behind the shock front on the time at which the halo forms; the dashed line corresponds to the minimum halo mass needed in order for the layer behind the shock front to be unstable according to (9), and the solid curve, to the mass of 3σ perturbations. It is seen that, at the given redshift, the Jeans mass in the layer formed behind the shock front decreases with increasing halo mass: the conditions are favorable for forming fragments with lower masses in more massive halos. For example, in the formation of a halo with mass $10^7 \, M_\odot$ (which corresponds to the minimum halo mass) at redshift $z = 15$, the layer behind the shock front becomes unstable, and the baryonic objects formed there can have masses of $M_b \sim 3 \times 10^3 \, M_\odot$, while $M_b \sim 300 \, M_\odot$ for a halo with mass $10^8 \, M_\odot$. Fragments with masses as low as $M_b \sim 100 \, M_\odot$ will be unstable in a halo with mass $3 \times 10^9 \, M_\odot$ formed at $z = 10$. The possible masses of gravitationally unstable baryonic condensations lie between the above limits which widen with decreasing z.

During the formation of a massive halo, the gas behind the shock front cools and is compressed more intensely, due to the more efficient formation of H_2 molecules behind more intense shock fronts. Figure 4 presents the dependence of the gas density on the mass of the halo and the redshift at which it begins to form. The variation of the slope of the $n(M)$ curves with increasing M reflects the fact that, by the end of the compression phase, the temperature behind the shock front reaches the temperature of the CMB, due to the efficient formation of HD molecules and energy exchange between the gas and CMB radiation via the absorption of background photons by HD molecules followed by collisional de-excitation. The final density at the given z depends only on the initial temperature of the gas behind the shock front, which, in turn, is a function of the halo mass: $n = n_0(T_f/T_{CMB})$.
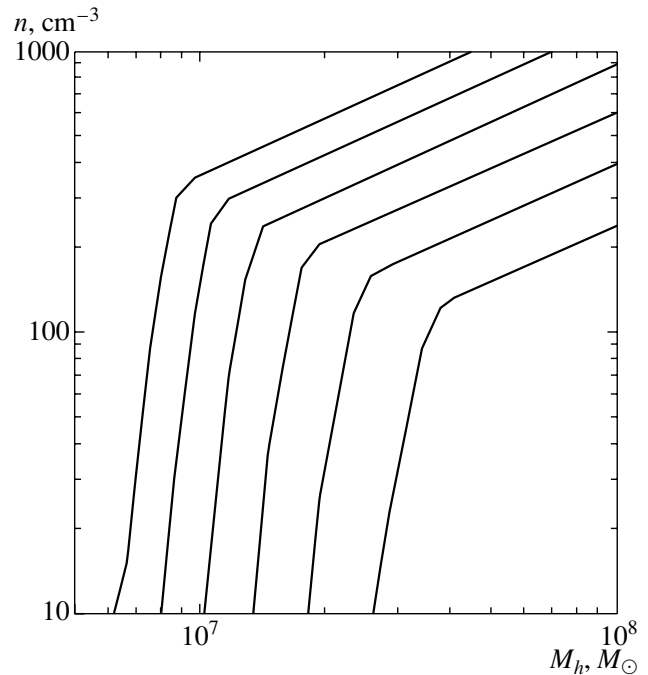

Fig. 4. Dependence of the final density of the gas behind the shock fronts on the halo mass and the initial redshift at which the halo forms for $z = 20, 18, 16, 14, 12, 10$ (curves from left to right, respectively).

3.2. The Minimum Mass of Baryonic Objects

Baryonic fragments with masses $10^2 - 10^4 \, M_\odot$ (Fig. 3) formed behind the shock front in an unstable layer are optically transparent to radiation in H_2 and HD lines. Compression of the fragments occurs in an isothermal regime at a temperature close to $2.73(1 + z)$ K, as is mainly ensured by the HD molecules.

In the isothermal regime, a characteristic self-similar density profile is formed, $\rho \sim r^{-2}$ [30, 31], and, as a result, the optical depth in the HD lines increases in the central regions, and the Jeans mass gradually decreases. The critical point is the formation of an optically thick ($\tau \geq 1$) core region with a mass equal to or exceeding the Jeans mass. In the central regions when $\tau \sim 1$, the gas density is high ($\sim 10^9 - 10^{10}$ cm^{-3}), and all the hydrogen is transformed into the molecular phase via three-particle reactions. However, due to the low temperature of the gas, the contribution of H_2 molecules to the cooling in optically transparent regions—and, therefore, their impact on the dynamics of the compression—will be unimportant. The formation of an opaque core will be fully determined by the HD molecules in the gas. After the formation of an opaque, gravitationally unstable core, the compression changes from being isothermal to being adiabatic. Figure 5 presents the depen-
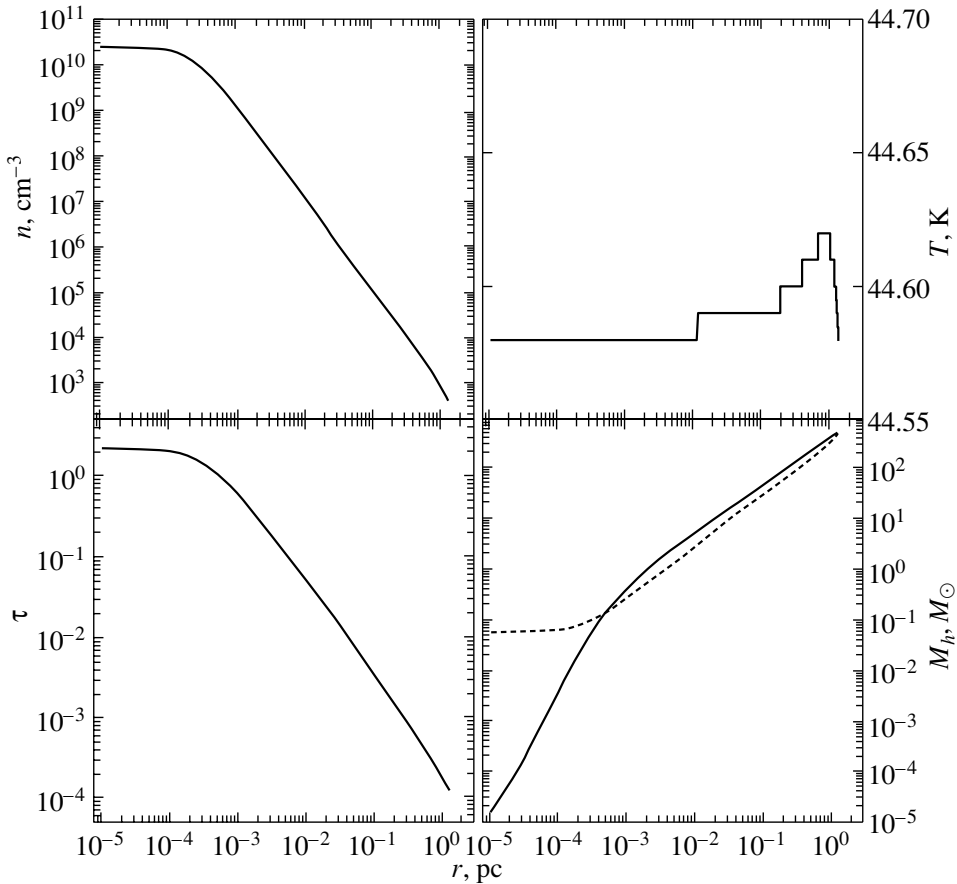
Fig. 5. Density, optical depth, temperature, and Jeans mass inside a cloud at the time of formation of an opaque gravitationally unstable core ($z \simeq 15.3$) for a fragment with mass $M = 500\,M_\odot$ formed due to instability of the layer during the formation of a halo with $M_h = 2 \times 10^7\,M_\odot$ at $z_v = 17$. The dashed line in the lower right graph indicates the mass inside the corresponding radius.

dences of the density, optical depth, temperature, and Jeans mass inside the cloud when this state has been reached by $z \simeq 15.3$, for a fragment with mass $M = 500\,M_\odot$ formed as a result of instability in the layer during the formation of a halo with $M_h = 2 \times 10^7\,M_\odot$ at $z_v = 17$. It is clear that the temperature is almost constant and close to that of the CMB throughout the cloud. The formed opaque core has a density of $\geq 10^{10}$ cm^{-3} and a mass of $\sim 0.15\,M_\odot$.

Figure 6 presents the dependence of the mass of the opaque core on the mass of the halo for the limiting masses: the minimum mass $M_{\min}$ needed for the layer behind the shock fronts to be unstable according to (9) for both models of the initial conditions and the mass corresponding to 3σ perturbations, $M_{3\sigma}$. The numbers in Fig. 6 denote the redshift at which the halo is virialized. It is seen that the mass of the opaque core decreases as the mass of the halo increases: the mass of the core for a halo mass of $M_h \simeq 2 \times 10^7\,M_\odot$ at $z = 17$ is $M(\tau \geq 1) \sim 0.15\,M_\odot$, while the mass of the core is $M(\tau \geq 1) \sim 0.06\,M_\odot$ for a halo mass $M_h \simeq 10^9\,M_\odot$ at $z = 12$.

Thus, the formation of dark halos results in the formation of intense shock fronts in the baryonic component, substantial cooling, and an increase in the gas density. The cold postshock gas layer is unstable against formation of baryonic condensations in which protostellar cores in turn be formed. The higher the mass of the initial halo, the lower the characteristic mass of the gravitationally unstable fragment—a possible protostellar cluster or opaque protostellar core.

3.3. Thermal Instability

The development of thermal instability becomes possible in a radiatively cooling gas under certain cooling regimes [32]. In the non-steady-state case, when radiative losses are not balanced by heating, the condition for this is

$$\frac{d\ln\Lambda}{d\ln T} < 2, \tag{10}$$

where Λ is the effective radiative-loss function (see, for example, [33]). Figure 7 shows the temperature
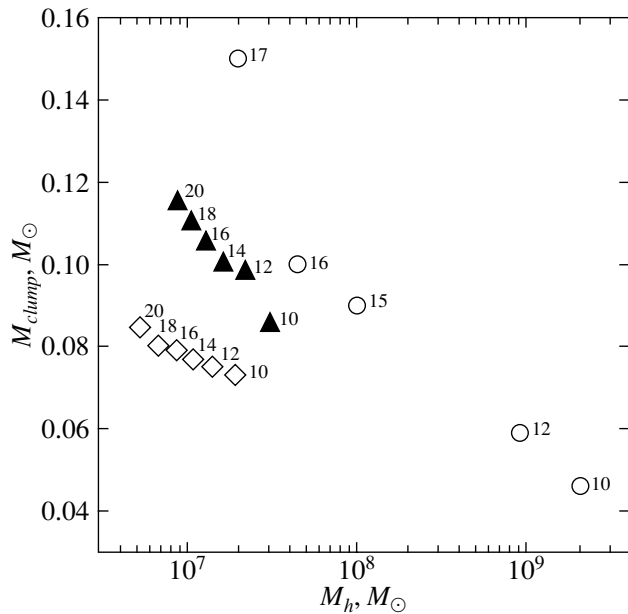
Fig. 6. Dependence of the mass of the opaque core on the mass of the halo. The case of the minimum mass M_{min} needed for the layer behind the shock front to be unstable according to (9) for the first and second models for the initial conditions are indicated by hollow diamonds and filled triangles, respectively. The case of the mass corresponding to 3σ perturbations, $M_{3\sigma}$, is indicated by hollow circles. The numbers near the symbols denote the redshift of the halo virialization.

Fig. 7. Temperature dependence of the cooling rate and its logarithmic derivative $\Phi = d\ln\Lambda/d\ln T$ for various approximations for the radiative losses in HD lines.

dependence of the function $d\ln\Lambda/d\ln T$ in the whole range of temperatures encompassed by the gas cooling behind the shock front. It is obvious that thermal instability can develop at all points where the cooling is determined by H_2 and HD molecules, i.e., when $70\ \mathrm{K} < T < 8000\ \mathrm{K}$. To order of magnitude, the characteristic size of the region of instability coincides with the size of the region behind the shock front where the gas is cooled from $T = T_f$. Therefore, it is obvious that the forming condensations will facilitate the subsequent gravitation fragmentation of the compressed layer.

4. DISCUSSION AND CONCLUSIONS

Collisions of baryonic flows during the formation of the first protogalaxies are accompanied by intense cooling of the gas, which promotes the fragmentation of the gas into condensations with characteristic masses that are close to the masses of the present-day molecular clouds. The subsequent cooling and compression of such condensations can be accompanied by star formation. The higher the mass of the galaxy, the more intense the gas cooling and the smaller the mass of the molecular cloud and the minimum mass of stars formed as a result of its
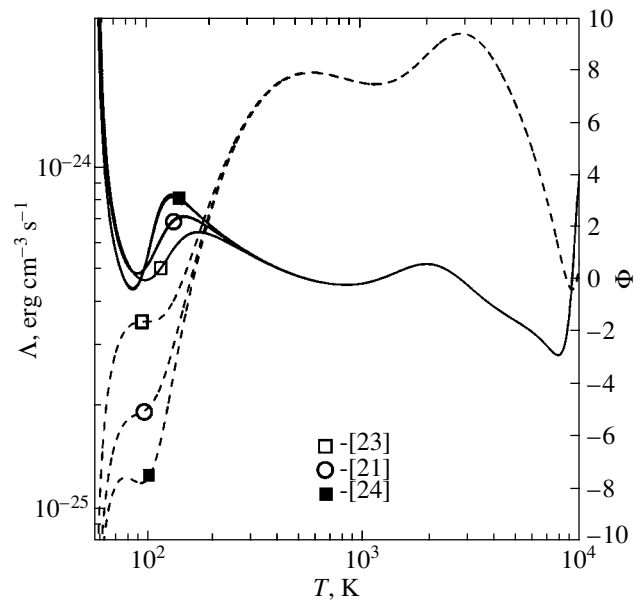
subsequent fragmentation. Thus, in this picture, we expect that the initial mass function of stars and initial luminosity function of stellar clusters will be shifted towards smaller values in more massive galaxies.

The formed low-mass fragments cool rapidly, due to the high abundances of H_2 and HD molecules. Starting from some time, a protostellar cloud switches to an isothermal compression phase, since the gas temperature is maintained near the temperature of the cosmic microwave background, due to efficient heat exchange between the CMB radiation and baryons, which absorb background photons and then transfer the excitation energy to the gas in collisional processes [24, 26, 27]. Further, as the density in the central regions of the fragment increases, an opaque core with a mass of $(10^{-1}-10^{-2})\ M_{\odot}$ is formed. This core evolves then into a hydrostatic protostellar core with a lower mass of the order of $10^{-3}\ M_{\odot}$ [30, 31], onto which matter gradually accretes, with the accretion rate determining the final mass of the star [34]. In this process, practically all of the total gravitational energy of the accreted gas is radiated in H_2 and HD lines. A possibility to observe H_2 line emission was discussed in [35–38]; it would be extremely difficult to detect this emission with either currently operating or planned telescopes [38]. For example, if ~ 2000 protostellar objects are formed behind the shock front during the virialization of a halo with mass $3 \times 10^7\ M_{\odot}$ (which corresponds to approximately 10% of the mass of baryons being transformed into

protostellar fragments), the luminosity of this cluster in the H_2 2.34-μm line will be $\sim 10^{38}$ erg/s. With a spectral resolution of $R \sim 1000$, this corresponds to a flux from the object of $\sim 10^{-2}$ μJy at redshift $z = 15$, while the sensitivity of the next planned SAFIR space telescope[1] is 1 μJy.

ACKNOWLEDGMENTS

REFERENCES

1. T. Abel, P. Anninos, Y. Zhang, and M. L. Norman, New Astron. **2**, 181 (1997).
2. V. Bromm, P. Coppi, and R. Larson, Astrophys. J. **564**, 23 (2002).
3. M. Tegmark, J. Silk, M. J. Rees, *et al.*, Astrophys. J. **474**, 1 (1997).
4. L. P. Grishchuk and Ya. B. Zel'dovich, Astron. Zh. **58**, 472 (1981) [Sov. Astron. **25**, 267 (1981)].
5. P. J. E. Peebles, Astrophys. J. **263**, L1 (1982).
6. E. Kolb and M. Turner, *The Early Universe* (Addison-Wesley, Readwood City, 1990).
7. P. J. E. Peebles, *Principles of Physical Cosmology* (Princeton Univ. Press, Princeton, 1993).
8. C. Lin, L. Mestel, and F. Shu, Astrophys. J. **142**, 1431 (1965).
9. Ya. B. Zel'dovich, Astrofiz. **6**, 119 (1970).
10. J. Smith, Astrophys. J. **238**, 842 (1980).
11. A. A. Suchkov, Yu. A. Shchekinov, and M. A. Édel'man, Astrofiz. **18**, 629 (1982) [Astrophys. **18**, 360 (1982)].
12. C. Struck-Marcell, Astrophys. J. **259**, 116 (1982).
13. C. Struck-Marcell, Astrophys. J. **259**, 127 (1982).
14. P. R. Shapiro and H. Kang, Astrophys. J. **318**, 32 (1987).
15. Yu. A. Shchekinov, Astrophys. Space Sci. **175**, 57 (1991).
16. M. Yamada and R. Nishi, Astrophys. J. **505**, 148 (1998).
17. R. Cen, Astrophys. J. (2005) (in press); astro-ph/0311329 (2003).
18. D. N. Spergel, L. Verde, H. V. Peiris, *et al.*, Astrophys. J., Suppl. Ser. **148**, 175 (2003).
19. D. Lynden-Bell, Mon. Not. R. Astron. Soc. **136**, 101 (1967).
20. F. Miniati, T. W. Jones, A. Ferrara, and D. Ryu, Astrophys. J. **491**, 216 (1997).
21. D. Galli and F. Palla, Astron. Astrophys. **335**, 403 (1998).
22. D. Hollenbach and C. F. McKee, Astrophys. J., Suppl. Ser. **41**, 555 (1979).
23. D. Flower, Mon. Not. R. Astron. Soc. **318**, 875 (2000).
24. D. Puy and M. Signore, New Astron. **3**, 247 (1998).
25. R. Barkana and A. Loeb, Phys. Rep. **349**, 125 (2001).
26. D. A. Varshalovich and V. K. Khersonskii, Pis'ma Astron. Zh. **2**, 574 (1976) [Sov. Astron. Lett. **2**, 227 (1976)].
27. D. Galli and F. Palla, Planet. Space Sci. **12−13**, 1197 (2002).
28. M. Stone, Astrophys. J. **159**, 277 (1970).
29. D. Gilden, Astrophys. J. **279**, 335 (1984).
30. K. Omukai and R. Nishi, Astrophys. J. **508**, 141 (1998).
31. E. Ripamonti, F. Haardt, A. Ferrara, *et al.*, Mon. Not. R. Astron. Soc. **334**, 401 (2002).
32. G. B. Field, Astrophys. J. **142**, 531 (1965).
33. Yu. A. Shchekinov, Astron. Zh. **55**, 311 (1978) [Sov. Astron. **22**, 182 (1978)].
34. K. Omukai and F. Palla, Astrophys. J. **589**, 677 (2003).
35. H. Kamaya and J. Silk, Mon. Not. R. Astron. Soc. **332**, 251 (2002).
36. H. Kamaya and J. Silk, Mon. Not. R. Astron. Soc. **339**, 1256 (2003).
37. K. Omukai and T. Kitayama, Astrophys. J. **599**, 738 (2003).
38. H. Mizusawa, R. Nishi, and K. Omukai, Publ. Astron. Soc. Jpn. (2005) (in press); astro-ph/0404333 (2004).

Translated by K. Maslennikov

[1]http://safir.jpl.nasa.gov

Determination of the Chemical Composition in HII Regions of Blue Compact Dwarf Galaxies Using an Optimized Photoionization Modeling Method

V. V. Holovatyi and B. Ya. Melekh

Chair of Astrophysics, Ivan Franko L'vov National University, ul. Kyryla i Mefodiya 8, L'vov, 79005 Ukraine
Received August 1, 2004; in final form, February 17, 2005

Abstract—Optimized photoionization models of selected HII regions in Blue Compact Dwarf Galaxies are calculated with the aim of determining their chemical composition. The stability of the optimized-photoionization calculations and means of deriving the most accurate chemical abundances are checked. Initialization of the free physical parameters using different data sets, deviations from spherical geometry, and the introduction of gas-density fluctuations in an HII region do not affect the accuracy of the derived chemical compositions. The chemical abundances found for 12 HII regions using the optimized-photoionization models are used to derive the primordial helium abundance Y_p and its enrichment dY/dZ. The obtained values of Y_p and dY/dZ are close to values found previously using ionization-correction factors. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In our first work in this series [1], we determined the elemental abundances for 43 HII regions in Blue Compact Dwarf Galaxies (BCDGs) using ionization-correction factors found from a grid of photoionization models for the emission of these objects [2]. We derived the primordial helium abundance $Y_p = 0.2440 \pm 0.0024$ and rate of its enrichment $dY/dZ = -4.02 \pm 2.46$. However, we adopted certain simplifications in the calculation of the model grid: (a) the He/H abundance was assumed to be constant; (b) the spectral energy distribution of the ionizing sources at $\lambda \leq 912$ Å was taken to be the minimum, mean, or maximum value on the energy scale for three objects; (c) we neglected the decrements in the abundances of different elements. This gave rise to the need to check the effect of these assumptions on the resulting Y_p and dY/dZ values.

Given the importance of determining the helium abundance, we further considered ways to take into account various factors affecting the derived He/H abundances [3]. As a result, we redetermined the He/H abundances for 28 HII regions, obtaining the values $Y_p = 0.244 \pm 0.004$ and $dY/dZ = 8.8 \pm 4.6$. The abundances of heavy elements were taken from [1].

In the current paper, we propose a new method for determining the chemical compositions of HII regions in BCDGs based on the calculation of optimized-photoionization (OP) models for the emission of these objects. The technique and algorithm used in the OP calculations, procedure used to select the HII regions, and verification of the stability of the solutions are described in Section 2. The derived Y_p and dY/dZ values and a comparison with other such results are given in Sections 3 and 4.

2. TECHNIQUE FOR CALCULATING OPTIMIZED-PHOTOIONIZATION MODELS OF HII REGIONS IN BLUE COMPACT DWARF GALAXIES

The goal of the OP calculations is to search for values of the free parameters such that the calculated model spectrum of an HII region is as close as possible to the observed spectrum (the relative intensities of the lines and Hβ line luminosity). The criterion for agreement between the model and observed spectra is the so-called χ^2 function, which, for the ith quantity in the photoionization model is [4]

$$\chi_i^2 = \left(\frac{\text{Obs}_i - \text{Mod}_i}{\sigma} \right)^2, \qquad (1)$$

where Obs_i and Mod_i are the observed and model values of the corresponding quantity, and σ is the standard deviation of the observed value. As in [5], we will use the average values of χ^2 for the relative intensities of emission lines in the OP calculations. The σ for the logarithm of the Hβ line luminosity of an HII region was always taken to be 0.05 (5%). The essence of the optimization algorithm (or χ^2 minimization) consists in varying the values of the free

Table 1. Comparison of free parameters obtained from the OP calculations applying the optimization to all selected line intensities (see text) and to only one line for each ion

Parameter	All lines in OP model	One line per ion in OP model
n_H, cm^{-3}	176^{+39}_{-19}	161^{+40}_{-14}
$\log Q_{\text{tot}}$	53.87	53.87
ϵ	$0.0059^{+0.0003}_{-0.0004}$	$0.0064^{+0.0036}_{-0.0008}$
He/H, 10^{-2}	$8.42^{+0.57}_{-0.72}$	$8.47^{+0.67}_{-0.71}$
O/H, 10^{-5}	$9.53^{+0.48}_{-0.46}$	$9.57^{+0.31}_{-1.67}$
N/H, 10^{-6}	$4.39^{+0.79}_{-0.67}$	$4.49^{+0.16}_{-1.09}$
Ne/H, 10^{-5}	$1.64^{+0.09}_{-0.12}$	$1.63^{+0.47}_{-0.06}$
S/H, 10^{-6}	$3.21^{+0.22}_{-0.16}$	$3.10^{+0.22}_{-0.80}$
Ar/H, 10^{-7}	$4.44^{+0.80}_{-1.08}$	$4.37^{+0.42}_{-0.39}$
Fe/H, 10^{-6}	$1.25^{+0.73}_{-0.46}$	$1.28^{+6.65}_{-0.47}$

parameters of the model until the χ^2 function reaches a minimum.

We selected the HII regions 0917+527, 0926+606, 0940+544N, 0948+532, 1054+365, 1135+581, 1152+579, 1211+540, 1256+351, 1533+574A for the OP calculations, taking their spectra from [6, 7]. The selection of these objects was determined by the fact that their elemental abundances (see [1]) reproduce the same values of Y_p and dY/dZ as for all the objects in the sample within the errors (Fig. 1). We chose four HeI lines for the analysis.

Note that we used the same observed spectra in the OP calculations as we did when finding the chemical compositions in [1]. For each of the objects, we recalculated the L_c spectrum using a modified version of our NLEHII code [8].
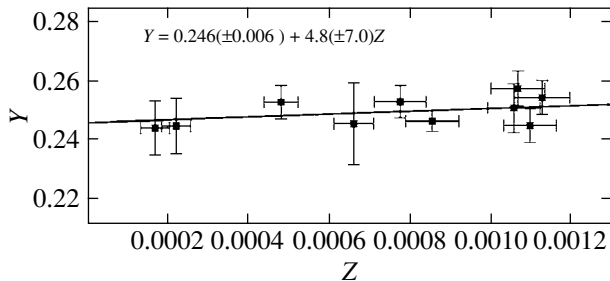

Fig. 1. $Y-Z$ dependence for data calculated using a photoionization-model grid for 10 objects selected for the OP calculations. Both Y_p and dY/dZ coincide with the corresponding values obtained for the entire sample of objects within the errors ($Y_p = 0.244 \pm 0.004$, $dY/dZ = 8.8 \pm 4.6$).

We optimized the resulting L_c spectra to reproduce the relative intensities of the HeII $\lambda4686$/Hβ lines using the method described in [2]. This yielded the L_c spectra for the ten selected and two additional (see below) HII regions (Fig. 2), which we used in the OP calculations for these objects.

We calculated the photoionization models for the OP calculations using the CLOUDY 96 photoionization code [4] (http://www.pa.uky.edu/~gary/cloudy) and carried out the optimization using the PHYMIR code [5], which is included as a function in CLOUDY 96.

We chose as free parameters in the OP calculations

(1) the total number of ionizing photons Q_{tot},
(2) the hydrogen density n_H (50−1000 cm^{-3}),
(3) the filling factor ϵ (0.00001−0.1),
(4) He/H (0.05−0.11),
(5) O/H (8.0×10^{-6}−1.4×10^{-4}),
(6) N/H (3.0×10^{-7}−6.5×10^{-6}),
(7) Ne/H (2.0×10^{-6}−4.0×10^{-5}),
(8) S/H (6.0×10^{-7}−4.0×10^{-6}),
(9) Ar/H (8.0×10^{-8}−8.0×10^{-7}),
(10) Fe/H.

The ranges of the variations of these parameters are given in parentheses. We adopted average values between the maximum and minimum as the initial values.

The parameters used to calculate χ^2 were the Hβ line luminosity and the relative intensities of the [OII]
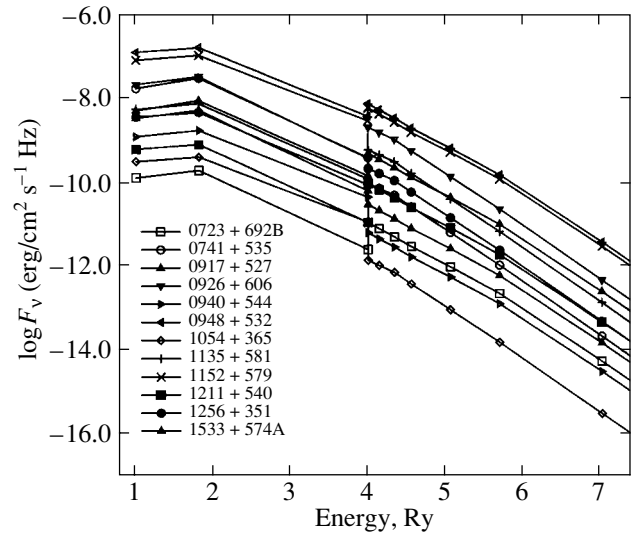

Fig. 2. L_c spectra obtained for the OP calculations for 12 HII regions in BCDGs. All fluxes are given at the inner boundary of the model HII region.

Table 2. Free parameters obtained from the OP calculations

Parameter	Objects					
	0917+527	0926+606	0940+544N	0948+532	1054+365	0723+692B*
n_H, cm^{-3}	116^{+17}_{-5}	$81.5^{+2.3}_{-1.9}$	235^{+89}_{-5}	179.7^{+24}_{-22}	494^{+15}_{-11}	291^{+33}_{-13}
$\log Q_{\mathrm{tot}}$	52.37	53.18	51.89	53.87	51.28	50.94
ϵ	$0.0255^{+0.0005}_{-0.0017}$	0.0147^{+0}_{-0}	$0.099^{+0.001}_{-0.012}$	$0.0058^{+0.0004}_{-0.0003}$	$0.0997^{+0}_{-0.0022}$	$0.088^{+0.004}_{-0.006}$
He/H, 10^{-2}	$8.31^{+0.53}_{-0.48}$	$8.55^{+0.35}_{-0.40}$	$8.32^{+0.00}_{-0.11}$	$8.40^{+0.59}_{-0.71}$	$8.37^{+0.14}_{-0.01}$	$8.22^{+0.46}_{-0.61}$
O/H, 10^{-5}	$9.43^{+0.18}_{-0.20}$	$9.56^{+0.20}_{-0.01}$	$4.31^{+0.00}_{-0.08}$	$9.52^{+0.43}_{-0.45}$	$8.85^{+0.00}_{-0.01}$	$7.25^{+0.18}_{-0.17}$
N/H, 10^{-6}	$2.78^{+0.08}_{-0.09}$	$4.38^{+0.22}_{-0.24}$	$1.27^{+0.22}_{-0.26}$	$4.40^{+0.78}_{-0.68}$	$3.62^{+0.44}_{-0.39}$	$2.47^{+0.46}_{-0.39}$
Ne/H, 10^{-5}	$1.99^{+0.06}_{-0.06}$	$1.80^{+0.06}_{-0.07}$	$0.773^{+0.001}_{-0.001}$	$1.65^{+0.09}_{-0.13}$	$1.50^{+0.07}_{-0.07}$	$1.06^{+0.05}_{-0.07}$
S/H, 10^{-6}	$3.32^{+0.13}_{-0.14}$	$4.00^{+0.00}_{-0.08}$	$0.99^{+0.13}_{-0.10}$	$3.22^{+0.22}_{-0.17}$	$2.56^{+0.26}_{-0.16}$	$2.34^{+0.16}_{-0.14}$
Ar/H, 10^{-7}	$2.60^{+0.42}_{-0.52}$	$3.59^{+0.35}_{-0.20}$	$1.28^{+0.20}_{-0.24}$	$4.37^{+0.86}_{-1.06}$	$4.16^{+0.25}_{-0.27}$	$3.12^{+0.19}_{-0.21}$
Fe/H, 10^{-6}	$1.20^{+0.71}_{-0.44}$	$1.33^{+0.78}_{-0.58}$	$0.75^{+0.46}_{-0.27}$	$1.26^{+0.71}_{-0.47}$	$0.18^{+0.03}_{-0.07}$	$0.12^{+0.03}_{-0.04}$
Parameter	Objects					
	1135+581	1152+579	1211+540	1256+351	1533+574A	0741+535*
n_H, cm^{-3}	$124^{+5.8}_{-6.6}$	$143^{+11}_{-3.36}$	$498^{+0.0}_{-19}$	$127^{+2.7}_{-0.1}$	$176^{+9.36}_{-8.89}$	284^{+32}_{-30}
$\log Q_{\mathrm{tot}}$	52.54	53.92	51.58	52.33	53.10	53.14
ϵ	$0.0321^{+0.0003}_{-0.0017}$	$0.0129^{+0.0010}_{-0.0002}$	$0.1000^{+0}_{-0.0015}$	$0.0769^{+0.00004}_{-0.00081}$	$0.0041^{+0.0001}_{-0}$	$0.0033^{+0.0002}_{-0.0002}$
He/H, 10^{-2}	$8.64^{+0.22}_{-0.21}$	$8.71^{+0.27}_{-0.27}$	$7.95^{+0.21}_{-0.003}$	$8.55^{+0.11}_{-0.14}$	$7.94^{+0.20}_{-0.20}$	$7.99^{+0.99}_{-1.04}$
O/H, 10^{-5}	$8.35^{+0.18}_{-0.04}$	$8.54^{+0.26}_{-0.13}$	$5.71^{+0.00}_{-0.00}$	$9.37^{+0.06}_{-0.01}$	$8.99^{+0.17}_{-0.04}$	$12.1^{+0.42}_{-0.48}$
N/H, 10^{-6}	$3.95^{+0.10}_{-0.09}$	$3.72^{+0.50}_{-0.48}$	$1.52^{+0.10}_{-0.11}$	$4.02^{+0.09}_{-0.03}$	$4.18^{+0.10}_{-0.11}$	$4.56^{+0.44}_{-0.53}$
Ne/H, 10^{-5}	$1.58^{+0.04}_{-0.04}$	$1.55^{+0.08}_{-0.05}$	$0.96^{+0.05}_{-0.06}$	$1.47^{+0.01}_{-0.02}$	$1.46^{+0.04}_{-0.04}$	$1.90^{+0.14}_{-0.15}$
S/H, 10^{-6}	$2.67^{+0.07}_{-0.06}$	$2.06^{+0.15}_{-0.11}$	$1.51^{+0.14}_{-0.11}$	$3.25^{+0.06}_{-0.04}$	$3.51^{+0.07}_{-0.10}$	$4.00^{+0.00}_{-0.14}$
Ar/H, 10^{-7}	$4.47^{+0.24}_{-0.34}$	$2.94^{+0.21}_{-0.37}$	$1.85^{+0.12}_{-0.22}$	$4.27^{+0.06}_{-0.08}$	$3.92^{+0.39}_{-0.35}$	$4.43^{+0.64}_{-0.91}$
Fe/H, 10^{-6}	$2.05^{+0.23}_{-0.73}$	$1.61^{+0.96}_{-0.59}$	$0.18^{+0.02}_{-0.01}$	$2.74^{+0.05}_{-0.02}$	$1.58^{+0.97}_{-0.61}$	$1.53^{+0.88}_{-0.57}$

* Additional objects that are not included in the sample of 10 selected HII regions.

$\lambda 3727/\mathrm{H}\beta$, [NeIII] $\lambda 3869/\mathrm{H}\beta$, [OIII] $\lambda 4363/\mathrm{H}\beta$, HeI $\lambda 4471/\mathrm{H}\beta$, HeII $\lambda 4686/\mathrm{H}\beta$, [FeIII] $\lambda 4658/\mathrm{H}\beta$, [OIII] $\lambda 4959/\mathrm{H}\beta$, [OIII] $\lambda 5007/\mathrm{H}\beta$, HeI $\lambda 5876/\mathrm{H}\beta$, HeI $\lambda 6678/\mathrm{H}\beta$, HeI $\lambda 7065/\mathrm{H}\beta$, [SIII] $\lambda 6312/\mathrm{H}\beta$, [NII] $\lambda 6584/\mathrm{H}\beta$, [SII] $\lambda 6717/\mathrm{H}\beta$, [SII] $\lambda 6731/\mathrm{H}\beta$, and [ArIII] $\lambda 7136/\mathrm{H}\beta$ lines (i.e., 17 parameters).

To determine the errors in the best-fit parameter values, we must find the value of $\Delta\chi^2$, which, as is known, is determined by the number of degrees of freedom of the fit [9]. Our fits have $17 - 10 = 7$ degrees of freedom. However, in the calculations, the intensity of one of the lines of an ion is related to intensities of all other lines of this ion. Thus, the number of degrees of freedom of the problem decreases, and becomes equal to one. In this case, if we calculate two OP models with identical initial values of the free

parameters but calculate the χ^2 value using all the above-listed lines in one and using the intensity of only one line for each ion in the other, the obtained best-fit free-parameter values in the two OP models should be the same within the errors. The results of such calculations are listed in Table 1, from which we can see that the optimal values of the free parameters obtained from the calculations of the two OP models coincide within the errors. Therefore, $\Delta\chi^2 = 1$ [9]. Thus, the errors of the free parameters obtained in our OP calculations will be determined from the maximum deviation of each free parameter from its final value using the photoionization-model analysis with $\chi^2 \leq \chi^2_{\min} + 1$.

The OP calculations for each of the ten objects yielded the best-fit values of the various free param-

Table 3. Observed [6, 7] and model relative line intensities and H_β line luminosities for the HII regions ($I(H_\beta) = 100$). The observed HeI lines have been corrected for the background stellar absorption

Line	0917+527 Observations	OP model	0926+606 Observations	OP model	0940+544N Observations	OP model	0948+532 Observations	OP model	1054+365 Observations	OP model	0723+692B* Observations	OP model
[O II]λ3727 Å	188.8 ± 1.4	190.5	178.5 ± 1.2	182.1	57.0 ± 1.3	60.3	136.7 ± 2.4	141.6	106.6 ± 1.0	108.6	157.6 ± 1.5	159.5
[Ne III]λ3869 Å	50.5 ± 0.6	50.5	44.9 ± 0.5	45.0	37.1 ± 0.8	36.9	48.0 ± 1.1	48.3	50.2 ± 0.6	50.2	32.5 ± 0.6	32.5
[O III]λ4363	9.2 ± 0.4	7.2	8.3 ± 0.9	7.1	13.5 ± 0.4	9.3	8.3 ± 0.4	8.5	9.5 ± 0.3	10.0	8.2 ± 0.4	7.7
He I λ4471	4.1 ± 0.3	3.8	4.1 ± 0.2	3.9	4.1 ± 0.2	3.7	4.1 ± 0.3	3.9	4.2 ± 0.3	3.4	4.3 ± 0.3	3.7
He II λ4686 Å	2.3 ± 0.3	2.3	1.6 ± 0.2	1.6	−	0.1	1.0 ± 0.1	1.1	−	0.1	1.4 ± 0.4	1.4
[O III]λ4959	158.3 ± 1.1	156.4	162.8 ± 1.0	158.8	133.3 ± 1.4	135.1	181.0 ± 2.4	184.6	200.7 ± 1.5	197.7	142.5 ± 1.2	142.4
[O III]λ5007 Å	468.0 ± 2.8	470.8	477.2 ± 2.6	477.8	398.1 ± 3.6	406.5	563.2 ± 6.4	555.7	594.8 ± 4.0	595.2	430.3 ± 3.1	428.7
He I λ5876 Å	10.5 ± 0.3	10.4	10.8 ± 0.2	10.8	10.7 ± 0.3	10.6	11.1 ± 0.4	10.8	11.1 ± 0.3	9.9	11.0 ± 0.3	10.6
He I λ6678 Å	3.0 ± 0.2	2.9	3.4 ± 0.2	3.0	2.8 ± 0.1	2.8	2.9 ± 0.2	3.0	3.0 ± 0.2	2.6	3.4 ± 0.2	2.9
He I λ7065 Å	2.1 ± 0.2	2.4	2.5 ± 0.1	2.4	3.0 ± 0.2	3.4	2.6 ± 0.3	2.5	2.3 ± 0.2	3.6	1.7 ± 0.2	3.0
[S III]λ6312 Å	1.5 ± 0.2	3.6	1.9 ± 0.2	4.3	1.1 ± 0.1	1.6	1.9 ± 0.2	3.5	2.1 ± 0.2	3.2	2.2 ± 0.2	3.1
[NII]λ6584 Å	5.8 ± 0.2	5.8	8.3 ± 0.2	8.4	1.6 ± 0.1	1.6	6.8 ± 0.3	6.8	4.8 ± 0.2	4.7	5.5 ± 0.3	5.4
[S II]λ6716 Å	16.4 ± 0.3	15.2	18.2 ± 0.3	16.6	4.0 ± 0.2	3.1	13.8 ± 0.5	12.4	9.4 ± 0.3	7.8	11.4 ± 0.3	10.3
[S II]λ6731 Å	11.4 ± 0.3	11.6	14.6 ± 0.3	12.4	3.2 ± 0.2	2.5	10.3 ± 0.4	9.3	6.5 ± 0.2	7.3	8.2 ± 0.3	8.8
[Ar III]λ7135 Å	4.8 ± 0.3	4.8	6.6 ± 0.2	6.5	2.9 ± 0.2	3.0	8.0 ± 0.5	8.1	8.0 ± 0.2	8.0	6.4 ± 0.2	6.4
[Fe III]λ4658 Å	0.9 ± 0.3	0.9	1.0 ± 0.2	0.9	0.4 ± 0.1	0.4	0.7 ± 0.1	0.7	−	0.09	−	0.09
log[$L(H_\beta)$]	39.44	39.44	40.24	40.24	39.99	39.99	40.98	40.97	38.37	38.38	38.02	38.02
$\chi^2(1)$	−	10.47	−	16.11	−	12.04	−	5.65	−	9.69	−	5.83
$\chi^2(2)$**	−	1.92	−	8.54	−	3.88	−	1.79	−	6.43	−	2.34

Line	1135+581 Observations	OP model	1152+579 Observations	OP model	1211+540 Observations	OP model	1256+351 Observations	OP model	1533+574A Observations	OP model	0741+535* Observations	OP model
[O II]λ3727 Å	134.6 ± 1.3	140.5	88.7 ± 1.4	90.9	64.6 ± 1.3	74.4	110.2 ± 0.2	110.4	245.9 ± 0.2	245.5	291.0 ± 4.0	296.5
[Ne III]λ3869 Å	47.1 ± 0.6	47.4	55.2 ± 0.9	55.2	38.5 ± 0.7	38.7	44.9 ± 0.2	45.1	35.4 ± 0.2	35.1	37.8 ± 1.0	38.1
[O III]λ4363	6.9 ± 0.2	8.2	14.0 ± 0.3	10.9	11.6 ± 0.3	9.5	8.9 ± 0.1	9.3	6.7 ± 0.1	5.9	6.8 ± 0.8	5.9
He I λ4471	4.1 ± 0.1	3.9	4.0 ± 0.2	3.9	3.9 ± 0.2	3.5	4.1 ± 0.1	3.9	4.1 ± 0.1	3.7	4.0 ± 0.8	3.8
He II λ4686 Å	1.9 ± 0.1	1.9	1.2 ± 0.1	1.3	2.6 ± 0.1	2.5	1.1 ± 0.1	1.1	−	0.1	−	0.1
[O III]λ4959	179.1 ± 1.0	165.6	211.2 ± 1.8	205.8	153.7 ± 1.3	154.6	198.2 ± 0.3	195.7	129.7 ± 0.3	127.0	146.7 ± 2.0	145.5
[O III]λ5007 Å	476.5 ± 2.4	498.5	604.3 ± 4.5	619.4	464.8 ± 3.4	465.4	580.2 ± 0.9	589.1	380.4 ± 0.9	382.1	440.4 ± 5.4	438.0
He I λ5876 Å	11.4 ± 0.2	10.8	11.2 ± 0.2	11.0	10.1 ± 0.2	10.3	10.8 ± 0.1	10.8	10.4 ± 0.1	10.4	10.7 ± 0.6	10.6
He I λ6678 Å	3.0 ± 0.1	3.0	2.8 ± 0.1	3.0	2.9 ± 0.2	2.7	3.2 ± 0.1	3.0	3.0 ± 0.1	2.9	3.1 ± 0.4	2.9
He I λ7065 Å	2.4 ± 0.1	2.6	3.2 ± 0.2	3.0	2.5 ± 0.2	3.8	2.7 ± 0.1	2.8	2.1 ± 0.1	2.4	2.1 ± 0.5	2.7
[S III]λ6312 Å	1.5 ± 0.1	3.2	1.6 ± 0.1	2.4	1.5 ± 0.1	2.3	1.7 ± 0.1	3.5	1.8 ± 0.1	3.9	2.2 ± 0.4	3.8
[NII]λ6584 Å	6.7 ± 0.1	6.7	4.0 ± 0.2	4.0	2.0 ± 0.1	2.0	4.6 ± 0.1	4.7	12.2 ± 0.1	12.0	12.2 ± 0.5	12.4
[S II]λ6716 Å	11.7 ± 0.2	10.4	6.8 ± 0.2	5.8	5.3 ± 0.2	3.9	9.5 ± 0.1	9.1	23.4 ± 0.1	22.3	21.7 ± 0.7	21.6
[S II]λ6731 Å	8.3 ± 0.2	8.0	5.4 ± 0.2	4.6	4.2 ± 0.2	3.7	7.1 ± 0.1	7.1	17.0 ± 0.1	17.7	21.1 ± 0.6	18.2
[Ar III]λ7135 Å	8.4 ± 0.2	8.5	5.4 ± 0.2	5.4	3.6 ± 0.2	3.8	7.6 ± 0.1	7.7	7.6 ± 0.1	7.5	7.9 ± 0.5	7.9
[Fe III]λ4658 Å	1.1 ± 0.1	1.2	0.7 ± 0.1	0.6	−	0.09	0.5 ± 0.1	1.2	1.7 ± 0.1	1.6	1.6 ± 0.7	1.4
log[$L(H_\beta)$]	39.63	39.62	41.04	41.04	38.67	38.68	39.43	39.42	40.15	40.15	40.18	40.18
$\chi^2(1)$	−	42.55	−	15.49	−	16.76	−	36.93	−	49.93	−	2.86
$\chi^2(2)$**	−	26.96	−	5.28	−	9.33	−	18.17	−	22.11	−	2.07

* Additional objects not included in the sample of 10 selected HII regions.

** The χ^2 value has been calculated without the [OIII] λ4363 Å, HeI λ7065 Å, and [SII] λ6312 Å lines.

Table 4. Optimum values of the free parameters found from OP models I−III with different sets of initial values of the free parameters

Parameters	Models		
	I	II	III
n_H, cm^{-3}	230^{+43}_{-30}	179^{+24}_{-22}	80^{+11}_{-17}
$\log Q_{\text{tot}}$	53.87	53.87	53.87
ϵ	$0.0050^{+0.0010}_{-0.0003}$	$0.0058^{+0.0004}_{-0.0003}$	$0.0092^{+0.0011}_{-0.0006}$
He/H, 10^{-2}	$8.36^{+0.80}_{-0.70}$	$8.40^{+0.59}_{-0.71}$	$8.57^{+0.52}_{-0.66}$
O/H, 10^{-5}	$9.54^{+0.30}_{-0.41}$	$9.52^{+0.43}_{-0.45}$	$9.58^{+0.31}_{-0.37}$
N/H, 10^{-6}	$4.34^{+0.42}_{-0.44}$	$4.40^{+0.78}_{-0.68}$	$4.53^{+0.28}_{-0.77}$
Ne/H, 10^{-5}	$1.65^{+0.07}_{-0.10}$	$1.65^{+0.09}_{-0.13}$	$1.66^{+0.10}_{-0.13}$
S/H, 10^{-6}	$3.20^{+0.29}_{-0.27}$	$3.22^{+0.22}_{-0.17}$	$3.28^{+0.20}_{-0.28}$
Ar/H, 10^{-7}	$4.37^{+0.74}_{-1.07}$	$4.37^{+0.86}_{-1.06}$	$4.42^{+0.91}_{-0.88}$
Fe/H, 10^{-6}	$1.26^{+0.50}_{-0.68}$	$1.26^{+0.71}_{-0.47}$	$1.33^{+0.64}_{-0.49}$

eters (Table 2). A comparison of the model parameters and the observed [6, 7] parameters on which the optimization was based is given in Table 3. We have corrected the observed values for the HeI lines for the background stellar absorption. We can see that, for most of the objects, the model and observed relative intensities of the HeI, HeII, [OIII], [NeIII], and [ArIII] lines are consistent, whereas the agreement is poor for the [SII], [SIII], and [OIII] λ4363 Å lines. To estimate the effect of the "unreliable" (from the viewpoint of atomic data and other factors) [OIII] λ4363 Å, HeI λ7065 Å, and [SIII] λ6312 Å lines, we list in Table 3 the values $\chi^2(1)$ including these lines and $\chi^2(2)$ excluding them. In most of the objects, precisely these lines systematically increase the value of χ^2.

When calculating any type of optimized models, there also arises the question of the uniqueness (or stability) of the obtained solution. We checked the uniqueness of the χ^2 minimum using the object 0948+532 as an example by calculating three OP models with three different sets of initial values for the free parameters. In model I, model II, and model III, the free parameters had initial values values close to their minimum, mean, and maximum values. The only initial parameter value that was the same in all three OP models was Q_{tot}, which was determined from the calculation of the L_c spectrum of this object. It would certainly be possible to calculate OP models with a multitude of other sets of initial free-parameter values; however, in our opinion, the sets we have chosen correspond to the most natural values, since, in an actual BCDG, an increase in the abundance of one element should be accompanied by an increase

Table 5. Observed [6, 7] and model relative line intensities and H$_\beta$ line luminosities for an HII region in the BCDG 0948+532 ($I(\text{H}_\beta) = 100$). The observed HeI lines have been corrected for the background stellar absorption

Line	Observations	OP model		
		I	II	III
[OII] λ 3727 Å	136.7 ± 2.4	141.8	141.5	141.3
[NeIII] λ 3869 Å	48.0 ± 1.1	48.1	48.0	48.2
[OIII] λ 4363 Å	8.3 ± 0.4	8.6	8.5	8.5
HeI λ 4471 Å	4.1 ± 0.3	3.8	3.8	3.9
HeII λ 4686 Å	1.0 ± 0.1	1.1	1.1	1.1
[OIII] λ 4959 Å	181.0 ± 2.4	182.9	182.7	183.8
[OIII] λ 5007 Å	563.2 ± 6.4	550.5	550.1	553.3
HeI λ 5876 Å	11.1 ± 0.4	10.8	10.8	10.8
HeI λ 6678 Å	2.9 ± 0.2	3.0	3.0	3.0
HeI λ 7065 Å	2.6 ± 0.3	2.9	2.8	2.4
[SIII] λ 6312 Å	1.9 ± 0.2	3.5	3.5	3.5
[NII] λ 6584 Å	6.8 ± 0.3	6.8	6.8	6.8
[SII] λ 6716 Å	13.8 ± 0.5	11.7	12.0	12.4
[SII] λ 6731 Å	10.3 ± 0.4	9.6	9.5	9.3
[ArIII] λ 7135 Å	8.0 ± 0.5	8.0	8.0	8.0
[FeIII] λ 4658 Å	0.7 ± 0.1	0.7	0.7	0.7
$\log[L(\text{H}\beta)]$	40.98	40.98	40.98	40.98
χ^2	−	5.99	5.80	5.61

in the abundances of other elements, in accordance with most theories of stellar nucleosynthesis and chemical evolution. It is difficult, if not impossible, to find a BCDG that has an extremely high abundance of, e.g., oxygen and simultaneously an extremely low abundance of nitrogen. Therefore, in our view, initializing the OP models using such extremely different abundances for different elements has no physical meaning.

The calculation of OP models I, II, and III yielded the parameter values listed in Table 4. A comparison of the model parameter and the observed parameters used to calculate the χ^2 values in these OP models is given in Table 5. We can see that the best-fit values of the free parameters obtained in OP models I and II coincide within the errors, whereas the results of OP model III differ in n_{H} and ϵ. Nevertheless, all the other free parameters for OP model III coincide with the corresponding values for OP models I and II within the errors. This testifies to the insignificance of the presence of two χ^2 minima when determining the

Table 6. Optimal values of the free parameters for an HII region in the BCDG 0948+532 obtained in various geometries using the OP calculations

Parameter	Models				
	Norm	Ri100	CF0.3	Cyl-h100	SinDensFluct
n_H, cm^{-3}	179^{+24}_{-22}	160^{+39}_{-10}	158^{+52}_{-11}	259^{+18}_{-26}	$-$
$\log Q_{\text{tot}}$	53.87	53.87	53.89	53.80	53.87
ϵ	$0.0058^{+0.0004}_{-0.0003}$	$0.0064^{+0.0004}_{-0.0008}$	$0.0059^{+0.0002}_{-0.0007}$	$0.0037^{+0.0007}_{-0.0001}$	$0.0071^{+0.0006}_{-0.0005}$
He/H, 10^{-2}	$8.40^{+0.59}_{-0.71}$	$8.44^{+0.69}_{-0.72}$	$8.49^{+0.59}_{-0.66}$	$8.31^{+0.59}_{-0.55}$	$8.45^{+0.94}_{-0.84}$
O/H, 10^{-5}	$9.52^{+0.43}_{-0.45}$	$9.51^{+0.18}_{-0.40}$	$9.10^{+0.42}_{-0.30}$	$9.10^{+0.32}_{-0.25}$	$9.46^{+0.35}_{-0.52}$
N/H, 10^{-6}	$4.40^{+0.78}_{-0.68}$	$4.39^{+0.39}_{-0.68}$	$4.27^{+0.45}_{-0.71}$	$4.16^{+0.44}_{-0.27}$	$4.24^{+0.44}_{-1.07}$
Ne/H, 10^{-5}	$1.65^{+0.09}_{-0.13}$	$1.65^{+0.14}_{-0.09}$	$1.56^{+0.11}_{-0.10}$	$1.55^{+0.11}_{-0.10}$	$1.63^{+0.11}_{-0.15}$
S/H, 10^{-6}	$3.22^{+0.22}_{-0.17}$	$3.22^{+0.19}_{-0.26}$	$3.10^{+0.23}_{-0.20}$	$3.11^{+0.30}_{-0.20}$	$3.13^{+0.37}_{-0.36}$
Ar/H, 10^{-7}	$4.37^{+0.86}_{-1.06}$	$4.35^{+0.84}_{-1.06}$	$4.24^{+0.83}_{-0.67}$	$4.28^{+0.43}_{-0.28}$	$4.37^{+1.48}_{-1.61}$
Fe/H, 10^{-6}	$1.26^{+0.71}_{-0.47}$	$1.26^{+0.73}_{-0.47}$	$1.20^{+0.70}_{-0.47}$	$1.31^{+0.24}_{-0.48}$	$1.25^{+0.73}_{-0.46}$

chemical composition, and, accordingly, the Y_p and dY/dZ values, from the OP calculations for these objects.

We also studied the effect of deviations of the HII regions from spherical geometry on the results of the OP calculations using the same object, 0948+532, as an example. We adopted the corresponding values for model II (see above) as the initial free parameters in the OP models. We calculated the Hβ line luminosity (which was compared with the observed luminosity) separately for OP models with covering factors differing from unity and in a cylindrical geometry for this HII region (see Table 7 below).

Earlier, we assumed in the OP calculations an inner radius for the envelope of $R_{\text{in}} = 0.01$ pc. Therefore, we first calculated an OP model with $R_{\text{in}} = 100$ pc. The results are listed in Tables 6 and 7 in the columns labeled "Ri100." For comparison, these tables also list the results of the OP calculations for this object obtained above (the column "Norm"). We can see that the free parameters obtained in the Ri100 OP calculation coincide with those for the Norm calculation within the errors. This means that an increase in the inner radius essentially does not affect the results of the OP calculations.

The effect of deviations from a spherical geometry can be studied by specifying a value for the so-called covering factor (CF) of the ionizing source by the envelope of the nebula that differs from unity (CF $= \Omega/4\pi$, where Ω is the solid angle of the HII region as seen from its center). Tables 6 and 7 list in the columns "CF0.3" the results of the OP calculation with CF $= 0.3$. As in the previous case, we obtained optimal values for most of the free parameters that are close to the data of the Norm model within the errors. The fact that the value of χ^2 is greater than in the Norm model suggests that the value CF $= 1$ is, nevertheless, more realistic for this object.

In addition, we calculated OP models for the HII region 0948+532 in a cylindrical geometry, with the cylinder half-height being $h = 100$ pc. For this case, we also redetermined the jump at $\lambda = 228$ Å in the L_c spectrum, since the HeII λ4686-Å line is very sensitive to changes in the geometry. In this OP model, we also fixed the outer radius of the HII envelope, since it was necessary to specify an Hβ line luminosity. The results for this OP model are listed in Tables 6 and 7 in the columns "Cyl-h100." The calculations for this OP models yielded optimal values of the free parameters n_{H}, ϵ, and Q_{tot} that differed from those for the Norm OP caluclation (corresponding to a different minimum of χ^2). However, the elemental abundances coincide within the errors with those of the Norm and all the other above-mentioned OP calculations. This indicates that deviation of the geometries of the actual HII regions from a spherical geometry will not lead to important changes in the values of Y_p and dY/dZ. Nevertheless, the geometry is, of course, important in precision modeling of actual objects.

Finally, the volume filling factor of the nebula ϵ is responsible for the gas density fluctuations in our

Table 7. Observed [6, 7] and model (OP models in various geometries) relative line intensities ($I(\mathrm{H}_\beta) = 100$) and $\mathrm{H}\beta$ line luminosities for an HII region in the BCDG 0948+532. The observed HeI lines have been corrected for the background stellar absorption

Lines	Observations	OP model				
		Norm	Ri100	CF0.3	Cyl-h100	SinDensFluct
[OII] λ 3727 Å	136.7 ± 2.4	141.5	141.7	141.7	141.1	140.8
[NeIII] λ 3869 Å	48.0 ± 1.1	48.0	48.2	48.2	48.2	48.2
[OIII] λ 4363 Å	8.3 ± 0.4	8.5	8.6	9.0	9.2	8.7
HeI λ 4471 Å	4.1 ± 0.3	3.8	3.9	3.9	3.8	3.9
HeII λ 4686 Å	1.0 ± 0.1	1.1	1.1	1.1	1.1	1.1
[OIII] λ 4959	181.0 ± 2.4	182.7	183.5	183.4	183.9	184.0
[OIII] λ 5007 Å	563.2 ± 6.4	550.1	552.4	552.1	553.5	553.9
HeI λ 5876 Å	11.1 ± 0.4	10.8	10.8	10.8	10.7	10.8
HeI λ 6678 Å	2.9 ± 0.2	3.0	3.0	3.0	2.9	3.0
HeI λ 7065 Å	2.6 ± 0.3	2.8	2.7	2.7	3.0	2.8
[SIII] λ 6312 Å	1.9 ± 0.2	3.5	3.6	3.6	3.4	3.4
[NII] λ 6584 Å	6.8 ± 0.3	6.8	6.8	6.8	6.8	6.8
[SII] λ 6716 Å	13.8 ± 0.5	12.0	12.0	12.1	11.9	12.3
[SII] λ 6731 Å	10.3 ± 0.4	9.5	9.5	9.5	9.9	9.7
[ArIII] λ 7135 Å	8.0 ± 0.5	8.0	8.0	8.0	8.0	8.0
[FeIII] λ 4658 Å	0.7 ± 0.1	0.7	0.7	0.7	0.7	0.7
$\log[L(\mathrm{H}\beta)]$	$(40.98, 40.45, 40.43)^*$	40.98	40.98	40.45	40.43	40.98
χ^2	–	5.80	6.02	6.11	5.78	5.10

* Values of $\log[L(\mathrm{H}\beta)]$ calculated from the observed $\mathrm{H}\beta$ line flux fitting the Norm+Ri100+SinDensFluct, CF0.3, and Cyl-h100 OP models, respectively.

OP models. Therefore, we decided to check how another representation for the density inhomogeneities would affect the results of the OP calculations. The columns "SinDensFluct" in Tables 6 and 7 list the results of OP calculations in which the hydrogen density n_H changes in the radial direction in accordance with a sine law, from 100 to 200 cm^{-3} with a period of 87.8 pc and zero phase shift relative to the inner radius of the nebula. It is clear that the value of n_H is no longer a free parameter in this case, and the number of degrees of freedom becomes equal to two. Tables 6 and 7 show that, most of the best-fit parameter values in this OP model coincide with those for the Norm OP model within the errors, testifying to the insignificant effect of density fluctuations on the final OP model results. The only exception is the filling

factor ϵ. The final value of χ^2 in this OP model is lower than in the Norm OP model, suggesting the existence of density fluctuations of this type (or similar) in 0948+532. Nevertheless, gas-density fluctuations do not significantly affect the final chemical composition or, accordingly, Y_p and dY/dZ, in contrast to Norm-type OP models, in which the filling factor alone is responsible for density fluctuations.

3. PRIMORDIAL HELIUM ABUNDANCE AND THE RATE OF ITS ENRICHMENT

We used the chemical composition yielded by the OP calculations (Table 2) to analyze the $Y-Z$ dependences. In contrast to the chemical compositions derived from our earlier calculations using a
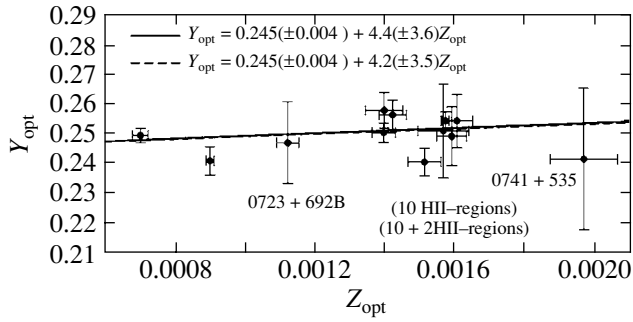
Fig. 3. $Y_{opt}-Z_{opt}$ dependence for the OP model calculations for the ten selected objects plus two additional objects (the latter are marked on the graph).

photoionization-model grid [1, 2], we also obtained the Fe/H abundance via the optimization modeling. We designate the values of Y and Z obtained from the OP modeling as Y_{opt} and Z_{opt}. Here, we adopted the following formulas for the calculation of Y_{opt} and Z_{opt}:

$$Y_{opt} = \frac{4\,He/H(1 - Z_{opt})}{1 + 4\,He/H}, \qquad (2)$$

$$Z_{opt} = \frac{Za_{opt}}{1 + 4\,He/H + Za_{opt}},$$

$$Za_{opt} = 14\,N/H + 16\,O/H + 20\,Ne/H$$

$$+ 32\,S/H + 40\,Ar/H + 56\,Fe/H,$$

where the elemental abundances correspond to those obtained in the OP calculations.

We derived errors for Y_{opt} and Z_{opt} by considering the maximum errors in the abundances listed in Table 2. The $Y_{opt}-Z_{opt}$ dependence is shown in Fig. 3 ("10 HII regions"), where the corresponding values of Y_p and dY/dZ obtained by fitting a linear approximation to this dependence are plotted together with their errors in both coordinates. The resulting parameter values coincide within the errors with those obtained earlier based on ionization-correction factors:

$$Y_p = 0.245 \pm 0.004, \qquad (3)$$

$$dY/dZ = 4.4 \pm 3.6. \qquad (4)$$

To analyze the accuracy of this approximation, we calculated OP models for two additional HII regions, 0723+692B and 0741+535. If we add the Y and Z values for these objects, which were obtained based on ionization-correction factors, to the data on the selected ten objects, we find $Y_p = 0.249 \pm 0.006$, $dY/dZ = 3.0 \pm 7.0$. Thus, adding the Y and Z for these two objects to the sample of ten HII regions worsens the reproduction of the Y_p value obtained from the entire sample of the objects ($Y_p = 0.244 \pm$

0.004, $dY/dZ = 8.8 \pm 4.6$). The free parameters obtained from OP calculations for the two objects are listed in Table 2, and the observed and model relative line intensities and $H\beta$ line luminosities are given in Table 3 (the two additional objects are labeled in these tables with asterisks). Here, $\chi^2(1)$ was calculated for all the listed observed parameters used in the fitting, and $\chi^2(2)$ does not take into account the so-called unreliable (from the viewpoint of atomic data and for other physical reasons) [OIII] $\lambda 4363$ Å, HeI $\lambda 7065$ Å, [SIII] $\lambda 6312$ Å lines. We can see that, for most of the HII regions, large values of $\chi^2(1)$ can be explained by poor fits to precisely these lines.

A linear fit to the $Y_{opt}-Z_{opt}$ dependence including the OP results for the HII regions 0723+692B and 0741+535 ("10 + 2 HII regions" in Fig. 3) yields

$$Y_p = 0.245 \pm 0.004, \qquad (5)$$

$$dY/dZ = 4.2 \pm 3.5, \qquad (6)$$

which coincide within the errors with the OP results for the ten selected objects, as well as with those obtained based on ionization-correction factors. This confirms the accuracy of the linear approximation for the $Y_{opt}-Z_{opt}$ relationship and suggests that the results of the OP calculations for other HII regions in BCDGs will not substantially change the values of Y_p and dY/dZ. Therefore, we have not calculated OP models for other HII regions in BCDGs (as this is a demanding task even for modern computers).

4. CONCLUSIONS

We have demonstrated the need to carry out optimized-photoionization (OP) calculations for HII regions in BCDGs. We have described our calculation technique and analyzed the results obtained. The advantage of OP models over the use of photoionization-model grids is that OP models are based on reproducing the observed spectrum of each of the considered objects and do not impose the assumptions on which calculations of photoionization-model grids are based. Therefore, it was necessary to calculate OP models for at least a few HII regions in BCDGs if we wished to refine elemental abundances and the values of Y_p and dY/dZ. However, the OP calculations are quite demanding (OP calculations for one object require calculating ≈ 2000 photoionization models). Therefore, we have thus far calculated OP models for only ten selected objects. The selection criterion was the requirement that the $Y-Z$ relationship for these objects reproduce within the errors the Y_p and dY/dZ values obtained from the $Y-Z$ dependence for all the analyzed objects [2]. The selected objects were 0917+527, 0926+606, 0940+544N,

0948+532, 1054+365, 1135+581, 1152+579, 1211+540, 1256+351, 1533+574A.

We used spectra obtained by Izotov *et al.* [6, 7] in the OP calculations for these objects. We refined their L_c spectra by deriving an approximation relationship for determining H^+/H^0 in these objects. A best-fit value for the jump at $\lambda = 228$ Å was determined from the calculated L_c spectra. This enabled us to derive the spectral energy distributions of the ionizing radiation of the cores of the selected HII regions, which were used in the OP calculations for these objects.

There were ten free input parameters in the OP calculations. A range of values was specified for each of the free parameters in order to adequately reproduce the spectra of the low-metallicity HII regions in BCDGs. The parameters used to calculate the χ^2 values were the Hβ line luminosity and the relative intensities of 16 emission lines. For this type of OP model, the principle of detailed balance operates, according to which we can use the intensity of one reliable line of one ion to calculate the intensities of all other lines of this ion. We have studied the uniqueness of the χ^2 minimum in OP models of this type. Although there can be several χ^2 minima, the elemental abundances are always determined unambiguously by the OP models (they are the same within the errors). This testifies to the insignificance of the presence of several χ^2 minima on the final Y_p and dY/dZ values derived using the OP models. We also analyzed the effect of deviations from spherical geometry and of density fluctuations on the final best-fit values of the free parameters in the OP calculations and concluded that these deviations are not important for the final derived Y_p and dY/dZ values.

We determined the primordial helium abundance, Y_p, and rate of its enrichment, dY/dZ, using the OP model calculations for the selected objects. A linear approximation of the $Y_{opt}-Z_{opt}$ relationship yields the values $Y_p = 0.245 \pm 0.004$ and $dY/dZ = 4.4 \pm 3.6$, which coincide within the errors with the values obtained earlier from a photoionization-model grid for the HII regions. To check the accuracy of this approximation, we calculated OP models for the two other HII regions 0723+692B and 0741+535, which do not satisfy the indicated selection criterion. However, taking into account the relevant data for their OP models in the $Y_{opt}-Z_{opt}$ approximation yielded $Y_p = 0.245 \pm 0.004$ and $dY/dZ = 4.2 \pm 3.5$, which coincide within the errors with the results of the OP models for the other ten objects. This shows that including other HII regions in the sample used for the OP calculations will not considerably change the values of Y_p and dY/dZ obtained in this paper.

ACKNOWLEDGMENTS

REFERENCES

1. V. V. Golovatyi and B. Ya. Melekh, Astron. Zh. **79**, 867 (2002) [Astron. Rep. **46**, 779 (2002)].
2. V. V. Golovatyi and B. Ya. Melekh, Kin. Fiz. Neb. Tel **18** (4), 362 (2002).
3. V. V. Golovatyi and B. Ya. Melekh, Astron. Zh. **82** (2005) (in press).
4. G. J. Ferland, *Hazy, a Brief Introduction to Cloudy 96*, Department of Physics and Astronomy Internal Report (Univ. of Kentucky, 2002).
5. P. A. M. van Hoof, PhD Thesis (Rijskuniversiteit Groningen, 1997).
6. Yu. I. Izotov, T. X. Thuan, and V. A. Lipovetsky, Astrophys. J. **435**, 10 (1994).
7. Yu. I. Izotov, T. X. Thuan, and V. A. Lipovetsky, Astrophys. J., Suppl. Ser. **108**, 1 (1997).
8. B. Ya. Melekh, J. Phys. Stud. **4** (2), 225 (2000).
9. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd ed. (Cambridge Univ. Press, Cambridge, 1992).
10. G. J. Ferland, *Hazy, a Brief Introduction to Cloudy 94*, Physics Department Internal Report (Univ. of Kentucky, 1999).

Translated by G. Rudnitskiĭ

The Motion of Halo Stars in Dynamical Numerical Models of Open Clusters

V. M. Danilov

Astronomical Observatory, Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia
Received November 4, 2004; in final form, February 17, 2005

Abstract—Equations of motion containing a small parameter μ are derived for stars at the peripheries of open star clusters. The parameter μ is obtained for six numerical open-cluster models. The general analytical solution of these equations of motion for $\mu = 0$ is found. An iterative method is used to derive the frequencies of the stellar motions for first-order expansions in μ of the solutions of the equations of motion for stars at the cluster periphery. Applications of the results are discussed. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

According to Danilov and Leskov [1], the trajectories of halo stars in dynamical numerical models of open clusters with various degrees of nonstationarity in the regular field have small, positive, maximum characteristic Lyapunov exponents λ and relatively simple Fourier spectra. According to [1], the stellar trajectories at the cluster peripheries have Lyapunov time scales $t_\lambda = \lambda^{-1}$ on the order of $1-10$ Myr, which are comparable to Laskar's [2] estimates for planetary orbits in the solar system. The three highest amplitude sinusoidal components in the time dependences of the clustercentric distances $r(t)$ of the stars contain $50-95\%$ of the total energy of the Fourier spectra for these dependences. The oscillations of the regular potential in the halos of the models of nonstationary open clusters of [3] are small and do not exceed 6% of the regular potential averaged over the oscillation period. These properties of the stellar trajectories and gravitational potential in the open-cluster halos make it possible to apply the methods of perturbation theory to obtain an approximate description of the motions of stars at the peripheries of a nonstationary open cluster. According to Danilov and Leskov [1], the distributions of the periods for the stellar trajectories in open-cluster models have sharp peaks at periods commensurable with the period of oscillations of the regular field. A comparison of the periods of these peaks with those derived by applying a small-parameter analysis to the stellar trajectories in such systems can be used to identify the main resonances in the motions of halo stars in open-cluster models.

Periodic oscillations of the size, density, and regular field of an open cluster are, to a large extent, due to the motions of halo stars. These oscillations result in a number of complex processes, which are observed in numerical simulations of open-cluster dynamics (transfer of energy from large-scale to small-scale stellar motions, oscillations of the entropy of the system, etc. [4]). It is, therefore, important to understand the mechanisms forming the trajectories of halo stars in open clusters.

The aims of this paper are to introduce a small parameter into the equations of motion for stars in an open cluster, to analyze the frequencies of the stellar motions computed in a first-order of expansion on small parameter of the solutions of the equations of stellar motion at the cluster periphery, and to identify the main resonances in the motions of halo stars in open-cluster models.

2. DESCRIPTION OF OPEN-CLUSTER MODELS

In this paper, we analyze an open cluster moving in a circular orbit in the plane of the Galaxy using the equations of stellar motion [5] written for a rotating coordinate system ξ, η, ζ. Like Chandrasekhar [5], we expand the regular potential of the Galaxy in a power series to quadratic terms in the coordinates ξ, η, ζ. The integrals of motion of the cluster center of mass are written in the usual way.

We construct an initial approximation for the trajectories of the halo stars using the following simplifying assumptions.

(1) To estimate the size of the critical zero-velocity surface of the cluster along the ξ, η, and ζ axes, we model the cluster as a binary system consisting of a star of mass $m_1 = 1\,m_\odot$ and a point mass $m_2 = M - m_1$, where M is the mass of the cluster model (as in our previous paper [1], $M = 500\,m_\odot$). The integrals of motion of the cluster center of mass

allow the sets of equations of motion of the bodies with masses m_1 and m_2 to be decoupled and analyzed separately. The equations of motion of the stars with mass m_1 yield three critical points $\xi = \eta = \zeta = 0$ and $\xi = \pm|\xi_t|, \eta = \zeta = 0$ (the latter two are saddle points and lie on the critical zero-velocity surface [6, p. 197]), where $|\xi_t| = (-Gm_2/(p^2\alpha_1))^{1/3}, p = 1 + m_1/m_2$, and the Jacobi integral is

$$\varepsilon = \frac{V^2 + \alpha_1\xi^2 + \alpha_3\zeta^2}{2} - \frac{Gm_2}{p^2 r} = \text{const}, \quad (1)$$

where V and r are the velocity and clustercentric distance of the star of mass m_1, respectively, $\alpha_1 = \left(\frac{1}{R}\frac{\partial\Phi}{\partial R} - \frac{\partial^2\Phi}{\partial R^2}\right)_0 < 0$, and $\alpha_3 = -\left(\frac{\partial^2\Phi}{\partial Z^2}\right)_0 > 0$. Here, R and Z are the cylindrical coordinates of the point. The subscripts "0" indicate that the corresponding derivatives of the Galactic potential Φ are calculated at the point with coordinates $R = R_0 = 8200$ pc and $Z = 0$. The formula for $\xi = \pm|\xi_t|$ as $m_1/m_2 \to 0$ transforms into formula (11.13) for the tidal radius of the cluster in the Galactic field in Chandrasekhar's book [6, p. 198].

We now substitute $V = 0$ into (1) to derive the equation for the zero-velocity surface, then substitute $\xi = |\xi_t|$, $\eta = \zeta = 0$ into this equation to find the critical value of the Jacobi integral that corresponds to the critical zero-velocity surface: $\varepsilon_t = -3Gm_2/(2p^2|\xi_t|)$. The distances from the coordinate origin to this surface measured along the η and ζ axes are $|\eta_t| = \frac{2}{3}|\xi_t|$ and $|\zeta_t| = |\xi_t|q^{1/3}[(\sqrt{1+q}+1)^{1/3} - (\sqrt{1+q}-1)^{1/3}] \simeq 0.5016$, respectively. Here, $q = -\alpha_1/\alpha_3$ and α_1 and α_3 are estimated using the model Galactic potential of Kutuzov and Osipkov [7]. The relation $|\eta_t| = \frac{2}{3}|\xi_t|$ can also be found in the book of Spitzer [8, p. 111], where he discusses the size of the critical zero-velocity surface of the cluster in the case when both the Galaxy and the cluster are modeled as point masses. The numerical simulations of Terlevich [9] and Danilov [10] point toward strong flattening of the open-cluster models in the ζ direction and a triaxial shape of the equal density and equipotential surfaces of the cluster at its periphery (the cluster is most extended along the ξ axis and least extended along the ζ axis). Heggier and Ramamani [11] obtained the same results when constructing a stationary collisionless cluster model with a tidal cutoff of the density and potential at the critical value of the Jacobi integral. We thus assume that the adopted model can be used to derive the size of the critical zero-velocity surface of the cluster along the ξ, η, and ζ axes.

(2) The cluster model defined in item (1) and equation (1) provides only an approximate description of the stellar motions at the cluster periphery, since most of the stars are located at the cluster periphery [9, 10]. We therefore assume that the critical zero-velocity surface of the open-cluster model has the shape of an ellipsoid with semiaxes $a = |\xi_t|$, $b = |\eta_t|$, and $c = |\zeta_t| \simeq a/2$, and propose to search for the unperturbed cluster potential in the form

$$U = U_0 - (U_0 + U_1)r^2 + S, \quad (2)$$

$$r^2 = \left(\frac{\xi}{a}\right)^2 + \left(\frac{\eta}{b}\right)^2 + \left(\frac{\zeta}{c}\right)^2, \quad S = \frac{\alpha_1\xi^2 + \alpha_3\zeta^2}{2}.$$

Here, U_0 is the potential at the cluster center ($r = 0$) and U_1 is the cluster potential at the critical zero-velocity surface at the points $\eta = \pm b$ and $\xi = \zeta = 0$. The unperturbed potential (2) corresponds to the initial approximation for the cluster potential that we use below to construct the unperturbed trajectories of the stars in the cluster halo. The potential (2) is equal to the potential at an inner point of some uniform ellipsoid of density ρ:

$$\rho = -\frac{\Delta U}{4\pi G} \quad (3)$$

$$= \frac{2(U_0 + U_1)(a^{-2} + b^{-2} + c^{-2}) - \alpha_1 - \alpha_3}{4\pi G}.$$

The equation of the equipotential surface for U given by formula (2) has the form

$$\gamma^2 = \left(\frac{\xi}{a_u}\right)^2 + \left(\frac{\eta}{b_u}\right)^2 + \left(\frac{\zeta}{c_u}\right)^2 = \text{const}, \quad (4)$$

where $a_u = a/\sqrt{1 - \alpha_1 a^2/s}$, $b_u = b$, $c_u = c/\sqrt{1 - \alpha_3 c^2/s}$, and $s = 2(U_0 + U_1)$. Here, $a_u < a$ and $c_u > c$, since $\alpha_1 < 0$ and $\alpha_3 > 0$. The surfaces defined by (4) are closer to spherical than the surfaces corresponding to $r^2 = 1$.

Using the potential U defined by (2), the equations of stellar motion in the combined force field of the cluster and the linearized Galactic field [5, formulas (5.517)–(5.519)] acquire the form

$$\ddot{\xi} = -\beta_1^2\xi + 2\omega\dot{\eta}, \quad \ddot{\eta} = -\beta_2^2\eta - 2\omega\dot{\xi}, \quad (5)$$

$$\ddot{\zeta} = -\beta_3^2\zeta,$$

where $\dot{\xi} = \frac{d\xi}{dt}$, $\ddot{\xi} = \frac{d\dot{\xi}}{dt}$ and $\dot{\eta}$, $\dot{\zeta}$, $\ddot{\eta}$, and $\ddot{\zeta}$ are defined analogously; $\beta_1^2 = 2(U_0 + U_1)/a^2$, $\beta_2^2 = \beta_1^2 a^2/b^2$, and $\beta_3^2 = \beta_1^2 a^2/c^2$; ω is the angular velocity of the cluster motion relative to the Galactic center; and $\beta_i^2 > 0$, since $(U_0 + U_1) > 0$ in the open-cluster models we consider here (see below).

Parameters of the open-cluster models

No.	U_0	U_1	ν_1	ν_2	β_3	μ
1	0.43618 ± 0.00684	-0.26109 ± 0.00496	0.04461	0.10789	0.11329	0.03709
2	0.45341 ± 0.00854	-0.27200 ± 0.00619	0.04568	0.10918	0.11532	-0.00982
3	0.48275 ± 0.00952	-0.29354 ± 0.00690	0.04696	0.11075	0.11777	-0.06771
4	0.49263 ± 0.00997	-0.30194 ± 0.00723	0.04721	0.11104	0.11823	-0.07870
5	0.49656 ± 0.01066	-0.30204 ± 0.00773	0.04783	0.11180	0.11941	-0.10713
6	0.48760 ± 0.00857	-0.29949 ± 0.00621	0.04678	0.11053	0.11743	-0.05955

We now use the usual notation to write the Jacobi integral for Eqs. (5) in the form

$$\frac{V^2 + \beta_1^2 \xi^2 + \beta_2^2 \eta^2 + \beta_3^2 \zeta^2}{2} = \epsilon = \text{const}, \quad \epsilon \geq 0. \tag{6}$$

We then substitute $V = 0$ into (6) to derive the equation of the zero-velocity surface for the uniform ellipsoid considered here: $(U_0 + U_1)r^2 = \epsilon$. We can set $r^2 = 1$ to obtain the critical energy of motion for a star in the field of a uniform ellipsoid and the Galaxy: $\epsilon_t = U_0 + U_1$. Thus, the surface of zero stellar velocity in the combined field of the model potentials of the cluster and the Galaxy, indeed, has an ellipsoidal shape and can be defined by the equation $r^2 = 1$ when $\epsilon = \epsilon_t$. The mass of the ellipsoid can therefore be estimated as $M_0 = 4\pi abc\rho/3$. In this case, the cluster stars completely fill the entire volume below the critical zero-velocity surface, as can be seen in the numerical simulations of [9, 10] and in the model of [11].

When $\epsilon = \epsilon_t$, in view of (2), the equation $(U_0 + U_1)r^2 = \epsilon$ can be simplified to the form $U(r = 1) = -U_1 + S$. Here, $U(r = 1)$ denotes the $U(\xi, \eta, \zeta)$ values at the surface $r^2 = 1$. Thus, a boundary condition of the same type as the boundary condition in model [11] is satisfied at the critical zero-velocity surface for the potential U.

We apply here the method of Marquardt [12] to determine U_0 and U_1 for the six open-cluster models of [3]. To this end, we use (2) to approximate the cluster potential $U(\xi, \eta, \zeta)$ averaged over the period of oscillations of the regular field for a number of points at the cluster periphery ($r \in [0.5, 1.2]$). See [10] for a description of the method used to select such points (ξ, η, ζ). The table lists the calculated model parameters. Column 1 gives numbers identifying the open-cluster models considered in [3], while columns 2 and 3 give the U_0 and U_1 values for these models in $(\text{pc/Myr})^2$.

We used the data listed in the table to calculate the mean semiaxes a_u, b_u, and c_u averaged over the six models [3]. The ratios of these mean semiaxes, $\langle a_u \rangle$, $\langle b_u \rangle$, and $\langle c_u \rangle$, are $\langle a_u \rangle : \langle b_u \rangle : \langle c_u \rangle \sim 1 : 0.84 : 0.98$. Thus, the equipotential surfaces of the potential U in the open-cluster models [3] are close to spheroids flattened along the η axis (i.e., along the direction of the cluster motion in the Galaxy).

The third equation in (5) can be integrated independently of the first two equations. The general solution of (5) has the form

$$\xi = C_1 \cos[\nu_1(t - t_1)] + C_2 \cos[\nu_2(t - t_2)],$$

$$\eta = AC_1 \sin[\nu_1(t - t_1)] + BC_2 \sin[\nu_2(t - t_2)], \tag{7}$$

$$\zeta = C_3 \cos[\beta_3(t - t_3)],$$

where $A = (\beta_1^2 - \nu_1^2)/(2\omega\nu_1)$, $B = (\beta_1^2 - \nu_2^2)/(2\omega\nu_2)$; ν_1, ν_2, and β_3 are the eigenfrequencies of (5); and C_i, t_i are constants $(i = 1, 2, 3)$. The constants ν_1 and ν_2 can be determined from the condition that the first two equations of (5) must have nonzero solutions:

$$\nu_{1,2} = \sqrt{\frac{P}{2}(1 \pm f^2)},$$

$$P = \beta_1^2 + \beta_2^2 + 4\omega^2, \quad f^2 = \sqrt{1 - \frac{4\beta_1^2\beta_2^2}{P^2}}.$$

Thus, the solution (7) for the coordinates ξ and η is a two-frequency function (with frequencies ν_1 and ν_2), while the solution for ζ is a single-frequency function (with frequency β_3). In our models, $\nu_1 < \nu_2 < \beta_3$. Columns 4, 5, and 6 of the table list the estimated values of these frequencies (in $(\text{Myr})^{-1}$) corresponding to the estimated parameters U_0 and U_1 from the same table.

Note that the U_0 and U_1 values derived by approximating the open-cluster gravitational potential considered in [3] using formula (2) depend on the adopted values of α_1 and α_3. The quantities a and b depend on α_1, and c depends on α_1 and α_3 (see the formulas for $|\xi_t|$, $|\eta_t|$, and $|\zeta_t|$), implying that the β_i^2 in (5) depend on the constants α_1 and α_3.

3. THE SMALL PARAMETER
IN THE STELLAR EQUATIONS
OF MOTION

The equations of motion of a star in the combined gravitational field of the cluster and Galaxy can be written in Hamiltonian form. We now use the cluster Lagrangian, L, written with allowance for the Lagrangian [5] for a star moving in the combined gravitational field of the cluster and Galaxy, to determine the generalized momenta of the ith cluster star:

$$p_{\xi_i} = \frac{\partial L}{\partial \dot{\xi}_i} = m_i(\dot{\xi}_i - \omega \eta_i), \tag{8}$$

$$p_{\eta_i} = \frac{\partial L}{\partial \dot{\eta}_i} = m_i(\dot{\eta}_i + \omega(R_0 + \xi_i)),$$

$$p_{\zeta_i} = \frac{\partial L}{\partial \dot{\zeta}_i} = m_i \dot{\zeta}_i,$$

where R_0 is the distance between the centers of mass of the cluster and Galaxy; unlike Chandrasekhar [5], we chose here signs that ensure that $\alpha_1 < 0$, $\alpha_3 > 0$, $\Phi > 0$, and $U > 0$.

We now use (8) and the cluster Lagrangian L to write the Hamiltonian of the cluster:

$$H = \frac{1}{2} \sum_{i=1}^{N} m_i \left\{ \left(\frac{p_{\xi_i}}{m_i} + \omega \eta_i \right)^2 \right. \tag{9}$$

$$+ \left[\frac{p_{\eta_i}}{m_i} - \omega(R_0 + \xi_i) \right]^2 + \left(\frac{p_{\zeta_i}}{m_i} \right)^2$$

$$- \omega^2 \left[(R_0 + \xi_i)^2 + \eta_i^2 \right] \left. \right\} + W - \sum_{i=1}^{N} m_i \Phi_i.$$

Here, N and W are the number of stars and potential energy of the cluster, respectively, and Φ_i is the Galactic potential at the point (ξ_i, η_i, ζ_i). The first sum in (9) is twice the kinetic energy of the cluster T. In this case, the unperturbed Hamiltonian of an ellipsoidal cluster can be written in the form $H_0 = T_0 + W_0 - \rho \int_Q \Phi dQ$, where the potential Φ is integrated over the volume Q of the cluster ellipsoid, and T_0 and W_0 are the kinetic and potential energy of the ellipsoidal cluster, respectively.

Let us now estimate μ. To this end, we determine the quantity T_0 from the condition of virial equilibrium for the cluster. When calculating T_0, we also use the formula for the integral of the cluster angular momentum relative to the ζ axis, L_ζ. According to Chandrasekhar [5], this integral of motion exists for clusters whose mass distributions have their symmetry defined by (2). As in the open-cluster models of [3], we set $L_\zeta = 0$ here (in this case, the cluster does not rotate with respect to external galaxies). We use

the formula $\Phi \simeq \Phi_0 - \omega^2 R_0 \xi - \dfrac{\alpha_1 + \omega^2}{2}\xi^2 - \dfrac{\omega^2}{2}\eta^2 - \dfrac{\alpha_3}{2}\zeta^2$ to compute $\int_Q \Phi dQ$. Here, Φ_0 is the value of the potential Φ at the Galactocentric distance R_0. Differentiating this formula with respect to ξ, η, ζ yields the corresponding components of the Galactic gravitational field obtained by Chandrasekhar [5].

The lack of the integral of motion L_ζ in [5] for a star cluster in the general case (i.e., without the restrictions imposed on the cluster symmetry in [5]) is due to the absence of rotational symmetry in the formula for the potential Φ in cylindrical coordinates. When computed using the exact equations of stellar motion taking into account the rotational symmetry of the potential Φ in cylindrical coordinates, and without imposing any constraints on the cluster symmetry, L_ζ can be shown to be an integral of the motion that does not change with time t and coincides with the value found by Chandrasekhar [5] when constraints are imposed on the cluster symmetry.

We now calculate $W_0 = -\dfrac{1}{2}\rho \int_Q U dQ$ using the quantity U from (2). We use the resulting formulas for T_0, W_0, and $\int_Q \Phi dQ$ to derive an expression for the constant H_0:

$$H_0 = -M_0 \left(\Phi_0 + \frac{\omega^2 R_0^2 + U_0}{2} - \frac{(3\beta_1^2 + 4\omega^2 + \alpha_1)a^2 + (3\beta_2^2 + 4\omega^2)b^2 + (3\beta_3^2 + \alpha_3)c^2}{20} \right). \tag{10}$$

We now introduce the small parameter $\mu = (H - H_0)/H$. In this case, $H = H_0 + \mu H$. The second term on the right-hand side can be viewed as a small perturbation of the Hamiltonian H. We then use formulas (9)–(10) and the parameters U_0 and U_1 for the six open-cluster models of [3] listed in the table to calculate the corresponding μ values. The parameters μ for the cluster models considered are listed in Column 7 of the table. The $|\mu|$ values are small compared to unity for these open-cluster models ($|\mu| \sim 0.01$–0.1), and μ can have different signs for different models.

The equations of motion of the ith star with the Hamiltonian H written in the form $H = H_0 + \mu H$ and using formula (9) have the form

$$\dot{p}_{\xi_i} = -\frac{\partial H}{\partial \xi_i} = (1+\mu)\left(\omega p_{\eta_i} + m_i \frac{\partial \Phi}{\partial \xi_i}\right) \quad (11)$$
$$+ m_i \frac{\partial U}{\partial \xi_i} - \mu \frac{\partial W}{\partial \xi_i},$$

$$\dot{p}_{\eta_i} = -\frac{\partial H}{\partial \eta_i} = (1+\mu)\left(-\omega p_{\xi_i} + m_i \frac{\partial \Phi}{\partial \eta_i}\right)$$
$$+ m_i \frac{\partial U}{\partial \eta_i} - \mu \frac{\partial W}{\partial \eta_i},$$

$$\dot{p}_{\zeta_i} = -\frac{\partial H}{\partial \zeta_i} = (1+\mu)\left(m_i \frac{\partial \Phi}{\partial \zeta_i}\right) + m_i \frac{\partial U}{\partial \zeta_i} - \mu \frac{\partial W}{\partial \zeta_i};$$

$$\dot{\xi}_i = \frac{\partial H}{\partial p_{\xi_i}} = (1+\mu)\left(\frac{p_{\xi_i}}{m_i} + \omega \eta_i\right), \quad (12)$$

$$\dot{\eta}_i = \frac{\partial H}{\partial p_{\eta_i}} = (1+\mu)\left(\frac{p_{\eta_i}}{m_i} - \omega(R_0 + \xi_i)\right),$$

$$\dot{\zeta}_i = \frac{\partial H}{\partial p_{\zeta_i}} = (1+\mu)\left(\frac{p_{\zeta_i}}{m_i}\right).$$

Differentiating (12) with respect to time, passing to the Lagrangian equations, and substituting the formulas for Φ, U and W into these equations, we find

$$\ddot{\xi}_i = (1+\mu) \quad (13)$$

$$\times \left(2\omega \dot{\eta}_i - (\beta_1^2 + \mu\alpha_1)\xi_i + \mu G \sum_{j \neq i} f_{ij}(\xi_j - \xi_i)\right),$$

$$\ddot{\eta}_i = (1+\mu)$$

$$\times \left(-2\omega \dot{\xi}_i - \beta_2^2 \eta_i + \mu G \sum_{j \neq i} f_{ij}(\eta_j - \eta_i)\right),$$

$$\ddot{\zeta}_i = (1+\mu)$$

$$\times \left(-(\beta_3^2 + \mu\alpha_3)\zeta_i + \mu G \sum_{j \neq i} f_{ij}(\zeta_j - \zeta_i)\right),$$

where $f_{ij} = m_j/(r_{ij}^2 - e^2)^{3/2}$, $i,j = 1,..,N$, $r_{ij}^2 = (\xi_i - \xi_j)^2 + (\eta_i - \eta_j)^2 + (\zeta_i - \zeta_j)^2$, and $e^2 = \text{const}$ is a small smoothing term added to r_{ij}^2, which we used in [3] when analyzing the open-cluster models.

Equations (13) become equal to Eqs. (5) when $\mu = 0$. Introducing the notation $\dot{\xi} = u$, $\dot{\eta} = v$, and $\dot{\zeta} = w$, we can write the solution of the system (13) in the form

$$\xi_i(t,\mu) = \xi_i(0,0) + \int_0^t u_i(t,0)dt, \quad (14)$$

$$\eta_i(t,\mu) = \eta_i(0,0) + \int_0^t v_i(t,0)dt,$$

$$\zeta_i(t,\mu) = \zeta_i(0,0) + \int_0^t w_i(t,0)dt,$$

$$u_i(t,\mu) = u_i(0,0) + \int_0^t \dot{u}_i(t,0)dt,$$

$$v_i(t,\mu) = v_i(0,0) + \int_0^t \dot{v}_i(t,0)dt,$$

$$w_i(t,\mu) = w_i(0,0) + \int_0^t \dot{w}_i(t,0)dt.$$

Here, the second argument of $\xi, \eta, \zeta, u, v, w$ is the μ value used. We then drop terms proportional to μ^2 in (13) to find the solution of system (13) in a first-order approximation in μ using the method of Picard. We use the unperturbed solution (7) as a zeroth-order approximation for $\xi, \eta, \zeta, u, v, w$. To find the solution (14), we expand the function f_{ij} in a power series in $\Delta_{ij} = r_{ij}^2 - r_{ij}^2(0)$ and limit it to terms $\sim \Delta_{ij}^3$:

$$f_{ij} \simeq m_j \left(s_{ij}^{-3} - \frac{3\Delta_{ij}}{2s_{ij}^5} + \frac{15\Delta_{ij}^2}{8s_{ij}^7} - \frac{35\Delta_{ij}^3}{16s_{ij}^9} + \cdots\right),$$
$$(15)$$

where $s_{ij}^2 = r_{ij}^2(0) + e^2$, $r_{ij}(0)$ is the initial distance between the ith and jth stars.

4. FREQUENCIES OF THE STELLAR MOTIONS

We applied the scheme (14) to solve the system (13) in a first-order approximation in μ. Due to the large number of computations, we could include in the solution (14) only the zeroth-, first-, and second-order terms of the power-series expansion of f_{ij} in Δ_{ij}. Since the solution has a very unwieldy form and is valid only over short time intervals, we do not present it here. However, we briefly describe below the frequencies of the harmonic oscillations corresponding to the solution.

(1) The frequencies we find for the stellar motions in ζ and w retaining the zeroth- and first-order terms in the power-series expansion of f_{ij} in Δ_{ij} are $\pm\beta_3, \pm3\beta_3, \pm\beta_3 \pm \nu_1 \pm \nu_2, \pm\beta_3 \pm 2\nu_1, \pm\beta_3 \pm 2\nu_2$ (20 frequencies). The frequencies we find for the motion along ξ, u and η, v are $\pm\nu_1, \pm\nu_2, \pm3\nu_1, \pm3\nu_2,$

$\pm\nu_1 \pm 2\beta_3$, $\pm\nu_2 \pm 2\beta_3$, $\pm\nu_1 \pm 2\nu_2$, $\pm\nu_2 \pm 2\nu_1$ (24 frequencies).

(2) For the stellar motions in ζ and w computed using the zeroth-, first-, and second-order terms of the power-series expansion of f_{ij} in Δ_{ij}, we find in addition the frequencies $\pm\beta_3 \pm 2\nu_1 \pm 2\nu_2$, $\pm\beta_3 \pm 3\nu_1 \pm \nu_2$, $\pm\beta_3 \pm \nu_1 \pm 3\nu_2$, $\pm3\beta_3 \pm 2\nu_1$, $\pm3\beta_3 \pm 2\nu_2$, $\pm5\beta_3$, $\pm\beta_3 \pm 4\nu_1$, $\pm\beta_3 \pm 4\nu_2$, $\pm3\beta_3 \pm \nu_1 \pm \nu_2$ (50 frequencies). For the corresponding stellar motions in ξ, u and η, v, we find in addition the frequencies $\pm5\nu_1$, $\pm5\nu_2$, $\pm4\nu_1 \pm \nu_2$, $\pm\nu_1 \pm 4\nu_2$, $\pm2\beta_3 \pm 3\nu_1$, $\pm2\beta_3 \pm 3\nu_2$, $\pm2\nu_1 \pm 3\nu_2$, $\pm3\nu_1 \pm 2\nu_2$, $\pm4\beta_3 \pm \nu_1$, $\pm4\beta_3 \pm \nu_2$, $\pm2\beta_3 \pm \nu_1 \pm 2\nu_2$, $\pm2\beta_3 \pm 2\nu_1 \pm \nu_2$ (52 frequencies).

The derived frequencies are linear combinations of the three eigenfrequencies of the stellar motions in the combined field $\mathbf{F}$ of the potential (2) and the Galactic potential. In this paper and in [1], we compared these frequencies with those derived via a Fourier analysis of the trajectories of the halo stars [1, 13] in open-cluster model 1 from [3]. This comparison enables us to identify several of the most significant resonances in the motions of the halo stars in ζ and w, at the frequencies $2\beta_3 - \nu_2 \simeq \frac{3}{5}\omega_r$, $3\beta_3 - 2\nu_2 \simeq \frac{5}{8}\omega_r$, $\beta_3 + 3\nu_1 - \nu_2 \simeq \frac{5}{7}\omega_r$, $\frac{4}{7}\omega_r$. The main resonances in the motions of the halo stars in ξ, u in cluster model 1 from [3] have the frequencies $\beta_3 - 3\nu_1 + \nu_2 \simeq \frac{2}{5}\omega_r$, $\frac{4}{9}\omega_r$, $\nu_2 \simeq \frac{9}{17}\omega_r$, $2\nu_2 - 3\nu_1 \simeq \frac{7}{17}\omega_r$, $\frac{3}{8}\omega_r$, whereas the motions in η, v have their main resonances at the frequencies $-\beta_3 + 2\nu_2 \simeq \frac{1}{2}\omega_r$, $2\beta_3 - 3\nu_1 \simeq \frac{7}{15}\omega_r$, $\beta_3 - 3\nu_1 + \nu_2 \simeq \frac{2}{5}\omega_r$, $3\beta_3 - 2\nu_2 \simeq \frac{3}{5}\omega_r$. Here, ω_r is the oscillation frequency of the regular field of the cluster. Six to ten (and sometimes up to 15) halo-star trajectories in cluster model 1 of [3] are grouped at (or near) these frequencies. The frequencies reported here are rational multiples of the oscillation frequency of the regular cluster field. The formation of such groups is most likely due to the effect of "pulling" or "frequency shifting" during the synchronization of the oscillations, as is described in the literature for oscillatory systems with small or large numbers of degrees of freedom [14, p. 290; 15, p. 348]. As a result of such synchronization, the system develops a set of frequencies that are commensurable with the oscillation frequency of the regular field. Synchronous motion of the stars along their trajectories (motions with the same frequency) is observed at some of these frequencies. At the same time, the system has a set of clustercentric distance intervals that are preferred by halo stars moving in the presence of oscillations of the regular field.

5. CONCLUSIONS

(1) We have constructed a model for the potential (2), which approximates the gravitational potential at the peripheries of the numerical dynamic open-cluster models of [3] and allows analytical solution of the equations of stellar motion in the combined force field $\mathbf{F}$ produced by the potential (2) and the Galactic potential. The parameters of (2) and the shapes of its equipotential surfaces and of the zero-velocity surfaces in the force field $\mathbf{F}$ were determined for six open-cluster models from [3].

(2) We derived equations of motion for the stars at the periphery of the open clusters containing the small parameter μ. The μ values were determined for the six cluster models from [3] using the parameters of the potential (2) and the Galactic potential. A general analytical solution for the equations of the stellar motions for $\mu = 0$ was derived. The motion of a star in the force field $\mathbf{F}$ is a triple-frequency motion (single-frequency motion in ζ and two-frequency motion in ξ and η). These frequencies are the eigenfrequencies of the stellar motions in the combined force field produced by the potential (2) and the Galactic potential.

(3) A convergence method was used to analyze the frequencies of the stellar motions in first-order expansions of the solutions of the equations of the stellar motions at the cluster peripheries in terms of the small parameter μ. The resulting frequencies are linear combinations of the eigenfrequencies of the stellar motions in the force field $\mathbf{F}$. The coefficients of these linear combinations are integers. Several main resonances in the halo-star motions in cluster model 1 from [3] were found. The frequencies of these resonances are commensurable (or close to commensurable) with the frequency of oscillations of the regular field of the cluster.

(4) The formation of small groups of halo-star trajectories at (or near) these frequencies, which was pointed out by Danilov and Leskov [1] for cluster model 1, indicates that the system develops a set of intervals of clustercentric distances that are preferred by halo stars moving in the presence of oscillations of the regular cluster field.

REFERENCES

1. V. M. Danilov and E. V. Leskov, Astron. Zh. **82**, 219 (2005) [Astron. Rep. **49**, 190 (2005)].
2. J. Laskar, Nature **338**, 237 (1989).
3. V. M. Danilov and E. V. Dorogantseva, Astron. Zh. **80**, 526 (2003) [Astron. Rep. **47**, 483 (2003)].
4. V. M. Danilov, Astron. Zh. **79**, 986 (2002) [Astron. Rep. **46**, 887 (2002)].
5. S. Chandrasekhar, *Principles of Stellar Dynamics* (Univ. Chicago Press, Chicago, 1942; Inostrannaya Literatura, Moscow, 1948).

6. I. R. King, *An Introduction to Classical Stellar Dynamics* (URSS, Moscow, 2002) [in Russian].

7. S. A. Kutuzov and L. P. Osipkov, Astron. Zh. **57**, 28 (1980) [Sov. Astron. **24**, 17 (1980)].

8. L. Spitzer, Jr., *Dynamical Evolution of Globular Clusters* (Princeton Univ. Press, Princeton, 1987; Mir, Moscow, 1990).

9. E. Terlevich, Mon. Not. R. Astron. Soc. **224**, 193 (1987).

10. V. M. Danilov, Astron. Zh. **74**, 188 (1997) [Astron. Rep. **41**, 163 (1997)].

11. D. C. Heggier and N. Ramamani, Mon. Not. R. Astron. Soc. **272**, 317 (1995).

12. D. Marquardt, J. Soc. Ind. Appl. Math. **11**, 1 (1963).

13. V. M. Danilov and E. V. Leskov, in *Order and Chaos in Stellar and Planetary Systems*, Ed. by G. G. Byrd, K. V. Kholshevnikov, A. A. Myllri, I. I. Nikiforov, and V. V. Orlov (Astron. Soc. Pac., San Francisco, 2004); Astron. Soc. Pac. Conf. Ser. **316**, 183 (2004).

14. V. V. Migulin, V. I. Medvedev, E. R. Mustel', and V. N. Parygin, *Foundations of the Theory of Oscillations* (Fiz.-Mat. Literatura, Moscow, 1988) [in Russian].

15. M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Theory of Oscillations and Waves* (Nauchno-izdatel'skii Tsentr "Regulyarnaya i khaoticheskaya dinamika," Moscow−Izhevsk, 2000).

Translated by A. Dambis

The Afterglow of a Dense Molecular Cloud after the Passage of a Cosmological Gamma-Ray Burst

M. V. Barkov and G. S. Bisnovatyĭ-Kogan

Space Research Institute, Moscow, Russia
Received November 5, 2004; in final form, February 17, 2005

Abstract—Bolometric light curves for the afterglow resulting from the passage of a gamma-ray burst through a molecular cloud are computed. The profile and duration of the afterglow light curve depend strongly on the distribution of matter in the cloud, the degree of collimation of the gamma-ray radiation, and the observing conditions. The peak can be reached as soon as seven days (the gamma-ray burst is located at some distance from the center of a molecular cloud with small-scale density enhancements), or as long as one to three years (the gamma-ray burst is located at the center of a uniform molecular cloud) after the burst. The bolometric luminosity of the re-radiated signal can reach 6.5×10^{42} erg/s.
© 2005 Pleiades Publishing, Inc.

1. INTRODUCTION

Studies of the afterglows of prolonged gamma-ray bursts (GRBs) in the X-ray, optical, and ultraviolet made it possible to establish their cosmological nature and to estimate the gigantic scale of the associated energy released. If the radiation is isotropic, the gamma-ray energy exceeds the rest energy of the Sun (GRB 990123 [1]), and supposing collimation of the gamma-ray beam provides the only means to avoid the need for such a large-scale, rapid energy release, which cannot be satisfactorily explained in any existing model.

Two stages can be traced in observations of optical afterglows. The first stage, in which there is a comparatively bright optical transient (to 9^m in the case of GRB 990123) with a duration comparable to that of the GRB itself, is followed by a much more prolonged (lasting up to a year) and weak stage, which clearly has a different origin. Several models have been proposed to explain the light curve in the prolonged stage, of which the most popular is the collimated-fireball model [2] (which we will refer to as the standard model). This phenomenological model postulates the formation of a fireball by some unknown mechanism and is able to explain many, but not all, of the properties of the post-flare radiation. Examples are provided by GRB 011211 and GRB 030227, whose XMM X-ray observations cannot be explained by the standard model [3]. The thermal spectra that were obtained, which correspond to an optically thin plasma, should not arise in the fireball model. In addition, observations of X-ray lines imply high concentrations of nuclei that should be split into nucleons in a relativistic plasma.

Analysis of the light curves of the afterglows of various GRBs led to the idea that there might be a connection between GRBs and very energetic supernovae (energies $>3 \times 10^{51}$ erg, hypernovae) corresponding to type SN 1c. The first such example was provided by SN 1998bw, which was associated with the anomalously weak burst GRB 980425. The best manifestation of a connection between a GRB and a supernova is probably the bright burst GRB 030329, in whose afterglow was detected supernova SN 2003du. The basis for this identification is the similarity of the spectrum of the afterglow of GRB 030329 and the spectrum of SN 1998bw, whose presence was noted a month after the onset of optical observations of this burst [4, 5].

A connection between GRBs and star-forming regions was noted based on spectra obtained immediately after the detection of the X-ray and optical afterglows [6, 7].

Dense material surrounding the region of a GRB will be heated by the GRB radiation and reradiate the absorbed radiation in softer spectral regions, giving rise to prolonged optical and ultraviolet afterglows.

The luminosity and duration of such afterglows and prospects for their detection depend strongly on both the properties of the surrounding medium and the properties of the GRB itself: its power, its collimation, and, to a lesser extent, its spectrum. The first

computations of heating of a cloud and the resulting afterglow were carried out in [8, 9].

In these papers we have considered a short, powerful gamma-ray impulse propagating from the center of a simple, one-dimensional model with a homogeneous, spherical cloud. This problem will have a self-similar solution if we do not include the effects of cooling of the heated gas [10]. A detailed discussion of the results of this numerical modeling are presented in [11]. Computed spectra of the afterglows associated with the cooling of the gas heated by the GRB radiation are presented in [12].

It is shown in [11] that cooling of a homogeneous molecular cloud with a density of 10^4-10^5 cm^{-3} gives rise to an afterglow in the optical and ultraviolet with a characteristic plateau lasting $5-20$ yrs (for the cloud parameters used and clouds with sizes of $1.5-10$ pc). In this case, with a density of 10^5 cm^{-3}, the visual magnitude over ~ 5 yrs reaches 22^m for an isotropic burst with an energy of 10^{52} erg and a flux at the Earth of 10^{-4} erg/cm^2.

Such prolonged, comparatively bright optical transients are not observed for powerful GRBs: their optical afterglows become undetectable (weaker than $\sim 25^m$) within several weeks or months [13, 14].

One reason for this discrepancy could be that the model of [11] is overidealized. In reality, the burst might occur at the edge of an inhomogeneous molecular cloud, and the radiation might be anisotropic, propagating within a fairly narrow cone, as in the standard model. Studies of a more realistic picture of the interaction between a GRB and a dense cloud require at least two-dimensional models; one example is the two-dimensional numerical simulations of [15]. Computations of the interaction of a GRB and a dense cloud based on the equations of radiative hydrodynamics were carried out for various density distributions, cone opening angles, and GRB energies. Rapid heating of the cloud disrupts the hydrostatic equilibrium and gives rise to motions that lead to the formation of a shock wave [10]. The role of the shock in the heated gas is important only in the vicinity of the GRB itself, while the main process at the periphery of the cloud is cooling of the gas and the production of the afterglow.

The main results of computations of the formation of gas-dynamical flows initiated by a GRB for various parameters of the GRB and dense cloud are presented and discussed in [16].

Here, we present the results of computations of the afterglow due to the heated gas in a formulation that is more precise than that used in [15]: we have used the Klein–Nishina cross section to describe Compton heating, taken into account more accurately (although still only approximately) radiative transfer in lines during the cooling of the gas, and carried out computations for more realistic molecular-cloud and GRB parameters. The resulting light curves are much more varied than those obtained in [11] and can be meaningfully compared with some observational results for the optical afterglow of GRB 030329 [14, 17].

2. FORMULATION OF THE PROBLEM AND MAIN EQUATIONS

The system of radiative hydrodynamical equations that were solved for the interaction between a GRB and a dense cloud is presented in [16]. Let us consider in more detail the heating and cooling processes in the cloud, which play an important role in the formation of the spectrum and afterglow light curve. Since the main thermal processes and the formation of the afterglow occur in comparatively distant regions, where the hydrodynamical motions are negligibly small ($v \ll v_s$, where v_s is the sound speed behind the heating front [16]), for the purpose of studying the light curves, we can restrict our analysis to solving the thermal-balance equation

$$\frac{\partial}{\partial t}\left(\rho\varepsilon\right) = \rho H_\gamma - \rho C_\gamma. \qquad (1)$$

Here, ρ is the density, ε the internal energy, and H_γ and C_γ the rates of heating and cooling per unit mass. We take the matter to be fully ionized. This is justified by the fact that the matter is heated to high enough temperatures during the passage of the gamma-ray front that it will be fully ionized. Because the speed of sound is small compared to the speed of light (the speed at which the heating and cooling fronts propagate; see below), the matter at temperature jumps does not have time to attain high velocities ($v/v_s \sim v_s/c$).

Since the time scales of interest ($\sim 10^5-10^8$ s) are much longer than the duration of the burst ($\lesssim 100$ s), the GRB itself can be represented as an instantaneous impulse with energy Γ:

$$L = \Gamma\delta\left(t - \frac{r}{c}\right). \qquad (2)$$

For gamma-rays with $h\nu \gg B_e^{a,i}$, where $B_e^{a,i}$ is the binding energy of an electron in an atom or ion, the cross section for interactions between these gamma-ray photons and electrons will be essentially the same whether the electrons are bound or free. The GRB spectrum is specified in the form

$$\frac{dL}{dE} = \frac{L}{E_{\max}}e^{-E/E_{\max}}; \qquad (3)$$

we also introduce the notation $W(E, E_{\max}) = \dfrac{dL}{LdE}$.

The rate of heating by the gamma-ray radiation per unit mass of gas, H_γ, was taken from [18−20] in the form

$$H_\gamma = \frac{L}{4\pi r^2} \frac{\mu_e \sigma_T}{m_u} \frac{E_{\max} f_h(E_{\max}) - 4kT f_c(E_{\max})}{m_e c^2},$$
(4)

where E is the photon energy, m_e is the electron mass, m_u is the nucleon mass, μ_e is the number of electrons per nucleon, and σ_T is the Thomson cross section for interactions between electrons and photons. The functions

$$f_h(E_{\max}) = \frac{1}{E_{\max}} \int_0^\infty W(E, E_{\max}) s(E) E dE \quad (5)$$

and

$$f_c(E_{\max}) = \int_0^\infty W(E, E_{\max}) q(E) dE \quad (6)$$

characterize the spectral-averaged corrections to the heating (f_h) and cooling (f_c) of the gas due to Compton interactions with the GRB radiation. These are obtained by integrating over the spectrum taking into account the dependence of the cross section for the relativistic scattering of a photon on an electron on the photon energy (the Klein−Nishina cross section, σ_{KN}).

Functions taking into account the relativistic corrections for the heating $s(E)$ and cooling $q(E)$ are calculated in [18]:

$$s(E) = \frac{3}{8E} \left[\left(\frac{1}{E} - \frac{2}{E^2} - \frac{3}{E^3} \right) \ln(1 + 2E) \right.$$
(7)

$$\left. + \frac{2 \left(-10E^4 + 51E^3 + 93E^2 + 51E + 9 \right)}{3E^2 (1 + 2E)^3} \right]$$

and

$$q(E) = \frac{3}{32} \left[\left(\frac{2}{3E} - \frac{2}{E^2} - \frac{13}{3E^2} \right) \ln(1 + 2E) \right.$$
(8)

$$\left. + \frac{-216E^6 + 476E^5 + 2066E^4 + 2429E^3 + 1353E^2 + 363E + 39}{3E^2 (1 + 2E)^5} \right].$$

The normalization here was chosen so that $f_h = f_c = 1$ when $E_{\max} \ll m_e c^2$. After substituting (3), (7), and (8) into (5) and (6), we integrated numerically for various $E_{\max}$ values and approximated the results using the formulas ($E_{\max}$ is in MeV)

$$f_h(E_{\max}) = (6.22\, E_{\max} + 1)^{-1.07} \quad (9)$$

and

$$f_c(E_{\max}) = (7.5\, E_{\max} + 1)^{-0.65} \quad (10)$$

in the energy interval $0.01\,\text{MeV} \leq E_{\max} \leq 5\,\text{MeV}$, which yielded an accuracy better than 3%. At large distances from the GRB source, the matter is not heated to such high temperatures, so that $E_{\max} \gg$

$4kT$, and cooling due to inverse Compton scattering is negligible. In accordance with (9), the rate of Compton heating in (4) will be virtually independent of the hardness of the spectrum if $E_{\max} \gtrsim 1/6$ MeV.

The rate of cooling due to line and continuum radiation per unit mass of optically thin plasma, C_γ, was calculated in [21, 22]. The results of these calculations were approximated by the analytical formula

$$C_\gamma = \frac{\Lambda(T)n^2}{\rho}, \quad (11)$$

where

$$\Lambda(T) = \begin{cases} 0, & T < 10^4 \text{ K}, \\ 10^{-48.8}\, T^{6.4}, & 10^4 \text{ K} < T < 10^{4.25} \text{ K}, \\ 10^{-16.5}\, T^{-1.2}, & 10^{4.25} \text{ K} < T < 10^{4.5} \text{ K}, \\ 10^{-27.48}\, T^{1.24}, & 10^{4.5} \text{ K} < T < 10^5 \text{ K}, \\ 10^{-21.03}\, T^{-0.05}, & 10^5 \text{ K} < T < 10^{5.4} \text{ K}, \\ 10^{-13.6698}\, T^{-1.413}, & 10^{5.4} \text{ K} < T < 10^{5.86} \text{ K}, \\ 10^{-22.8378}\, T^{0.1515}, & 10^{5.86} \text{ K} < T < 10^{6.19} \text{ K}, \\ 10^{-13.1969}\, T^{-1.406}, & 10^{6.19} \text{ K} < T < 10^{6.83} \text{ K}, \\ 10^{-22.2877}\, T^{-0.075}, & 10^{6.83} \text{ K} < T < 10^{7.5} \text{ K}, \\ 10^{-26.6}\, T^{0.5}, & 10^{7.5} \text{ K} < T. \end{cases} \tag{12}$$

The solar chemical composition was used for the interstellar medium (ISM) [23], and it was assumed that all heavy elements are in gaseous form.

3. MAIN PHYSICAL PROCESSES AND CONDITIONS

Let us consider the simplifications that we have adopted and the physical processes that are important in the formation of the spectrum of a cooling cloud that was heated by the radiation of a GRB.

3.1. Approximation of a Fully Thermalized, One-Temperature Plasma

The calculations carried out in [24] showed that, at each moment during the thermal relaxation, the ionized component of the medium is close to a stationary ionization state corresponding to some temperature. This result is illustrated in Figs. 1 and 2, which are reproduced from [24]. Figure 2 shows the distribution of the number densities of some ions in a thermal plasma at various temperatures, and Fig. 1 the time dependence of the number densities of these same ions in a plasma that is heated from a temperature of $T_{in} = 10^5$ K to a temperature of $T_{fin} = 10^9$ K. It also follows from the results of [24] that, for a density of the interstellar material of $n \sim 10^5$ cm^{-3} and a temperature on the order of $T \sim 10^7$ K, the time required to establish ionization balance is close to 10^6 s, which is comparable to the time scale we are interested in here.

Let us demonstrate the justification of the one-temperature approximation for the plasma. We considered the action of gamma-rays with a flat spectrum
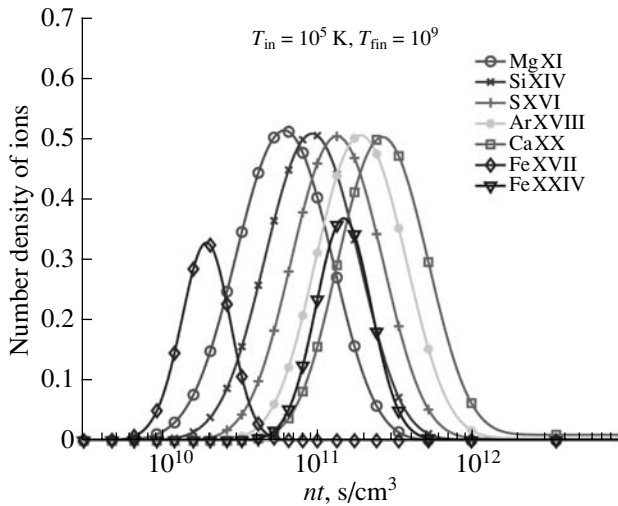

Fig. 1. Relative number density of ions as a function of time during the thermalization of a plasma that is rapidly heated to a temperature of $T_{fin} = 10^9$ K from an initial temperature of $T_{in} = 10^5$ K.
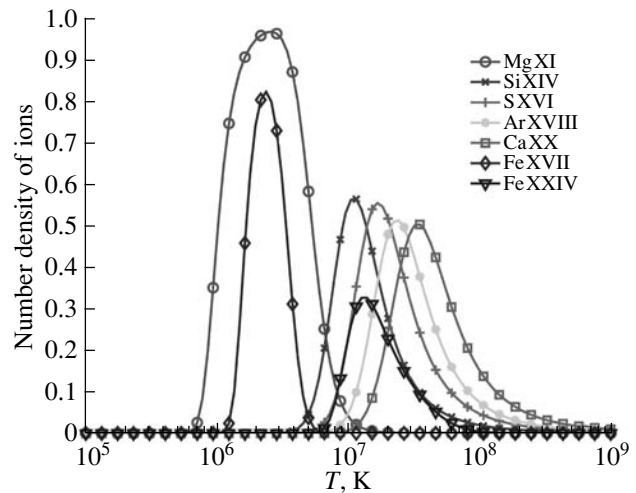

Fig. 2. Relative number density of ions as a function of temperature for a thermalized plasma.
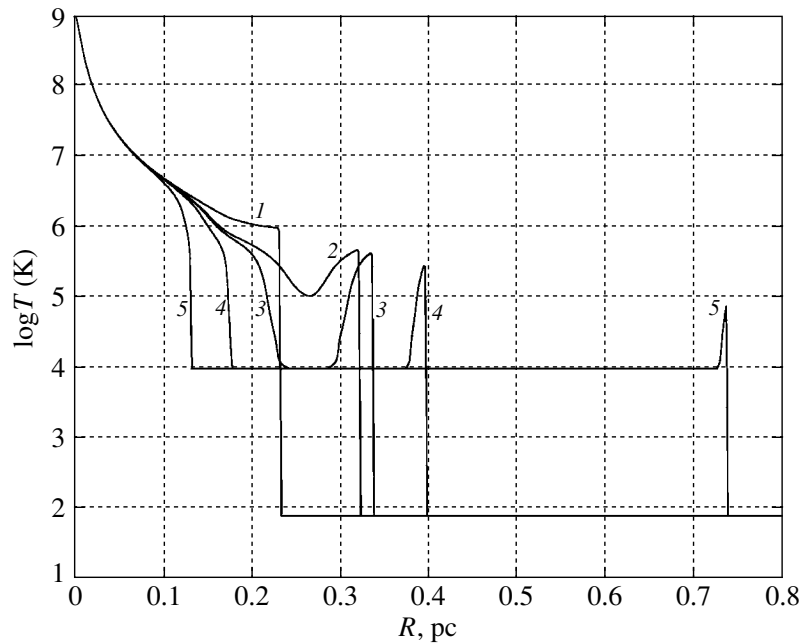
Fig. 3. Dependence of the temperature in the cloud on distance from the GRB source at various times for case 1. The numbered curves correspond to times after the GRB of (1) 0.76, (2) 1.054, (3) 1.103, (4) 1.30, (5) 2.40 yr.

on an ISM with the normal chemical composition. For this spectrum and with $E_{\max} \sim 0.6$ MeV, the number density of photons with energies of 4 keV is approximately a factor of 150 higher than the number of photons with energies of 0.6 MeV, which means that the efficiency of primary ionization by the 4 keV gamma-rays is a factor of $\ln 150 \approx 5$ higher than the efficiency of ionization by gamma-rays at the hard end of the spectrum. We estimate the degree of ionization for photon energies of $E_{\min} = 4$ keV, since the cloud remains optically thin to these photons, and interactions between the photons and atoms are not important; i.e., the electrons can be taken to be free and the interaction cross section is equal to the Thomson cross section. This enables us to obtain a lower limit for the degree of ionization of the gas immediately after the passage of the gamma-ray impulse:

$$\frac{n_e}{n} \sim \min \left\{ 1, \frac{\Gamma \ln(E_{\max}/E_{\min})}{4\pi r^2 E_{\max}} \sigma_T \right\}. \qquad (13)$$

For the radiation fluxes considered, this degree of ionization in the numerical models in Section 4 is $\sim 10^{-4}$ and ~ 1 at the edge and in the central region of the cloud, respectively.

Since we have a directed flux of gamma-rays, after Compton scattering on electrons in the medium, we obtain a directed flux of fast electrons with energies of several hundreds of keV moving parallel to the flux of gamma-rays against the background of relatively slow electrons with energies of the order of several keV. The flux of energetic electrons moving

throgh the region of low-energy electrons rapidly becomes isotropic, and loses energy to the excitation of plasma turbulence, leading to efficient thermalization of the plasma. The decay increment for the plasma oscillations is given by

$$\gamma = \frac{\sqrt{3}}{2} \left(\frac{\alpha}{2} \right)^{1/3} \omega_{pe}, \quad \omega_{pe} = \left(\frac{4\pi e^2 n_e}{m_e} \right)^{1/2}, \quad (14)$$

where $\alpha = n_h/n_e \sim 1/100$ is the relative number density of hot electrons with energies of several hundreds of keV in the beam, n_h is the number density of these hot electrons, and ω_{pe} is the plasma frequency. The characteristic time scale for this thermalization is

$$\tau \sim \frac{2\pi}{\gamma} \ll 1 \text{ s} \qquad (15)$$

for the densities considered [25–28]. The electrons thermalized in this way will have temperatures of 10^9 K, which coincides with the temperature of the cloud near the GRB source but is one to two orders of magnitude higher than the temperature given by (1), (4) at a distance of about 0.1–1 pc from the GRB (Fig. 3).

The cross section for the ionization of atomic hydrogen by these electrons is [29]

$$\sigma(\varepsilon) = \frac{\pi a_0^2 (1.07 \ln \varepsilon + 5.56)(\varepsilon - 1)}{\varepsilon^2 + 1.67\varepsilon + 3.57}, \qquad (16)$$
$$E_e \sim 100 \text{ keV},$$

where a_0 is the Bohr radius of the hydrogen atom, $\pi a_0^2 = 0.88 \times 10^{-16}$ cm^2, and $\varepsilon = E_e/13.6$ eV.

Knowing the electron-collisional ionization cross section for hydrogen and the energy of the electrons, we can estimate the time required for the gas to become fully ionized; this turns out to be $t_{ion} = 1/\sigma n_e v_e \sim 4 \times 10^3$ s for a density of 10^5 cm^{-3}. In the case of complete ionization ($n_e = n$), the time between Coulomb collisions between electrons [30],

$$\tau_{ee} = \frac{3\sqrt{3m_e}(kT)^{3/2}}{8\pi nLe^4} \simeq \frac{T_8^{3/2}}{n} \times 10^{10} \text{ s} \qquad (17)$$

(T_8 is the temperature in units of 10^8 K and $L \simeq 30$ is the Coulomb logarithm), becomes so short that, when $n = 10^5$ cm^{-3}, the electron component is thermalized over a time of the order of 10^5 s, and the temperature is lowered to the values given by (1), (4).

Coulomb collisions are able to fully thermalize the plasma over a time [30]

$$\tau_{ie} = \frac{3m_i(kT)^{3/2}}{8\sqrt{2\pi m_e}nLe^4} \simeq \frac{T_8^{3/2}}{n} \times 10^{13} \text{ s.} \qquad (18)$$

For temperatures of $T \sim 10^6$ K (where the cooling rate is highest) and densities of 10^5 cm^{-3}, this time is about 10^5 s. Starting from this time, the approximation of a one-temperature plasma can be applied. The error brought about by the delay in the thermalization does not exceed 10%. The times presented are appropriate for gas densities $>10^3$ cm^{-3}. According to the computations of [8], the action of the GRB on a low-density ISM does not lead to the formation of a double that could be observed at cosmological distances using modern telescopes.

3.2. Role of Interstellar Dust

Results of computations of heating of the ISM by a GRB taking into account the ionization of atoms, dissociation of molecular hydrogen, and the destruction of dust due to heating and electrostatic forces are presented in [31]. In this last case, the dust grains are torn apart by repulsive Coulomb forces arising due to the ejection of electrons from the grain during impacts with energetic photons. For a GRB energy of the order of 10^{52} erg, dust can be evaporated out to a distance of 15 pc from the GRB source [31]. The cooling of the ISM due to interactions with dust at a high temperature on the order of 10^8 K is considered in [29]. In this case, the dust is rapidly destroyed. Since the radiated energy is approximately equal to the binding energy of the dust (1 eV/atom), the energy radiated by dust is negligible in our model due to the small number of heavy elements in the ISM and the low energy output of this process per unit mass.

The peak of the radiation intensity by dust grains is in the infrared. The IR energy radiated by dust is two to three orders of magnitude lower than the energy radiated by the gas in the ultraviolet and optical. For the reasons described above, the influence of dust on the radiation was disregarded.

3.3. Approximation of an Optically Thin Plasma

When computing the luminosity of the cloud, we assumed that the radiation is produced by an optically thin plasma. As was shown in [8, 11, 16], at the densities typical for a molecular cloud, the topology of the radiating fronts is such that nearly all the radiation emitted by the heated cloud passes through a region of thermal material with a temperature of the order of 10^4 K. In our computations, this region ($r_0 = 0.2$ pc) is opaque to Lyman line and continuum radiation right to 1.2−1.5 keV. The most energetic photons will first collide with heavy ions and be reprocessed into softer photons. Further, they will be absorbed by hydrogen atoms and be reradiated as continuum radiation (an important role is played by two-photon transitions from the second level) and line radiation corresponding to transitions to the second and higher levels. A more detailed analysis can be found in [11, 15]. Thus, a large fraction of the afterglow radiation will arrive at the Earth in the form of photons with energies below 13.6 eV. We assumed that all radiation arrives at the Earth in the form of optical and ultraviolet radiation.

Let us consider what volume of cool material ($T_e \sim 10^4$ K) can be ionized by the hot central region. In our computations (see case 1 below), the radius of this hot, central region does not exceed $r_{hot} \sim 10^{18}$ cm, or 0.3 pc. The number of photons emitted in this region with energies higher than the Lyman continuum energy, $\nu_c = 3.29 \times 10^{15}$ Hz, will be [32]

$$N_\nu = 4\pi \int\limits_0^{r_{hot}} \int\limits_{\nu_c}^{\infty} r^2 \frac{\varepsilon_\nu^{ff}}{h\nu} d\nu dr, \qquad (19)$$

where

$$\varepsilon_\nu^{ff} = n_e n^+ \frac{2^5 \pi^2}{(6\pi)^{3/2}} \frac{e^6}{m_e^2 c^3} \left(\frac{m_e}{kT}\right)^{1/2} g_\nu e^{-\frac{h\nu}{kT}} \qquad (20)$$

$$= 1.72 \times 10^{-42} n_e n^+ \left(\frac{1}{T_7}\right)^{1/2} e^{-\frac{h\nu}{kT}}$$

is the volume emission coefficient due to free−free transitions. It follows from the computations of [11, 15] and our current analysis that the luminosity of the central region changes only slightly for a long time, remaining in the cases we have considered of the order of or lower than 10^{40} erg/s. The bulk of the ionizing photons have energies of about 20 eV, and the total number of photons capable of ionizing hydrogen emitted per second turns out to be

$N_\nu \sim 3.4 \times 10^{50}$ s^{-1}. We can write the hydrogen recombination rate Θ in the cloud to all levels higher than the first in the form

$$\Theta = 4\pi \sum_2^\infty C_i \int_{r_{hot}}^{r_{hot}+\delta r} n_e n^+ r^2 dr, \qquad (21)$$

where δr is the thickness of the layer of ionized hydrogen. The hydrogen recombination coefficient at a temperature of $T_e \sim 10^4$ K is $\sum_2^\infty C_i = 2.17 \times 10^{-13}$ cm^3/s.

Assuming that $n_H \approx n_i = n_e$, the thickness of the layer δr, in which hydrogen in the intermediate cool zone (Fig. 3) will be highly ionized by radiation from the cooling hot gas in the central region (the Strömgren zone for an electron temperature of $T_e \sim 10^4$ K) proves to be much smaller than the size of the region of hot gas, where the temperature is $T \gtrsim 3 \times 10^6$ K:

$$\delta r = \frac{N_\nu}{4\pi r_{hot}^2 n_i n_e \sum_2^\infty C_i} \approx \frac{0,0057}{n_5^2} \text{ pc}, \qquad (22)$$

where $n_5 = n/10^5$ cm^{-3}. We find that the zone of ionized hydrogen is directly adjacent to the cooling front of the inner hot region, and essentially merges with this region.

4. COMPUTATIONAL RESULTS

As we noted above, the computations were carried out in an optically thin approximation; i.e., we neglected the opacity of the gaseous cloud to its own radiation, and the observed luminosity was calculated by integrating the radiative losses over the volume of the cloud, taking into account the finite speed of light. Measuring the time $\tilde{t}$ from the time when the GRB is detected, we obtain the following expression for the observed time dependences of the afterglow luminosity, $L(\tilde{t})$:

$$L(\tilde{t}) = 2c \int_{\max\left(\frac{\tilde{t}}{2}, \tilde{t} - \frac{R}{c}\right)}^{\tilde{t}+\frac{R}{c}} dt \int_0^\pi d\phi \qquad (23)$$

$$\times \int_0^{\min\left(c\sqrt{2t\tilde{t}-\tilde{t}^2}, \sqrt{R^2-c^2(t-\tilde{t})^2}\right)} \varepsilon(r,z,\phi,t,\alpha) z dz,$$

where $\varepsilon(r,z,\phi,t,\alpha)$ is the power of the radiation per unit volume at the point with cylindrical coordinates ($r = \sqrt{c^2(t-\tilde{t})^2 + z^2}$, z, ϕ) at the time t after the GRB, and α is the angle between the symmetry axis for the problem (a line passing through the

GRB source and the center of the molecular cloud) and the direction toward the observer. Formula (23) generalizes the corresponding expression obtained in [11] to the case when the observer is not located along the line joining the GRB and the center of the molecular cloud. When $\alpha \neq 0$, it is necessary to carry out a rotation of the coordinate system. We denote Π to be the plane containing the lines joining the GRB with the observer and the GRB with the center of the molecular cloud. We introduce the variable $\theta = \arccos\left(c(t-\tilde{t})/r\right)$, which describes the angle between the ray from the GRB to the observer and the projection onto the Π plane of the ray from the GRB to the integration point. The coordinates of this point on the Π plane are $x = r\cos\theta, y = r\sin\theta\cos\phi$. After performing a rotation through the angle α, projecting the point onto the GRB−cloud center axis, finding the distance from the point at the base of the projection to the GRB source, $x'(\alpha) = x\cos(\alpha) + y\sin(\alpha)$, and obtaining the angular coordinate in the frame fixed to the GRB and the center of the molecular cloud, $\theta'(\alpha) = \arccos(x'(\alpha)/r)$, we obtain

$$\varepsilon(r,z,\phi,t,\alpha) = \Lambda\left(T(r,\theta'(\alpha),t)\right) n(r,\theta'(\alpha),t)^2, \qquad (24)$$

where $\Lambda(T)$ is given by (12).

The main results for all the cases we computed are given in Table 1, which presents (1) a number identifying each case computed; (2) the logarithm of the GRB energy, $\log E$, if the GRB is isotropic; (3) the total energy radiated by the GRB, E_{tot}, taking into account the collimation of the gamma-ray impulse; (4) the distance from the GRB source to the center of the molecular cloud, R_0; (5) the radius of the quasi-homogeneous central region of the molecular cloud, r_0; (6) a parameter determining the collimation opening angle, ϑ_0; (7) the characteristic size of inhomogeneities, r_1; (8) the peak bolometric luminosity, $L_{\max}$; (9) the time to reach the peak luminosity after the arrival of the gamma-ray impulse, $t_{\max}$; (10) the duration of the phase during which the luminosity exceeds half the peak luminosity, $t_{1/2}$; (11) the duration of the phase during which the luminosity is lower than one-tenth the peak luminosity, $t_{1/10}$; and (12) the duration of the phase during which the luminosity exceeds one-hundredth of the peak luminosity, $t_{1/100}$. Everywhere, we assumed $E_{\max} \geqslant 1/6$ MeV, so that the heating rate depends only on the energy of the GRB pulse and not on $E_{\max}$.

4.1. Characteristics of the Cases Computed

Case 1 corresponds to an isotropic GRB that occurs in a homogeneous cloud. The density in the

Table 1. Main parameters of the cases computed

Case number	$\log E_0$ (erg)	E_{tot}, 10^{49} erg	R_0, pc	r_0, pc	ϑ_0, rad	r_1, pc	L_{max}, 10^{41} erg/s	t_{max}, days	$t_{1/2}$, days	$t_{1/10}$, days	$t_{1/100}$, days
1	52	10^3	0	1.5	∞	−	43	401	1081	2670	3397
2	52	2.5	1	0.5	0.1	−	5.1	11.7	10.2	16.8	43.8
3	52	2.5	1	0.2	0.1	−	2.16	8.5	9.5	18.8	40.1
4	51	0.25	1	0.5	0.1	−	0.0414	7.5	9.1	25.7	96.4
5	53	25	1	0.5	0.1	−	36.5	14.1	17.9	31.5	128
6	53	6.25	1	0.5	0.05	−	12	8.5	12.2	29.2	115
7	53	10^2	1	0.5	0.2	−	65	27.3	33.9	76.0	245
8	53	25	1	0.5	0.1	0.03	38.7	14.1	17.9	34.8	128
9	52	2.5	2	0.2	0.1	−	0.224	11.	12.6	21.7	47
10	52	2.5	0.5	0.2	0.1	−	2.74	8.2	10.7	21.4	55.9
11	52	2.5	0.5	0.2	0.1	0.02	2.88	6.9	10.7	21.4	55.9
12	52	2.5	0.5	0.2	0.1	0.02	3.02	7.8	12.6	22.0	55.0
13	52	2.5	0.5	0.2	0.1	−	0.0425	19.6	62.8	162	437

cloud is $n = 10^5$ particles/cm^3, and the GRB energy is $E = 10^{52}$ erg.

In cases 2−12, the GRB is anisotropic and occurs at some distance from the center of the molecular cloud. The angular distribution of the energy in the GRB impulse is specified as $E = E_o e^{-(\vartheta/\vartheta_0)^2}$ (E_o is the isotropic energy of the GRB; Table 1), and the density distribution as $n = 10^5 e^{-(r/r_0)^2}$ cm^{-3}, where r is the distance from the center of the molecular cloud. The GRB axis is directed from the cloud center.

In cases 8, 11, and 12, various density inhomogeneities are added.

Case 8 adds four regions of enhanced density with the distribution $n = 3 \times 10^5 e^{-\left(\frac{|\bar{r}-\bar{r}_i|}{r_1}\right)^2}$ cm^{-3}, where $\bar{r}_i$ is the radius vector to the center of each region. Three of these are spherical and located along the line of symmetry of the GRB−cloud-center at distances of 0.8, 1.0, and 1.2 pc from the GRB source. The fourth region of enhanced density is located 1.0 pc from the GRB and 0.2 pc from the symmetry axis. In our two-dimensional formulation, this last density enhancement has the form of a torus.

Case 11 adds one spherical region of enhanced density with its center at the center of the molecular cloud and with the density distribution $n = 3 \times 10^5 e^{-(r/r_1)^2}$ cm^{-3}, where r is the distance from the cloud center.

Case 12 adds five spherical regions of enhanced density with the density distribution

$n = 3 \times 10^5 e^{-\left(\frac{|\bar{r}-\bar{r}_i|}{r_1}\right)^2}$ cm^{-3}, where $\bar{r}_i$ is the radius vector of the center of each region. The dense regions are located along the GRB−cloud axis at distances of 0.4, 0.45, 0.5, 0.55, and 0.6 pc from the GRB.

Case 13 is precisely like case 10, but with the molecular-cloud density an order of magnitude lower.

Let us now consider these individual cases in more detail.

4.2. An Isotropic GRB in a Homogeneous Molecular Cloud

This case is topologically fully analogous to the case considered in [11]. At the initial time, the passing GRB impulse heats the central part of the molecular cloud to temperatures of the order of 10^9 K; heating to higher temperatures is prevented by inverse Compton scattering. Since the material in the central region is heated to high temperatures ($T > 10^7$−10^8 K), it rapidly becomes fully ionized, and the dominant radiative losses are associated with free−free transitions in the plasma. The rate at which energy is lost per cubic centimeter of the optically thin plasma due to free−free radiation is proportional to $T^{1/2}$; at such temperatures, the gas can be considered ideal, and its heat capacity does not depend on the temperature. We find that, the stronger the heating of the ISM plasma, the longer its cooling time. Therefore, cooling of the gas just after the GRB can be neglected, and the temperature profile will repeat the profile of the

energy from the GRB impulse absorbed by the material; i.e., the temperature goes as $T \sim R^{-2}$ (Fig. 3, curve *1*).

As soon as the GRB impulse has traveled to a distance where it heats the matter only to temperatures of the order of several million Kelvin or less, the cooling rate per volume grows sharply due to the increase in the efficiency of atomic-line radiation [formula (12)], and the cooling time becomes less than the time for the GRB impulse to travel to this radius: $t_{cool} < R/c$ (Fig. 3, curves *2–5*). This leads to the following picture. At the initial time, the temperature profile formed by the GRB pulse obeys the dependence $T \sim R^{-2}$. After the impulse has reached some radius (which depends on the GRB energy and the number density in the molecular cloud), a cutoff appears in the temperature profile (the temperature of the matter falls to $\sim 10^4$ K, so that the matter becomes virtually completely neutral, and the cooling rate falls sharply). A heating wave (due to the GRB) and two cooling waves (one in the direction of increasing and one in the direction of decreasing distance from the GRB source) travel radially. As a result, at large distances from the GRB source, the gamma-ray impulse has become so weak that it is not able to heat the matter to 10^4 K, the interstellar material remains neutral, and the cooling wave catches up with the heating front. Energetic gamma-rays can ionize them matter, but, in this case, our approach is not applicable—we are not considering such a developed stage in this analysis.

We should note an important detail here: in the case of a homogeneous density, the phase velocity of the cooling wave traveling outward is always higher than the speed of light. The phase velocity of the wave traveling toward the center rapidly falls, and this wave moves through the cloud with a velocity of $\sim 10^4$ km/s, much lower than the speed of light. The duration of the afterglow is determined by the geometrical dimensions of the radiating region. Hence, we find the characteristic time for the peak of the "optical" radiation, $\Delta t \approx 2R/c \approx 9$ yrs (Fig. 4, Table 1).

4.3. An Anisotropic GRB at Some Distance from the Center of an Inhomogeneous Molecular Cloud

In the most probable scenario, the GRB source is anisotropic and is not located precisely at the center of an inhomogeneous molecular cloud. Let us now consider various cases for this general situation (cases *2–11*).

Immediately after the GRB, the temperature distribution in the cloud close to the GRB axis is fully determined by the angular distribution of the energy
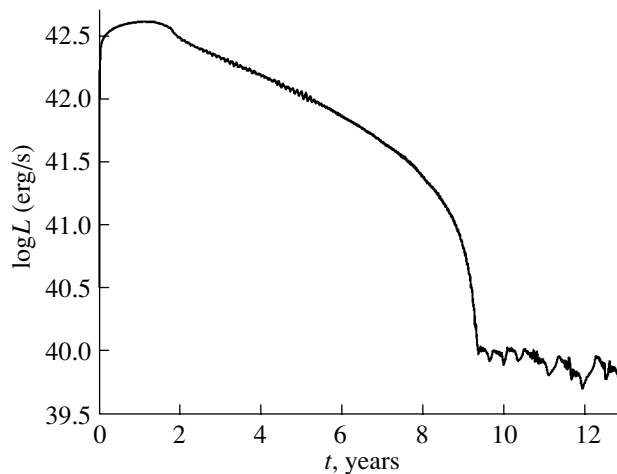

Fig. 4. Time dependence of the optical/ultraviolet luminosity. The time is measured in years.

in the GRB and the distance from the source. At the edges of the GRB impulse (far from its axis), the temperature does not reach high values exceeding several million Kelvin, and the matter rapidly cools to temperatures of the order of 10^4 K (Fig. 5a). As the gamma-ray front moves into denser regions of the cloud, the temperature to which it heats the interstellar material falls and, as a consequence of the increased density, the cooling rate grows, leading to a thinning of the region of hot gas (Fig. 5b). At this time, the region of heated gas divides into two parts. One region, which follows behind the gamma-ray impulse, decreases its geometrical thickness and takes on the shape of a thin meniscus. The second region has the form of a cylinder with a variable cross section; this region slowly cools, decreasing its geometrical size as it does so (Fig. 5c). After the gamma-ray impulse passes through the center of the molecular cloud, it continues into regions of decreasing density. The decreasing density causes the cooling rate to drop, leading to an increase in the geometrical thickness of the region of heated gas following behind the gamma-ray impulse (Fig. 5d).

The dominant flux of radiation from a molecular cloud heated by a nearby GRB is emitted by regions directly adjacent to the gamma-ray front. As we showed above, the region that gives rise to the bulk of the radiation (the region with high density) is geometrically thin, and there is only a modest increase in the duration of the impulse of optical radiation as a consequence of the curvature of the radiating region, due to the narrow cone of the GRB impulse. All this means that, for an observer located on the GRB axis, the duration of the optical impulse at half its peak intensity is only $t_{1/2} = 10$ days, with the peak intensity being reached 8.5 days after the GRB (Fig. 6, curve A; Table 1, case 3). Even in the case
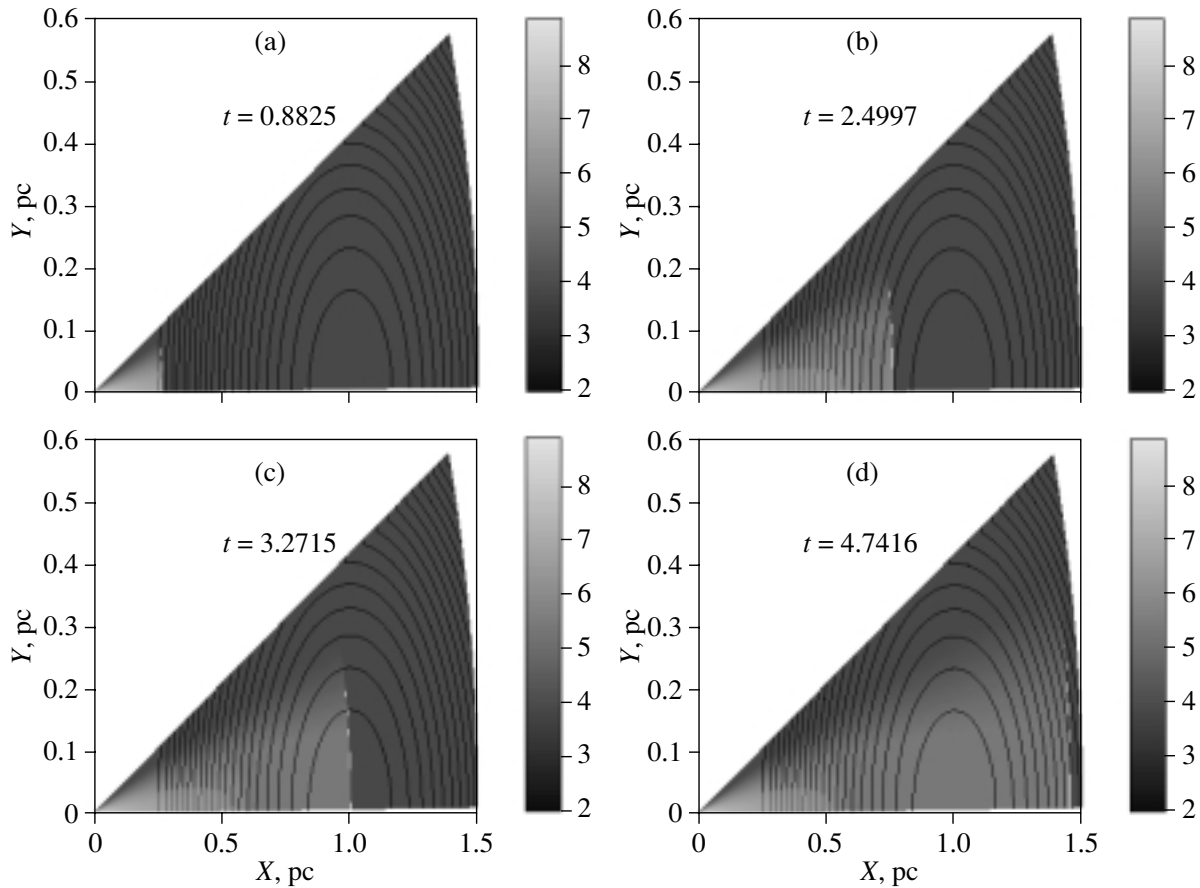
Fig. 5. Distribution of the temperature (gray scale) and contours of equal density for a cloud at various times for case 3. The distance from the GRB source is indicated in parsecs, and the time is measured in years after the GRB. Columns show the correspondence between the gray scale and $\log T$.

of comparable peak fluxes in the optical afterglow, the duration of the peak is two orders of magnitude shorter when the GRB is offset from the center of the molecular cloud than when the GRB is in the center of a spherically symmetrical cloud with a homogeneous density. The total energy of an isotropic GRB exceeds that of an anisotropic GRB by a factor of 40.

Let us consider the main factors influencing the light curve of the optical afterglow.

1. Let us first compare cases 2 and 3 in Table 1, which differ in the size of the central region of the molecular cloud: its mass is nearly a factor of 16 higher in case 2 than in case 3. The maximum luminosity of the afterglow in case 2 exceeds that in case 3 by a factor of 2.5. This is due to the collimation of the GRB and the fact that the GRB impulse heats only a cylindrical part of the molecular cloud. Accordingly, the volume of heated gas differs due to the height of the cylinder, which cannot be larger than the diameter of the cloud. Regions of the molecular cloud that are far from the GRB−cloud axis do not contribute to the light curve.

2. Now compare cases 3, 9, and 10 in Table 1. The main characteristics of the GRB and molecular cloud coincide in these cases, and only the distance of the center of the molecular cloud from the GRB source differs.

The peak luminosities in cases 3 and 10 differ by only 25%, and the temporal parameters differ only slightly. This is due to the fact that the center of the molecular cloud is located half as far from the GRB in case 10 as in case 3. Accordingly, the matter is heated to temperatures that are a factor of four higher. However, due to the collimation of the GRB, not all the cloud participates in this heating; only a cylindrical volume lying near the GRB axis is heated, with the radius of the base being half as large in case 10 as in case 3. The influences of these two effects nearly precisely cancel, giving rise to similar light curves. The peak luminosity in case 3 is an order of magnitude higher than the peak luminosity in case 9; the duration of the light curve is about 25% longer in case 9, and the net result is that the integrated radiated energy is a factor of eight lower in case 9. This is due primarily to the large distance of the GRB
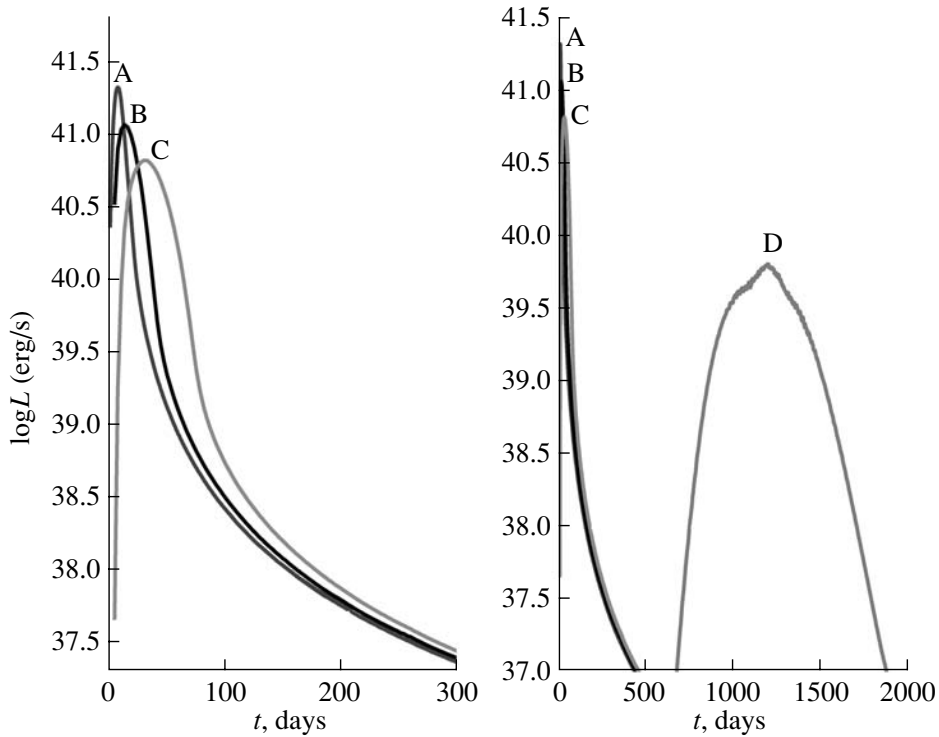
Fig. 6. Light curves for case 3 for an observer located on the GRB–cloud axis (curve A), at an angle of $\alpha = 0.1$ rad (curve B), at an angle of $\alpha = 0.2$ rad (curve C), and at an angle of $\alpha = \pi/2$ rad (curve D). The time is measured in days after the GRB.

from the center of the molecular cloud, which means that dense regions of the molecular cloud are heated and reradiate much more weakly.

3. Let us now turn to cases 2, 4, and 5. All the main characteristics of the molecular cloud coincide in these cases, but the energy of the GRBs differ.

The GRB energy is an order of magnitude higher in case 5 than in case 2, and the peak luminosities of the afterglows differ by a factor of seven. Taking into account the fact that the afterglow is more prolonged in case 5, the total energy radiated is a factor of 12 higher in case 5 than in case 2. The longer duration in case 5 comes about because the minimum cooling time until the peak luminosity is reached is longer in case 5 than in case 2. In addition, the curvature of the radiating layer also extends the light-curve plateau near the maximum luminosity.

The GRB energy is an order of magnitude lower in case 4 than in case 2; the peak luminosity is a factor of 120 higher and the total energy radiated a factor of 100 higher in case 2. This difference of two orders of magnitude in the total radiated energy is due to the fact that the material in the molecular cloud is heated to lower temperatures in case 4 than in case 2; since matter radiates efficiently only to 10^4 K, the radiated fraction of the thermal energy is lower in case 4 than in case 2. In addition, a smaller part of the cloud is heated to temperatures above 10^4 K in case 4.

4. Let us now consider cases 5, 6, and 7, which have the same peak luminosities for the gamma-ray impulse but different collimation angles. We can see from Table 1 that the ratio of the maximum luminosities for cases 5 and 6 is three, while this ratio for cases 7 and 5 is two, although the ratio of the GRB energies for each pair is equal to four. This is due to the increase in the size of the radiating layer with increasing collimation angle, which extends the time when the heated material radiates. The duration of the luminosity at the half-peak intensity level can be estimated as

$$t_{1/2} \simeq (1 - \cos \vartheta_0) R_0/c + t_{cool}, \qquad (25)$$

where t_{cool} is the cooling time for the molecular-cloud material (in our case, $t_{cool} \sim 7$ days).

5. As in cases 10, 11, and 12, the difference between cases 5 and 8 is that cases 8, 11, and 12 include inhomogeneity in the molecular-cloud density.

The added dense regions in the molecular cloud lead to a modest growth in the luminosity of the afterglow and an insignificant shortening of the time to reach the light-curve maximum.

6. Cases 10 and 13 have different densities and, accordingly, masses for the molecular cloud. The density is an order of magnitude lower in case 13 than in case 10, and the maximum luminosity is lower by a factor of 64. This is due to the fact that the mass of

Table 2. Main parameters of case 3 from Table 1 for various viewing angles

Case number	$\log E_0$ (erg)	$E_{\rm tot}$, 10^{49} erg	R_0, pc	r_0, pc	ϑ_0, rad	α	$L_{\max}$, 10^{41} erg/s	$t_{\max}$, days	$t_{1/2}$, days	$t_{1/10}$, days	$t_{1/100}$, days
3A	52	2.5	1	0.2	0.1	0	2.16	8.5	9.5	18.8	40.1
3B	52	2.5	1	0.2	0.1	0.1	1.2	14.1	18.6	31	59
3C	52	2.5	1	0.2	0.1	0.2	0.677	31	36	59	87
3D	52	2.5	1	0.2	0.1	$\pi/2$	0.066	1200	376	724	1000

heated gas is an order of magnitude lower in case 13; in addition, the radiative time scale for a volume element of gas is inversely proportional to its density (the cooling time in case 13 is $t_{\rm cool} \sim 70$ days), which leads to an additional drop of nearly a factor of ten in the maximum luminosity. The nonstrict adherence to a quadratic dependence is also associated with the kinematic extension of the width of the phase of maximum luminosity.

If the line of sight of the observer is not directed along the line passing through the GRB source and the center of the molecular cloud, the shape of the light curve will change (Fig. 6, left and right plots; Table 2). Figure 6 presents four light curves for the cases when the observer is located (A) on the GRB–cloud-center axis and at angles to this axis of (B) $\alpha = 0.1$ rad, (C) $\alpha = 0.2$ rad, and (D) $\alpha = \pi/2$ rad. The peak luminosity for case A is reached 8.5 days after the arrival of the GRB; the duration of the impulse at the half-peak level is $t_{1/2} = 9.5$ days, and the luminosity reaches 2.16×10^{41} erg/s. Note that the peak luminosity falls by approximately a factor of two from case A to case B and from case B to case C, but the duration of the peak grows by approximately a factor of two from case A to case B and from case B to case C. Thus, the total radiated energy is the same in all cases. In case D, the peak intensity is reached only 3.5 yrs after the GRB; the duration of the afterglow at the half-peak level is $t_{1/2} = 376$ days, and the flux itself is a factor of 32 lower than when the observer is located along the GRB–cloud-center axis.

5. DISCUSSION AND CONCLUSION

The interaction of a cosmological GRB with the dense material of the host galaxy leads to a number of interesting effects, some of which can be observed and verified in practice. The computations show a strong dependence of the light curve on the parameters of the GRB and molecular cloud, and on the geometry and the viewing angle in anisotropic cases. The shape and duration of the light curve vary over wide intervals. The peak luminosity can be reached as soon as 7 days after the GRB (case 11, the GRB is located at some

distance from the center of a molecular cloud with small-scale density enhancements) or as long as one to three years after the GRB (case 1, the GRB is at the center of a homogeneous molecular cloud). The bolometric luminosity can reach values of 6.5×10^{42} erg/s (case 7).

Our analysis has revealed a number of important patterns in the behavior of the afterglow light curves. In the case of strongly collimated gamma-ray impulses, we find a linear growth of the maximum afterglow luminosity with increase in the linear size and a quadratic growth with increase in the density of the cloud. The maximum luminosity initially depends only weakly on the distance of the GRB source from the center of the molecular cloud, but a sharp decrease in the peak luminosity occurs when the GRB is not able to heat an appreciable fraction of the molecular cloud to temperatures above 10^4 K. The dependence of the maximum afterglow luminosity on the opening angle for the gamma-ray impulse is initially quadratic, then reaches a plateau when $\theta_0 > 0.1$ rad. The maximum afterglow luminosity for a given opening angle depends linearly on the GRB energy, with a cutoff at low energies. The GRB energy at which this cutoff occurs depends approximately quadratically on the distance R between the GRB and the center of the molecular cloud, $E_{cut,52} \sim 2.2R^2$ pc ($E_{52} \equiv E/10^{52}$ erg).

When the observer does not detect the gamma-ray impulse itself (the GRB is anisotropic and the viewing angle does not coincide with the GRB axis; case D of case 3), it is still possible to observe an optical transient (a so-called "orphan"). However, this is complicated by a number of factors that decrease the probability of observing orphan optical transients of this type.

(1) For this to happen, the GRB peak luminosity must be appreciably weaker (by up to a factor of $\sim$32) than when the observer is located along the GRB axis.

(2) The maximum luminosity is reached after about four years, and the duration of the afterglow is about one to three years. Since the observer does not detect the gamma-ray impulse, there is no

natural starting time, although the long duration at the maximum luminosity might help facilitate the detection of orphan optical transients.

(3) When a GRB occurs in a molecular cloud, the gamma-ray impulse does not vaporize the dust, and the cloud remains optically thick in the direction of the observer, making it virtually impossible to observe such an optical afterglow at cosmological distances. In this case, the appearance of orphan infrared transients associated with reradiation by dust is possible.

(4) In view of the fact that dust does not absorb X-ray radiation [29], observations of the X-ray afterglows of GRBs without detection of the associated gamma-ray impulses may be possible if the GRB source is located in a dense part of the cloud.

Recent observations of GRB 030329 [17, 33] display a plateau in the R light curve from 64 to 94 days after the GRB (for technical reasons, it has not been possible to continue the observations). The existence of this plateau can be explained by the model considered here. These observations do not contradict the temporal or energetic constraints derived in the computations.

ACKNOWLEDGMENTS

REFERENCES

1. C. Akerlof *et al.*, Nature **398**, 400 (1999).
2. R. Sari, T. Piran, and J. P. Halpern, Astrophys. J. Lett. **519**, L17 (1999).
3. P. T. O'Brien *et al.*, astro-ph/0312602.
4. K. Z. Stanek *et al.*, Astrophys. J. Lett. **591**, L17 (2003).
5. J. Hjorth *et al.*, astro-ph/0306347.
6. B. Paczynski, Astrophys. J. **494**, 45 (1998).
7. S. R. Kulkarni *et al.*, Nature **393**, 35 (1998).
8. A. N. Timokhin and G. S. Bisnovatyi-Kogan, Astrophys. Space Sci. **231**, 323 (1995).
9. A. N. Timokhin and G. S. Bisnovatyi-Kogan, Astrophys. Space Sci. **235**, 59 (1996).
10. G. S. Bisnovatyi-Kogan and M. V. A. Murzina, Astron. Astrophys. **307**, 686 (1996).
11. G. S. Bisnovatyi-Kogan and A. N. Timokhin, Astron. Zh. **74**, 483 (1997) [Astron. Rep. **41**, 423 (1997)].
12. R. Perna, J. Raymond, and A. Loeb, Astrophys. J. **533**, 658 (2000).
13. J. S. Bloom, Astrophys. J. **518**, L1 (1999).
14. Y. M. Lipkin *et al.*, Astrophys. J. **606**, 381 (2004).
15. M. V. Barkov and G. S. Bisnovatyi-Kogan, Preprint No. 2091, IKI (Institute for Space Research, Russian Academy of Sciences, Moscow, 2003).
16. M. V. Barkov and G. S. Bisnovatyi-Kogan, Astron. Zh. **82**, 29 (2005) [Astron. Rep. **49**, 24 (2005)].
17. A. S. Pozanenko, private communication (2003).
18. A. M. Beloborodov and A. F. Illarionov, Astrophys. J. **450**, 64 (1995).
19. G. S. Bisnovatyi-Kogan and S. I. Blinnikov, Mon. Not. R. Astron. Soc. **191**, 711 (1980).
20. L. L. Cowie, J. P. Ostriker, and A. A. Stark, Astrophys. J. **226**, 1041 (1978).
21. A. B. Kirienko, Pis'ma Astron. Zh. **19**, 27 (1993) [Astron. Lett. **19**, 11 (1993)].
22. J. Raymond, D. Cox, and B. Smith, Astrophys. J. **204**, 290 (1976).
23. E. Anders and N. Grevesse, Geochim. Cosmochim. Acta **53**, 197 (1989).
24. D. I. Kosenko *et al.*, Pis'ma Astron. Zh. **29**, 243 (2003) [Astron. Lett. **29**, 205 (2003)].
25. N. A. Krall and A. W. Trivelpiece, *Principles of Plasma Physics* (Academic, New York, 1973; Mir, Moscow, 1975).
26. S. A. Kaplan and V. N. Tsytovich, *Plasma Astrophysics* (Nauka, Moscow, 1972; Pergamon, Oxford, 1974).
27. A. A. Galeev and R. Z. Sagdeev, *Problems in Plasma Theory* (Atomizdat, Moscow, 1973), Vol. 7 [in Russian].
28. A. B. Mikhailovskii, *Theory of Plasma Instabilities* (Atomizdat, Moscow, 1975; Consultants Bureau, New York, 1974), Vol. 1.
29. N. G. Bochkarev, *Fundamentals of the Physics of the Interstellar Medium* (Mosk. Gos. Univ., Moscow, 1992) [in Russian].
30. D. V. Sivukhin, *Problems in Plasma Theory* (Atomizdat, Moscow, 1964), Vol. 4 [in Russian].
31. B. T. Draine and Lei Hao, Astrophys. J. **569**, 780 (2002).
32. V. V. Sobolev, *Course in Theoretical Astrophysics* (Nauka, Moscow, 1985) [in Russian].
33. M. A. Ibrahimov, I. M. Asfandiyarov, B. B. Kahharov, *et al.*, GRB Coordinates Network, No. 2219, 1 (2003).

Translated by D. Gabuzda

Masers in the Cool Molecular Cloud L 379

M. I. Pashchenko[1] and E. E. Lekht[2]

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

[2]*Instituto Nacional de Astrofísica, Óptica y Electrónica, Apartado Postal 51 y 216,*
Tonantzintla, Puebla, México

Received August 25, 2004; in final form, February 17, 2005

Abstract—We present the results of monitoring the H_2O masers in the IR sources IRAS 18265−1517 and IRAS 18277−1516 associated with the cool molecular cloud L 379, which contains high-velocity bipolar molecular jets. The sources were observed in the 1.35 cm H_2O line using the 22-m radio telescope of the Pushchino Radio Astronomy Observatory (Russia) during 1991−2004. We detected H_2O maser emission from IRAS 18265−1517 at radial velocities of 17.8 and 18.4 km/s, virtually coincident with the velocity of the molecular cloud derived from CO-line observations (18.4 km/s). The maser emission towards the other source, IRAS 18277−1516, was at higher velocities than the central velocity of the CO molecular cloud. The H_2O maser spots are most likely associated with a redshifted region of CO emission. Cyclic variability of the integrated H_2O maser emission that may be related to cyclic activity of the central star was detected for IRAS 18277−1516. The strongest and most long-lived component ($V_{LSR} \approx 20.6$ km/s) displays a radial-velocity drift, which could be due to deceleration of a dense clump of matter (maser condensation) in the circumstellar medium during the descending branch of a strong flare. We found numerous emission features for both IRAS 18265−1517 and IRAS 18277−1516, providing evidence for fragmentation of the medium surrounding their central objects. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

During the early stages of their formation, stars are embedded in dense, cool gas and dust clouds, where the protostar is invisible in the optical due to the very strong extinction. Its optical light is absorbed in the protostellar cloud and reradiated in the far IR. Stars in early evolutionary stages undergo episodes of very strong ejection, accompanied by the formation of strong high-velocity bipolar flows of cool molecular gas. At this stage, CH_3OH, H_2O, and later OH masers associated with compact far-IR sources appear in the surrounding cloud, in the vicinity of the star.

The high-velocity bipolar flows originate in sources associated with cool IRAS objects located in dense disks. The IRAS sources emit strongly at 60 and 100 μm. In the radio, they either do not emit at all or are weak emitters at millimeter wavelengths.

L 379 is an isolated dark molecular cloud that is not detected at visible wavelengths. Seventeen IRAS objects were revealed in the cloud, within $\pm 21'$ in right ascension and $\pm 10'$ in declination from the point with coordinates $\alpha_{1950} = 18^h 27^m 30^s$, $\delta_{1950} = -15°19'00''$ [1]. Only three of these have fluxes exceeding 300 Jy at 100 μm. Maser emission was found towards two sources: IRS 3 and IRS 2 (in later papers, IRS 3 is designated IRS 1). These

sources also bear the names IRAS 18265−1517 and IRAS 18277−1516, which we use below. The distance to the cloud L 379 is estimated to be 2 kpc [2].

The region of IRAS 18265−1517 was mapped by Hilton *et al.* [1] in CO line emission with $80''$ angular resolution. They detected a bipolar flow to the northwest and southeast, within a region $5' \times 7'$ in size, with the IR source spatially coincident with the maximum of the redshifted emission. Hilton *et al.* [1] observed a velocity gradient along the bipolar flow, corresponding to a velocity change by 10 km/s over $10'$. The velocity of the CO cloud is about 18.5 km/s. Hilton *et al.* [1] present a map of the central source and its immediate vicinity, together with a plot of contours of equal velocity. In this velocity range, OH maser emission was found towards IRAS 18265−1517 [3, 4], and Kalenskii *et al.* [5] detected CH_3OH (methanol) maser emission.

The bipolar flow's central source is IRAS 18265−1517. Hilton *et al.* [1] found a "hot spot" $\sim 2'$ in diameter, with its maximum approximately coincident with IRAS 18265−1517. The spot is probably a dense disk around the central source. The methanol masers are situated near the central source of the bipolar flow, and are probably projected against the dense disk [5].

IRAS 18265−1517 was mapped with a higher angular resolution ($28''$) in the $^{12}CO\,(J = 2 \to 1)$ line
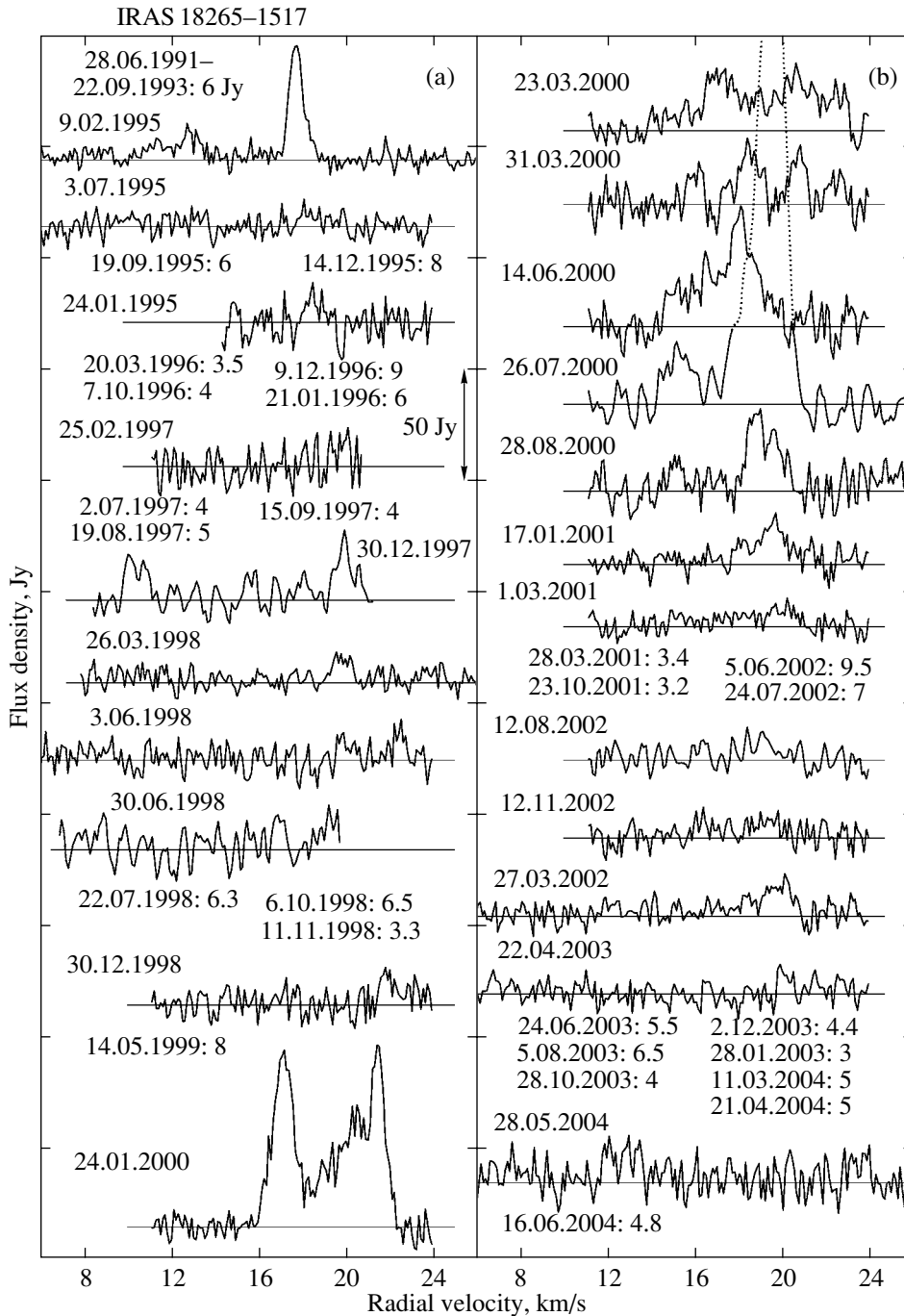
IRAS 18265−1517

Fig. 1. H$_2$O maser spectra for the IR source IRAS 18265−1517 obtained in 1991−2004. The vertical arrow indicates the scale in Jy. The velocity relative to the Local Standard of Rest is plotted along the horizontal axis. For spectra with possible emission below the noise level, we give the dates of the observations together with upper limits for the possible fluxes in Jy, separated by a colon.

by Wilking *et al.* [6], who found that the maxima of the redshifted and blueshifted emission coincided and were centered on the IR source. Later, IRAS 18265−1517 was studied by Kelly and MacDonald [7] in the $J = 2 \rightarrow 1$ transitions of the ^{12}CO, C^{18}O, and ^{13}CO lines with $22''-23''$ resolution. These observations demonstrated that two matter flows were present

in this region, associated with the northern and southern components of IRS 1 (IRAS 18265−1517) and with different alignments. The dust temperatures in the northern and southern clumps are 22 and 24 K, respectively, and the gas masses (M_{H_2}) are 1000 and 680 $M_\odot$, respectively. Each clump contains a B0−5 star with a mass of about 15 $M_\odot$. It has been
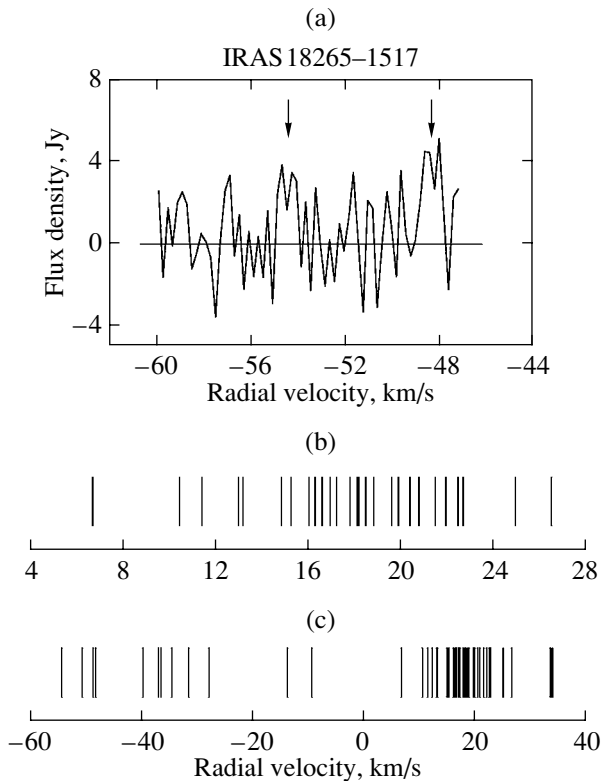
Fig. 2. (a) H_2O spectrum acquired on January 28, 2004 in the direction of IRAS 18265−1517 at negative velocities. (b) Positions on the V_{LSR} scale of the spectral components detected for the H_2O maser in IRAS 18265−1517 at $V_{LSR} > 0$. (c) All the components found in the vicinity of the source (see text).

suggested that a binary star system is forming in IRAS 18265−1517 [7].

IRAS 18277−1516 is also located in L 379 and is associated with the IR source IRS 2. The field was mapped in the CO $(J = 1 \rightarrow 0)$ line by Schwartz *et al.* [8]. The positions of the maxima of the red-shifted and blueshifted emission differ in declination by $\approx 50''$. A molecular flow with a central velocity of 18.5 km/s is present, but it is not bipolar. The velocity range for the CO emission is 17 km/s—a factor of three narrower than for IRAS 18265−1517. The kinematic value of 2 kpc is adopted for the distance to IRAS 18277−1516 [8].

The water masers in the cool molecular cloud L 379 are associated with strong matter outflows and have wider ranges of emission velocities than the molecular-gas outflows themselves. To explain the observed phenomena and devise models for sources of this type, regular observations covering a wide velocity range obtained over many years are needed. This paper presents the results of our many-year monitoring of the H_2O masers in IRAS 18265−1517 and IRAS 18277−1516 in the molecular cloud L 379.

2. OBSERVATIONS AND DATA

Our observations of the H_2O maser emission towards IRAS 18265−1517 ($\alpha_{1950} = 18^h26^m33^s$, $\delta_{1950} = -15°17'51''$) and IRAS 18277−1516 ($\alpha_{1950} = 18^h27^m43^s$, $\delta_{1950} = -15°16'14''$) were acquired with the 22-m radio telescope of the Pushchino Radio Astronomy Observatory, from June 1991 and December 1996, respectively, until April 2004. The mean intervals between our observations were about two and three months for the two sources.

The system included a cooled transistor amplifier yielding system temperatures of 150−230 K. The radiometer was upgraded in 2000, lowering the system noise temperature to 100−130 K. The signal was analyzed using a 96-channel filter spectrum analyzer (128-channel after June 1997) with a resolution of 7.5 kHz, corresponding to 0.101 km/s in radial velocity in the 1.35 cm line. For an unpolarized point source, an antenna temperature of 1 K corresponds to a flux density of 25 Jy.

A catalog of the IRAS 18265−1517 spectra is presented in Fig. 1. The vertical arrow shows the scale in Jy. The horizontal axis plots the velocity relative to the Local Standard of Rest (LSR), and the scale is the same for all the panels. Spectra for which possible emission is below the noise level are not shown. The dates of such observations and upper limits for the emission are given.

We also observed IRAS 18265−1517 from time to time at radial velocities near −50 km/s. The spectrum observed on January 28, 2004 is shown in the upper part of Fig. 2. To improve the sensitivity (at the cost of halving the spectral resolution), we averaged pairs of adjacent channels. The arrows mark weak emission at −54.4 and −48.3 km/s with fluxes of 2.6 and 3.5 Jy.

Despite the maser's low activity, we were able to identify 27 individual spectral components with $V_{LSR} > 0$ that were manifest in some way between 1991 and 2004. Their positions in radial velocity are shown in Fig. 2b. The lower panel shows all the components found in the region of IRAS 18265−1517, at any radial velocity, in the present study and in the studies of Xiang and Turner [9] and Furuya *et al.* [10].

Figure 3 presents a catalog of H_2O spectra for IRAS 18277−1516, which we observed mainly in the velocity range 14−27 km/s. Because the flux variations are considerable, the figures have different vertical scales. A superposition of all the spectra shows that there are three velocity intervals in which H_2O maser emission is observed: 17.6−21.5, 21.5−23.3, and 23.3−27.2 km/s. For convenience, we will call these the left, central, and right groups of emission features.

We calculated integrated fluxes for the total spectra and the three individual sections, whose variability
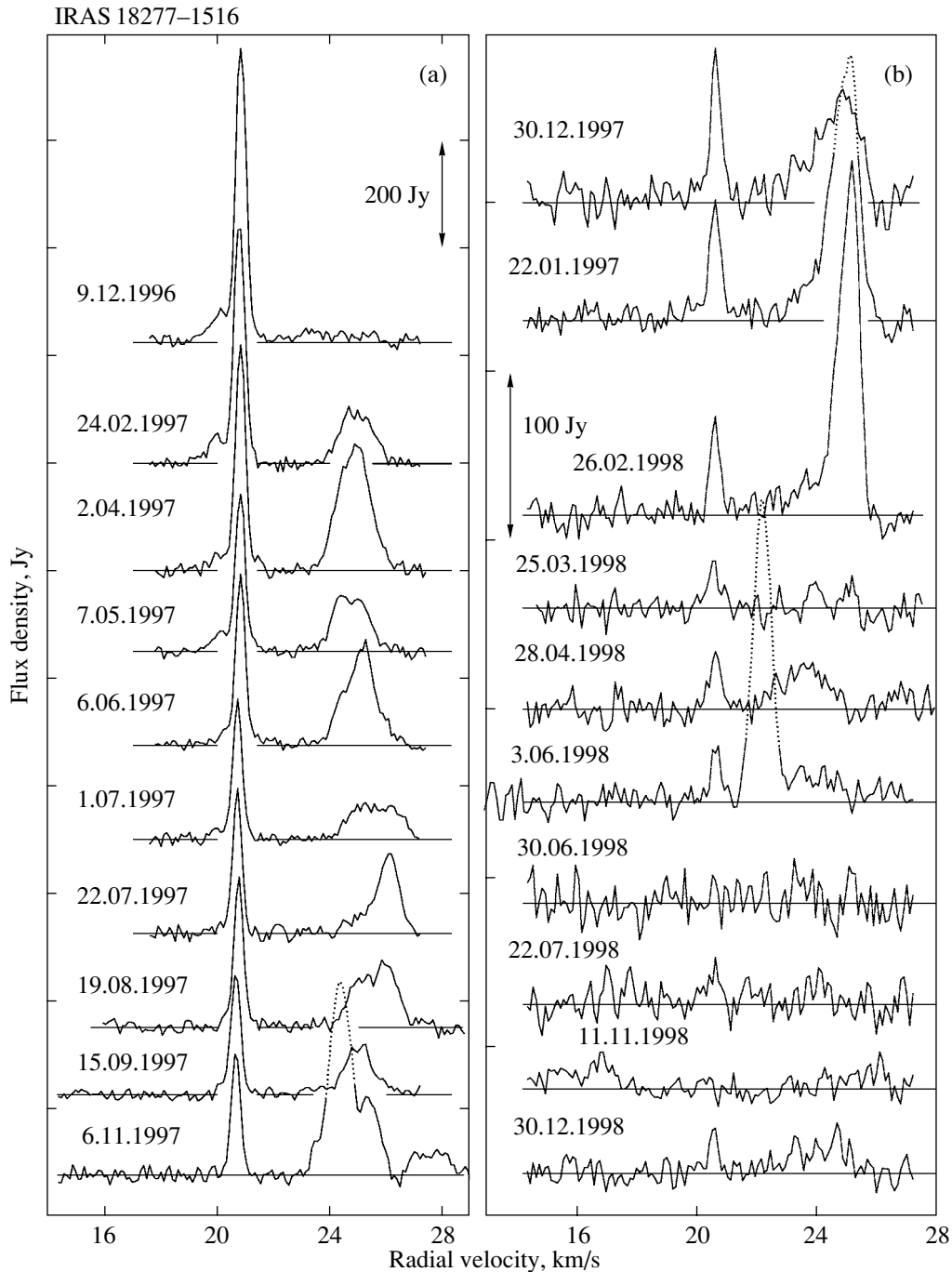
IRAS 18277−1516

Fig. 3. H_2O maser spectra in the IR source IRAS 18277−1516 obtained in 1996−2003. The scale in Jy is shown by the vertical arrows. The velocity relative to the Local Standard of Rest is plotted along the horizontal axis.

curves are shown in Fig. 4. The dash−dotted curve in the upper panel is a polynomial describing the slow integrated variations for the total spectrum, i.e., for the whole radial-velocity range. Three deep minima reaching the zero level are visible. The insert in the upper panel includes data obtained prior to our monitoring. The data point for 1990 was taken from Felli *et al.* [11], and we computed the next two points

using the spectra of Xiang and Turner [9]. We indicate the radial velocities of the components giving the largest contribution to the corresponding emission near some of the peaks in Fig. 3b.

Based on the variability of the integrated flux, we subdivided the whole period covered by our observations into four time intervals, which we take to represent four activity cycles. The boundaries between in-
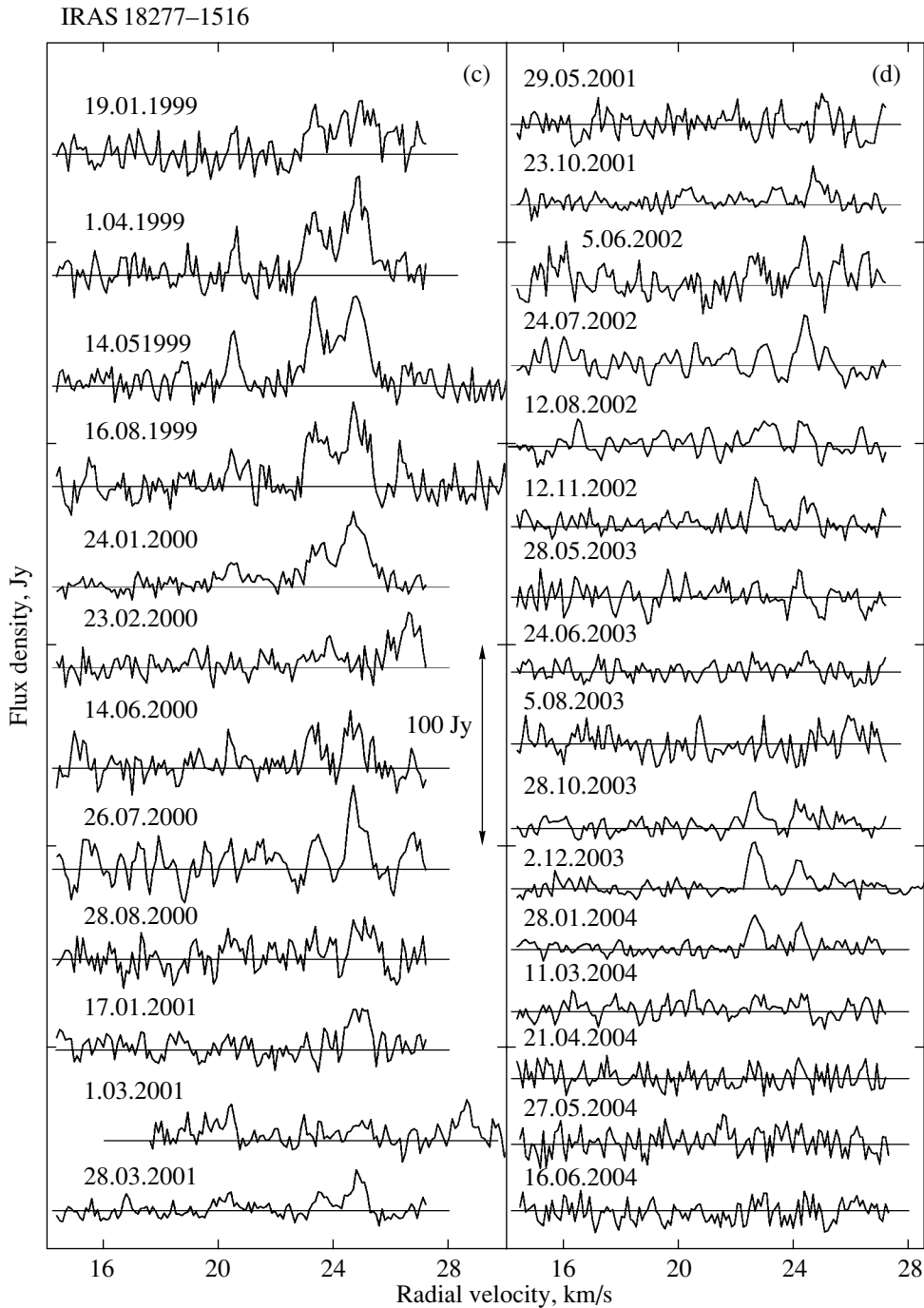
IRAS 18277–1516

Fig. 3. (Contd.)

tervals are shown as vertical arrows. The first cycle is incomplete; our monitoring started on the descending branch of the H_2O maser activity. We calculated the mean spectra (Fig. 5) for each of the activity cycles, except for the last one.

We then identified individual emission features using all the spectra from our monitoring (Fig. 3) as well as the mean spectra (Fig. 5). Analysis of the mean spectra enabled us to identify several emission features that were not evident in the individual spectra. This provides evidence for the presence of faint, long-lived components. The total number of identified spectral features is 22. Their positions in the spectrum are displayed in the bottom part of Fig. 5. The solid bars identify the two components that demonstrated a radial-velocity drift. For this reason, we give their mean radial velocities. The dotted line indicates the radial velocity of the CO molecular cloud.
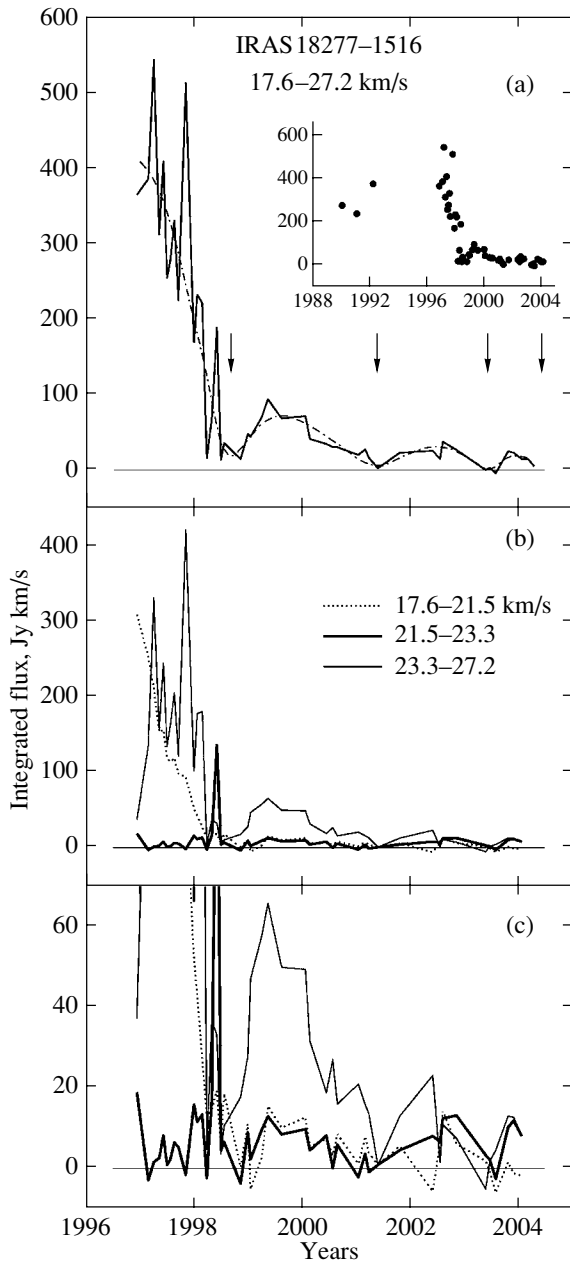
Fig. 4. Variations of the integrated H_2O maser emission from IRAS 18277−1516 for (a) the combined spectra and, on different scales, (b), (c) three groups of emission features. The dash-dotted curve displays a polynomial describing slow variations of the integrated flux. The vertical arrows mark the positions of minima. The insert in the upper panel presents data for observations prior to our monitoring (see text).

A radial-velocity drift was found for the strongest emission feature in IRAS 18277−1516. Figure 6 shows variability curves for its flux and V_{LSR}. Because of the wide range covered by the flux variations, an insert for 1999−2001 is presented in the upper panel. The flux variations can be approximated with
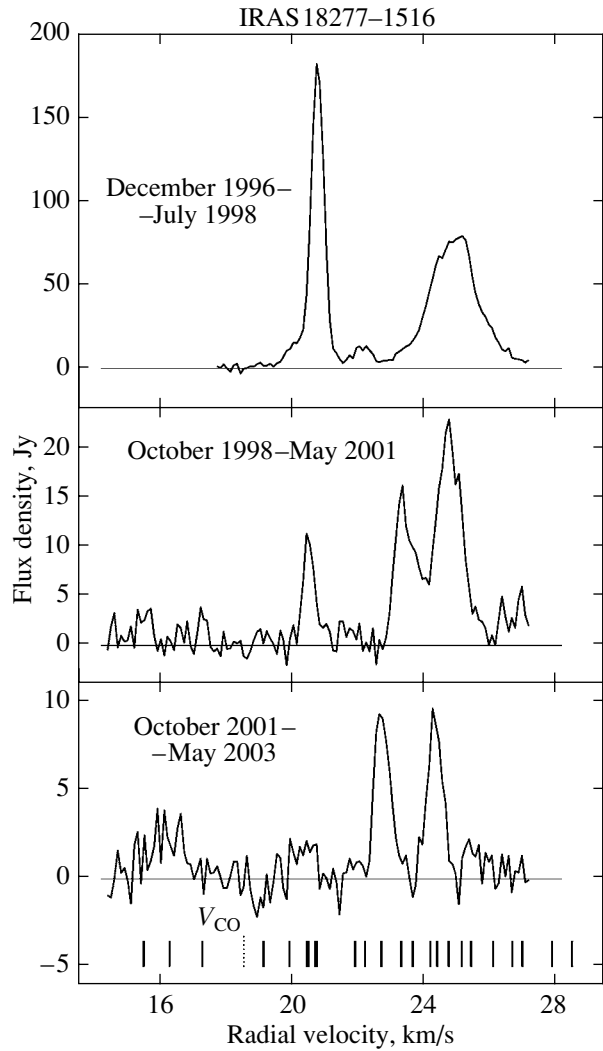

Fig. 5. Mean H_2O spectra of IRAS 18277−1516 for various time intervals. The vertical segments in the bottom panel mark the positions of individual spectral components, and the dotted line shows the position of the CO molecular cloud.

two straight line segments. Their intersection point, corresponding to the observations of March 25, 1998, is plotted as a large circle. The radial-velocity variations were also approximated with a second-order polynomial (the dashed curve). In addition, two individual intervals, also separated by March 25, 1998, were fitted by two second-order polynomials (dotted curves), which, together with the straight segments in Fig. 6a, provide the best description of the emission variability. The variability of the emission features in the right group is displayed in Fig. 7. Radial velocities are indicated for most emission peaks.
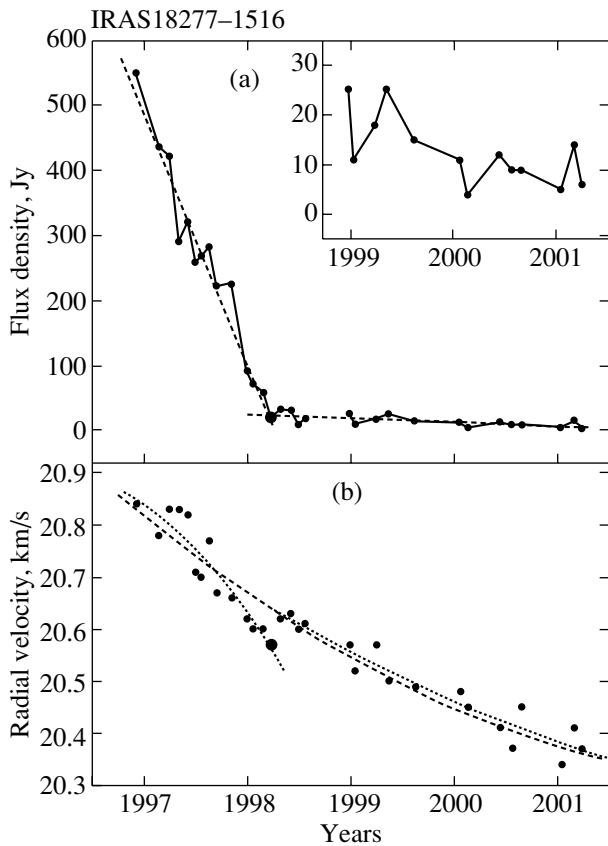
IRAS18277−1516

(a)

(b)

Fig. 6. Variability of the flux and radial velocity for the main emission feature in IRAS 18277−1516. The dashed curves are smooth approximations.

3. IRAS 18265−1517

3.1. The Region Surrounding IRAS 18265−1517

IRAS 1265−1517 is associated with the cool molecular cloud L 379, whose temperature is about 40 K. The source is a strong emitter in the far IR. The flux density is 445 Jy at 60 μm and rises to 1297 Jy at 100 μm. The radio continuum emission is present only at short wavelengths, to $\lambda_{\max} = 2.7$ mm. The 2.7 mm flux is 471 mJy [12].

The OH maser emission was discovered by Pashchenko and Le Squeren [3, 4] during a survey of IRAS sources in June 1991 using the Large Radio Telescope at Nancay (France). Emission in the 1665 and 1667 MHz lines was observed in right-circular polarization (RCP), at $V_{LSR} = 15.5$ and 16 km/s, respectively. There was no strong maser emission in left-circular polarization (LCP). This made it possible to observe hydroxyl absorption at the same velocities (15−16 km/s). The absorption was not obvious in the RCP spectrum due to interference from the strong maser emission. Unpolarized emission at 18−20 km/s was observed in the satellite OH 1612 MHz line [6]. Thus, the velocities where
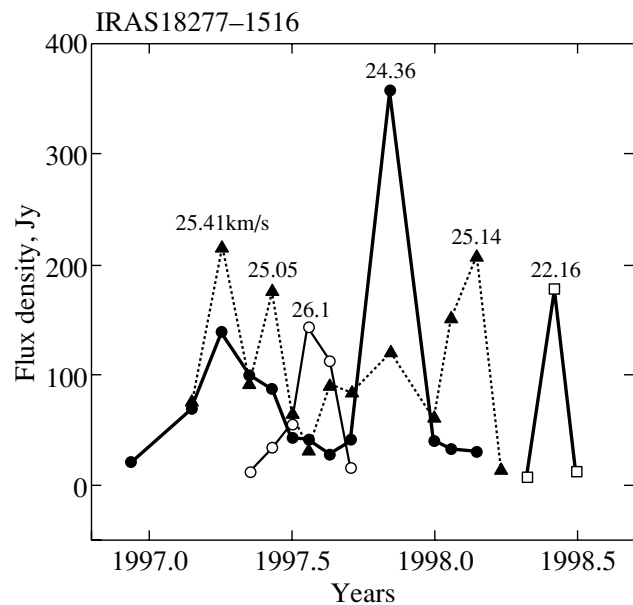
IRAS18277−1516

Fig. 7. Variability of the emission features in IRAS 18277−1516 at radial velocities of 22−27 km/s.

the OH maser emission was observed are close to the velocity of the molecular cloud, suggesting that the OH masers are embedded in the molecular cloud [6].

Attempts to detect H_2O maser emission towards L 379 IRS 1 (IRAS 18265−1517) were undertaken in 1990−1992. Felli et al. [11] observed this source in February 1990 at a wide range of radial velocities, from −118 to 218 km/s. They found no emission with fluxes exceeding 3 Jy. H_2O emission in a direction close to the source (offset by $34''$ in declination) was detected by Xiang and Turner [9], whose observations of February 25, 1990, demonstrated the presence of three spectral components, at −50.7, −36.6, and −28 km/s with fluxes of 9.6, 3.3, and 3.1 Jy, respectively. These velocities differ strongly from the velocities of the molecular cloud in both CO and OH.

We began our monitoring in June 1991 and found H_2O emission at a velocity close to that of the cloud on February 9, 1995. The spectrum consisted of two very close components, at $V_{LSR} = 17.8$ and 18.4 km/s, with fluxes of 50 and 10 Jy. Weaker emission was observed at 11.5−14 km/s. In 2003, Furuya et al. [10] published a catalog of water-maser spectra associated with young massive stellar objects. The catalog contains spectra of the IRAS 18265−1517 maser for December 31, 1997 and June 2, 1998. They observed emission in a wide velocity range, from −54 to 34 km/s. Furuya et al. [10] adopted a distance of 2 kpc and assumed that the emission they had

detected belonged to L 379−IRS 1 (IRAS 18265−1517). However, according to Wilking *et al.* [6], the bipolar flow in the CO cloud has negative velocities only at −9 km/s.

3.2. Analysis of the H_2O Monitoring Data

The many-year observations of the H_2O maser show that its individual components are very strongly variable and have low intensities. The exception is emission observed in February 1995 at 17.8 km/s, close to the velocity of the molecular cloud, with a flux of about 50 Jy. We observed strong emission only in the first half of 2000, when the flux reached 270 Jy at $V_{LSR} = 19.7$ km/s. In January of the same year, we observed strong emission at 16−22 km/s, with two dominant peaks with similar fluxes, near 80 Jy. The flux at other velocities did not exceed 20 Jy.

Despite the low general emission level at velocities of 6−27 km/s, we identified 27 individual emission features with $V_{LSR} > 0$ that were manifest in 1991−2003 (Fig. 2b). Most of the components were short-lived. The intervals between our observations (one to two months) were not short enough to enable us to accurately determine the component lifetimes, but we estimate these lifetimes to be one to six months. The lower value here may be somewhat too long. With all this in mind, it is likely that the real number of spectral components that were present in 1991−2003 is somewhat larger. This provides evidence for strong fragmentation of the medium in which the H_2O maser emission originates.

3.3. Model of the H_2O Maser

Examination of all available observations in the neighborhood of IRAS 18265−1517 indicates that H_2O maser emission has been observed at velocities from −55 to 34 km/s (Fig. 2c). The highest concentration of emission features (28 components) was observed at 6−34 km/s, near the radial velocity of the molecular cloud. The OH maser emission is in the same velocity range [4]. Like the hydroxyl and methanol masers, these water-maser emission features ("spots") are associated with the central double source in IRAS 18265−1517. In particular, the masers could be associated with one of the binary system's components. The emission at −9 km/s [10] is at the edge of the CO line [6], and it is possible that it is associated with IRAS 18265−1517. The velocity scatter of ≈40 km/s (from −9 to 34 km/s) could be due to the influence of the high-velocity molecular flow on the corresponding maser spots.

The remaining emission features, with $V_{LSR} < -9$ km/s, are difficult to relate to the bipolar flow.

However, the velocity scatter—for example, in HH1, where a high-velocity matter outflow is present—can reach near 100 km/s. It is quite possible that the high-velocity features in IRAS 18265−1517 are related to some other component. The question of the identification of the emission in IRAS 18265−1517 at $V_{LSR} < -9$ km/s can be addressed by high-angular-resolution observations or monitoring over the entire radial-velocity range, which would make it possible to detect (or refute) a correlation between the emission in the two velocity ranges, $V_{LSR} > -9$ km/s and $V_{LSR} < -9$ km/s.

4. IRAS 18277−1516

4.1. The Region of IRAS 18277−1516

The source IRAS 18277−1516 emits strongly in the far IR. No radio continuum emission has been detected. The water-maser emission from this source was first discovered by Felli *et al.* [11]. On February 16, 1990, the spectrum consisted of two emission peaks at 20.6 and 25.6 km/s, with fluxes of about 400 and 25 Jy, respectively. Later (January 25, 1991), Xiang and Turner [9] also observed a double structure in the spectrum, but the second component's radial velocity was 24 km/s and its flux was 90 Jy. By April 1992, the flux in the main peak reached almost 800 Jy, while the second peak had disappeared ($F < 20$ Jy). In December 1997, in addition to two components at 20.2 and 24.5 km/s, Furuya *et al.* [10] observed emission at a negative velocity, −23.2 km/s. Thus, the range of the H_2O maser emission in IRAS 18277−1516 became considerably wider.

4.2. Analysis of the H_2O Monitoring

Our analysis of the monitoring results reveals considerable differences between the maser emission of IRAS 18277−1516 and IRAS 18265−1517, despite the fact that the two sources are located in the same molecular cloud, L 379:

(1) all the emission features, except for three weak ones, have velocities exceeding the central velocity of the CO molecular cloud;

(2) it is clear that there are two radial-velocity ranges where the maser emission is predominantly found (17.6−21.5 and 23.3−27.2 km/s);

(3) the integrated flux displays deep minima that divide our monitoring interval into cycles of maser activity of unequal duration;

(4) the lifetimes of some features are as long as several years;

(5) the strongest and apparently most long-lived component at $V_{LSR} \approx 20.6$ km/s displayed a drift in radial velocity;

(6) a series of flares of individual components was observed between early 1997 and mid-1998 at velocities of 22–26.3 km/s.

The maser-activity cycles derived from the variability of the integrated flux differ in both the intensity of their maser emission and their duration. The higher the integrated flux, the longer the activity cycle. In addition, large differences in the structure of the mean spectra derived for each activity cycle were found. There was no emission from the central group during the first cycle. An exception was late 1996, when weak emission ($F \approx 15$ Jy) was observed at 23.3 km/s. In addition, a strong flare with a flux of 175 Jy occurred at 22.2 km/s towards the end of the first cycle.

4.3. Individual Components

Our monitoring of IRAS 18277−1516 began during the descending branch of the evolution of the strongest emission feature at $V_{LSR} \approx 20.8$ km/s. The decrease in the the flux from 550 to 90 Jy was almost linear (Fig. 6a). There was no emission at all in the second half of 1998. Then, beginning in early 1999, emission with a flux of about 25 Jy again appeared at 20.6 km/s, which then disappeared by mid-2001. The flux variations were accompanied by a radial-velocity drift of this component from 20.85 to 20.35 km/s (Fig. 6b). During the entire period when this emission was observed (from late 1996 to mid-2001), a correlation between the flux and radial-velocity variations was present. Since the flux was decreasing, it is natural to suppose that the maser spot was decelerating in the circumstellar medium after the maximum of a strong flare.

The variability curves for the flux and V_{LSR} do not provide unambiguous information about the number of components of the flare (one or two). Our analysis suggests that there were probably two components present, with the second initially not observed because its emission was essentially blocked by the first component prior to February–March 1998.

The emission maxima for the components in the right group are delayed in time relative to the main feature at 20.6 km/s. However, the epoch of the main feature's maximum is not known, and we can only estimate lower limits for the delays. Thus, a series of consecutive flares of the maser emission of IRAS 18277−1516 was observed, which displayed a global character. This series of flares probably reflected enhanced activity of the central star.

4.4. The Model

According to Schwartz *et al.* [8], the velocity of the molecular cloud in the CO line towards IRAS 18277−1516 is 18.5 km/s, the same as towards IRAS 18265−1517 (18.4 km/s [6]). The velocities of all the H_2O and CO emission features are higher than the mean velocity of the L 379 cloud. The only exceptions are three short-lived features with fluxes below 30 Jy (Fig. 6). It is natural to suppose that the maser spots are spatially related to the redshifted CO emission region. The large number of emission features provides evidence for fragmentation of the medium surrounding the central star. Some fragments (clumps of matter) are stable, and their lifetimes in the active emission stage could reach several years. The stability of the component with $V_{LSR} \approx 20.8$ km/s was displayed during its interaction with the surrounding medium, via a radial-velocity drift that was observed over a long time.

The cyclic variability of the integrated flux and the sequence of strong flares could be associated with cyclic activity of the central star. The duration of a cycle depends strongly on its intensity.

As for IRAS 18265−1517, the emission at −23 km/s [10] is outside the velocity limits for the CO molecular flow. To elucidate the nature of this emission, it is desirable to carry out further monitoring of IRAS 18277−1516 over a wider velocity range than was previously covered.

5. MAIN RESULTS

Let us summarize the most important results of this study.

1. We have presented a catalog of 1.35-cm H_2O maser spectra for the region of the IR sources IRAS 18265−1517 (during 1991−2004) and IRAS 18277−1516 (during 1996−2004).

2. The emission features in IRAS 18265−1517 were mainly short-lived, being observed from one to six months.

3. We detected 27 individual spectral components for IRAS 18265−1517 and 22 components for IRAS 18277−1516. This suggests the presence of a strongly fragmented medium in the vicinity of the central objects.

4. The water-maser emission in the vicinity of IRAS 18265−1517 was observed at radial velocities from −55 to 34 km/s. The main group of features is concentrated at 9−34 km/s, near the molecular cloud's velocity. It is natural to suppose that the maser spots responsible for this emission, and probably for the emission at −9 km/s as well, are related to the central sources of the binary system in IRAS 18265−1517. The velocity scattered of

≈40 km/s displayed by the various components is probably due to the action of the bipolar outflow on the corresponding maser spots in the source.

5. The variability of the integrated H_2O flux in IRAS 18277−1516 is cyclic. The cycle durations are one to six years, and depend strongly on the intensity of the emission.

6. The maser spots responsible for the emission in IRAS 18277−1516 are most likely associated with the redshifted region of the CO emission.

7. We observed a radial-velocity drift for the strongest and most long-lived component ($V_{LSR} \approx$ 20.6 km/s). The highest drift rate, 0.2 km/s per year, occurred in 1997. This drift may be associated with deceleration of a clump of material in the circumstellar medium.

ACKNOWLEDGMENTS

REFERENCES

1. J. Hilton, G. J. White, N. J. Cronin, and R. Rainey, Astron. Astrophys. **154**, 274 (1986).
2. J. Hilton and J. F. Lahulla, Astron. J. **106**, 672 (1993).
3. M. I. Pashchenko and A. M. Le Squeren, Astron. Tsirk, No. 1550, 17 (1991).
4. M. I. Pashchenko and A. M. Le Squeren, Pis'ma Astron. Zh. **20**, 85 (1994) [Astron. Lett. **20**, 69 (1994)].
5. S. V. Kalenskii, R. Bachiller, I. I. Berulis, *et al.*, Astron. Zh. **69**, 1002 (1992) [Sov. Astron. **36**, 517 (1992)].
6. B. A. Wilking, J. H. Blackwell, and L. G. Mundy, Astron. J. **100**, 758 (1990).
7. M. L. Kelly and G. H. Macdonald, Mon. Not. R. Astron. Soc. **282**, 401 (1996).
8. P. R. Schwartz, G. Gee, and Y. L. Huang, Astrophys. J. **327**, 350 (1988).
9. D. Xiang and B. E. Turner, Astrophys. J., Suppl. Ser. **99**, 121 (1995).
10. R. S. Furuya, Y. Kitamura, A. Wootten, M. J. Claussen, and R. Kawabe, Astrophys. J., Suppl. Ser. **144**, 71 (2003).
11. M. Felli, F. Palagi, and G. Tofani, Astron. Astrophys. **255**, 293 (1992).
12. B. A. Wilking, L. G. Mundy, J. H. Blackwell, and J. E. Howe, Astrophys. J. **345**, 257 (1989).

Translated by N. Samus'

Meridional Circulation in Young Massive Stars with Laminar Rotation

E. I. Staritsin

Astronomical Observatory, Ural State University, pr. Lenina 51, Ekaterinburg, 620083 Russia
Received August 1, 2004; in final form, February 17, 2005

Abstract—We study the rotation of a chemically homogeneous star with a mass of 16 $M_\odot$, assuming that the angular-momentum distribution in its radiative envelope is determined by hydrodynamical processes—flows and turbulent diffusion. Meridional circulation and horizontal shear turbulence are the main hydrodynamical processes forming the radial distribution of the angular momentum in young massive stars in the absence of magnetic fields. The rotation of such stars is close to steady-state. The angular velocity of rotation of the convective core can be ~5−20% higher than the surface value. Under these conditions, the characteristic time for the radial transport of angular momentum by meridional flows and shear turbulence is comparable to the nuclear time scale. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

The current standard theory of stellar structure and evolution assumes spherical symmetry and the absence of rotation and magnetic fields. This theory can explain a number of the observed characteristics of stars [1, 2]. However, detailed studies of the chemical compositions of stellar atmospheres have revealed some new properties that cannot be explained without taking rotation into account. These include, first and foremost, the helium, nitrogen, and carbon abundances in the atmospheres of massive ($M \geq 10\, M_\odot$), rapidly rotating main-sequence stars [3−6]. The helium and nitrogen abundances are enhanced and the carbon abundance reduced compared with the solar values. These differences may be related to the fact that products of the CNO hydrogen-burning cycle can be partly carried out from the convective core to the radiative envelope of the star, right up to its surface.

The injection of the products of nuclear burning into the radiative envelopes of stars can be associated with both rotation and the presence of a magnetic field. Rotation is common to all stars; the equatorial rotational velocities of massive main-sequence stars are 100−300 km/s [7]. The direct detection of magnetic fields in OB main-sequence stars is very rare (one example is β Cep [8]). Rotation-related hydrodynamical processes leading to the movement of chemical elements, such as turbulent diffusion, can bring about the transport of some products of hydrogen burning from the convective core to the radiative envelope of the star. The intensity of this transport will be determined by the rotation of the star at the beginning of its nuclear evolution and the intensity of turbulent diffusion in the horizontal direction [9]. Here, we consider the role of hydrodynamical angular-momentum transport in the rotation of young massive stars.

Rotation stimulates a number of processes that are not present in a spherically symmetrical nonrotating star. If the angular velocity of the rotation depends only on the distance to the rotation axis, the surfaces of constant temperature and pressure coincide with the equipotential surfaces. The condition of local radiative equilibrium is not satisfied in a rotating star. Due to radiative energy transport, some areas of each equipotential surface should experience small temperature increases, and others small decreases. In areas where excess heat energy is accumulated, the matter will rise and be cooled by expansion; in the areas where there is cooling, the matter will sink and be heated by compression. Meridional circulation, i.e., the large-scale motion of matter in a meridional plane, is established in rotating stars [10, 11]. In the steady-state case, the meridional circulation maintains a constant temperature on an equipotential surface. On average, meridional circulation does not bring about any energy transport through the equipotential surface [12]. Thus, the condition of radiative equilibrium is on average satisfied in a rotating star.

In the steady-state case, the velocity of the meridional circulation can be specified for any law describing the star's rotation. In general, it has been assumed that the rotational angular velocity is constant or depends only on the distance from the rotation axis. In the radiative envelopes of stars rotating like rigid bodies, the circulation of matter occurs in two cells with opposite directions of the circulation separated by an equipotential surface [13, 14]. The sign of the radial component of the meridional circulation velocity changes in the transition through the equipotential

surface due to the Gratton–Opik term [13, 14], which is inversely proportional to the matter density. If the rotational angular velocity depends on the distance from the rotation axis, the shapes of cells with different circulation directions become more complex [15].

The motion of gas in a meridional plane should be accompanied by a radial transport of angular momentum. Therefore, the circulation of matter should alter the rotational angular velocity. In the steady-state case, the influence of the circulation on the rotation law is not taken into account. As a result, the velocity fields for the meridional circulation and the rotation are not self-consistent [16]. The specified rotation law and velocity field for the meridional circulation derived for this law do not satisfy the equation of motion in the steady-state case.

In the current study, we analyze the steady-state rotation of a chemically homogeneous star with a mass of $16\,M_\odot$ and with laminar rotation in its layers [17–20]. We obtained self-consistent velocity fields for the rotation and meridional circulation for this state. The possible character of the rotation of young massive stars is considered.

2. BASIC EQUATIONS

Stellar rotation is manifest in two ways compared to spherically symmetrical, nonrotating models. First, rotation alters the mechanical structure of the star. In a one-dimensional approximation, the impact of rotation can be taken into account if the mass inside a constant-pressure ("horizontal") surface is chosen as the independent variable [16, 21]. Second, the gas in a rotating star is in motion due to the rotation. In addition, the motion of matter in meridional planes maintains the local thermal equilibrium in a rotating star. When gas moves, certain instabilities—most notably, the shear instability—can develop, which do not have analogs in gas in rest. In a medium with a homogeneous density, the shear instability develops on the dynamical time scale.

Meridional circulation that transports angular momentum facilitates the establishment of a gas flow in which the rotational angular velocity varies with the latitude along the constant-pressure surfaces [17]:

$$\Omega_*(r,\theta) = \Omega(r) + \Omega_2(r)P_2(\theta),$$

where $\Omega_*(r,\theta)$ is the rotational angular velocity, r the average distance to a constant-pressure surface, θ the latitude in spherical coordinates, $\Omega(r)$ the average rotational angular velocity on the constant-pressure surface, $\Omega_2(r)$ the amplitude of the variations of the rotational angular velocity with latitude on the constant-pressure surface, and $P_2(\theta)$ the second-order Legendre polynomial. The shear flow is unstable in the direction of latitude. A latitude shift of an

arbitrary element along a constant-pressure surface may be countered only by viscosity of the gas. Since the microscopic viscosity is relatively small in stellar interiors [9], shear instability that leads to turbulent flow should develop on the dynamical time scale. Since the turbulent transport of angular momentum along a constant-pressure surface equilizes the rotational angular velocities on this surface, the rotation of a constant-pressure surface should be close to rigid-body. Small deviations from rigid-body rotation, $\Omega_2(r) \ll \Omega(r)$, will come about due to the transport of angular momentum through this surface, driven by meridional circulation [17]. These small deviations from rigid-body rotation maintain the turbulent flow along the constant-pressure surface.

In the presence of a sufficiently high angular-velocity gradient, the shear flow in the vertical direction (perpendicular to the constant-pressure surfaces) may also be turbulent. The turbulent-viscosity coefficient should be much smaller in the vertical than in the horizontal direction, since a vertical shift of an element is countered by the Archimedes buoyance force, which substantially exceeds the microscopic viscosity in stellar interiors [9, 17]. The turbulent gas flows in the radiative envelopes of massive stars are strongly anisotropic. Therefore, the rotational angular velocity may depend only on the average distance from the constant-pressure surface. This is called "laminar rotation" [17].

The total velocity of an arbitrary fluid element $\mathbf{V}$ in an averaged turbulent flow is a combination of the rotational velocity $\mathbf{V}_{rot}$ and the velocity of the meridional circulation $\mathbf{u}$; note that $|\mathbf{u}| \ll |\mathbf{V}_{rot}|$. Assuming, in accordance with the Boussinesque hypothesis, that the turbulent stress tensor is proportional to the velocity tensor for deformations of the averaged turbulent flow, we can write the ϕ component of the equation of motion in spherical coordinates:

$$\frac{\partial(\rho\varpi^2\Omega)}{\partial t} + \text{div}(\rho\varpi^2\Omega\mathbf{u}) = \text{div}(\rho\nu_v\varpi^2\text{grad}\Omega). \quad (1)$$

Equation (1) describes the transport of angular momentum in the vertical direction, driven by meridional circulation and turbulent diffusion. Here, ρ is the density, ϖ the distance to the rotation axis, t the time, and ν_v the turbulent-viscosity coefficient in the vertical direction.

Early studies of meridional circulation considered the steady-state case without viscosity [16], so that (1) takes the form

$$\mathbf{u}\text{grad}(\varpi^2\Omega) = 0. \quad (2)$$

Because the distribution of the rotational angular velocity was specified arbitrarily (for example, rigid-body rotation was assumed), Eq. (2) was not used to determine the velocity of the meridional circulation.

The derived distribution of this velocity and the specified distribution of the rotational angular velocity do not satisfy (2). In general, (2) can be satisfied only for dynamically unstable types of rotation [16]. Here, we derive the distributions of the rotational velocity and the meridional circulation velocity of a star from the solution of Eq. (1).

The meridional circulation velocity is determined from the energy equation

$$\rho T \frac{ds}{dt} = \rho\varepsilon + \mathrm{div}(\chi \mathrm{grad}T),$$

where T is the temperature, s the specific entropy, ε the rate of energy release per unit mass, and χ the temperature-conductivity coefficient. In the steady-state case, this equation takes the form

$$\rho T \mathbf{u}\mathrm{grad}s = \rho\varepsilon + \mathrm{div}(\chi \mathrm{grad}T).$$

In the case of laminar rotation, the vertical component of the meridional circulation velocity can be written [17]

$$U_r(r,\theta) = U(r)\mathrm{P}_2(\theta),$$

where $U(r)$ is the amplitude of the vertical component of the meridional circulation velocity. The density and temperature vary only slightly along constant-pressure surfaces [17]:

$$\rho_*(r,\theta) = \rho(r) + \rho_2(r)P_2(\theta),$$

$$T_*(r,\theta) = T(r) + T_2(r)P_2(\theta).$$

The vertical component of the meridional circulation velocity in a chemically homogeneous star is then [17]:

$$U(r) = \frac{P}{\rho g c_P T(\nabla_a - \nabla)}\frac{L}{M_*}E_\Omega. \qquad (3)$$

The quantity E_Ω is determined by the relation [17, 19]

$$E_\Omega = 2\left(1 - \frac{\Omega^2}{2\pi G\rho}\right)\frac{g_2}{g} \qquad (4)$$
$$- \frac{\rho_m}{\rho}\left\{\frac{r}{3}\frac{d}{dr}\left[H_T\frac{d}{dr}\left(\frac{\Theta}{\delta}\right)\right.\right.$$
$$\left. - \chi_T\Theta + \left(1 - \frac{1}{\delta}\right)\Theta\right]$$
$$- \frac{2H_T}{r}\left(1 + \frac{\nu_h}{K}\right)\frac{\Theta}{\delta} + \frac{2}{3}\Theta\right\}.$$

Equation (4) takes into account the fact that the energy release in massive stars is negligible beyond the convective core. In (4), G is the gravitational constant, g is the average gravitational acceleration on the constant-pressure surface, g_2 is the magnitude of latitude variations of the gravitational acceleration,

$\Theta = \rho_2/\rho$, H_T is the temperature scale height, ∇ is the temperature gradient, ∇_a is the adiabatic temperature gradient, c_P is the specific heat capacity at constant pressure, L is the energy flowing through the constant-pressure surface per unit time, ρ_m is the average density in the volume of the star beneath the constant-pressure surface, ν_h is the turbulent-viscosity coefficient in the horizontal direction, $\delta = (4 - 3\beta)/\beta$, β is the ratio of the gas pressure and the total pressure, $\chi_T = (\partial \ln \chi/\partial \ln T)_{P,\mu}$, and μ is the mean molar mass of the matter. M_* is given by the expression

$$M_* = M\left(1 - \frac{\Omega^2}{2\pi G\rho_m}\right),$$

where M is the mass of matter inside the constant-pressure surface. The following relations are valid for Θ and g_2/g [17]:

$$\Theta = \frac{1}{3}\frac{r^2}{g}\frac{d\Omega^2}{dr}, \qquad (5)$$

$$\frac{g_2}{g} = \frac{4}{3}\frac{\Omega^2 r^3}{GM}. \qquad (6)$$

Slight variations of density and temperature with latitude along the constant-pressure surface result in the appearance of additional (so-called baroclinic) terms in the expression for the vertical component of the meridional circulation velocity [in curly brackets in (4)], compared to the case of rigid-body rotation.

The turbulent-viscosity coefficient in the horizontal direction (along the constant-pressure surface) and deviations of the rotation of this surface from rigid-body rotation can be derived by analyzing the angular-momentum transport and energy dissipation in the turbulent flow. The deviations of the constant-pressure surface from rigid-body rotation, $\Omega_2(r)$, are specified by the balance between the angular-momentum transport by meridional circulation, which generates differential rotation in latitude, and the turbulent diffusion in the horizontal direction, which tends to re-establish rigid-body rotation [17]:

$$\nu_h\Omega_2(r) = \frac{1}{5}\Omega(r)rU(r) \qquad (7)$$
$$\times\left(\frac{1}{3}\frac{d\ln\left(\rho r^2 U\right)}{d\ln r} - \frac{1}{2}\frac{d\ln\left(r^2\Omega\right)}{d\ln r}\right).$$

Another relation for ν_h and $\Omega_2(r)$ can be obtained from the rate of dissipation of excess kinetic energy in the turbulent flow. On the one hand, this dissipation rate is determined by the relation [22]

$$\varepsilon_{turb} = \nu_h\left(\frac{\Delta v_\lambda}{\lambda}\right)^2,$$

where Δv_λ is the characteristic velocity variation on the scale λ. On the other hand,

$$\varepsilon_{turb} = \frac{\Delta E_\lambda}{\Delta t},$$

where ΔE_λ is the excess kinetic energy and Δt the characteristic time for dissipation of this excess. Δv_λ and ΔE_λ are specified by the value of $\Omega_2(r)$. Differential rotation in latitude inevitably gives rise to differential rotation in azimuth. Assuming the characteristic time for the re-establishment of uniform rotation in azimuth is of the order of one revolution of the star, we can write [20]

$$\nu_h = \left(\frac{1}{160\pi} \frac{r^2 \Omega_2}{\rho} \frac{d\left(\rho r^2 U \right)}{dr} \right)^3. \qquad (8)$$

Eliminating $\Omega_2(r)$ from (7) and (8), we obtain an expression for the horizontal turbulent-viscosity coefficient [20]:

$$\nu_h = \left(\frac{3}{400\pi} \right)^{\frac{1}{3}} r \left(r\Omega(r)U^2(r) \right)^{\frac{1}{3}} Y, \qquad (9)$$

$$Y^3 = \frac{1}{6} \frac{d\ln\left(\rho r^2 U \right)}{d\ln r}$$

$$\times \left[\frac{1}{3} \frac{d\ln\left(\rho r^2 U \right)}{d\ln r} - \frac{1}{2} \frac{d\ln\left(r^2 \Omega \right)}{d\ln r} \right].$$

In a similar way, we find for the magnitude of the differential rotation on the constant-pressure surface $\Omega_2(r)$, to order of magnitude,

$$\frac{\Omega_2}{\Omega} \approx \frac{80\pi}{3} \frac{\nu_h^2}{r^3 \Omega U}. \qquad (10)$$

The turbulent-viscosity coefficient in the vertical direction is determined by equating the work done by the gravitational force when two arbitrary fluid elements exchange places and the kinetic-energy excess in the flow. The destabilizing role of the exchange of energy between the element and surrounding matter due to the reradiation of photons and the horizontal transport of chemical elements due to the anisotropy of the turbulence is taken into account. The vertical turbulent-viscosity coefficient is [18]

$$\nu_V = 2\mathrm{Ri}_c \frac{K + \nu_h}{N_T^2} \left(\frac{dV_{rot}}{dz} \right)^2, \qquad (11)$$

where

$$K = \frac{4acT^3}{3c_P \rho^2 \kappa},$$

the square of the local Brunt−Väisälä frequency is given by
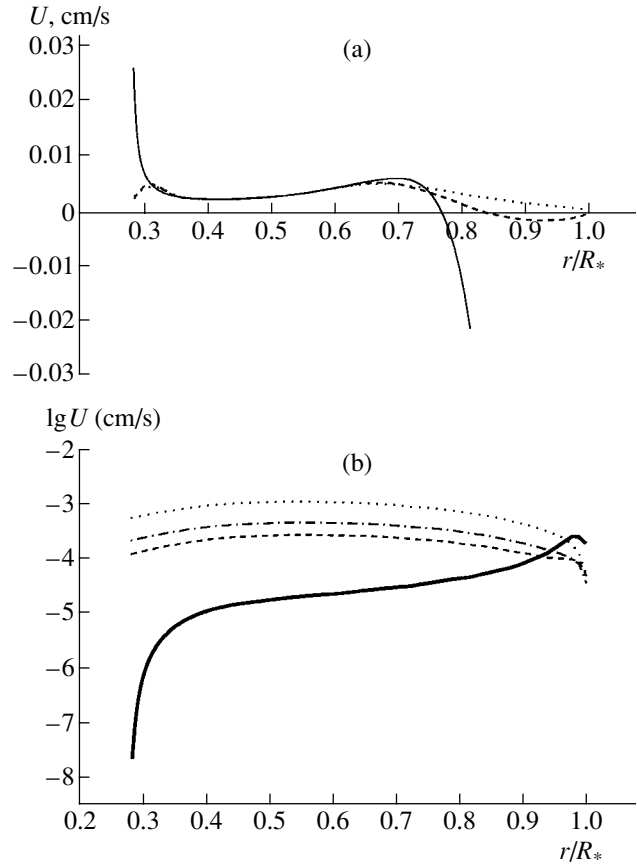
$$N_T^2 = \frac{g\delta}{H_p}(\nabla_a - \nabla),$$

Fig. 1. The meridional circulation velocity U in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r in the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$ for times (a) $t = 0$ yrs (solid curve), $t = 60$ yrs (dashed curve), $t = 100$ yrs (dotted curve), (b) $t = 10^5$ yrs (dotted curve), $t = 3 \times 10^5$ yrs (dash-dotted curve), and $t = 5 \times 10^5$ yrs (dashed curve). The meridional circulation velocity for steady-state rotation is plotted by the solid, thick curve.

where Ri_c is the critical Richardson number, the z axis is in the vertical direction, a is Stefan's constant, c is the speed of light, κ is the opacity of the matter, and H_p is the pressure scale height.

Substituting (3) into (1), we obtain a fourth-order equation in the angular velocity $\Omega(r,t)$. The horizontal and vertical turbulent-viscosity coefficients are specified by (9) and (11). The solution of this equation can be used to trace the time variations of the vertical distributions of four parameters: the angular velocity $\Omega(r,t)$, the vertical component of the meridional circulation velocity $U(r,t)$, and the horizontal and vertical turbulent-viscosity coefficients $\nu_h(r,t)$ and $\nu_V(r,t)$. It is also possible to obtain a steady-state solution.

In the steady-state solution, the advective and turbulent angular-momentum fluxes are exactly equal at
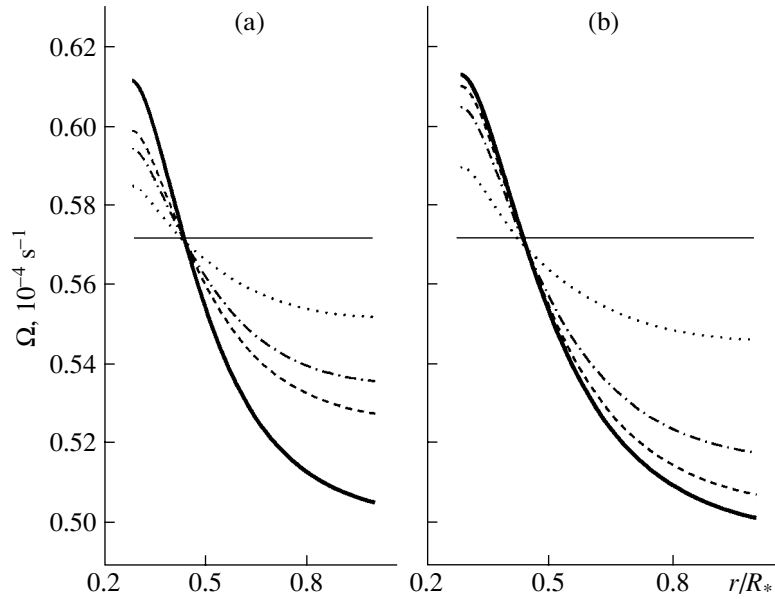
Fig. 2. Rotational angular velocity Ω in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r in the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$ for times $t = 0$ yrs (solid, thin curve), $t = 10^5$ yrs (dotted curve), $t = 3 \times 10^5$ yrs (dash-dotted curve), and $t = 5 \times 10^5$ yrs (dashed curve). The case of steady-state rotation is plotted by the solid, thick curve. The turbulent-viscosity coefficients in the horizontal direction are determined from (a) Eq. (9) and (b) Eq. (12).

each point in the radiative envelope. In this case, the rotational and meridional circulation velocity fields are consistent with each other. The deviation of the rotation of the constant-pressure surface from rigid-body rotation is specified by (10). An arbitrary vertical distribution for the rotational angular velocity of the star is adopted for the initial data. The advective and turbulent angular-momentum fluxes at the surface of the star are assumed to be zero.

It is assumed that the convective core and an adjacent thin radiative layer rotate as a rigid body, due to the penetration of convective elements beyond the boundary of the convective core (overshooting) [23]. The thickness of this layer is specified to be much smaller than the pressure scale height. Thus, the angular momentum of the convective core can vary with time in accordance with the advective and turbulent angular-momentum fluxes at the base of the radiative envelope. The advective and turbulent angular-momentum fluxes at the center of the star are assumed to be zero. All calculations were made for the critical Richardson number $Ri_c = 1/4$ [16].

3. STEADY-STATE ROTATION

We studied the evolution of the rotational-velocity profile in a chemically homogeneous star with mass $16\,M_\odot$. A linear variation of the rotational angular velocity with the current mass was specified as an initial $\Omega(r)$ distribution in the radiative envelope. The solution of (1) was obtained until

steady-state rotation had been established. We studied the cases $((\Omega_c/\Omega_s)_0, J) = \{(1, 1.83 \times 10^{52}), (2, 1.83 \times 10^{52}), (1, 2.75 \times 10^{52}), (1.3, 2.75 \times 10^{52})\}$, where $(\Omega_c/\Omega_s)_0$ is the initial ratio of the rotational angular velocities of the convective core and the stellar surface and J is the angular momentum of the star in cgs units.

In the case of initial rigid-body rotation, the circulation of matter in the meridional plane of the radiative envelope of a star with mass $16\,M_\odot$ initially occurs in two cells divided by an equipotential surface, in agreement with the classical results [13, 14]. The vertical component of the meridional circulation velocity reaches its maximum in layers immediately beneath the stellar surface and above the convective core (Fig. 1a). It is precisely in these layers that the angular-momentum fluxes are highest, and the angular-velocity distribution varies rapidly.

When the rotation begins to deviate from rigid-body rotation, the vertical component of the meridional circulation velocity is determined more and more by the baroclinic terms. The vertical component of the meridional circulation velocity decreases in both the layer above the convective core and the sub-surface layer. Very soon, baroclinic terms associated with the density and temperature variations along the constant-pressure surface begin to dominate in the relatively rarefied subsurface layers of the star. These terms first become comparable to, then exceed, the Gratton–Opik term, and determine both the speed
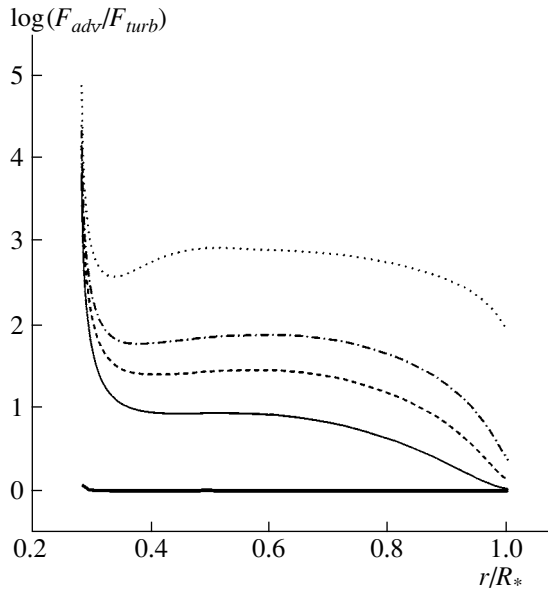
$\log(F_{adv}/F_{turb})$

Fig. 3. Ratio of the densities of the advective and turbulent angular-momentum fluxes in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r for the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$ for times $t = 10^5$ yrs (dotted curve), $t = 3 \times 10^5$ yrs (dash-dotted curve), $t = 5 \times 10^5$ yrs (dashed curve), and $t = 10^6$ yrs (solid thin curve). The flux-density ratio for steady-state rotation is plotted by the thick solid curve.

Ω_c/Ω_s

Fig. 4. Evolution of the ratio of the rotational angular velocities of the convective core and the surface for a star with mass $16\,M_\odot$ as a function of time t. The calculation results for $J = 1.83 \times 10^{52}$ are plotted by the solid curves (upper for $(\Omega_c/\Omega_s)_0 = 2$, lower for $(\Omega_c/\Omega_s)_0 = 1$) when the horizontal turbulent-viscosity coefficient ν_h is determined using (9), and by the dotted curve when ν_h is determined using (12). The calculation results for $J = 2.75 \times 10^{52}$ are plotted by the dashed curves (upper for $(\Omega_c/\Omega_s)_0 = 1.3$, lower for $(\Omega_c/\Omega_s)_0 = 1$) when ν_h is determined using (9), and by the dash-dotted curve when ν_h is determined using (12).

and the direction of the circulation in subsurface layers of the radiative envelope. The size of the outer cell, where matter ascends in the equatorial plane and descends in circumpolar regions, rapidly decreases, disappearing after ~ 100 yrs (Fig. 1a).

A single circulation direction is established in the entire radiative envelope, from the stellar surface to the boundary of the convective core: matter descends in the equatorial plane and ascends along the rotation axis. In the course of this motion, the angular momentum of the gas is transported from the outer to the inner parts of the star. The rotational angular velocity of the convective core and the adjacent part of the radiative envelope increases with time, while the rotational angular velocity of the outer layers of the radiative envelope decreases until a steady state has been reached (Fig. 2a). The direction of the circulation in the radiative envelope remains single and unchanged. The meridional circulation velocity decreases with time (Fig. 1b). In the envelope, the density of the advective angular-momentum flux appreciably exceeds the density of the turbulent flux, and only in the case of steady-state rotation are these flux densities equalized (Fig. 3). Thus, the meridional gas flows dominate in the formation of the star's rotational state.

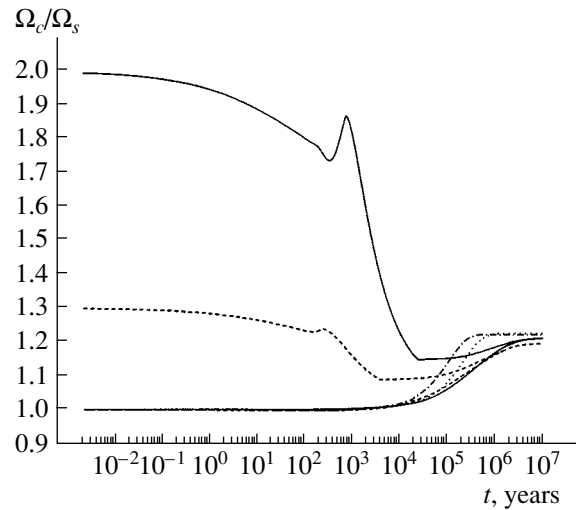When the rotation is initially rigid-body, the state of rotation of the star varies in the same way for the two considered angular-momentum values. The ratio of the rotational angular velocities of the convective core and the stellar surface increases as angular momentum is transported by the meridional circulation (Fig. 4). After $\sim 5-7$ Myrs, the angular-momentum fluxes for the turbulent and advective flows become essentially equal at all points of the radiative envelope. In the resulting steady state, the rotational angular velocity and the velocity of the meridional circulation are correlated. The rotational angular velocity of the convective core is $\sim 20\%$ higher than the surface value. The vertical component of the meridional circulation velocity increases from $\sim 10^{-6}$ cm/s near the convective core to $\sim 10^{-4}$ cm/s in the subsurface layers (Fig. 1b). The linear rotational velocity at the equator is ~ 180 km/s when $J = 1.83 \times 10^{52}$ g cm^2 s^{-1} and ~ 280 km/s when $J = 2.75 \times 10^{52}$ g cm^2 s^{-1}.

The typical vertical and horizontal turbulent-viscosity coefficients are 10^6-10^8 cm^2/s and 10^9-10^{10} cm^2/s (Fig. 5). The turbulence is anisotropic. For the specified angular momentum of the star, the variations of the gravitational acceleration on the constant-pressure surface do not exceed 10% (6) (Fig. 6). The density variations (5)
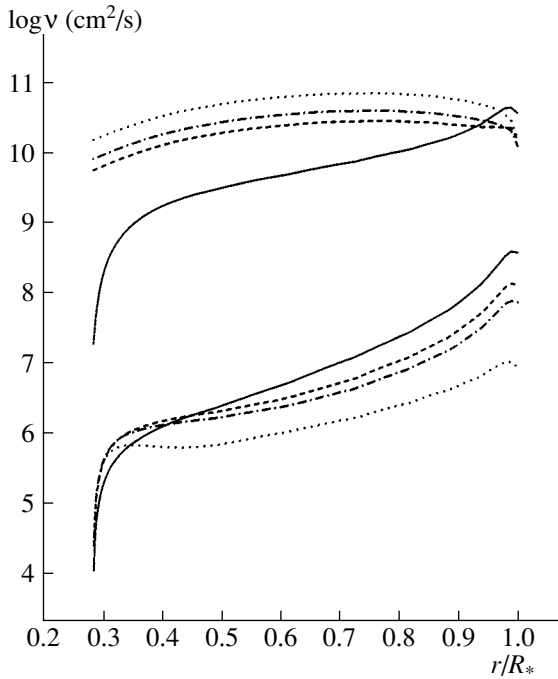
log ν (cm²/s)

Fig. 5. The turbulent-viscosity coefficient ν in the horizontal (four top curves) and vertical (four bottom curves) directions in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r for the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$ for times $t = 10^5$ yrs (dotted curve), $t = 3 \times 10^5$ yrs (dash-dotted curve), and $t = 5 \times 10^5$ yrs (dashed curve). The turbulent-viscosity coefficients for steady-state rotation are plotted by the solid curves.

log ζ

Fig. 6. Relative amplitudes ζ of the latitude variations of several physical parameters at the constant-pressure surface in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r in the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$. Shown are values for $\zeta = g_2/g$ (solid curve), $\zeta = \Omega_2/\Omega$ (dotted curve), and $\zeta = -\rho_2/\rho$ (dashed curve). All distributions are given for an age of $t = 10^5$ yrs.

on the constant-pressure surface are small (Fig. 6). The variations of the rotational angular velocity with latitude (10) are $\sim 0.1\%$. Thus, the stellar rotation is laminar.

If we assume $\sim 20\%$ variation of the rotational angular velocity at the constant-pressure surface, we can use (7) to obtain for the horizontal turbulent-viscosity coefficient [17]

$$\nu_h = r\,|U(r)|\left[\frac{1}{3}\frac{d\ln\left(\rho r^2 U\right)}{d\ln r} - \frac{1}{2}\frac{d\ln\left(r^2\Omega\right)}{d\ln r}\right].\tag{12}$$

This expression was used in the calculations for the rotational angular velocity in models of stars with masses of $9-30\,M_\odot$ [24, 25]. When (12) is used, steady-state rotation is reached over a shorter time (Fig. 2b) and is achieved more rapidly in stars with higher angular momentum (Fig. 4). The horizontal turbulent-viscosity coefficient calculated using (12) is substantially smaller than the value calculated using (9). With lower ν_h and the same drop between the rotational angular velocities of the convective core
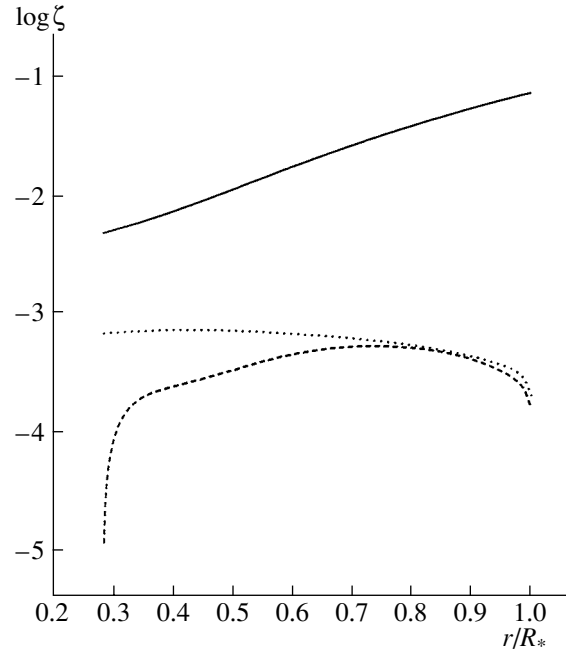
and stellar surface, the velocity of the meridional circulation is higher [the dependence of $U(r)$ on ν_h is given by (4)]. Accordingly, the density of the advective angular-momentum flux is higher, while the characteristic time for the transport of the angular momentum via meridional circulation decreases. However, the turbulence is not appreciably anisotropic in the case of steady-state rotation: the turbulent-viscosity coefficients in horizontal (12) and vertical directions are roughly equal. The anisotropy condition is violated even before the equilibrium solution is reached.

In the case $((\Omega_c/\Omega_s)_0, J) = (2, 1.83 \times 10^{52})$, the flow in the meridional plane initially occurs in two cells with different directions of motion. In contrast to the case of initial rigid-body rotation, in the inner cell, which is adjacent to the convective core, the matter ascends in the equatorial plane and descends along the rotation axis. The velocity of the flow is ~ 500 m/s and substantially exceeds the circulation velocity in the outer cell. Meridional flows transport angular momentum from the convective core to the central part of the radiative envelope. Variation of the rotational angular velocity at the base of the radiative envelope results in a rapid decrease of the circulation velocity. As soon as after ~ 0.2 yrs, the maximum
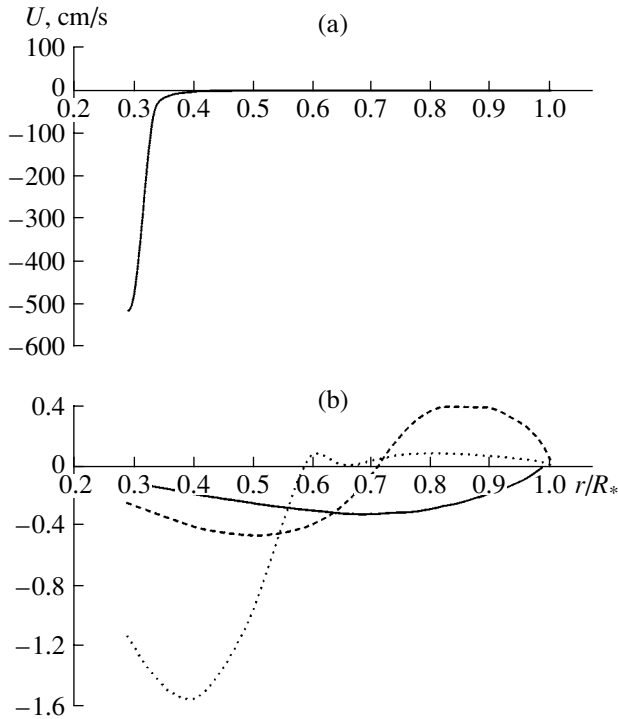
Fig. 7. Velocity of meridional circulation U in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r in the case $((\Omega_c/\Omega_s)_0, J) = (2, 1.83 \times 10^{52})$ for times (a) $t = 0.2$ yrs (solid curve), (b) $t = 100$ yrs (dotted curve), $t = 550$ yrs (dashed curve), and $t = 1300$ yrs (solid curve).

Fig. 8. The rotational angular velocity Ω in the radiative envelope of a star with mass $16\,M_\odot$ and radius R_* as a function of r in the case $((\Omega_c/\Omega_s)_0, J) = (2, 1.83 \times 10^{52})$ for times $t = 0$ yrs (solid, thin curve), $t = 100$ yrs (dotted curve), $t = 1000$ yrs (dash-dotted curve), $t = 10\,000$ yrs (dashed curve). The steady-state rotation is plotted by the solid, thick curve.

meridional circulation velocity decreases by two orders of magnitude (Fig. 7a). As angular momentum is transported from the convective core to the central part of the radiative envelope, the boundary dividing the two circulation cells shifts outwards. The velocity of the meridional flows decreases in the inner cell and increases in the outer cell (Fig. 7b). At this evolutionary stage, the outflow of angular momentum from the outer parts of the star inside the radiative envelope (i.e., the local increase of Ω_c/Ω_s with time (Fig. 4)) becomes appreciable. By an age of ~ 1300 years, the outer circulation cell has disappeared, and a single direction for the circulation in the radiative envelope has been established: matter ascends in the equatorial plane and descends along the polar axis.

Meridional flows transport angular momentum from inner layers of the star outwards. As soon as after $\sim 10\,000$ yrs, the vertical distribution of the rotational angular velocity becomes close to steady state (Fig. 8). The direction of the meridional circulation changes at the base of the radiative envelope. An inner cell is formed where matter descends in the equatorial plane and ascends along the rotation axis. The characteristic meridional circulation velocity decreases to $\sim 10^{-3}$ cm/s in the outer circulation
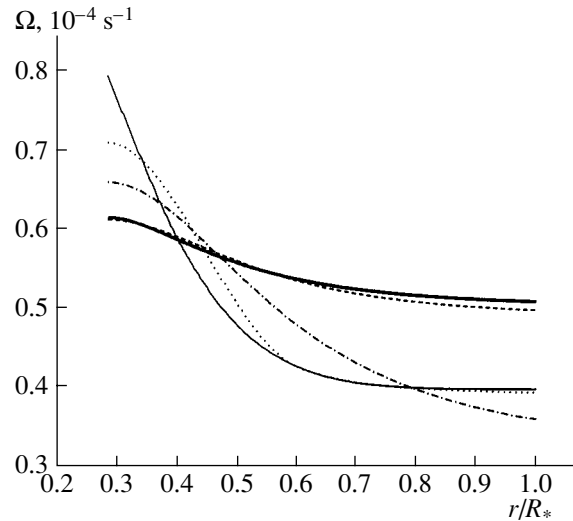
cell. In the inner cell, the characteristic meridional circulation velocity does not exceed $\sim 10^{-4}$ cm/s. The characteristic time for the angular-momentum transport increases accordingly, and further variations of the rotational angular velocity occur substantially more slowly (Fig. 4). The boundary separating the two circulation cells shifts outwards, and the outer cell disappears. The steady-state rotation is the same as in the case $((\Omega_c/\Omega_s)_0, J) = (1, 1.83 \times 10^{52})$.

In the case $((\Omega_c/\Omega_s)_0, J) = (1.3, 2.75 \times 10^{52})$, the angular-momentum transport occurs qualitatively in the same way as in the case $((\Omega_c/\Omega_s)_0, J) = (2, 1.83 \times 10^{52})$. As soon as after ~ 300 yrs, a single circulation direction is established in the radiative envelope, with matter ascending in the equatorial plane and descending in circumpolar regions. Meridional flows transport angular momentum outward from the inner layers of the star. After ~ 3000 years, the excess of the rotational angular velocity of the convective core compared to the surface value decreases to $\sim 10\%$. The direction of the meridional circulation changes in the layer adjacent to the convective core, where an inner circulation cell is formed, in which matter descends in the equatorial plane and ascends along the rotation axis. In this cell, meridional flows transport angular momentum inside the star. The velocities of the meridional flow are $\sim 10^{-4}$ cm/s in the inner cell and $\sim 10^{-3}$ cm/s in the outer cell. The boundary separating the two cells shifts outwards and the outer cell disappears. The steady-state rotation

appears the same as in the case $((\Omega_c/\Omega_s)_0, J) = (1, 2.75 \times 10^{52})$.

The steady rotational states obtained for two different values of the angular momentum of the star vary only a little.

4. THE STATE OF ROTATION OF YOUNG MASSIVE STARS

Let us consider the characteristic time scale for angular-momentum transport in the radiative envelope due to the meridional flows,

$$t_{adv} = \Delta r/|U|,$$

where Δr is the thickness of the radiative envelope and U is the characteristic velocity of the meridional circulation. All variations in the distribution of the rotational angular velocity can be tentatively divided into two types: rapid variations occurring on time scales smaller than the characteristic time scale for the nuclear evolution ($t_{adv} < t_{nuc}$), and slow variations occurring on time scales comparable to or exceeding the nuclear time scale ($t_{adv} \geq t_{nuc}$). The latter condition is fulfilled when $|U| \leq 10^{-3}$ cm/s. In turn, the characteristic meridional circulation velocity depends on the ratio of the rotational angular velocities of the convective core and stellar surface. The more this ratio deviates from its equilibrium value, the larger the meridional circulation velocity and the smaller the characteristic time scale for the angular-momentum transport. When Ω_c/Ω_s approaches its equilibrium value, the meridional circulation velocity decreases, and the characteristic time scale for the angular-momentum transport increases. When considering slow variations of the distribution of the rotational angular velocity, it is important to take into account variations in the star's structure associated with variations of the chemical composition of the matter produced in nuclear reactions. When the variations are fast, there is not sufficient time for nuclear reactions to influence the star's structure. Precisely such variations determine the vertical angular-momentum distribution in young massive stars. Thus, young stars may be characterized by rotation that is close to equilibrium, so that the characteristic time scale for the angular-momentum transport is comparable to the nuclear time scale.

In the case of initial rigid-body rotation, the meridional circulation velocity decreases to 10^{-3} cm/s over $\sim 100\,000$ yrs (Fig. 1b), which is 1% of the lifetime of a star with mass $16\,M_\odot$ on the main sequence. Over this time, the rotational angular velocity of the convective core increases by $\sim 5\%$ compared with the surface value. Further variations of Ω_c/Ω_s occur on the nuclear time scale (Fig. 4). When the initial rotational angular velocity of the convective

core exceeds the surface value by factors of 1.3 and 2, the meridional circulation velocity decreases to $\sim 10^{-4}$–10^{-3} cm/s after ~ 3000 and $\sim 10\,000$ years, respectively. In these cases, the excess of the rotational angular velocity of the convective core over the surface value decreases to $\sim 10\%$ and $\sim 18\%$, respectively. Thus, the rotational angular velocities of the convective cores in young massive stars can exceed the surface values by ~ 5–20%. In this case, the characteristic time scale for angular-momentum transport via meridional flows and shear turbulence is comparable to the nuclear time scale. If the excess of the rotational angular velocity of the convective core over the surface value exceeds these limits, the meridional flows alter the rotation of the star before nuclear reactions have time to appreciably change the stellar structure.

5. CONCLUSION

We have analyzed the rotation of a chemically homogeneous star with a mass of $16\,M_\odot$ assuming that the angular-momentum distribution in its radiative envelope is determined by hydrodynamical processes—flows and turbulent diffusion. Assuming laminar rotation of the star [17–20], we have calculated the vertical component of the meridional circulation velocity, rotational angular velocity, deviations of the rotation of the constant-pressure surface from rigid-body rotation, and the horizontal and vertical turbulent-viscosity coefficients as functions of the average distance of the constant-pressure surface from the center of the star and of time.

In the case of steady-state rotation of a massive star, the advective and turbulent flux densities of the angular momentum are equal at each point of the radiative envelope and only weakly depend on the total angular momentum of the star. Meridional flows transport angular momentum inside the star. The advective angular-momentum transport is balanced by the outward component of the turbulent flow. In this state, the rotational and meridional circulation velocity fields are self-consistent. The constant-pressure surfaces rotate essentially as rigid bodies; the rotational angular velocity of the convective core is $\sim 20\%$ higher than the surface value. The vertical component of the meridional circulation velocity increases from $\sim 10^{-6}$ cm/s near the convective core to $\sim 10^{-4}$ cm/s in subsurface layers.

The turbulence in flows occurring in the radiative envelopes of young massive stars is highly anisotropic; the horizontal turbulent-viscosity coefficient exceeds the vertical coefficient by three to four orders of magnitude (Fig. 5). Due to the high efficiency of turbulent angular-momentum transport in the horizontal direction, the rotational velocity on

the constant-pressure surface remains essentially constant. Deviations of this rotation from rigid-body rotation do not exceed 0.1% (Fig. 6).

If the rotation of the star deviates from steady-state, the basic means of angular-momentum transport in the vertical direction is flows in the meridional plane. When deviations from the steady-state rotation are appreciable, the velocity of the meridional flows increases dramatically. For example, if the rotational angular velocity of the convective core is twice the surface value, the circulation velocity can reach 500 m/s. With such velocities, the characteristic time scale for angular-momentum transport in the vertical direction of the circulation is substantially smaller than the nuclear time scale.

When the rotation of the star deviates appreciably from steady-state, the contribution of turbulent angular-momentum transport to the vertical distribution of the angular velocity is negligible compared to the contribution of advective angular-momentum transport (Fig. 3). Thus, the vertical distribution of the rotational angular velocities of young massive stars is determined by the meridional flows, on characteristic time scales that are small compared with the nuclear time scale. Shear turbulence in the horizontal direction ensures an almost constant rotational angular velocity the constant-pressure surface.

The rotational angular velocity of the convective core of a young massive star can exceed the surface value by $\sim 5-20\%$. In this case, the characteristic velocity of the meridional circulation is $10^{-3}-10^{-4}$ cm/s. This rotation is non-steady-state, but for such a drop between the rotational angular velocities of the convective core and stellar surface, the characteristic time for the transport of angular momentum in the vertical direction via meridional flows and shear turbulence is comparable to or exceeds the nuclear time scale.

The models for young massive stars considered here can be used as initial models in studies of stellar evolution on the main sequence taking into account the hydrodynamical transport of angular momentum and the concentrations of various chemical elements.

ACKNOWLEDGMENTS

REFERENCES

1. A. G. Massevich and A. V. Tutukov, *Stellar Evolution: Theory and Observations* (Nauka, Moscow, 1988) [in Russian].
2. G. S. Bisnovatyi-Kogan, *Physical Problems of the Theory of Stellar Evolution* (Nauka, Moscow, 1989) [in Russian].
3. L. S. Lyubimkov, *Chemical Composition of Stars: Methods and Analysis* (Astroprint, Odessa, 1995) [in Russian].
4. A. Herrero, R. P. Kudritzki, J. M. Vilchez, *et al.*, Astron. Astrophys. **261**, 209 (1992).
5. A. Herrero, R. P. Kudritzki, R. Gabler, *et al.*, Astron. Astrophys. **297**, 556 (1995).
6. A. Herrero, L. J. Corral, M. R. Villamariz, *et al.*, Astron. Astrophys. **348**, 542 (1999).
7. C. de Jager, *The Brightest Stars* (Reidel, Dordrecht, 1980; Mir, Moscow, 1984).
8. J.-F. Donati, G. A. Wade, J. Babel, *et al.*, Mon. Not. R. Astron. Soc. **326**, 1265 (2001).
9. E. I. Staritsin, Astron. Zh. **78**, 541 (2001) [Astron. Rep. **45**, 467 (2001)].
10. H. Vogt, Astron. Nachr. **223**, 229 (1925).
11. A. S. Eddington, Observatory **48**, 73 (1925).
12. L. Mestel, Mon. Not. R. Astron. Soc. **122**, 473 (1961).
13. L. Gratton, Mem. Soc. Astron. Ital. **17**, 5 (1945).
14. E. J. Opik, Mon. Not. R. Astron. Soc. **111**, 278 (1951).
15. G. G. Pavlov and D. G. Yakovlev, Astron. Zh. **55**, 1043 (1978) [Sov. Astron. **22**, 595 (1978)].
16. J.-L. Tassoul, *Theory of Rotating Stars* (Princeton Univ. Press, Princeton, 1979; Mir, Moscow, 1982).
17. J.-P. Zahn, Astron. Astrophys. **265**, 115 (1992).
18. S. Talon and J.-P. Zahn, Astron. Astrophys. **317**, 749 (1997).
19. A. Maeder and J.-P. Zahn, Astron. Astrophys. **334**, 1000 (1998).
20. A. Maeder, Astron. Astrophys. **399**, 263 (2003).
21. E. I. Staritsin, Astron. Zh. **76**, 678 (1999) [Astron. Rep. **43**, 592 (1999)].
22. L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, Oxford, 1987).
23. J.-P. Zahn, Space Sci. Rev. **66**, 285 (1994).
24. S. Talon, J.-P. Zahn, A. Maeder, *et al.*, Astron. Astrophys. **322**, 209 (1997).
25. P. A. Denissenkov, N. S. Ivanova, and A. Weiss, Astron. Astrophys. **341**, 181 (1999).

Translated by K. Maslennikov

The Mass Spectrum of Black Holes in Close Binary Systems

A. I. Bogomazov[1], M. K. Abubekerov[2], and V. M. Lipunov[1,2]

[1]Moscow State University, Moscow, Russia

[2]Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, Russia

Received January 20, 2005; in final form, February 17, 2005

Abstract—We present the results of population syntheses obtained using our "scenario machine." The mass spectra of black holes in X-ray binary systems before and after the stage of accretion from an optical companion are obtained for various evolutionary scenarios. The results of the model computations are compared to observational data. The observational data are used to estimate the fraction of a pre-supernova's mass that collapses into a black hole. This model can explain the formation of low-mass $(2-4\,M_\odot)$ black holes in binary systems with optical companions. We show that the number of low-mass black holes in the Galaxy is sufficiently high for them to be detected. The population-synthesis results suggest that the vast majority of low-mass black holes are formed via the accretion-induced collapse of neutron stars. The percentage of low-mass black holes in binary systems that form due to accretion-induced collapse is $2-15\%$ of the total number of black holes in binaries, depending on the evolutionary scenario. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Currently, about a thousand X-ray sources have been detected in the Milky Way and nearby galaxies [1, 2], most of which are close binary systems in which an optical component supplies matter to a neutron star or black hole. However, mass estimates have been derived for only ~30 neutron stars and ~20 black holes. This number is not sufficient to enable firm conclusions about the properties of the mass spectrum of relativistic objects.

Thus, the mass spectrum of compact objects is not well known due to poor statistics and insufficient accuracy in the estimated masses of neutron stars and black holes. In addition, the lack of compact objects with masses between 2 and $6\,M_\odot$ in the observed mass distribution for neutron stars and black holes is striking. This gap is especially surprising in the light of new data on the masses of the CO cores of Wolf–Rayet stars at the end of their evolution [3], which are continuously distributed over a wide range, $M_{CO} = (1-2)-(20-44)\,M_\odot$. Since it is thought that Wolf–Rayet stars are progenitors of relativistic objects [4–6], such a large difference between the final masses of the CO cores of Wolf–Rayet stars and the masses of the relativistic objects that they are thought to produce requires an explanation. For this reason, with the aim of determining the possible masses of black holes and the shape of the mass spectrum of compact objects before and after accretion of matter from the optical companion, we carried out population syntheses using the "scenario machine" described in [7].

Accretion onto neutron stars and estimates of the mass accumulated on their surfaces have been analyzed in detail in [8]. The aim of the present paper is to model the mass spectrum of black holes in binary systems with optical components (BH + opt).

2. OBSERVATIONAL DATA

As we noted above, masses have been estimated for more than 30 neutron stars and about 20 black
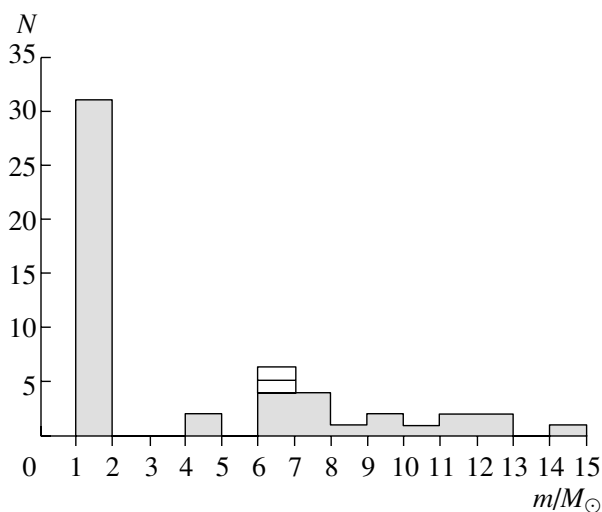

Fig. 1. Observed mass distribution of compact objects. The open rectangles show the masses of black holes derived from microlensing experiments [9].

Table 1. Masses of black holes in binary systems

Name	i, deg	$f(M)$, $M_\odot$	m_{BH}, $M_\odot$	$m_{\rm opt}$, $M_\odot$	References
Cyg X-1	$31-44$	0.2580 ± 0.0007	11.05 ± 2.55	22.0 ± 0.51	[10]
LMC X-1	~ 63	0.14 ± 0.05	7 ± 3	22 ± 4	[2, 11]
LMC X-3	67 ± 3	2.29 ± 0.32	$5.94-9.17$	$3-8$	[12−14]
SS 433	90	7.7 ± 1.1	11 ± 5	19 ± 7	[15]
A0620-00	40.75 ± 3	2.72 ± 0.06	11.0 ± 1.9	0.68 ± 0.18	[16]
V404 Cyg	$54-64$	5.819 ± 0.003	10.65 ± 1.95	0.64 ± 0.12	[17]
GRS 1124-683	54 ± 1.5	3.01 ± 0.15	6.95 ± 0.6	0.75 ± 0.05	[18, 19]
GRS 1915+105	70 ± 2	9.5 ± 3.0	14 ± 4	1.2 ± 0.2	[2, 11]
GS 2000+25	64 ± 1.3	5.01 ± 0.12	$7.15-7.78$	$0.25-0.41$	[11]
GRO J0422+32	45 ± 2	1.19 ± 0.02	3.97 ± 0.95	0.46 ± 0.31	[35]
GRO J1655-40	70.2 ± 1.9	2.73 ± 0.09	6.3 ± 0.5	2.4 ± 0.4	[20]
H 1705-250	70 ± 10	4.86 ± 0.13	$4.9-7.9$	0.26 ± 0.42	[21, 22]
4U 1543-47	20.7 ± 1.5	0.25 ± 0.01	$8.45-10.39$	$2.0-2.5$	[11]
GRS 1009-45	~ 78	3.17 ± 0.12	$4.4^{+0.34}_{-0.76}$	$0.6^{+0.05}_{-0.10}$	[34]
SAX J1819.3-25	75 ± 2	3.13 ± 0.13	$6.82-7.42$	$2.35-3.34$	[11]
XTE J1118+480	81 ± 2	6.1 ± 0.3	$6.0-7.7$	$0.09-0.5$	[23]
XTE J1550-564	$67-77.4$	6.86 ± 0.71	$9.41^{+1.35}_{-1.05}$	<0.79	[24]
XTE J1859+226	$-$	7.4 ± 1.1	$7.6-12.0$	$-$	[25]
GX 339-4	$-$	5.8 ± 0.5	$-$	$-$	[26]
XTE 1650-500*	$-$	$-$	~ 8.2	$-$	[27]

* Mass estimate obtained from high-frequency quasi-periodic oscillations in the X-ray.

holes. The mass distribution of these relativistic objects is shown in Fig. 1. The black-hole masses are listed in Tables 1 and 2. The neutron-star masses presented in Fig. 1 are already given in [8], and we do not list them here.

Figure 1 shows that this distribution is bimodal [3, 28, 29]. The masses of the neutron stars are confined to a narrow range, with the average mass being $1.35 \pm 0.15\,M_\odot$. The black-hole masses are distributed over a relatively wide range: $m_{BH} = 4-15\,M_\odot$. The average black-hole mass is $6.64 \pm 0.77\,M_\odot$.

No candidate black holes with masses of $2-4\,M_\odot$ have been discovered. Only the central masses of the compact components of the close binary systems Vela X-1, 4U 1700−37, and J0751+1807 fall in this range. The compact objects in these systems are neutron stars with masses close to $\sim 2\,M_\odot$. As recent studies have shown [30−32], the estimated masses of the compact objects in Vela X-1, 4U 1700−37, and J0751+1807 are not firm enough to be confident that massive neutron stars ($m_{NS} > 1.8\,M_\odot$) are present in these systems. For this reason, we placed the masses of the neutron stars in Vela X-1, 4U 1700−37, and J0751+1807 in the $1-2\,M_\odot$ bin in Fig. 1. The method used to estimate the masses of the neutron stars in these binaries and results based on these estimates are considered in detail in [33].

Table 2. Masses of black holes based on observations of microlensing

Name	m_{BH}, $M_\odot$	References
MACHO-96-BLG-5	6^{+10}_{-3}	[9]
MACHO-98-BLG-6	6^{+7}_{-3}	[9]

The central mass estimates of the compact components in the X-ray binaries GRS 1009-45 and GRO J0422+32 are in the range $4-6 \, M_\odot$ (Fig. 1), making the relativistic objects in these binary systems candidate low-mass black holes. According to [34], the mass of the GRS 1009-45 black hole is $4.4^{+0.34}_{-0.76} \, M_\odot$; the mass of the GRO J0422+32 black hole is $3.97 \pm 0.95 \, M_\odot$ [35]. Note that this latter estimate is not entirely firm. In particular, according to [36], the mass of this black hole exceeds $9 \, M_\odot$. We have adopted the more recent mass estimate obtained in [35].

The central mass estimates for the remaining candidate black holes exceed $6 \, M_\odot$ (Table 1). The mass distribution for the candidate black holes with $m_{BH} \geq 6 \, M_\odot$ has a broad peak at $6-8 \, M_\odot$ and a uniform tail extending to $14 \, M_\odot$.

We are predominantly considering mass estimates for compact objects obtained using dynamical methods (i.e., from observations of the radial-velocity curve of the optical companion), which are the most accurate and reliable. The masses of the candidate black holes MACHO-96-BLG-5 and MACHO-98-BLG-6 derived from microlensing observations are listed in Table 2. Due to their lower accuracy, these two mass estimates are shown as unfilled entries in Fig. 1.

Due to the meagre statistics, it is not possible to determine the shape of the mass spectrum for the observed black holes with confidence (Fig. 1). However, we can use the existing black-hole mass estimates to find the ratio of the numbers of low- and high-mass black holes in binary systems:

$$R = N(m_{BH} \leq m_{BH}^{\min})/N(m_{BH} > m_{BH}^{\min}). \quad (1)$$

Based on the mass estimates listed in Table 1, we took $m_{BH}^{\min}$ to be $\sim 4 \, M_\odot$. Accordingly, we took black holes with masses lower than $\sim 4 \, M_\odot$ to be "low-mass" and black holes with masses greater than $\sim 4 \, M_\odot$ to be "high-mass" black holes. The observational estimates of the black-hole masses in Table 1 yield a ratio of the numbers of low- and high-mass black holes in binary systems $R_{obs} \lesssim 1/10$. The parameter R served as a criterion of the adequacy of the various models in terms of the observational statistics of the masses of black holes in close binaries.

Another such criterion was that an adequate model should have at least one Cyg X-1-like system per Galaxy (per $\sim 10^{11}$ stars). Following [37], we suppose that Cyg X-1 is not a statistical outlier and that there may be several such systems in the Galaxy. Support for this hypothesis is provided by the existence of similar candidate systems in the Large Magellanic Cloud: LMC X-1 and LMC X-3. In the model computations, we considered a binary to be a Cyg X-1-like system if it contains a massive optical star ($m_{opt} \geq 10 \, M_\odot$) close to Roche-lobe overflow and a massive black hole ($m_{BH} > 4 \, M_\odot$) that is accreting from a disk (for more details see [38]).

The last important requirement of the population-synthesis calculations was the virtual absence of black holes in pairs with radio pulsars (BH + PSR systems). To exclude observational selection effects, we used the ratio of the numbers of BH + PSR systems and of all radio pulsars to evaluate this criterion $N(\text{BH} + \text{PSR})/N(\text{PSR})$. According to the observational data, there are no radio pulsars in binaries with black holes among the ~ 1500 detected radio pulsars. Thus, in the model Galaxy, the ratio $N(\text{BH} + \text{PSR})/N(\text{PSR})$ should not exceed $1/1500$.

The above criteria based on observational information enabled us to restrict in advance the model parameters used in the "scenario machine."

3. POPULATION SYNTHESIS

General Description of the Model

With the aim of determining the shape of the black-hole mass spectrum before and after the accretion stage, we carried out population-synthesis computations for 10^6 binary systems under various assumptions about their evolution using the "scenario machine" [7]. The computations assumed a Salpeter initial mass function for the binary components:

$$f(m) = m^{-2.35}. \quad (2)$$

The initial component masses were varied from $10 \, M_\odot$ to $120 \, M_\odot$. Population-synthesis computations were carried out for two distributions of the initial component-mass ratio $f(q) = q^{\alpha_q}$: uniform ($\alpha_q = 0$) and quadratic ($\alpha_q = 2$), where $q = m_2/m_1$. We considered masses of the secondary m_2 and primary m_1 such that $m_1 > m_2$. The distribution of the binaries over the initial component separation $f(a)$ was taken to obey the function

$$f(a) = 1/a \quad (3)$$

(see [39] for more details). The initial semimajor axis of the orbit could have any value in the range $(10 - 10^6) \, R_\odot$.

We selected from the computation results only black holes with optical companions (BH + opt). Among the many parameters of the BH + opt systems that form, we are interested in the masses m_{BH} and lifetimes of the black holes. Since black holes undergo an accretion stage during their evolution, we considered the black-hole mass spectrum in BH + opt systems both before and after this stage.

We assumed that the strengths of the anisotropic kicks that the nascent black holes receive in supernova explosions have a Maxwellian distribution:

$$f(v) \sim \frac{v^2}{v_0^2} e^{-\frac{v^2}{v_0^2}}, \qquad (4)$$

with all directions having equal probability. However, the magnitude of the kick after the supernova v_0 remains an important but poorly studied parameter. The results of the population synthesis are very sensitive to v_0. A sharp increase of v_0 results in a sharp decrease of the number of systems with relativistic companions. The v_0 value for neutron stars is $100-180$ km/s [40], but the typical value of v_0 for black holes formed via collapse is not currently known.

Taking the typical anisotropic kick velocity for neutron stars to be $v_0 = 180$ km/s [40] and assuming that the magnitude of the kick acquired by the neutron star during the collapse depends on the mass of the ejected envelope, we specified the value of v_0 for black holes using the relation

$$v_0 = 180 \frac{m_{preSN} - m_{BH}}{m_{BH}} \text{ km/s}, \qquad (5)$$

where m_{preSN} is the mass of the presupernova and m_{BH} the mass of the nascent black hole.

The mass-loss rate of the optical star $\dot{m}$ is also a relatively poorly studied parameter. For this reason, we carried out model computations for three mass-loss scenarios during the evolution of a star, which we label A, B, and C. We also carried out a computation for the Woosley model of stellar evolution [41, 42]. The scenario based on the Woosley model is labeled W.

Evolutionary Scenario A

The mass-loss rate in the main-sequence (MS) stage is described by the classical formula of de Jager [43]:

$$\dot{m} \sim L/V_\infty, \qquad (6)$$

where L is the star's luminosity and V_∞ the terminal velocity of the stellar wind.

For giants, we used the larger of the values given by (6) and by the expression of Lamers [44]:

$$\dot{m} \sim L^{1.42} R^{0.61}/m^{0.99}, \qquad (7)$$

where R and m are the radius and mass of the star.

For the red supergiant stage, we used the larger of the values given by (6) and by the wind model of Kudritzki and Reimers [45]:

$$\dot{m} \sim LR/m. \qquad (8)$$

The change of the star's mass Δm during a single stage in the model with type-A wind does not exceed $0.1(m - m_{core})$, where m is the mass of the star at the beginning of this stage and m_{core} is the mass of the stellar core. The mass loss in the Wolf–Rayet stage was parameterized as $0.1 \, m_{WR}$, where m_{WR} is the maximum stellar mass in this stage. We used the core masses computed in [46–48] to calculate the parameters of the type-A wind.

In scenario A, the mass lost by a star did not exceed 30% of its initial mass m_{opt}.

Evolutionary Scenario B

In scenario B, we used the results of the evolutionary computations of [49], which indicate that a massive star loses up to $\sim$90% of its initial mass in the main-sequence, supergiant, and Wolf–Rayet stages via its stellar wind. Therefore, the presupernova mass in scenario B was $\sim$8–10 $M_\odot$, essentially independent of the mass of the parent star.

Evolutionary Scenario C

In scenario C, the mass-loss rate in the main-sequence, supergiant, and Wolf–Rayet stages was based on the computations of Vanbeveren [50], which reproduce the observed distribution of Galactic Wolf–Rayet stars and the stellar-wind mass loss by massive stars accurately. The computations made use of the relation

$$\Delta m = (m - m_{core}), \qquad (9)$$

where the stellar core mass m_{core} is defined as

$$m_{core} = \begin{cases} 1.62 m_{opt}^{0.83} & \text{for MS stars,} \\ 10^{-3.051 + 4.21 \log m_{opt} - 0.93 (\log m_{opt})^2} & \text{for supergiants,} \\ 0.83 m_{WR}^{0.36} & \text{for Wolf–Rayet stars with } m_{WR} < 2.5 \, M_\odot, \\ 1.3 + 0.65(m_{WR} - 2.4) & \text{for Wolf–Rayet stars with } m_{WR} > 2.5 \, M_\odot, \\ m_{core} = 3.03 m_{opt}^{0.342} & \text{for Wolf–Rayet stars with } m_{\max}^{opt} > 20 \, M_\odot. \end{cases} \qquad (10)$$
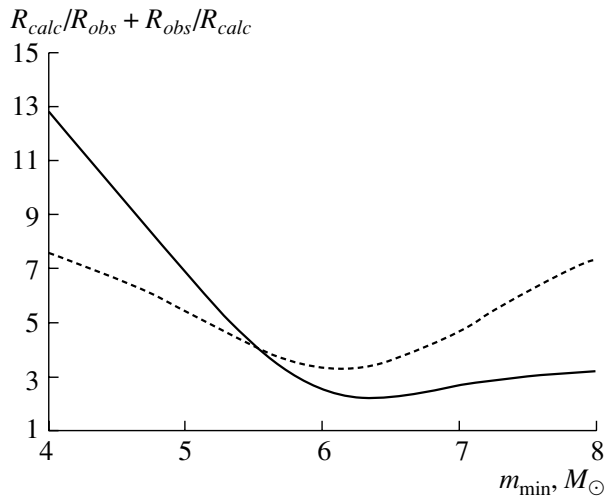
Fig. 2. Dependence of $R_{calc}/R_{obs} + R_{obs}/R_{calc}$ on $m_{\min}$ in scenarios A (solid curve) and C (dashed curve); see the text for more details. The parameter $k_{BH} = 0.5$.
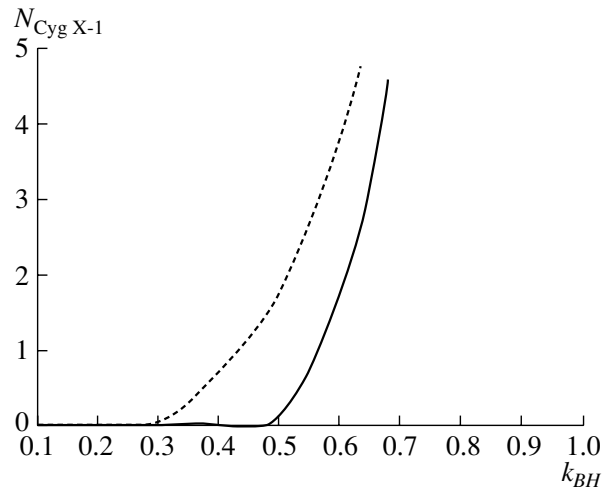

Fig. 3. Dependence of the number of Cyg X-1-like systems in the Galaxy on $k_{BH} = 0.5$ in model A (for $\alpha_q = 0$, dashed curve) and model C (for $\alpha_q = 2$, solid curve).

Evolutionary scenario C has moderate mass loss, and a mass-loss rate that is lower than in scenario B, but higher than in scenario A. For example, in scenario C, a star with an initial mass of $m_{opt} > 15\,M_{\odot}$ can lose up to 30% of its initial mass in the main-sequence, giant, and supergiant stages. We applied scenario A for the computations of mass loss by lower-mass stars with $m_{opt} < 15\,M_{\odot}$. A high mass loss in the Wolf–Rayet stage is typical of scenario C, with the Wolf–Rayet star losing up to $\sim$50% of its initial mass.

Evolutionary Scenario W

Evolutionary scenario W is based on the evolutionary diagram for stars of various masses published by Woosley [41, Fig. 16], which represents the relationship between the mass of the relativistic remnant and the initial mass of the star. We carried out population-synthesis computations for two models with W-type stellar winds, which we label Wb and Wc. In models Wb and Wc, the mass-loss rates were computed as in scenario B and scenario C, respectively. The use of these models to calculate the wind rate in a scenario based on Woosley's diagram [41, Fig. 16] is justified by the fact that scenarios B and C are based on the same numerical expressions for the mass-loss rates from [49–51] that were used by Woosley in his work (see [7] for more details).

4. THE MINIMUM BLACK-HOLE MASS

One of the important parameters of the population syntheses is the minimum mass of the black holes

formed in supernovae, $m_{\min}$. Since this parameter is not known precisely, we carried out population-synthesis computations for several values of $m_{\min}$ in the range $2.5-10\,M_{\odot}$. We found that the maximum of the black-hole mass spectrum corresponds to $m_{\min}$. We can see from the observational data (Table 1 and Fig. 1) that the maximum of the black-hole mass spectrum is in the range $6-8\,M_{\odot}$; i.e., $m_{\min}$ should be close to $7\,M_{\odot}$.

Additional evidence that $m_{\min}$ is close to $7\,M_{\odot}$ is provided by special model computations carried out for this purpose. In these computations, we calculated the parameter R_{calc} [the ratio of the numbers of low- and high-mass black holes; see Eq. (1)] for $m_{\min} = 2, 3, 4, \ldots, 10\,M_{\odot}$. The computations were carried out for scenarios A and C and assumed that half of the presupernova mass m_{preSN} collapses into a black hole (with mass m_{BH}). Figures (2a, 2b) show the dependence of $R_{calc}/R_{obs} + R_{obs}/R_{calc}$ on $m_{\min}$ in evolutionary scenarios A and C. Recall that we assumed that the observed ratio of the numbers of low- and high-mass black holes $R_{obs} = 1/10$ (Table 1). In both cases (Figs. 2a and 2b), the minimum of $R_{calc}/R_{obs} + R_{obs}/R_{calc}$, which is related to R_{calc}/R_{obs}, corresponds to $m_{\min} \simeq 6-6.5\,M_{\odot}$.

Given these results, in the subsequent model computations, we set the minimum mass of a black hole formed in a supernova to $m_{\min} = 7\,M_{\odot}$. This does not preclude the formation of black holes with masses below $7\,M_{\odot}$. For example, in all the evolutionary scenarios (A, B, C, and W) we allowed for the accretion-induced collapse of a neutron star into a black hole if the mass of the neutron star grew to the Oppenheimer–Volkoff limit, which we took to be $2\,M_{\odot}$.

5. THE MASS FRACTION OF PRESUPERNOVA THAT COLLAPSE INSIDE THE EVENT HORIZON

An important parameter that enters the population-synthesis algorithm is the fraction of presupernovae with masses of m_{preSN} that collapse into black holes with masses of m_{BH}. In scenarios A, B, and C, the mass of the black hole m_{BH} produced by the explosion of a presupernova with mass m_{preSN} was computed as

$$m_{BH} = k_{BH} m_{preSN}, \qquad (11)$$

where the constant coefficient k_{BH} is the fraction of the presupernova mass that collapses inside the event horizon. In scenario W, k_{BH} was taken to be variable and to have a value determined by the dependence of the mass of the compact remnant on the initial mass of the optical star [41, Fig. 16].

The best-fit value of k_{BH} in scenarios A, B, and C was searched for. Population syntheses were carried out for each value of k_{BH} in the range from 0.1 to 1.0 with steps of 0.1. The other population-synthesis parameters in this set of calculations were held fixed. When varying the value of k_{BH}, we monitored the number of Galactic Cyg X-1-like systems produced in the computations. The results are presented in Fig. 3. Following [37, 38], we assumed that Cyg X-1-like systems are not statistical outliers and are always present in the Galaxy. We assume that approximately one such system should exist in the Galaxy at any given time. This criterion yielded k_{BH} values of 0.43 for scenario A (Fig. 3) and 0.57 for scenario C (Fig. 3). In scenario B, no Cyg X-1-type systems were produced for any of the k_{BH} values in the range 0.1−1.0. For this reason, we rejected scenario B as being unrealistic. Further runs of the population synthesis code were carried out for scenarios A, C, and W.

6. BLACK-HOLE MASS SPECTRA IN MODEL A

For each scenario, a population synthesis was carried out for 10^6 initial binaries. For models A and C, we assumed that the initial distribution of the component-mass ratios in the binaries was flat ($\alpha_q = 0$). The results obtained for a quadratic distribution ($\alpha_q = 2$) did not satisfy the observational criteria.

The spectrum of initial black-hole masses obtained for scenario A, shown in Fig. 4a, is clearly bimodal. Most of the black-hole masses are concentrated in the range 7−12 $M_\odot$, although there are also low-mass black holes ($m_{BH} = 2$−3 $M_\odot$) present in
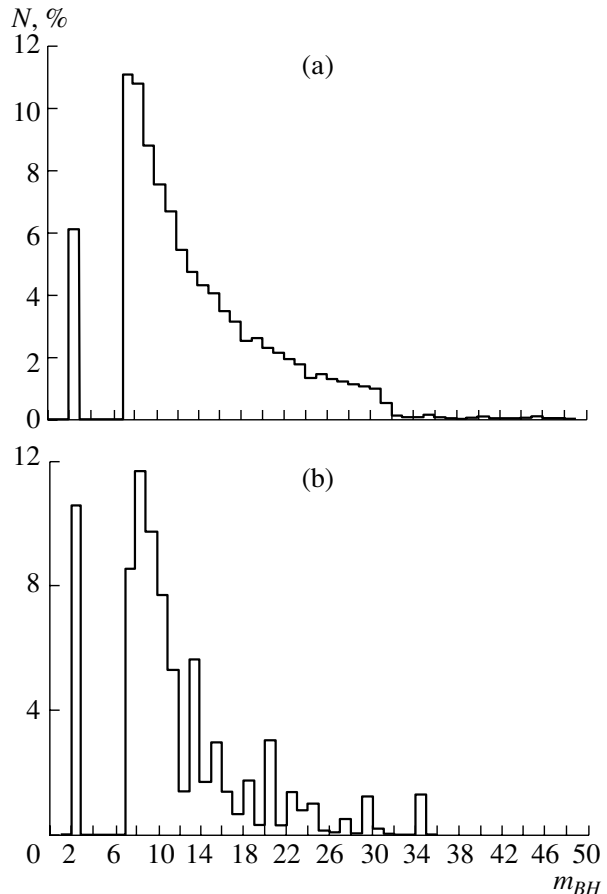
Fig. 4. (a) Initial model mass spectrum for black holes in BH + opt binaries for scenario A ($\alpha_q = 0$ and $k_{BH} = 0.43$; see the text for more details). (b) The same spectrum after taking into account observational selection effect due to the difference in the lifetimes of optical stars.

the spectrum. These low-mass black holes are produced by collapses of neutron stars whose masses grow to the Oppenheimer−Volkoff limit [8].

The overwhelming majority of black holes ($\sim$99.99% of the total number) do not increase their mass in the course of their evolution due to the accretion of matter from their optical companion. Only a negligible fraction ($\sim$0.01% of the total number) increase their mass by $\Delta m \simeq 1\,M_\odot$. For this reason, the shape of the black-hole mass spectrum in model A is the same before and after the accretion stage (therefore, we do not show the latter here). Figure 4b shows the black-hole mass spectrum in the final stage of the evolution of the optical component, corrected for selection effects due to differences in the lifetimes of different binaries in the BH + opt stage. These were taken into account using the formula

$$N(m_k) = \sum_{j=0}^{n_k} t_j \Big/ \sum_{i=0}^{N} t_i, \qquad (12)$$

Table 3. Observational parameters derived via the population syntheses for evolutionary scenarios A, C, and W

Observational parameter	Observational data*	A	C	Wc
Number of Cyg X-1-type systems in the Galaxy	1	≈ 1	≈ 1	≈ 1
$\dfrac{N(\text{BH} + \text{PSR})}{N(\text{PSR})}$	$< \dfrac{1}{1500}$	$\approx \dfrac{1}{1500}$	$\approx \dfrac{1.3}{1500}$	$\approx \dfrac{1.5}{1500}$
$R = \dfrac{N(m_{BH} < 4.0 M_\odot)}{N(m_{BH} \geq 4.0 M_\odot)}$	~ 0.1	≈ 0.15	≈ 0.03	≈ 0.03

* See the text for more details.

where $N(m_k)$ is the number of black holes with masses in the range $m_k + dm_k$ ($m_k = 1, 2, 3, 4 \ldots M_\odot$, and $dm_k = 1 M_\odot$), n_k the number of black holes in the bin for mass m_k, t_j the lifetime of a binary with a black-hole of mass in the interval $m_k + dm_k$, N the total number of model tracks ($N = 10^6$), and t_i the lifetime of a binary with a black hole. We can see that, after correcting for differences in the stellar lifetimes in the BH + opt stage, the bimodal structure of the mass spectrum

$N, \%$

(a)

(b)

m_{BH}

Fig. 5. Same as Fig. 4 for scenario C ($\alpha_q = 0$ and $k_{BH} = 0.57$).

becomes even more pronounced (Fig. 4). Thus, the population-synthesis results for scenario A suggest that the masses of the vast majority of black holes should be confined to $8-15 M_\odot$. However, the model computations confirm the possible formation of low-mass ($m_{BH} = 2-4 M_\odot$) black holes in binary systems. The number of binaries with low-mass black holes is fairly high: there should exist one low-mass ($m_{BH} \leq 4 M_\odot$) black-hole system for every ~ 7 binaries with high-mass ($m_{BH} > 4 M_\odot$) black holes. In scenario A, low-mass black holes are produced only by accretion-induced collapse, and they comprise $\sim 15\%$ of the total number of black holes in binaries. The numerical criteria characterizing the consistency of model A with the observational data are presented in Table 3.

7. BLACK-HOLE MASS SPECTRA IN MODEL C

Recall that the value of k_{BH} in evolutionary scenario C was set to 0.57, and that the minimum black-hole mass was set to $7 M_\odot$. The shape of the initial black-hole mass spectrum is shown in Fig. 5a. In contrast to scenario A, the scenario C spectrum has three peaks (Fig. 5a). Black holes with masses $2-4 M_\odot$ are formed only via the accretion-induced collapse of neutron stars. In contrast to scenario A, the fraction of low-mass black holes in scenario C is $\sim 5\%$ of the total number of all black holes that are formed in binary systems.

Figure 5a shows that, in scenario C, most ($\sim 54.5\%$) black holes have masses of $7-9 M_\odot$ at the time of formation of the BH + opt system. All black holes with masses of $7-9 M_\odot$ descend from Wolf−Rayet stars. Black holes with masses of $10-12 M_\odot$ comprise $\sim 40.5\%$ of the total number of black holes (Fig. 5a) and are produced by collapses of presupernovae that avoided the Wolf−Rayet stage.

Due to the high mass loss in the main-sequence, supergiant, and Wolf−Rayet stages (up to 50% of the initial mass), no very massive black holes are produced in scenario C. Thus, the maximum mass of black holes in this scenario is $12 M_\odot$ (Fig. 5).
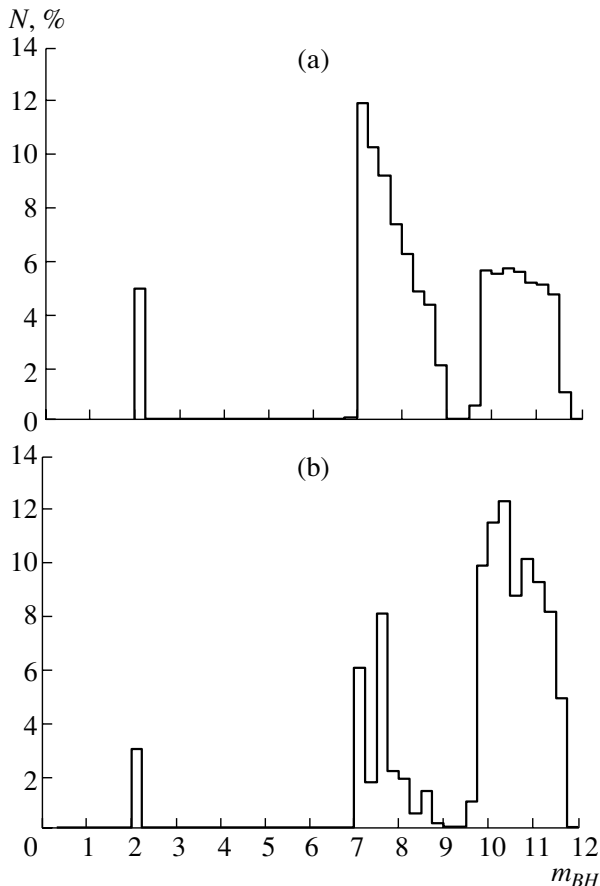
As in scenario A, only $\sim$0.01% of black holes increase their mass by $\Delta M \simeq 1\,M_\odot$ during the evolution of the binary system. The masses of most black holes remain unchanged $\sim$99.99% or increased by only $\Delta m \simeq 0.01\,M_\odot$ due to accretion. For this reason we do not show the mass spectrum of the post-accretion black holes. Figure 5b shows the black-hole mass spectrum at the end of the evolution of the optical component, corrected for differences in the lifetimes of binaries in the BH + opt stage using formula (12) with $dm_k = 0.25\,M_\odot$.

Selection effects due to differences in the lifetimes of binaries in the BH + opt stage had an appreciable influence on the initial black-hole mass distribution (Figs. 5a, 5b). In scenario C, black holes with masses exceeding 12 $M_\odot$ should not be observed, and a considerable fraction of black holes ($\sim$76%) should have masses $\sim$10$-$12 $M_\odot$. Black holes with masses $\sim$7$-$9 $M_\odot$ comprise $\sim$21% of the total number of observed black holes. Note that there are few low-mass black holes in the "observed" distribution of black holes produced in scenario C ($\sim$3% of the total number of black holes). In the total sample, the ratio of the numbers of low- and high-mass black holes is $R \simeq 0.03$. The numerical criteria characterizing the consistency of model C with the observational data are presented in Table 3.

8. BLACK-HOLE MASS SPECTRUM IN SCENARIO W

Scenario Wb did not satisfy the observational criteria: not a single Cyg X-1-like system was produced in the model Galaxy (10^{11} stars). For this reason, we considered scenario Wb to be unrealistic and did not analyze it further. Scenario Wc fit the required observational criteria sufficiently well (Table 3), and we, accordingly, carried out population-synthesis computations for this scenario.

As we noted above, scenario Wc is based on the relation of [41] between the mass of the optical star m_{opt} and the mass of the compact remnant, as well as a dependence of the mass loss on m_{opt} similar to that in scenario C. The maximum black-hole mass in the model of Woosley [41, Fig. 16] did not exceed $\sim$11 $M_\odot$. This model assumed that black holes are produced by stars with initial masses greater than 20.7 $M_\odot$. The initial distribution of the component-mass ratios was taken to be quadratic: $\alpha_q = 2$. The hypothesis of a flat initial mass-ratio distribution was rejected based on the observational criteria: the number of Cyg X-1-like systems was far greater than one per model Galaxy.

The initial black-hole mass spectrum obtained for evolutionary scenario Wc is shown in Fig. 6a. The masses of $\sim$10% of the produced black holes are in the
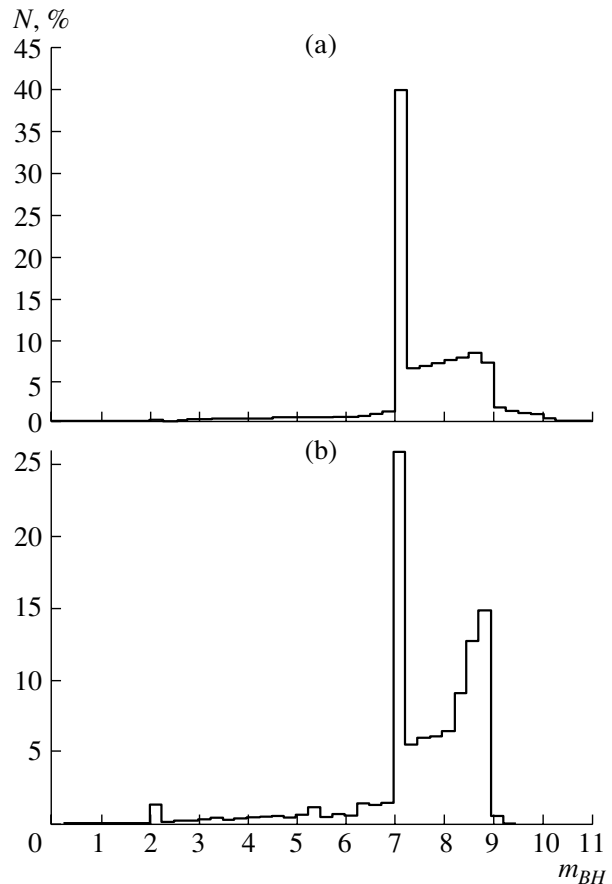

Fig. 6. Same as Fig. 4 for scenario Wc with $\alpha_q = 2$ and with k_{BH} values that depend on the initial mass of the star.

range 2$-$7 $M_\odot$, while about 45% of the black holes have masses of 7.1$-$7.5 $M_\odot$. The fraction of black holes with masses exceeding 7.5 $M_\odot$ is also $\sim$45%.

The peak of the histogram at $\sim$7 $M_\odot$ is due to the nature of the relation between the mass of the optical star m_{opt} and the mass of the black hole produced by this star [41, Fig. 16]. Stars with masses of 20.7$-$27 $M_\odot$ produced low-mass black holes with $m_{BH} \simeq 2-6\,M_\odot$. Due to the low mass of the nascent black holes and the considerable mass loss in the supernova explosions, the vast majority of binaries could not remain bound after the collapse. This explains the deficit of black holes with masses of $m_{BH} \simeq 2-6\,M_\odot$.

According to the evolutionary scheme of [41], optical stars with masses of 27$-$84 $M_\odot$ are the progenitors of $\sim$7 $M_\odot$ black holes. This feature of the scheme results in the peak in the initial black-hole mass distribution in the W model (Fig. 6a). Black holes with masses $m_{BH} > 7.5\,M_\odot$ are produced by optical stars with initial masses of $\gtrsim$84 $M_\odot$. We do not present the post-accretion black-hole mass spectrum, since it is very similar to the initial mass spectrum.

Figure 6b shows the black-hole mass spectrum after correcting for differences in the lifetimes of the BH + opt systems using formula (12) with $dm_k = 0.25\,M_\odot$. We can see that this selection effect strongly influences the mass distribution of the black holes. The group of low-mass black holes ($m_{BH} < 4\,M_\odot$) in BH + opt systems becomes much more prominent: such black holes comprise $\sim$3% of all "observed" black holes in scenario W (see Fig. 6b). About 2% of the total number of low-mass black holes in binary systems are produced by accretion-induced collapses of neutron stars in scenario Wc (Fig. 6). At the same time, all the low-mass ($m_{BH} \leq 4\,M_\odot$) black holes in scenarios A and C were produced by accretion-induced collapses of neutron stars.

The maximum of the black-hole mass distribution remained at $\sim$7 $M_\odot$, but the total number around this peak was reduced to $\sim$25% of the total number of black holes in BH + opt binaries. The fraction of black holes with masses $m_{BH} \simeq 7.5-9\,M_\odot$ in BH + opt binaries increased to $\sim$60% (Fig. 6b). Note that the maximum mass of an "observed" black hole in a pair with an optical companion in scenario W is $m_{BH} \simeq 9\,M_\odot$. The numerical criteria characterizing the consistency of the population-synthesis results for model Wc with the observational data are presented in Table 3.

9. CONCLUSION

The formation of BH + opt binaries with low-mass black holes ($m_{BH} = 2-4\,M_\odot$) in all three evolutionary scenarios (A, C, and W) is one of the most important results of our computations. The population-synthesis results indicate that $\sim$3−15% of the total number of black holes in binary systems with optical companions in scenarios A, C, and W are low-mass black holes. In other words, according to our computations, the number of low-mass black holes ($m_{BH} = 2-4\,M_\odot$) in binaries with optical companions is sufficiently high to enable their detection in the near future. It is not ruled out that the objects GRS 1009-45 and GRO J0422+32 (Table 1), which contain compact objects with estimated masses $m_{BH} = 2-4\,M_\odot$ (within current levels of accuracy), are black holes produced by accretion-induced collapse.

According to the population-synthesis results, accretion plays a negligible role in the formation of the mass spectrum of black holes in BH + opt systems. On the other hand, selection effects related to the lifetimes of binary systems considerably distort the intrinsic shape of the mass spectrum (Figs. 4−6).

The poor statistics of black-hole mass estimates and limited accuracy of these estimates prevent us from identifying one of the scenarios as being "correct." With the current accuracy of the observed black-hole mass spectrum, all three scenarios (A, C, and W) satisfy both the observational criteria (Table 1) and the observed mass spectrum (Fig. 1).

The population-synthesis results for scenarios A and W suggest that the masses of the vast majority of black holes that will be discovered in the future will be in the range 7−9 $M_\odot$. If scenario C is the most realistic, the masses of most "new" black holes will be 10−12 $M_\odot$. According to scenario A, black holes with masses up to $\sim$50 $M_\odot$ may be present in binary systems.

To conclude, we note the recent discovery of the compact object 2S 0921-630, with an estimated mass of 1.9−2.9 $M_\odot$ [53]. The nature of this object is not yet known, but, given our population-synthesis results, we cannot rule out the possibility that 2S 0921-630 is a low-mass black hole. Recall that, according to our computations, 3−15% of all the black holes in binaries may be low-mass black holes.

We stress that, according to the population-synthesis computations, the overwhelming majority of low-mass black holes (2−4 $M_\odot$) are formed via accretion-induced collapses of neutron stars. Therefore, we expect that the masses of a high fraction of low-mass black holes should be very close to the Oppenheimer−Volkoff limit ($\sim$2.5−3 $M_\odot$).

REFERENCES

1. A. M. Cherepashchuk, Usp. Fiz. Nauk **166**, 809 (1996) [Phys. Usp. **39**, 759 (1996)].
2. A. M. Cherepashchuk, Usp. Fiz. Nauk **173**, 345 (2003) [Phys. Usp. **46**, 335 (2003)].
3. A. M. Cherepashchuk, Astron. Zh. **78**, 145 (2001) [Astron. Rep. **45**, 120 (2001)].
4. A. V. Tutukov and L. R. Yungel'son, Nauchni Inform. Astrosovet Akad. Nauk SSSR **27**, 58 (1973).
5. E. P. J. van den Heuvel and J. Heise, Nature Phys. Sci. **239**, 67 (1972).
6. L. R. Yungel'son and A. M. Cherepashchuk, Astron. Zh. **80**, 419 (2003) [Astron. Rep. **47**, 386 (2003)].
7. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, *Astrophysics and Space Physics Reviews*, Ed. by R. A. Sunyaev (Harwood, Amsterdam, 1996), Vol. 9.

8. A. I. Bogomazov, M. K. Abubekerov, V. M. Lipunov, and A. M. Cherepashchuk, Astron. Zh. **82**, 331 (2005) [Astron. Rep. **49**, 295 (2005)].

9. D. P. Bennett, A. C. Becker, J. L. Quinn, *et al.*, Astrophys. J. **579**, 639 (2002).

10. M. K. Abubekerov, E. A. Antokhina, and A. M. Cherepashchuk, Astron. Zh. **81**, 606 (2004) [Astron. Rep. **48**, 550 (2004)].

11. J. A. Orosz, in *IAU Symposium No. 212: A Massive Star Odyssey: From Main Sequence to Supernova*, Ed. by K. van der Hucht, A. Herrero, and E. Cesar (Astron. Soc. Pac., San Francisco, 2003), p. 365; astro-ph/0209041.

12. M. van der Klis, M. Clausen, K. Jensen, *et al.*, Astron. Astrophys. **151**, 322 (1985).

13. A. P. Cowley, D. Crampton, J. B. Hutchings, *et al.*, Astrophys. J. **272**, 118 (1983).

14. L. Kuiper, M. van der Klis, and J. van Paradijs, Astron. Astrophys. **203**, 79 (1988).

15. D. R. Gies, W. Huang, and M. V. McSwain, Astrophys. J. **578**, L67 (2002).

16. D. M. Gelino, T. E. Harrison, and J. A. Orosz, Astron. J. **122**, 2668 (2001).

17. A. M. Cherepashchuk, N. V. Borisov, M. K. Abubekerov, *et al.*, Astron. Zh. **81**, 1119 (2004) [Astron. Rep. **48**, 1019 (2004)].

18. D. M. Gelino, T. E. Harrison, and B. J. McNamara, Astron. J. **122**, 971 (2001).

19. I. Baraffe, G. Chabrier, F. Allard, and P. H. Hauschildt, Astron. Astrophys. **337**, 403 (1998).

20. D. M. Greene, C. D. Bailyn, and J. A. Orosz, Astrophys. J. **554**, 1290 (2001).

21. R. A. Remillard, J. A. Orosz, J. E. McClintock, and J. E. C. D. Bailyn, Astrophys. J. **459**, 226 (1996).

22. E. T. Harlaftis, D. Steeghs, K. Horne, and J. E. A. V. Filippenko, Astron. J. **114**, 1170 (1997).

23. R. M. Wagner, C. B. Foltz, T. Shahbaz, *et al.*, Astrophys. J. **556**, 42 (2001).

24. J. A. Orosz, P. J. Groot, M. van der Klis, *et al.*, Astrophys. J. **568**, 845 (2002).

25. A. V. Filippenko and R. Chornock, IAU Circ., No. 7644, 2 (2001).

26. R. I. Hynes, D. Steeghs, J. Casares, *et al.*, Astrophys. J. **583**, L95 (2003).

27. J. Homan, M. Klein-Wolt, S. Rossi, *et al.*, Astrophys. J. **586**, 1262 (2003).

28. C. D. Bailyn, R. K. Jain, P. Coppi, and J. A. Orosz, Astrophys. J. **499**, 367 (1998).

29. A. M. Cherepashchuk, in *Modern Problems in Stellar Evolution,* Ed. by D. S. Wiebe (GEOS, Moscow, 1998), p. 198.

30. M. K. Abubekerov, E. A. Antokhina, and A. M. Cherepashchuk, Astron. Zh. **81**, 108 (2004) [Astron. Rep. **48**, 89 (2004)].

31. M. K. Abubekerov, Astron. Zh. **81**, 714 (2004) [Astron. Rep. **48**, 649 (2004)].

32. D. J. Nice, E. M. Splaver, and I. H. Stairs, astro-ph/0411207.

33. M. K. Abubekerov and A. M. Cherepashchuk, Astrofiz. (2005) (in press).

34. A. V. Filippenko, D. C. Leonard, T. Matheson, W. Li, E. S. Moran, and A. G. Riess, Publ. Astron. Soc. Pac. **111**, 969 (1999).

35. D. M. Gelino and T. E. Harrison, Astrophys. J. **599**, 1254 (2003).

36. A. M. Cherepashchuk, Space Sci. Rev. **93**, 473 (2000).

37. J. van Paradijs and J. E. McClintock, *X-ray Binaries*, Ed. by W. H. G. Lewin, J. van Paradijs, and E. P. J. van den Heuvel (Cambridge Univ. Press, Cambridge, 1993).

38. V. M. Lipunov, A. I. Bogomazov, and M. K. Abubekerov, Mon. Not. R. Astron. Soc. (2005) (in press).

39. Z. T. Kraĭcheva, E. I. Popova, A. V. Tutukov, and L. R. Yungel'son, Astron. Zh. **56**, 520 (1979) [Sov. Astron. **23**, 290 (1979)].

40. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Mon. Not. R. Astron. Soc. **288**, 245 (1997).

41. S. E. Woosley, A. Heger, and T. A. Weaver, Rev. Mod. Phys. **74**, 1015 (2002).

42. A. Heger *et al.*, astro-ph/0211062.

43. C. Jager, *The Brightest Stars* (Reidel, Dordrecht, 1980).

44. H. J. G. L. M. Lamers, Astrophys. J. **245**, 593 (1981).

45. B. P. Kudritzki and D. Reimers, Astron. Astrophys. **70**, 227 (1978).

46. V. I. Varshavskii and A. V. Tutukov, Astron. Zh. **52**, 227 (1975) [Sov. Astron. **19**, 142 (1975)].

47. I. Iben, Jr. and A. V. Tutukov, Astrophys. J., Suppl. Ser. **58**, 661 (1985).

48. I. Iben, Jr. and A. V. Tutukov, Astrophys. J. **313**, 727 (1987).

49. G. Schaller, D. Schaerer, G. Meynet, and A. Maeder, Astron. Astrophys., Suppl. Ser. **96**, 269 (1992).

50. D. Vanbeveren, New Astron. **3**, 443 (1998).

51. H. Nieuwenhuijzen and C. de Jager, Astron. Astrophys. **231**, 134 (1990).

52. W. R. Hamman and L. Koesterke, Astron. Astrophys. **335**, 1003 (1998).

53. P. G. Jonker, D. Steegh, G. Nelemans, and M. van der Klis, Mon. Not. R. Astron. Soc. **356**, 621 (2005).

Translated by L. Yungel'son

The Phase Matrix for Scattering in a Weak Magnetic Field. The Hanle Effect

D. N. Rachkovskiĭ

Crimean Astrophysical Observatory, Nauchnyĭ, Crimea, Ukraine
Received January 23, 2005; in final form, February 17, 2005

Abstract—A new formulation for the phase matrix for the scattering of radiation in a weak magnetic field is presented. The phase matrix is represented as a product of individual matrices that depend on the angle of incidence and the scattering angle. It is shown that the Hanle effect is absent in observations of scattered light in the solar atmosphere when the magnetic field is perpendicular to the atmosphere. In strong magnetic fields, the phase matrix depends only on the direction of the magnetic field relative to the basic coordinate system. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

In [1], we developed an algorithm for calculating the phase matrix for the scattering of radiation in a weak magnetic field. The analysis was based on the quantum-mechanical formula for the cross section for resonance scattering [2]. The presence of a magnetic field was taken into account by introducing magnetic splitting in the frequency factor for each possible transition between sublevels of a spectral line.

Analogous approaches to this problem are presented by Stenflo [3, 4], House [5–7], and others. We showed in [1] that, when calculating the phase matrix using this method, it is necessary to obey a certain sign rule for the various intraline transitions.

Omont *et al.* [8, 9] developed a rigorous quantum-mechanical theory for the phase matrix for scattering in a magnetic field. Landi Degl'Innocenti [10–13] and M. Landi Degl'Innocenti and E. Landi Degl'Innocenti [14] applied this theory to calculations of various types of phase matrices.

We also adopt the phase-matrix theory of Omont *et al.* [8, 9] as a basis for the present study. We have obtained a new representation of the phase matrix that is fully equivalent to the form given in [14]. Our formulation of the phase matrix has a block structure that partially separates the angle of incidence and the scattering angle. From a theoretical point of view, this representation of the phase matrix is preferable over the form given in [14]. In particular, it becomes evident that the polarization of the scattered radiation when the magnetic field is normal to the plane-parallel surface layers of an atmosphere is the same as the polarization in the absence of a magnetic field. This conclusion can be drawn in the framework of the theory of radiative transfer in a homogeneous medium.

We show that, when the magnetic field is strong, the phase matrix is independent of the field strength but depends on the field orientation. A similar conclusion is drawn in [13], but that work presents an expression for the phase matrix in a coordinate system that is determined by the magnetic field.

Here, we restrict our analysis to case A in the classification of Omont *et al.* [9]: the scattering of light averaged over the absorption and emission frequencies.

2. VECTOR REPRESENTATION OF THE PHASE MATRIX

In [8, 9], the relationship between the incident and scattered radiation is essentially determined by two equations:

$$\rho_Q^K = \sum_{n=0}^{3} B_{KQ,n}(\theta, \phi) S_n / \sqrt{2}, \qquad (1)$$

$$S_n / \sqrt{2} = \sum_{K,Q} B_{KQ,n}^*(\theta, \phi) \rho_Q^K. \qquad (2)$$

Here, K, Q can take the values $K = 0 \ldots 2, Q = -K \ldots K$, and S_n denotes a Stokes parameter, $S_0 = I_0, S_1 = Q, S_2 = U, S_3 = V$.

The functions $B_{KQ,n}(\theta, \phi)$ were derived in [9] and are tabulated there in the form of functions of the coordinates θ, φ. For our purposes, it is more convenient to express them in terms of the functions $P_Q^K(\mathbf{e}_1, \mathbf{e}_2)$,

which were also derived in [9]:

$$B_{KQ,0}(\theta,\phi) = (P_Q^K(\mathbf{e}_1,\mathbf{e}_1) + P_Q^K(\mathbf{e}_2,\mathbf{e}_2))/\sqrt{2}, \quad (3)$$

$$B_{KQ,1}(\theta,\phi) = (P_Q^K(\mathbf{e}_1,\mathbf{e}_1) - P_Q^K(\mathbf{e}_2,\mathbf{e}_2))/\sqrt{2},$$

$$B_{KQ,2}(\theta,\phi) = (P_Q^K(\mathbf{e}_1,\mathbf{e}_2) + P_Q^K(\mathbf{e}_1,\mathbf{e}_2))/\sqrt{2},$$

$$B_{KQ,3}(\theta,\phi) = i(P_Q^K(\mathbf{e}_1,\mathbf{e}_2) - P_Q^K(\mathbf{e}_2,\mathbf{e}_1))/\sqrt{2}.$$

In (3), we chose $\mathbf{e}_1 = \mathbf{x}, \mathbf{e}_2 = \mathbf{y}$, where $\mathbf{x}, \mathbf{y}$ are unit vectors perpendicular to the direction of propagation of the incident or scattered beam.

The functions P_Q^K are given by

$$P_Q^K(\mathbf{e}_1,\mathbf{e}_2) = -\sqrt{(2K+1)}(-1)^Q \quad (4)$$

$$\times \sum_{q,q'} \begin{pmatrix} 1 & 1 & K \\ q & q' & -Q \end{pmatrix} (e_1)_q (e_2)_{q'}.$$

Here, the subscripts q, q' can take the values $\pm 1, 0$. The quantities $(e_{1,2})_q$ are projections of the vectors $\mathbf{e}_{1,2}$ onto the vectors $\mathbf{v}_q$:

$$\mathbf{v}_{\pm 1} = (\mp\mathbf{i} - i\mathbf{j})/\sqrt{2}, \quad \mathbf{v}_0 = \mathbf{k}. \quad (5)$$

In (5), $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ are unit vectors in the basic coordinate system. The values of the $3j$ symbols in front of the summation sign in (4) can be found, for example, in [15].

To write the previous relations in matrix form, we introduce the vectors

$$\mathbf{P}_Q^K(\mathbf{R}) = \begin{pmatrix} P_Q^K(\mathbf{e}_1,\mathbf{e}_1) \\ P_Q^K(\mathbf{e}_2,\mathbf{e}_2) \\ P_Q^K(\mathbf{e}_1,\mathbf{e}_2) \\ P_Q^K(\mathbf{e}_2,\mathbf{e}_1) \end{pmatrix}, \quad \mathbf{I} = \begin{pmatrix} I_0 \\ Q \\ U \\ V \end{pmatrix}. \quad (6)$$

Here, $\mathbf{R}$ is a vector characterizing the direction of the radiation. Using the matrix $\mathbf{T}$,

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \end{pmatrix} \quad (7)$$

relations (1), (2) can be written in matrix form:

$$\rho_Q^K = \tilde{\mathbf{P}}_Q^K(\mathbf{R}) \cdot \mathbf{T} \cdot \mathbf{I}/2, \quad (8)$$

$$\mathbf{I} = \sum_{K,Q} \tilde{\mathbf{T}}^* \cdot \mathbf{P}_Q^{K*}(\mathbf{R}) \cdot \rho_Q^K.$$

Let us now turn to the derivation of a matrix formulation of the phase matrix for scattering in a weak magnetic field. Omont *et al.* [9] derived the following

equation for the case of averaging over the frequencies of the incident (ω_1) and scattered (ω_2) radiation:

$$\rho_Q^K(\omega_2) \sim W(J_f, J_i, J_e, K) f_A^{KQ} \rho_Q^K(\omega_1). \quad (9)$$

We can use (8) to write this equation in matrix form:

$$\mathbf{I}_s = \sum_{KQ} W_K \cdot \tilde{\mathbf{T}}^* \cdot \mathbf{P}_{-Q}^{K*}(\mathbf{R}_s) \quad (10)$$

$$\times f_A^{KQ} \cdot \tilde{\mathbf{P}}_{-Q}^K(\mathbf{R}_f) \cdot \mathbf{T} \cdot \mathbf{I}_f.$$

Let us now determine certain coefficients in this expression. $W_K = W(J_f, J_i, J_e, K)$ is the $6j$ symbol, which is given, for example, in tabular form in [12]. However, these quantities can also be expressed in terms of the Chandrasekhar coefficients E_1 and E_3 [16], namely:

$$W_0 = 1, \quad W_1 = W(J_f, J_i, J_e, 1) = E_3, \quad (11)$$

$$W_2 = W(J_f, J_i, J_e, 2) = E_1.$$

In (11), J_i, J_e, J_f are the principle quantum numbers for the initial, intermediate, and final levels, respectively.

The averaged frequency factor f_A^{KQ} for case A [9] is

$$f_A^{KQ} \sim (1 + i\gamma_Q)^{-1}, \quad \gamma_Q = Q\frac{\omega_B}{\gamma_e^K}. \quad (12)$$

Here, γ_e^K is the decay constant for the upper level, which we will assume to be independent of K. The quantity ω_B is

$$\omega_B = 2\pi 1.4 \times 10^6 g_e B, \quad (13)$$

where g_e is the Landé factor of the upper level and B is the magnetic-field strength.

Expression (12) can be written

$$f_A^{KQ} = \cos A_Q e^{-iA_Q}, \quad (14)$$

$$\sin A_Q = \gamma_Q/\sqrt{(1 + \gamma_Q^2)}.$$

3. TRANSFORMATION OF THE PHASE MATRIX INTO AN EXPLICIT DEPENDENCE ON THE COORDINATES OF THE INCIDENT AND SCATTERED RADIATION (WHEN THE MAGNETIC FIELD IS PARALLEL TO THE Z AXIS)

Expression (10) defines the phase matrix. We can use this to obtain a more explicit relation that will enable us relatively easily to make a change of coordinates and see more explicitly the dependence of the phase matrix on the coordinates of the incident and scattered beams.

Angular characteristics of the magnetic-field vector $(\mathbf{B}(\theta_b, 0))$ and the vectors for the incident $(\mathbf{F}(\theta_f, \chi_f))$ and scattered $(\mathbf{S}(\theta_s, \chi_s))$ radiation. The vectors $\mathbf{x}_f, \mathbf{y}_f$ and $\mathbf{x}_s, \mathbf{y}_s$ lie in planes perpendicular to the vectors $\mathbf{F}, \mathbf{S}$, and determine the polarization of the radiation.

The figure illustrates the relative positions of the magnetic-field vector $\mathbf{B}$ and the vectors for the incident and scattered radiation $\mathbf{R}_f = \mathbf{F}$ and $\mathbf{R}_s = \mathbf{S}$.

We will denote the polar coordinates (θ, χ) with the superscript b if the Z axis is directed along the vector $\mathbf{B}$: $(\theta_{f,s}^b, \chi_{f,s}^b)$. The subscripts (f, s) will everywhere denote the parameters of the incident or scattered radiation.

We can obtain for the components of the vectors $\mathbf{e}_{1,q}^{(f,s)}, \mathbf{e}_{2,q}^{(f,s)}$, i.e., the unit vectors characterizing the Stokes parameters of the incident or scattered radiation,

$$e_{1,+1}^{(n)} = -\cos\theta_n^b e^{i\chi_n^b}/\sqrt{2}, \qquad (15)$$

$$e_{1,-1}^{(n)} = \cos\theta_n^b e^{-i\chi_n^b}/\sqrt{2},$$

$$e_{1,0}^{(n)} = -\sin\theta_n^b, \quad e_{2,+1}^{(n)} = -ie^{i\chi_n^b}/\sqrt{2},$$

$$e_{2,-1}^{(n)} = -ie^{-i\chi_n^b}/\sqrt{2}, e_{2,0}^{(n)} = 0, \quad n = f, s.$$

Further, we consider all possible forms of the product $\mathbf{P}_{-Q}^{K*}(\mathbf{S}) \cdot \tilde{\mathbf{P}}_{-Q}^{K}(\mathbf{F})$ with all possible values of K, Q, which we then sum.

As a result, writing the three first Stokes parameters as a single vector, $\tilde{\mathbf{I}} = (I_0, Q, U)$, we obtain for

the phase matrix

$$\mathbf{P}(\theta_s^b \chi_s^b, \theta_f^b \chi_f^b) \qquad (16)$$

$$= \sum_{m=0}^{2} \mathbf{H}_m(\theta_s^b) \cdot \mathbf{D}_m(\chi_s^b, \chi_f^b, \mathbf{B}) \cdot \mathbf{H}_m(\theta_f^b) + \delta_{1,1}.$$

Here, $\mathbf{H}_m(\theta_{f,s}^b)$ is a diagonal matrix whose elements can be presented as follows for the sake of brevity:

$$\mathbf{H}_0(\theta_n) = [3\sin^2\theta_n - 2, 3\sin^2\theta_n, 0], \qquad (17)$$

$$\mathbf{H}_1(\theta_n) = [\sin\theta_n\cos\theta_n, \sin\theta_n\cos\theta_n, 1],$$

$$\mathbf{H}_2(\theta_n) = [\sin^2\theta_n, (1 + \cos^2\theta_n), 2\cos\theta_n],$$

$$n = f, s.$$

The matrices $\mathbf{D}_n(\chi_s^b, \chi_s^b, \mathbf{H})$ can be written

$$\mathbf{D}_0 = \frac{1}{8}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad (18)$$

$$\mathbf{D}_1 = 3W_2\cos A_1 \begin{pmatrix} cs_1 & cs_1 & sn_1 \\ cs_1 & cs_1 & sn_1 \\ -sn_1 & -sn_1 & cs_1 \end{pmatrix},$$

$$\mathbf{D}_2 = \frac{3}{8}W_2\cos A_2 \begin{pmatrix} cs_2 & -cs_2 & -sn_2 \\ -cs_2 & cs_2 & sn_2 \\ sn_2 & -sn_2 & 0 \end{pmatrix}.$$

Here,

$$cs_n = \cos\left[n(\chi_s^b - \chi_f^b) - A_n\right], \qquad (19)$$

$$sn_n = \sin\left[n(\chi_s^b - \chi_f^b) - A_n\right],$$

$$A_n = \arcsin\frac{n\gamma_n}{\sqrt{1 + n^2\gamma_n^2}}, \qquad n = 1, 2.$$

We have for the Stokes parameter V

$$V_s(\theta_s, \chi_s) = \frac{3}{2}W_1(\cos\theta_s^b\cos\theta_f^b \qquad (20)$$

$$+ \cos A_1 \sin\theta_s^b \sin\theta_f^b cs_1)V_f(\theta_f, \chi_f).$$

4. REPRESENTATION OF THE PHASE MATRIX IN AN ARBITRARY COORDINATE SYSTEM

Relations (16), (20) are valid for a coordinate system in which the Z axis is oriented along the magnetic field $\mathbf{B}$. We will now obtain corresponding expressions for an arbitrary coordinate system XYZ.

We denote the coordinates of the vector $\mathbf{B}$ in the XYZ system θ_b, χ_b. We find by considering the spherical triangles in the figure that

$$\cos\theta_m^b = \cos\theta_b \cos\theta_m + \sin\theta_b \sin\theta_m \cos\chi_m, \quad (21)$$

$$\sin\theta_m^b = \sqrt{1 - \cos\theta_m^b} \quad (\theta_m^b \leq \pi),$$

$$\cos\chi_m^b = -(\cos\theta_m - \cos\theta_b \cos\theta_m^b)/(\sin\theta_b \sin\theta_m^b),$$

$$\sin\chi_m^b = \sin\theta_m \sin\chi_m / \sin\theta_m^b,$$

$$m = f, s.$$

We will assume that the vector $\mathbf{e}_1$ determining the Stokes parameters is directed along a meridian of the coordinate system and that the vector $\mathbf{e}_2$ is directed in the direction of latitude. Therefore, to obtain the phase matrix in the XYZ coordinate system, in addition to expressing the coordinates θ_m^b, χ_m^b in terms of θ_m, χ_m [relations (21)], we must carry out a rotation of the reference vectors $\mathbf{e}_1, \mathbf{e}_2$ for the Stokes parameters of the incident and scattered radiation.

The transformation of the Stokes parameters $\mathbf{I}_1 = (I_0, Q, U)$ to some other system $\mathbf{I}_2$ with a rotation by some angle F that is clockwise for an observer looking opposite to the direction of propagation of the beam is realized as follows:

$$\mathbf{I}_2 = \mathbf{L}(F)\mathbf{I}_1,$$

where

$$\mathbf{L}(F) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos 2F & -\sin 2F \\ 0 & \sin 2F & \cos 2F \end{pmatrix}. \quad (22)$$

We conclude from examining the figure that the reference unit vectors for the incident radiation must be rotated by the angle F_1, while those for the scattered radiation must be rotated by the angle S_1. We can make this transformation clearer using the Stokes parameters of the scattered radiation as an example. In the coordinate system in which the vector $\mathbf{B}$ is directed along the Z axis, the unit vector $\mathbf{e}_1$ is directed along the arc BS, and the unit vector $\mathbf{e}_1$ in the XYZ system should be directed along the arc ZS. Consequently, the Stokes parameters $\mathbf{I}_s^{xyz}$ in the XYZ system and in the system in which $\mathbf{B}$ is parallel to the Z axis are related by the expression

$$\mathbf{I}_s^{xyz} = \mathbf{L}(S_1)\mathbf{I}_s^b.$$

Thus, the phase matrix and the Stokes parameters I_0, Q, U will have the form

$$\mathbf{P}(\theta_s\chi_s, \theta_f\chi_f) = \mathbf{L}(S_1) \cdot \mathbf{P}(\theta_s^b\chi_s^b, \theta_f^b\chi_f^b) \cdot \mathbf{L}(-F_1). \quad (23)$$

The angles S_1, F_1 are determined from the formulas

$$\sin M = \sin\theta_b \sin\chi_m / \sin\theta_m^b, \quad (24)$$

$$\cos M = (\cos\theta_b - \cos\theta_m^b \cos\theta_m)/(\sin\theta_m^b \sin\theta_m),$$

$$M = S_1, F_1, \quad m = s, f.$$

When deriving (24), we assumed that $\chi_b = 0$; i.e., the vector $\mathbf{B}$ lies in the ZX plane.

5. PHASE MATRIX IN THE SYSTEM WITH UNIT VECTORS ORIENTED RELATIVE TO THE SCATTERING PLANE

We will now obtain the phase matrix in a coordinate system that is related to the scattering plane. This could be useful when verifying the algorithm, since the phase matrix has a very simple form in the absence of a magnetic field [16]. Let the vector $\mathbf{e}_1$ be directed along the scattering plane and the vector $\mathbf{e}_2$ be perpendicular to this direction (the (l, r) subscripts in the notation of Chandrasekhar [16]). Let θ denote the angle between the incident and scattered beams.

We have from the definition of the angle θ

$$\cos\theta = \cos\theta_s \cos\theta_f + \sin\theta_s \sin\theta_f \cos(\chi_s - \chi_f), \quad (25)$$

$$\sin\theta = \sqrt{1 - \cos\theta^2}.$$

We have for the angles S_2, F_2, determining the transformation matrix for the Stokes parameters

$$\sin S_2 = \sin\theta_f^b \sin(\chi_f^b - \chi_s^b)/\sin\theta, \quad (26)$$

$$\cos S_2 = (\cos\theta_f^b - \cos\theta \cos\theta_s^b)/(\sin\theta_s^b \sin\theta),$$

$$\sin F_2 = \sin\theta_s^b \sin(\chi_f^b - \chi_s^b)/\sin\theta,$$

$$\cos F_2 = -(\cos\theta_s^b - \cos\theta \cos\theta_f^b)/(\sin\theta_f^b \sin\theta).$$

In this case, relation (23) acquires the form

$$\mathbf{P}(\theta_s\chi_s, \theta_f\chi_f) \quad (27)$$

$$= \mathbf{L}(-S_2) \cdot \mathbf{P}(\theta_s^b\chi_s^b, \theta_f^b\chi_f^b) \cdot \mathbf{L}(-(\pi - F_2)).$$

6. CONCLUSION

As one example of the application of this phase-matrix theory, let us consider the formation of a resonance line in the absence of a magnetic-field that is normal to the layers of a plane-parallel atmosphere. Due to the axial symmetry of the problem, the radiation incident onto a certain element of the atmosphere will also be axially symmetrical. Therefore, we can find the radiation scattered by this element by integrating (16) over the angle χ_f. The magnetic field B drops

out of the result. The scattered radiation is given by the expression

$$\mathbf{I}_s \sim \left[\frac{W_2}{8} \begin{pmatrix} 1 - 3\mu^2 \\ 3\mu^2 \end{pmatrix} \right. \quad (28)$$

$$\times \left(1 - 3\mu'^2, 3\mu'^2 \right) + \left. \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right] \mathbf{I}_f,$$

$$U_s = 0, \qquad V_s = \frac{3}{2} W_1 \mu\mu'.$$

Here, $\tilde{\mathbf{I}} = [I_0, Q]$ and $\mu = \cos\theta_s, \mu' = \cos\theta_f$.

Let us write the phase matrix that follows from (28) in the factored form

$$\mathbf{P}_0(\mu, \mu') = \frac{1}{8} \beta(\mu) \tilde{\beta}(\mu'). \quad (29)$$

Here,

$$\beta(\mu) = \begin{pmatrix} \sqrt{W_2}(1 - 3\mu^2) & 2\sqrt{2} & 0 & 0 \\ 3\sqrt{W_2}(1 - \mu^2) & 0 & 0 & 0 \\ 0 & & 0 & 0 & 0 \\ 0 & & 0 & 0 & \sqrt{3W_1}\mu \end{pmatrix}. \quad (30)$$

The magnetic field B is not present in the phase matrix (30), (31); with one exception, this is an ordinary resonance matrix: the Stokes parameter U vanishes during the scattering. Thus, we conclude that the Hanle effect does not affect the Stokes parameter Q when the magnetic field is normal to the surface of the solar atmosphere.

In strong magnetic fields, the magnetic field ceases to influence the polarization of the scattered radiation. This was first pointed out by M. Landi Degl'Innocenti and E. Landi Degl'Innocenti [14]. However, as before, the direction of the magnetic field will determine the polarization. Indeed, in a strong magnetic field, the angles A_1 and A_2 will be approximately equal to $\pi/2$ [see (12), (19)], so that the matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ will not appear in expressions (16) and (23). We thus have for the phase matrix in a strong magnetic field

$$\begin{pmatrix} I_0 \\ Q \end{pmatrix} = \left[\mathbf{L}(S_1) \mathbf{H}_0(\theta_s^b) \mathbf{D}_0 \mathbf{H}_0(\theta_f) \right. \quad (31)$$

$$\times \mathbf{L}(-F_1) + \delta_{1,1} \left] \begin{pmatrix} I_0 \\ Q \end{pmatrix}, \right.$$

$$U_s = 0, \qquad V_s = \frac{3}{2} W_1 \cos(\theta_s^b) \cos(\theta_f^b).$$

We can see that the magnetic field does not appear in this expression. The phase matrix depends only on the directions of the incident and scattered beams and the direction of the magnetic field in the chosen coordinate system. The unit vectors for the Stokes parameters are oriented along the meridians and parallels of the basic coordinate system.

We carried out a series of calculations in order to estimate the lowest value of the magnetic field for which the phase matrix will become independent of the magnetic-field strength. The criterion for this is the quantity γ_n [formula (19)]. When $\gamma_n > 15$, the phase matrix is essentially constant. If the decay constant for the upper level is $\gamma_n = 10^8$, we find $B_{\min} \sim 100$ G. Consequently, in the presence of magnetic fields higher than this value, point-to-point variations of the polarization can be due only to a rotation of the magnetic-field vector and do not depend on the magnitude of the magnetic field.

REFERENCES

1. D. N. Rachkovskiĭ, Astron. Zh. **82** (2005) (in press).
2. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory* (Nauka, Moscow, 1968; Pergamon Press, Oxford, 1971), Part 1.
3. J. O. Stenflo, Astron. Astrophys. **338**, 301 (1998).
4. J. O. Stenflo, *Solar Magnetic Fields—Polarized Radiation Diagnostic* (Kluwer, Dordrecht, 1994).
5. L. L. House, J. Quant. Spectrosc. Radiat. Transf. **10**, 909 (1970).
6. L. L. House, J. Quant. Spectrosc. Radiat. Transf. **10**, 1171 (1970).
7. L. L. House, J. Quant. Spectrosc. Radiat. Transf. **11**, 367 (1971).
8. A. Omont, E. W. Smith, and J. Cooper, Astrophys. J. **175**, 185 (1972).
9. A. Omont, E. W. Smith, and J. Cooper, Astrophys. J. **182**, 283 (1973).
10. E. Landi Degl'Innocenti, Sol. Phys. **85**, 3 (1983).
11. E. Landi Degl'Innocenti, Sol. Phys. **85**, 33 (1983).
12. E. Landi Degl'Innocenti, Sol. Phys. **91**, 1 (1984).
13. E. Landi Degl'Innocenti, Sol. Phys. **102**, 1 (1985).
14. M. Landi Degl'Innocenti and E. Landi Degl'Innocenti, Astron. Astrophys. **192**, 374 (1988).
15. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Fiz.-Mat. Literatura, Moscow, 1963; Pergamon Press, Oxford, 1977).
16. S. Chandrasekhar, *Radiative Transfer* (Clarendon, Oxford, 1950; Inostrannaya Literatura, Moscow, 1953).

Translated by D. Gabuzda

Quasibiennial Oscillations of the North−South Asymmetry

O. G. Badalyan[1], V. N. Obridko[1], J. Rybák[2], and J. Sýkora[2]

[1] *Institute of Terrestrial Magnetism, Ionosphere, and Radiowave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
[2] *Astronomical Institute, Slovak Academy of Sciences, Tatranská Lomnica, 05960 Slovak Republic*
Received December 25, 2004; in final form, February 17, 2005

Abstract—The north−south (N−S) asymmetry of the solar activity (A), which reflects differences in the behavior of the northern and southern hemispheres of the Sun, is studied using data on the brightness of the coronal green line, the total number and area of sunspots, and the net magnetic flux. The spatial and temporal distributions and correlations between the A values represented by these indices are considered. The characteristic time variations in A are similar for all the indices, on both long and short time scales. Quasibiennial oscillations (QBOs) can be traced in the asymmetries of all four indices. A detailed study of the QBOs is carried out based on spectral-variation and wavelet analyses. Long-term increases and decreases occur synchronously in the asymmetries of various indices and are much more pronounced in A than in the indices themselves. A negative correlation between the power of the QBOs and the asymmetry of A can be traced; it is most clearly manifest as a substantial diminishing of the QBOs during the mid-1960s, which coincided with an especially strong increase in A. Our analysis shows that the N−S asymmetry is probably a fundamental property that controls the coupling and degree of coincidence between the magnetic-field-generation mechanisms operating in the northern and southern hemispheres. © *2005 Pleiades Publishing, Inc.*

1. INTRODUCTION

Over the long period in which solar activity has been investigated, the Sun has usually been treated as a single object, without division into hemispheres. It has been assumed that solar processes occur in the same manner in the northern and southern hemispheres. Numerous solar-activity indices have been introduced for the Sun as a whole—the Wolf number, the number of polar faculae, the full radio flux, etc. This approach made it possible to reveal the basic properties of the solar cycle. Nearly synchronous cyclic variations were detected in the behavior of various indices reflecting manifestations of activity at all levels in the solar atmosphere, from the photosphere to corona. Similar cyclic variations were also traced in interplanetary space and in geomagnetic-activity indices. Various theoretical concepts, such as the theory of the differential rotation of the Sun, the dynamo theory, and helioseismology, were developed assuming that both solar hemispheres were identical.

However, it gradually became clear that the northern and southern solar hemispheres do not display quite the same behavior. The time variations in various solar-activity indices demonstrate phase and power mismatches between the two hemispheres over intervals of several months to several years. These differences are manifest in numerous solar-activity indices and are commonly described as the north−south (N−S) asymmetry. The differences represented by the N−S asymmetry are small compared to the indices themselves. It is noteworthy that, for a fairly long time, many investigators were inclined to consider this asymmetry an artifact resulting from observational errors and statistically insignificant fluctuations of measurable quantities. For this reason, even the recent study [1] opened its list of basic conclusions with the statement that the N−S asymmetry of sunspot areas is indeed statistically meaningful and represents a real phenomenon.

Although the asymmetry has been studied for a considerable time, its nature remains unclear, and the identification of new observational facts reflecting various aspects of this interesting phenomenon is still in progress. The sunspot asymmetry has been studied most thoroughly, using sunspot areas and some other sunspot indices over various time intervals (see, e.g., [1−3]). The asymmetries of many other solar-activity indices describing filaments, prominences, unsteady solar phenomena, and geomagnetic activity have also been analyzed. Knaack *et al.* [4] recently considered asymmetry in the solar magnetic field. A fairly detailed review of studies of the asymmetries in various indices can be found in [1, 3, 5, 6].

A number of characteristic features of the asymmetry have been revealed. In particular, a number

of studies have noted that the northern hemisphere was much more active over a long period during the second half of the 20th century. Quasiperiodic oscillations in the wide range of frequencies were detected in the asymmetries of various solar-activity indices (see, e.g., [1, 4, 7, 8]). Furthermore, due to the presence of this asymmetry, the active longitudes do not coincide in the northern and southern hemispheres, and observations even suggest that they are antipodal [9, p. 264].

Some studies have compared the asymmetries in different solar-activity indices [2, 10–12]. These results suggest that the asymmetry behaves similarly in various indices and on various temporal and spatial scales. One of our areas of study is the verification and investigation of this hypothesis.

We consider here the asymmetries in four different solar-activity indices over a relatively long time interval. We have chosen the brightness of the coronal green line I, the total sunspot area S_p, the total sunspot number Q, and the net magnetic flux Φ for this study.

We consider

(1) the spatial and temporal distributions of the N–S asymmetries for various solar-activity indices and correlations between these asymmetries on both short and long time scales;

(2) QBOs in the asymmetry of the activity indices and the spatial and temporal distributions of these QBOs; comparisons of the QBOs in the asymmetries with the QBOs in the activity indices themselves;

(3) the relationship between the power of QBOs in the asymmetry and the magnitude of asymmetry [itself].

Our results show that the N–S asymmetry is a very significant and informative parameter of solar activity. The Conclusion presents a discussion of problems related to this phenomenon.

2. CHARACTERIZATION OF THE CHOSEN ACTIVITY INDICES AND DATABASES

An important property common to the chosen indices is that they all describe the activity of the Sun at a specific point on its surface. Most other indices are either integrated and so related to the entire Sun (the total flux of solar radiation, the total radio flux, etc.) or characterize nonstationary processes (the number, intensity, and coordinates of solar flares and coronal mass ejections), which are beyond the scope of our consideration. Digitized filtergrams of the solar disk are not readily available, and high-resolution radio maps have become available only fairly recently. Another important property of the indices we have chosen is that the corresponding series of observational

data are long. The database of coronal green-line brightness measurements encompasses more than 5.5 activity cycles. Data on sunspot parameters and numbers are also available for this same period. Information on the total magnetic flux reduced to a uniform system is available starting from 1975 and so covers 2.5 cycles.

The brightness of the FeXIV λ 530.3-nm coronal green line, I, is a very convenient index that can be used to study solar activity and its asymmetry in the corona. An important advantage of the index I is that it can be determined virtually simultaneously for all heliographic latitudes. This index provides a uniform and contiguous digital field for all points of the disk over a long time interval based on regular observations. In this respect, it stands out from, for example, the Wolf numbers and sunspot areas (which emerge at low latitudes only) or the numbers of polar faculae (which reflect solar activity at high latitudes). The database used contains green-line intensities for each day in steps of $5°$ in position angle, referenced to a $60''$ height over the limb.

Sunspots can be used to study the solar activity at the photospheric level. The advantage of sunspot indices is their relatively long data series, which appreciably overlaps with the time covered by green-line observations. Naturally, these indices can be used to study the activity of the Sun and its asymmetry in the equatorial zone only, at latitudes of up to $\sim 30°$. We consider here two indices characterizing the sunspot activity the total area S_p and total number Q of the sunspots. We emphasize that we used the total number of sunspots rather than the traditional Wolf number. As will become clear below, the total sunspot number is an independent and very interesting index, which, in many cases, correlates better with the green-line asymmetry parameters than, for example, the sunspot areas or Wolf numbers. As follows from the studies of Kopecký and Kuklin (see the monograph [9] and references therein), the indices S_p and Q are related to the primary sunspot-formation indices in different ways.

Like the green-line brightness I, the magnetic flux Φ can, in principle, be determined for every point of the solar surface. In contrast to the sunspot-activity indices, this characteristic of the magnetic field is a continuous quantity. The magnetic flux is the main characteristic of the solar magnetic field, which determines all elements of solar activity, including coronal structure and sunspot formation.

We used the following databases in our study.

(1) The brightness of the λ 530.3-nm coronal green line. This database contains observations obtained at several coronal stations and currently covers the period 1939–2001. The data of individual stations
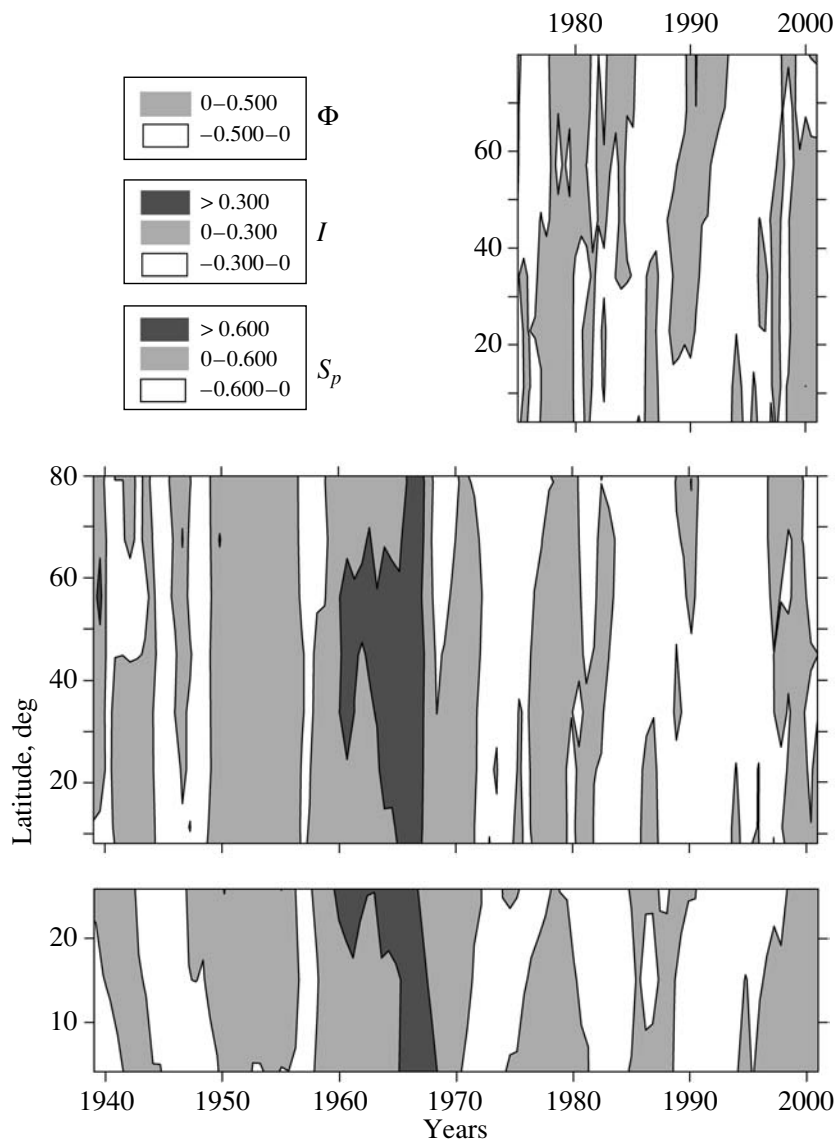
Fig. 1. Latitude–time diagram for the asymmetries in the net magnetic flux, coronal green-line brightness, and total sunspot area (from top to bottom). The gray scales for the maps are shown at the top left.

are reduced to a single photometric system; the technique used to unify the data is described in [13–15], and a description of the database is also given in [16, 17].

(2) The total sunspot area. The monthly averages were calculated from the data of the Greenwich Observatory obtained via the Internet.

(3) The total sunspot number was also calculated from data of the Greenwich Observatory available via the Internet.

(4) The net magnetic flux. Data on the magnetic flux at the photospheric level are obtained from magnetic-field observations carried out at the Kitt Peak Observatory,

(ftp://argo.tuc.noao.edu/kpvt/synoptic/mag/),

in the form of synoptic magnetic-field maps, which cover the interval 1975–2001,

(ftp://nsokp.nso.edu/kpvt/synoptic/).

Direct measurements yield the line-of-sight component of the magnetic field. Based on the assumption that the photospheric magnetic field is radial with respect to the solar surface, the observed values are divided by the cosine of the heliographic latitude. We used here the mean magnetic-flux values for each Carrington rotation and each 10° latitude zone.

Thus, the activity indices whose asymmetry we consider correspond to different manifestations of solar activity. Unified statistical-analysis techniques were applied to all these data. This makes it possible to compare the results obtained for objects resulting

Table 1. Correlation coefficients for the asymmetries of various pairs of indices at latitudes of $0°-30°$

Indices	$k_A(X,Y)$
S_p-Q	0.916 ± 0.006
S_p-I	0.800 ± 0.013
$S_p-\Phi$	0.768 ± 0.024
$Q-I$	0.860 ± 0.010
$Q-\Phi$	0.800 ± 0.021
$I-\Phi$	0.848 ± 0.016

from completely different magnetic field—matter interactions. Note that we did not include nonstationary processes in our analysis.
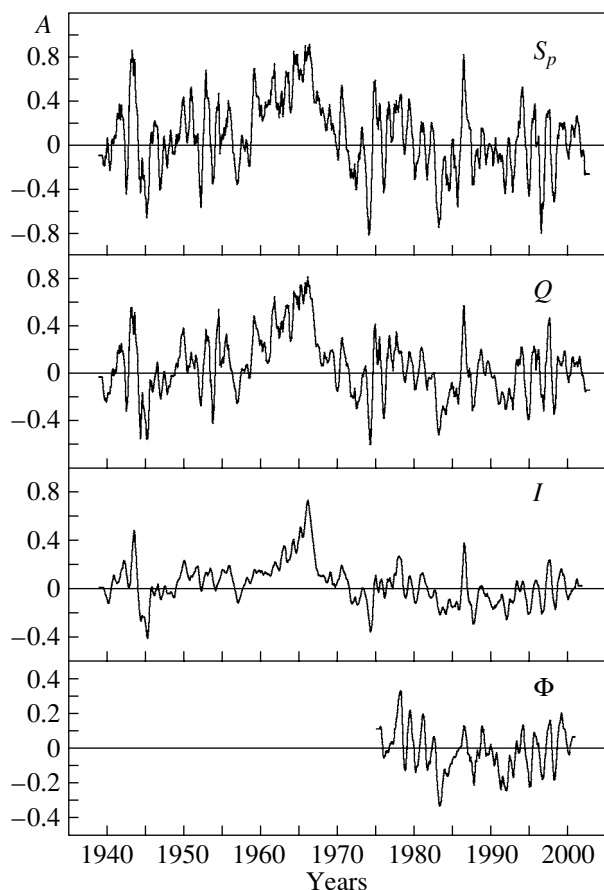
Fig. 2. Variation of the asymmetries in the four solar-activity indices in the zone of spot formation, $0°-30°$. Smoothing based on a 13th-order polynomial was done using the Savitzky–Golay technique. The time variations in the different indices are clearly correlated on short time scales.

3. VARIATIONS OF THE ASYMMETRY IN VARIOUS SOLAR-ACTIVITY INDICES ON SHORT AND LONG TIME SCALES

We determined the asymmetry indices in the standard way, $A = (N - S)/(N + S)$, where N and S are the corresponding activity indices for the northern and southern hemispheres. For our particular investigation, we needed the mean asymmetries for specific intervals of time and latitude. Since A depends nonlinearly on N and S, we must define the adopted procedure for averaging A. When deriving the mean asymmetries, we first found the monthly average (for the green-line brightness and net magnetic flux) or monthly summed (for areas or sunspot numbers) solar-activity indices for the chosen latitude zone in the northern and southern hemispheres separately, then calculated A according to the above formula. In some cases, we also smoothed the resulting asymmetry values; this will always be specially indicated in the text.

Normally, the N—S asymmetry refers to the difference between the entire northern and southern hemispheres. However, as is well known, the solar activity exhibits a zonal structure. For this reason, we found it expedient to use a latitude-dependent activity index. In doing so, we took N and S to be the values of a given index for solar latitude zones placed symmetrically about the equator.

Figure 1 shows latitude—time diagrams for the asymmetries in three of the activity indices considered the net magnetic flux, coronal green-line brightness, and the total sunspot area. The long-term, large-scale asymmetry variations are shown here. The upper two diagrams were constructed using the semiannual mean asymmetry values for 10-degree latitude zones. The lowest panel was constructed by calculating the semiannual mean asymmetries for the total sunspot numbers in the three latitude zones $0°-10°$, $10°-20°$, and $>20°$, then averaging using a ninth-order polynomial using the Savitzky—Golay method (see http://www.mathworks.com). This method suppresses noise but preserves high-frequency components. This averaging was needed to reveal large-scale variations in the asymmetry, since the asymmetries of areas and sunspot numbers undergo broader variations than the asymmetries of the other indices (for example, A can reach ± 1 in periods when there are no sunspots in one hemisphere); thus, the map would be fairly "noisy" without such averaging.

Figure 1 demonstrates that the asymmetries in these indices vary similarly over the entire latitude range (although the A values themselves are different). This can be seen from the alternation of dark and light strips, which is similar in all three diagrams in Fig. 1. Note that the asymmetry in the
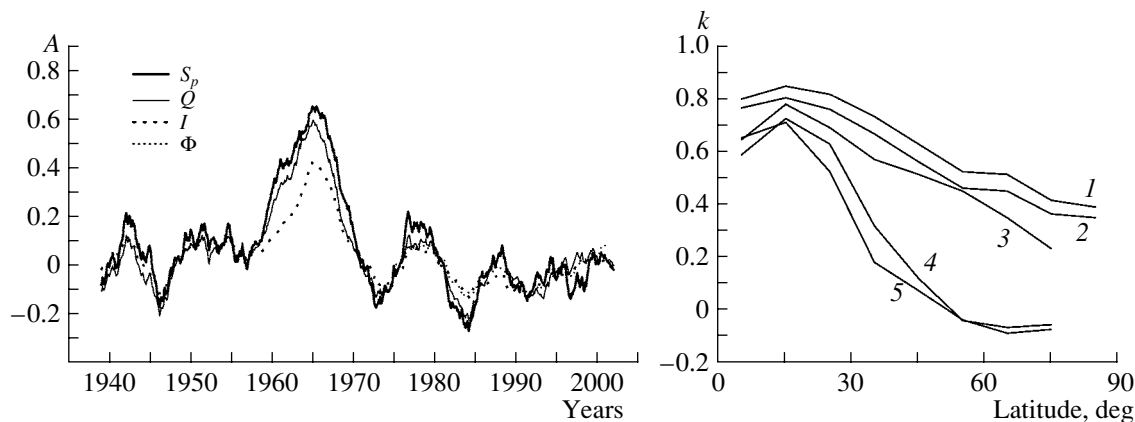
Fig. 3. Left: asymmetries in the four activity indices at latitudes $0°−30°$ subjected to a running-average with a 49 month window. A similarity can be noted in the asymmetry variations on long time scales. Right: correlations between the asymmetries for various pairs of solar-activity indices: (*1*) the green-line brightness I and the total sunspot number Q in 10-degree latitude zones, (*2*) I and the total sunspot area S_p, (*3*) I and the magnetic flux Φ in 10-degree zones, (*4*) Φ and Q, and (*5*) Φ and S_p.

total sunspot number (which is not presented here) is virtually identical to the asymmetry in the sunspot areas (shown in the lowest diagram). During the first half of the time interval considered, the corona was generally brighter in the northern than in the southern hemisphere. The opposite was true during the second half of the interval, although the asymmetry was considerably less pronounced. In other words, positive asymmetry (with the northern hemisphere exceeding the southern hemisphere in brightness) was predominant in 1939−2001. A substantial increase in the asymmetry was observed in 1965−1968, as has repeatedly been noted before [11]. The vertical strips in the latitude−time diagram result from nearly simultaneous changes in the asymmetry at all latitudes. We demonstrated this effect previously for the time variation of the N−S asymmetry of the green-line emission in 10-degree latitude zones [18, 19].

We also considered another sunspot-activity index: the traditional Wolf numbers [19, Fig. 3]. The cyclic variations in the sunspot number, green-line brightness, and sunspot area were considered for the northern and southern hemispheres separately. The asymmetry indices for these three parameters increase or decrease nearly simultaneously, and their variations correlate well in terms of power on both long and short time intervals. All three parameters demonstrate that the northern hemisphere dominated during the first half of the considered time interval and the southern hemisphere during the second half. This provides evidence for a long-period wave with a period of about 40 years in the variations, as has already been mentioned in some previous studies [1, 7].

Figure 2 shows the variations in the asymmetries calculated from the four activity indices averaged over the zone of spot formation, $0°−30°$. This figure was

constructed by first calculating monthly mean A values for all the indices, then smoothing them using the Savitzky−Golay technique with a 13th-order polynomial. Figure 2 shows that the resulting curves are very similar, even in specific small details. For example, an asymmetry increase in 1965−1968, a decrease near 1983, and a sharp asymmetry outburst in 1986 are clearly visible in all the indices. Table 1 presents the pair-correlation coefficients for the asymmetries in the indices; $k_A(X, Y)$ denotes the correlation coefficient for the asymmetries in indices X and Y.

Figure 3 (left) shows the same asymmetry indices subjected to a running-averaging procedure with a 49 month window. In contrast to Fig. 2, the high-frequency component is removed, leaving only the slowly varying component. We can see from Fig. 3 that the asymmetry curves are likewise similar in all their large-scale features. This similarity is reflected by the very high correlation coefficients for pairs of indices: $k_A(S_p, I) = 0.950$, $k_A(Q, I) = 0.960$, $k_A(S_p, Q) = 0.975$. The presence of several maxima and minima in the left-hand graph in Fig. 3 suggests the existence of a quasiperiodic wave with a period of ~12 yr. It is important to realize that, although this smoothing procedure reveals the 11-yr cyclic variations in the activity indices themselves, these variations do not display a correlation as high as that for the asymmetry. In particular, while $k_A(S_p, I) = 0.95$, the correlation coefficients for the smoothed indices themselves taken separately for the northern and southern hemispheres are 0.76.

The right-hand graph in Fig. 3 represents the pair-correlation coefficients for the asymmetries in the four indices as functions of latitude. The green-corona brightness and magnetic flux are taken for 10-degree latitude zones, while the area and total number of
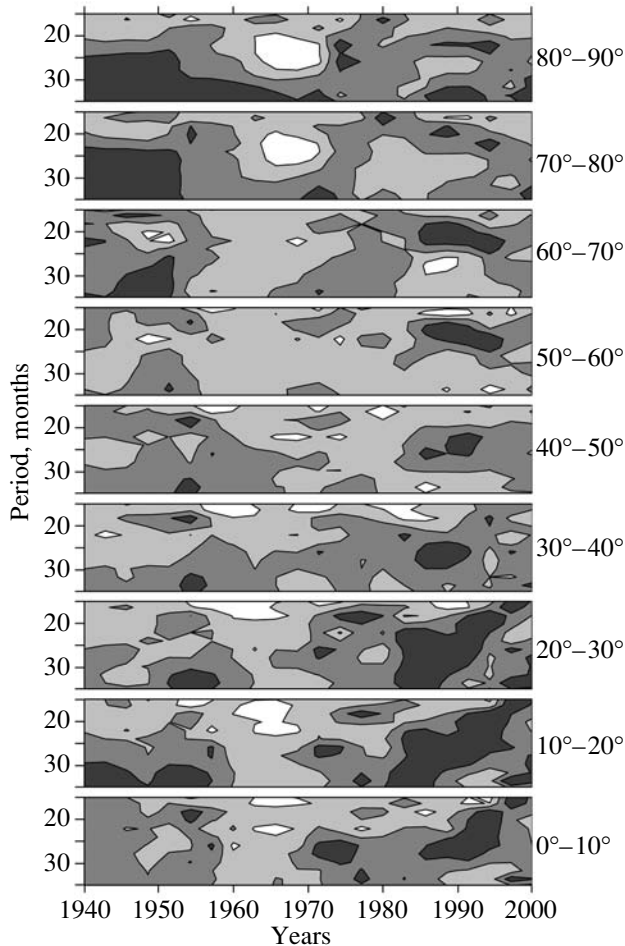
Fig. 4. SVAN diagrams for the asymmetry of the coronal green-line brightness in 10-degree latitude zones. The contour increment is 0.065, with black corresponding to amplitudes exceeding 0.195.
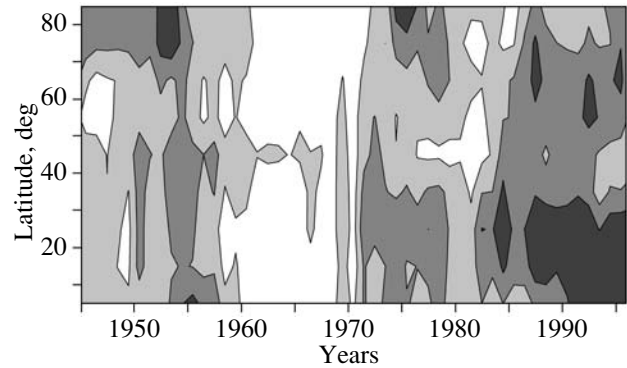

Fig. 5. Time and latitude distributions of the sum of the squared amplitudes of oscillations in the green-line asymmetry. The contour increment is 0.04, with black corresponding to values in excess of 0.12.

the polar zone is predominantly controlled by large-scale fields, while the green-corona brightness is substantially affected by local fields at all latitudes. It should, however, be kept in mind that we calculated the magnetic flux using observations of the line-of-sight magnetic field, which is essentially tangential at high latitudes.

Thus, the increases and decreases in A for all four activity indices here are essentially simultaneous and correlate well in magnitude on both long and short time intervals.

Our analysis in this section demonstrates that the asymmetry represents a general phenomenon in the Sun, which is manifest simultaneously and in nearly the same way in different indices. It is this property that substantially distinguishes the asymmetry from the activity indices themselves. We find evidence that the asymmetry is a fundamental characteristic of solar activity, which determines the degree of coincidence between the processes of magnetic-field generation in the two hemispheres.

4. QUASIBIENNIAL OSCILLATIONS IN THE N−S ASYMMETRIES OF VARIOUS INDICES

The QBOs are among the most interesting periodic oscillations observed in the solar activity. There have been fairly numerous studies devoted to the QBOs in recent years (see references in [20]). It has even been suggested that the QBO phenomenon is no less important than the 11-year solar cycle. There are grounds to believe that the QBOs are related to processes near the base of the convection zone with similar periodicities. Obridko and Gaziev [21] have shown that QBOs are clearly manifest in the asymmetry of magnetic fields inferred from Hα data. Oscillations in this frequency range were also detected by Knaack *et al.* [4].
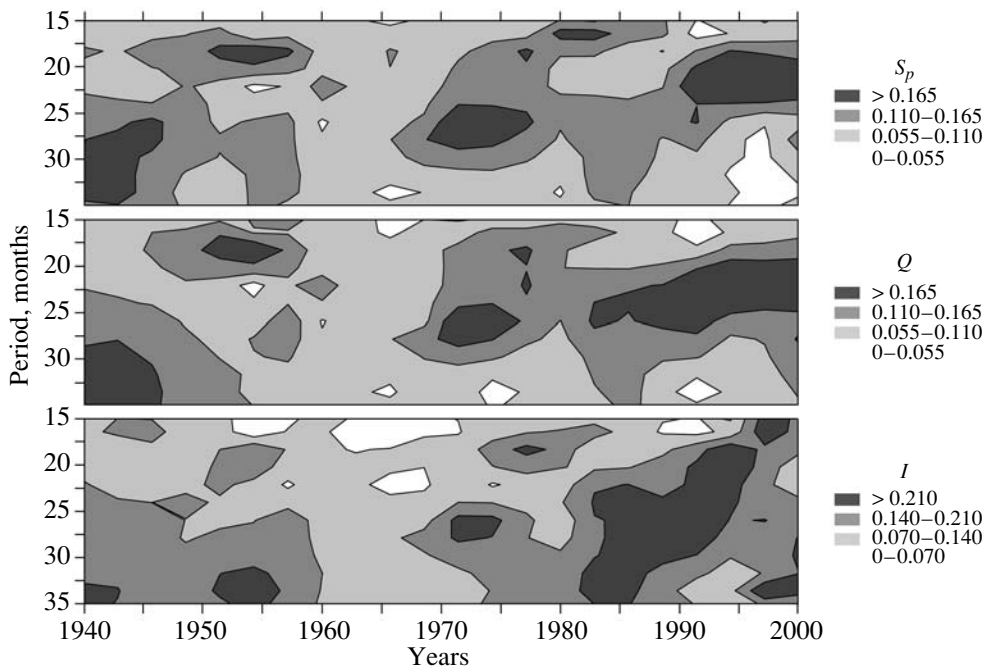
sunspots are calculated for the entire spot-formation zone. The semiannual mean asymmetry values are used. The A values for the coronal green-line brightness and the total sunspot number show the best correlation in all latitude zones; the correlation between the A values for the green-corona brightness and the total sunspot area is slightly lower, again in all latitude zones. The highest correlation of the asymmetries is observed at latitudes of $10°-20°$ for all the pairs of activity indices; the correlation is considerably lower in the polar region. At the same time, note the following interesting and not fully understood fact: the asymmetry in the green corona at middle latitudes above $30°$, where virtually no sunspots emerge, remains well correlated with the asymmetry in the sunspot areas, which is mainly contributed by latitudes below $30°$. Note also that the magnetic flux in the polar zone exhibits a negative, although weak, correlation with the total number and area of sunspots. This may suggest that the magnetic flux in

Fig. 6. SVAN diagrams for the asymmetries in the sunspot area, total sunspot number, and green-corona brightness over the spot-formation zone, $0°-30°$.

The time variations shown in Fig. 2 indicate that short-period oscillations are present in the asymmetry variations. Even without a more detailed analysis, we can see that the characteristic time scale of these oscillations is about 1.5−3 yr. A Fourier analysis testifies to the presence of quasibiennial variations in the spectrum of the asymmetry. Consideration of oscillations with periods ranging from 15 months to 4 yr for three latitude zones indicated that the oscillations in the spot-formation zone, $0°-30°$, are enhanced in the period range 2.2−2.5 yr. At high latitudes, $60°-90°$, two peaks can be identified, near 2.5 and 3 yr. At middle latitudes, $35°-55°$, the oscillations do not exhibit well-pronounced maxima, suggesting that, if QBOs are present in this latitude zone, they are much weaker than at lower and higher latitudes.

We used a spectral-variation analysis (SVAN) to study the time and latitude dependences of the QBOs. We employed an original SVAN code, which, in contrast to widely used codes described in the literature (see, e.g., [23]), includes normalization to a standard value (division by the standard deviation). This normalization reduces all oscillations to a single scale, with the sum of the squared amplitudes being unity in any spectrum. This enables us to compare the SVAN results for processes that are described quantitatively using different units.

A SVAN consists of successive Fourier decompositions over running time intervals. A Fourier decomposition is calculated for a time window of a chosen length, after which this window is shifted by some time interval and the entire procedure repeated. The resulting expansions are then used to construct a general map of the oscillation amplitudes (SVAN diagram) in time−(oscillation period) coordinates. We used the mean monthly asymmetries for our analysis, with a running window of 132 months and a time shift of 12 months. Normalization to the standard deviation was done in each window. We determined the amplitudes of oscillations with periods of 6−44 months. Although the relative power of any oscillation is characterized by its squared amplitude, analyzing the amplitudes themselves is more convenient in some cases. Obviously, such a representation of the results cannot affect the conclusions. We will always specify below which quantity is presented in figures and tables.

Figure 4 shows SVAN diagrams (oscillation amplitudes) for the asymmetry in the coronal green-line brightness in 10-degree latitude zones. The entire range of variations in the amplitude of the asymmetry oscillations is divided into four gradations. Darker regions correspond to larger amplitudes, with black used for amplitudes exceeding 0.195 (recall that the sum of the squared amplitudes over all periods in any running window is unity). Figure 4 demonstrates the presence of quasibiennial oscillations (periods of 20−35 months) over most of the time period studied. QBOs are especially pronounced in the equatorial zone after 1970. Longer period harmonics are manifest at high latitudes, especially in the 1940s
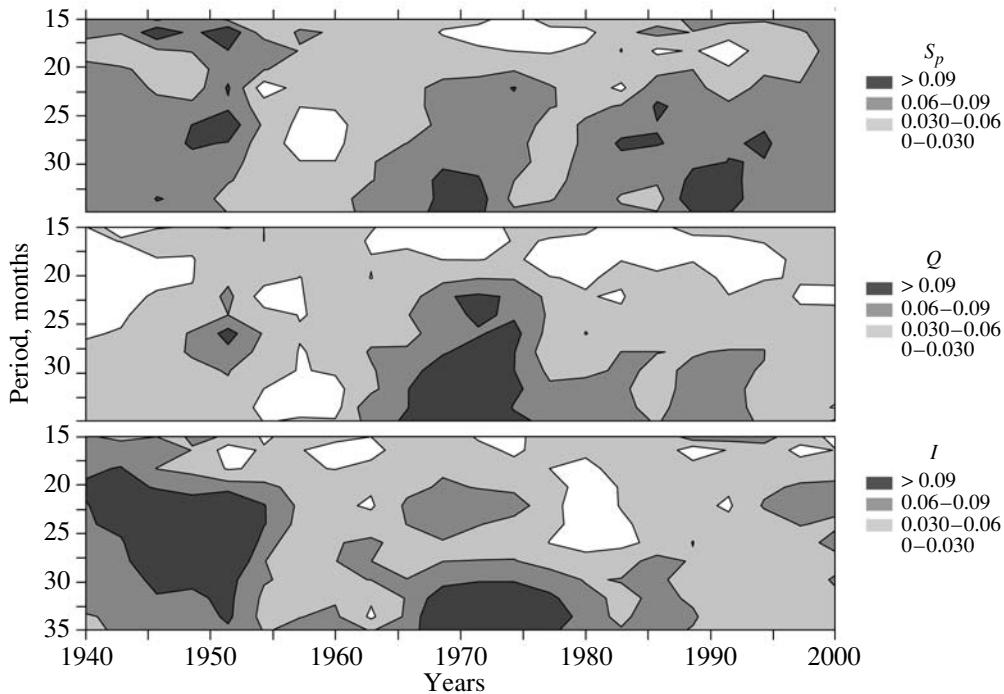
Fig. 7. Same as Fig. 6 for the indices S_p, Q, and I themselves.

and 1950s. On the other hand, the QBOs are always weak at latitudes of $50°-60°$, and QBOs were virtually absent at all latitudes in the 1960s. Note that the asymmetry was very strong precisely during this period.

Figure 5 shows maps of the time and latitude distributions of the squared amplitudes of the QBOs observed in the asymmetry of the coronal green-line brightness. We summed the squared amplitudes for three oscillation periods that roughly correspond to QBOs (18.86, 22.0, and 26.4 months). Darker areas correspond to higher amplitudes, with black denoting values exceeding 0.12. We can clearly see that (a) the QBOs are enhanced during the second half of the time interval studied (after 1970), (b) the QBOs are more pronounced at low than at high and, even more so, than at middle latitudes, and (c) we can identify several vertical strips indicating that the QBOs in the asymmetry are enhanced or weakened at all latitudes

Table 2. Correlation coefficients for the sums of the squared QBO amplitudes of the asymmetry and the activity indices themselves

Indices	$k_{As}(X,Y)$	$k_s(X,Y)$
$I-S_p$	0.693 ± 0.099	0.50 ± 0.20
$I-Q$	0.854 ± 0.071	0.29 ± 0.13
S_p-Q	0.889 ± 0.063	0.56 ± 0.11

during certain time intervals. A comparison of Figs. 5 and 1 reveals a general large-scale anticorrelation between the asymmetry and the power of its QBOs. This point will be considered in more detail in Section 5.

We carried out a wavelet analysis of the asymmetry in the coronal green-line brightness, again for 10-degree latitude zones [22, Fig. 5]. The wavelet diagrams provide higher time resolutions than the SVAN diagrams, but at the expense of spectral resolution. A gradual drift of the QBOs from the equatorial zone to high latitudes was observed during 1943–1948 (the growth phase of cycle 18) and 1984–1993 (the growth and maximum phases of cycle 22); this may somehow be related to the presence of the above-mentioned ∼40 yr wave in the asymmetry variation. In other cycles, QBOs were present during some time intervals over a fairly broad range of latitudes without any pronounced drift (for example, QBOs were clearly visible in 1973–1974 at latitudes of $0°-40°$). As in the SVAN diagrams, there is a decrease in the QBO power during the 1960s.

Quasibiennial oscillations are also clearly pronounced in the SVAN and wavelet diagrams for the magnetic flux. At low and middle latitudes, they are most intense near 1980 and 1995. By and large, the characteristic enhancements of the QBOs in the magnetic-flux asymmetry over the spot-formation zone, $0°-30°$, agree in time with the enhancements revealed in the other indices studied. A 4–6-year

Table 3. Sums of the squared oscillation amplitudes for various frequency ranges

Range	Periods, months	Asymmetry		Indices	
		S_p	I	S_p	I
I	25.2–12.6	0.06000	0.06278	0.01988	0.01606
II	18.9–37.8	0.06016	0.07270	0.01279	0.01517
III	94.5–189.0	0.03971	0.08942	0.32392	0.36761

($\sim$1500 day) period can also be identified in the wavelet diagrams, which is especially pronounced in the magnetic-flux values near the maximum of cycle 22 in 1990 [19, Fig. 5].

Figure 6 presents SVAN diagrams for the asymmetry of the indices S_p, Q, and I at latitudes of 0°–30°. The features detected in the asymmetry of the green-line brightness can also be seen in the asymmetry of these other activity indices. In particular, the decrease in the QBO amplitude during the 1960s and the substantial increase during and after the 1970s can be traced in the asymmetries of the sunspot area S_p and the total sunspot number Q. There are also other periods of QBO enhancements in the spot-formation zone. Of course, the similarity of the SVAN diagrams for the asymmetries in S_p and Q is quite natural, although, as is noted above, these indices are not equivalent, and the parameters relating them are time-dependent [24]. At the same time, the SVAN diagram for the green corona also exhibits a general similarity with the other two SVAN diagrams, although complete agreement could obviously hardly be expected in this case.

In addition to the SVAN diagrams for the activity-index asymmetries at latitudes of 0°–30°, we constructed SVAN diagrams for the indices themselves, the total sunspot area S_p, total sunspot number Q, and the green-corona brightness I (Fig. 7). First and foremost, we note that the SVAN diagrams for these three solar-activity indices are considerably less similar than the the SVAN diagrams for the asymmetries in these same indices (cf. Fig. 6). However, the most important difference is that, unlike Fig. 6, these diagrams do not reveal pronounced QBOs in the activity indices themselves. A comparison of Figs. 6 and 7 indicates that the QBOs are much more clearly manifest in the asymmetries than in the solar-activity indices themselves. In addition, the entire range of variation of the oscillation amplitudes, indicated on the right in Figs. 6 and 7, is a factor of 1.5–2 smaller for the indices than for their asymmetries.

Table 2 compares the correlation coefficients for the sums of the three squared amplitudes in the QBO range (periods of 18.86, 22.0, and 26.4 months).

A pair-wise comparison is presented for the correlation coefficients of the sums of the squared QBO amplitudes in the spot-formation zone, 0°–30°, obtained for the asymmetry, $k_{As}(X,Y)$, and for the indices themselves, $k_s(X,Y)$.

We can see from Table 2 that (i) the pair-correlation coefficients for the asymmetry are statistically significant, while the correlation coefficients for the indices exceed 3σ only in the last row, and (ii) the correlation coefficients for the asymmetry are definitely higher than for the activity indices themselves. This is especially clear when comparing the asymmetries of the coronal green-line brightness and the total sunspot number.

It has been repeatedly noted in the literature that oscillations observed on the Sun include those with characteristic time scales of one to two years, QBOs, and the 11-year cycle. The relative contributions of oscillations within these ranges to the overall spectrum for oscillations from 1 month to 63 yr is of interest. We calculated the sums of the squared amplitudes in the spot-formation zone for various frequency ranges in the Fourier spectrum (recall again that the sum of the squared amplitudes of all oscillations is unity), then isolated intervals in the observed spectrum corresponding to oscillations with periods of about 1–2 yr (range I), 1.5–3 yr (range II), and 8–16 yr (range III) (Table 3).

The 11-year cycle is predominant in the activity indices. The oscillation power in range III exceeds 30% of the total power of all the oscillations in the spectrum. At the same time, the oscillations in the quasibiennial (II) and high-frequency (I) ranges are a factor of 20–30 weaker. The situation changes radically when we consider the asymmetry spectrum. In range III, the relative power of oscillations in the asymmetry is a factor of 5–10 lower than the power of the oscillations in the activity indices. On the other hand, the oscillations with periods of 1.5–3 yr and 1–2 yr are enhanced by a factor of 3–5, and become comparably significant to the roughly 11-year oscillations.
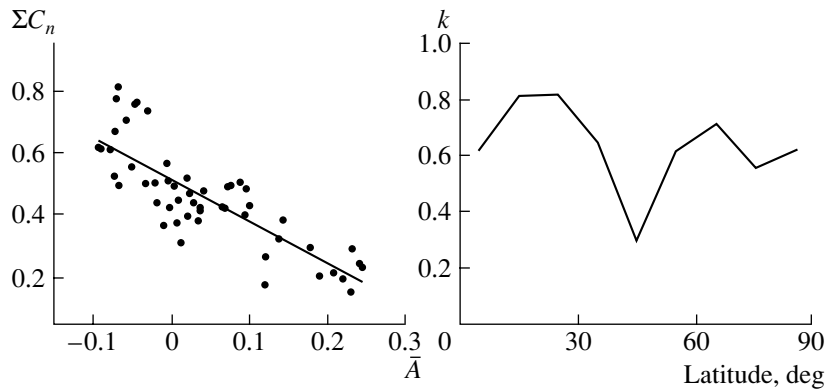
Fig. 8. Left: relationship between the sum of the squared amplitudes of oscillations with periods of 18.8−26.4 months and the mean asymmetry in a running window for latitudes 0°−30°. The correlation coefficient is 0.82. Right: latitude dependence of the correlation coefficient characterizing the relation between the sum of the squared QBO amplitudes and the amplitude of the asymmetry.

5. RELATION BETWEEN THE ASYMMETRY AND ITS QUASIBIENNIAL OSCILLATIONS

A comparison of Figs. 5 and 1 suggests the presence of a general large-scale anticorrelation between the asymmetry and the power of its QBOs. This is also supported by other tendencies that can be distinguished in these figures. In particular, the interval of weakened QBOs during the 1960s coincides in time with an increase in the asymmetries in the green-corona brightness and the areas and total number of sunspots. We conclude that a negative correlation between the QBO power and the asymmetry amplitude A can be traced over the entire time interval and over all latitudes.

The left-hand graph in Fig. 8 shows the relationship between the sum of the amplitudes of the QBOs observed in the green-line brightness asymmetry for latitudes 0°−30° and with periods from 18.8 to 26.4 months and the mean asymmetry in the running window of the SVAN analysis. The correlation coefficient for the negative correlation between these quantities at these latitudes is 0.82 ± 0.05.

The negative correlation between the relative sum of the QBO amplitudes and the asymmetry magnitude can also be traced in all other latitude zones. The latitude dependence of the correlation coefficient is shown in Fig. 8 (right). The highest correlation coefficients occur at 10°−20° and 60°−70°. These two regions are separated by a narrow zone (40°−50°) where the correlation coefficient is very low. Note that the asymmetry in the green line reaches its maximum in 1965−1968 precisely near this zone (Fig. 1). Furthermore, high-frequency oscillations disappear almost completely in this zone. Note as well that the boundary delimiting the regions of low-latitude and polar magnetic fields is located in this same zone.

The fact that the relationship between the relative QBO power and the asymmetry amplitude typified by the left-hand graph in Fig. 8 holds even in regions where the southern hemisphere dominates (i.e., the asymmetry is negative) appears somewhat strange. It would be more natural to see a similar relationship between the absolute value of A and the QBO power. Since there were virtually no intervals with large negative asymmetries during 1939−2001, we continued our analysis of this problem using a long series of data on the areas and total numbers of sunspots covering the period 1874−2002 [24]. We have shown that precisely the absolute value of the asymmetry for the index A is important, and that the QBO power also decreases with increase in this quantity, in regions of both negative and positive A values [24, Fig. 4]. Unfortunately, Figs. 2 and 3 in [24], which are similar to the upper two graphs in Figs. 6 and 7 here but refer to the longer period of 1874−2002, were spoiled in printing and cannot be used.

6. CONCLUSION

We have analyzed the north−south asymmetry A using four different indices characterizing the activity at various levels of the solar atmosphere: the brightness of the coronal green line, the total sunspot area, the total sunspot number, and the net magnetic flux. Our analysis demonstrates that the N−S asymmetry is a fundamental characteristic of the solar activity, which displays its own trends and is not directly controlled by the cyclic behavior of the solar activity. Thus, the N−S asymmetry is a specific, independent and very promising tool for analyses of solar-activity variations.

We have obtained the following main results.

(1) Similar behavior is visible in the time variations (Fig. 2; Fig. 3, left) of the N−S asymmetry of

all the solar-activity indices considered in the spot-formation zone, 0°−30°, on both short (1.5−3 yr) and long (~12 yr) time scales. The northern hemisphere dominates during the first half of the time interval studied (with a well-pronounced maximum in 1964−1966), and the southern hemisphere in the second half.

We emphasize that the solar-activity indices we have considered correspond to completely different solar phenomena associated with completely different mechanisms for the interaction between the magnetic field and matter. In one case, the magnetic field controls the convective motions in subphotospheric layers (sunspots), while, in another, the magnetic field controls the flux of magnetoacoustic waves (coronal-heating mechanism). Nevertheless, comparisons between the results for different formations on the Sun testify to a surprising similarity in their asymmetries.

(2) We also calculated the quantity A separately for narrower latitude zones, and considered correlations between pairs of asymmetry values for various activity indices in these zones. The highest correlation coefficient for any pair is reached at latitudes of 10°−20° (Fig. 3, right). The asymmetry in the green-corona brightness correlates well with the asymmetries in the total sunspot number and the sunspot area. This correlation remains significant at high latitudes, where sunspots are not observed. We conclude that the green-corona-brightness asymmetry is determined predominantly by a single parameter, which is related to the local magnetic fields. Conversely, the magnetic-flux asymmetry is determined by two magnetic-field components, as is reflected by the sign reversal in the corresponding curves for high latitudes in the right-hand graph of Fig. 3. It seems most reasonable to suppose that these components correspond to the local low-latitude fields and large-scale polar fields.

The high correlation of the high-latitude coronal brightness with the sunspot numbers may not by itself imply a direct physical relationship between these two phenomena. Many solar processes give the impression that the rhythm of the process is more important than direct energetic control. For example, in our case, the sunspot number may reflect the overall rhythm of the associated solar processes, which governs both the local magnetic fields and the coronal-brightness variations. Since the asymmetry is a difference quantity, it is potentially a sensitive tool for recognizing this overall rhythm.

(3) Quasibiennial oscillations (QBOs) were detected in the asymmetries of all the activity indices considered. We studied these QBOs using an original code for spectral-variation analysis (SVAN), which includes normalization to a standard value. We constructed SVAN diagrams for the green line and magnetic flux in 10-degree latitude zones, which demonstrate the existence of long periods in which there is a steady enhancement or weakening of the QBO power that is virtually simultaneous over a broad range of latitudes (Fig. 4).

(4) The SVAN diagrams for the spot-formation zone, 0°−30°, show that specific features of the quasibiennial variations displayed in the asymmetry of the green-line brightness can also be traced in the behavior of the other activity indices. Similar enhancements can be noted in the QBOs, and coincide in both time and frequency (Fig. 6). For example, a decrease in the QBO amplitude can be observed in the asymmetries of the sunspot areas and total numbers during the mid-1960s, followed by an appreciable increase during and after the 1970s. QBOs are continuously present in the interval 1970−2000. The correlation coefficients for the correlations between the sums of the squared QBO amplitudes for these three indices reach ~0.7−0.9. At the same time, the QBOs are not oscillations with a steady period but are instead a combination of pulses with a period of 1.5−3 yr.

(5) We carried out a spectral-variation analysis for the activity indices themselves and compared the results with the SVAN data for the asymmetries. The QBOs are much more pronounced in the N−S asymmetries than in the corresponding activity indices. The amplitudes of the QBOs in the indices are much more weakly correlated with one another than are the amplitudes of the QBOs in the asymmetries: the correlation coefficients for correlations between the QBO amplitudes are as low as ~0.3−0.55 in this case.

(6) We have discovered the unexpected effect that the manifestation of QBOs in the asymmetries of the studied parameters of solar activity, especially the green-corona brightness, appears to be in antiphase with the asymmetry itself. For example, the substantial increase in the asymmetry of all the activity indices in the mid-1960s is accompanied by an appreciable decrease in the QBO amplitude in all our SVAN diagrams. The negative correlation coefficient for the negative correlation between the sum of the three QBO amplitudes and the green-line asymmetry at latitudes of 0°−30° is 0.82 ± 0.05.

On the whole, the nature of the asymmetry of the solar indices remains unclear. As a rule, the most widespread dynamo theories do not include this asymmetry, instead assuming that both solar hemispheres are identical. The asymmetry could be the manifestation of a slowly varying relic magnetic field, but its existence has not been proven in any way, and is doubtful from the theoretical standpoint. To account for the asymmetry, the relic magnetic field should have fairly unusual properties: it should

not obey the Hale polarity law and must have a pronounced asymmetric component. A spherical-harmonic expansion of the large-scale magnetic field yields a coefficient that formally corresponds to a quasimonopole component. However, the nature of this effect is not understood, and it is quite possible that this coefficient results from observational errors (see, e.g., [25]).

We have the impression that the solar activity is generated largely independently in the two hemispheres, being governed by the differential rotation and meridional flows in each. However, since the cyclic variations in the two hemispheres exhibit a general temporal and energetic similarity, we suggest that some currently unknown mechanism determines the degree of similarity of the activity-generating processes in the two hemispheres. This mechanism should be superposed on and independent of the basic generation mechanism. The north−south asymmetry is apparently a quantitative measure of the properties of this mechanism. We should recall again in this context that the degree of similarity between the time variations and the manifestation of quasibiennial oscillations in the asymmetries of various indices is higher than in the original activity indices themselves. Thus, the asymmetry remains an unresolved problem in the question of the origin of the solar magnetic field.

REFERENCES

1. M. Carbonell, R. Oliver, and J. L. Ballester, Astron. Astrophys. **274**, 497 (1993).
2. M. Waldmeier, Sol. Phys. **29**, 332 (1971).
3. K. J. Li, J. X. Wang, S. Y. Xiong, *et al.*, Astron. Astrophys. **383**, 648 (2002).
4. R. Knaack, J. O. Stenflo, and S. V. Berdyugina, Astron. Astrophys. Lett. (2005) (in press).
5. G. Vizoso and J. L. Ballester, Astron. Astrophys. **229**, 540 (1990).
6. J. Mariş, M. D. Popescu, and M. Mierla, Rom. Astron. J. **12** (2), 131 (2002).
7. Yu. A. Nagovitsyn, Izv. Gl. Astron. Obs. Pulkovo, No. 212, 145 (1998).
8. P. I. Duchlev, Sol. Phys. **199**, 211 (2001).
9. Yu. I. Vitinskii, M. Kopetskii, and G. V. Kuklin, *Statistics of Spot-Formation Activity of the Sun* (Nauka, Moscow, 1986) [in Russian].
10. H. W. Newton and A. S. Milson, Mon. Not. R. Astron. Soc. **115**, 398 (1955).
11. J. Sýkora, *Solar and Interplanetary Dynamics*, Ed. by M. Dryer and E. Tandberg-Hanssen (Reidel, Dordrecht, 1980), p. 87.
12. V. Rušin, Bull. Astron. Inst. Czech. **31**, 9 (1980).
13. J. Sýkora, Bull. Astron. Inst. Czech. **22**, 12 (1971).
14. J. Sýkora, Contrib. Astron. Obs. Skalnaté Pleso **22**, 55 (1992).
15. J. Sýkora, Adv. Space Res. **14** (4), 73 (1994).
16. O. G. Badalyan, V. N. Obridko, and J. Sýkora, Sol. Phys. **199**, 421 (2001).
17. J. Sýkora and J. Rybák, Adv. Space Res. (2004) (in press).
18. J. Sýkora, O. G. Badalyan, and M. Storini, Adv. Space Res. **29** (12), 1975 (2002).
19. O. G. Badalyan, V. N. Obridko, J. Rybák, and J. Sýkora, in *Proceedings of the Second Solar Cycle and Space Weather Euroconference, Vico Equense, Italy, 2001,* Ed. by H. Sawaya-Lacoste (ESA Publications Division, Noordwijk, 2002), ESA SP-477, p. 201.
20. V. N. Obridko and B. D. Shel'ting, Astron. Zh. **78**, 1146 (2001) [Astron. Rep. **45**, 1012 (2001)].
21. V. N. Obridko and G. Gaziev, in *The Solar Cycle. Proceedings of the National Solar Observatory/Sacramento Peak 12th Summer Workshop,* Ed. by K. L. Harvey (Astron. Soc. Pac., San Francisco, 1992); Astron. Soc. Pac. Conf. Ser. **27**, 410 (1992).
22. O. G. Badalyan, V. N. Obridko, J. Rybák, and J. Sýkora, in *Solar Variability as an Input to the Earth's Environment,* Ed. by A. Wilson (ESA Publications Division, Noordwijk, 2003), ESA SP-535, p. 63.
23. A. Dzewonski, S. Block, and M. Landisman, Bull. Seismol. Soc. Am. **59**, 427 (1969).
24. O. G. Badalyan and V. N. Obridko, in *Climatic and Ecological Aspects of Solar Activity*, Ed. by V. I. Makarov and V. N. Obridko (Gl. Astron. Obs. Ross. Akad. Nauk, St. Petersburg, 2003), p. 33.
25. J. T. Hoeksema and P. H. Scherrer, *The Solar Magnetic Field—1976 through 1985*, WDCA Report UAG-94 (NGDC, Boulder, 1986).

Translated by A. Getling