The Stellar Epoch in the Evolution of the Galaxy

A. V. Tutukov, B. M. Shustov, and D. S. Wiebe

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

Received May 14, 1995

Abstract—We consider the astrophysical evolution of the Galaxy over large time scales, from early stages (an age of ~10^8 yrs) to the end of traditional stellar evolution (~10^{11} yrs). Despite the fact that the basic parameters of our stellar system (such as its size, mass, and general structure) have varied little over this time, variations in the characteristics of stars (their total luminosity, color, mass function, and chemical composition) are rather substantial. The interaction of the Galaxy with other stellar systems becomes an important factor in its evolution 100–1000 Gyr after its origin; however, we take the Galaxy to be isolated. In the model considered, the basic stages of Galactic evolution are as follows. The Galaxy forms as the result of the contraction (collapse) of a protogalactic cloud. The beginning of the Milky Way's life—the relaxation period, which lasts about 1−2 Gyr—is characterized by active star formation and final structurization. The luminosity and colors of the Galaxy are correlated to the star formation rate (SFR). The young Galaxy intensely radiates high-energy photons, which are mostly absorbed by dust and re-emitted at IR wavelengths. In the subsequent period of steady-state evolution, the gas content in the Galactic disk gradually decreases; accordingly, the SFR decreases, reaching 3–5$M_{\odot}$/yr at the present epoch and decreasing to 0.03$M_{\odot}$/yr by an age of 100 Gyr. Essentially all other basic parameters of the Galaxy vary little. Later, the decrease in the SFR accelerates, since the evolution of stars with masses exceeding 0.4$M_{\odot}$ (i.e., those able to lose matter and renew the supply of interstellar gas) comes to an end. The Galaxy enters a period of "dying", and becomes fainter and redder. The variation of its chemical composition is manifested most appreciably in a dramatic enrichment of the interstellar gas in iron. The final "stellar epoch" in the life of the Galaxy is completed ~10^{13} yrs after its formation, when the evolution of the least massive stars comes to an end. By this time, the supplies of interstellar and intergalactic gas are exhausted, the remaining stars become dark, compact remnants, there is no further formation of new stars, and the Galactic disk no longer radiates. Eventually, infrequent outbursts originating from collisions of stellar remnants in the densest central regions of the Galaxy will remain the only source of emission. © 2000 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

The study of the origin and evolution of the Universe is a strategic problem in astronomy. The history of our own stellar system—the Sun and solar system—constitutes an important part of this problem. These questions are closely related to the origin and development of life, which are vitally important for mankind. Until recently, as a rule, natural science (astrophysics, in this instance) was concerned only with the prior evolution of the Universe; recently, however, serious interest has also been expressed in the astrophysical aspects of the future. One example is the large international conference "The Future of the Universe" held in 1999 in Hungary.

The information stored in stellar physics and chemistry not only makes it possible to consider the future development of our stellar aggregate qualitatively, but also to study many aspects of its evolution quantitatively [1, 2]. Moreover, the luminosity function of galaxies in clusters indicates that Milky Way-type galaxies contain most of the light-emitting matter in the Universe [3], suggesting that predictions made for the Galaxy should be applicable to all "normal" galaxies.

Observations of objects at large redshifts provide evidence for an open and infinitely expanding Universe (see, for example, [4] and the recent estimates in [5]). Hence, when developing an evolutionary model for the future of the Galaxy, we are limited only by the time when fundamental physical processes (for example, the decay of protons and electrons) considerably change the structure of matter itself. For the Galaxy, however, this timescale is very long, and its parameters will vary radically long before such ages. Let us consider briefly the basic evolutionary landmarks in the history of a typical gaseous–stellar system.

The first stage in the history of the Galaxy is an epoch of large-scale star formation; i.e., the period over which the bulk of the gas from the protogalactic cloud is transformed into stars. It is commonly assumed that the star formation rate (SFR) in the Galaxy has remained essentially constant, and has differed from the present value (0.8–13$M_{\odot}$ (Gyr)$^{-1}$ pc^{-2} [6]) by no more than a factor of a few. For example, in their study of the evolution of the solar neighborhood, Pardi and Ferrini [7] derived the ratio of the present star formation rate and the SFR averaged over the lifetime of the Galaxy to be $\psi_{now}/\psi_{ave} \cong 0.2$. The maximum SFR in this

model is $20 M_\odot$ (Gyr)$^{-1}$ pc^{-2}, which is reached 1 Gyr after the formation of the Galaxy. Similar values are obtained in the model of [8].

Until recently, these results seemed to be consistent with the available observational data. Madau *et al.* [9] discovered that the global SFR derived from optical observations is about 30% of the present value at $z = 4$, reaches its maximum—which exceeds the present value by an order of magnitude—at $z = 1–2$, and then gradually decreases. This means that large-scale star formation in the Universe occurred when it was around 4 Gyr old (in the standard cosmological model). Some authors (for example, Pettini *et al.* [10]) have stressed that the "optical" SFR could be underestimated by a substantial factor due to absorption by dust.

However, small values for the maximum SFR in the Universe's past came into conflict with recent observations of galaxies at large redshifts at infrared and sub-millimeter wavelengths [11]. According to the data of [12], there are a large number of galaxies with high SFRs exceeding $100 M_\odot$/yr at $z > 5$. Taking into account the large numbers of such sources, it was concluded that they must be ordinary galaxies at early stages of their evolution, rather than some exotic objects.

Observations of these primordial galaxies suggest that they all experienced a short burst of star formation lasting no longer than 1 Gyr. The short duration of this interval is reflected in the chemical composition of old objects in the Galaxy, which display enhanced abundances of oxygen compared to iron. This is usually interpreted as evidence that these objects formed over a timescale shorter than the lifetimes for the progenitors of type Ia supernovae (the main producers of iron); i.e., shorter than $\sim 5 \times 10^8$ yrs. This value is also close to the gravitational timescale for a protogalactic cloud. Thus, we can distinguish an initial evolutionary timescale for the Galaxy of the order of 1 Gyr—the epoch for the initial formation of stars and basic Galactic structure.

The further global evolution of the Galaxy up to the present time apparently proceeded quite quietly. This is due to the fact that the most important time scales characterizing the main properties of the Galaxy—variations in its global structure and energy release—appreciably exceed the age of the Galaxy, estimated to be about 10 Gyr.

The energy-release timescale τ_* is determined by the course of nuclear reactions in the cores of the most numerous and long-lived low-mass stars. We can estimate this timescale from

$$\tau_* = \frac{10^{10} \text{ yrs}}{M_*^3}, \qquad (1)$$

where the mass of the star M_* is given in solar mass units. Relation (1) follows from calculations of the evolution of M dwarfs [13]. Assuming the minimum stellar mass for which thermonuclear reactions are possible

inside the star to be $M_{\text{low}} = 0.077 M_\odot$ [14], we obtain $\tau_* \sim 2 \times 10^{13}$ yrs. This time determines the total duration of the stellar stage in the life of the Galaxy, when its composition and the properties of its populations evolve. After completion of this stage, the Galaxy consists exclusively of cold, compact stellar remnants: white (or, to be more exact, by that time already black) dwarfs, neutron stars, and black holes.

The large-scale structure of the Galaxy in this time interval is essentially constant. An upper limit for the lifetime of the Galaxy as a whole is determined by the laws of stellar dynamics. An isolated system of gravitating particles (stars) dynamically relaxes on a timesclae [15, 16]

$$\tau_r = \frac{\nu R}{V^2} \frac{N}{4\pi \ln N}, \qquad (2)$$

where R is the size of the system, ν a typical random velocity, V the "regular" velocity, and N the total number of particles. Assuming for a spherical subsystem $\nu = V = 200$ km/s, $N = 10^{11}$, and $R = 10$ kpc, we obtain $\tau_r \sim 10^{17}$ yrs. The time for evaporation of particles from the system τ_e is $100 \tau_r$ to order of magnitude. For the disk population, $\nu \sim 40$ km/s and $\tau_r \sim 10^{16}$ yrs. The interaction of stars with clouds of interstellar gas can reduce this time by several orders of magnitude [17]. At the current age of the Galaxy, the structure of the stellar disk is determined to a considerable extent precisely by this relatively efficient process. Later, however, the density of the gas component decreases and the role of clouds becomes less important. It has been suggested that the general structure of the stellar disk will not undergo substantial changes for at least $10^{13}–10^{14}$ yrs [18]. Then, close approaches of stars (or, more precisely, their remnants) will "divide" the system into an extended halo and a dense inner core, possibly containing a supermassive black hole.

These estimates were obtained for an isolated galaxy. However, the Galaxy interacts with other members of the Local Group. Based on the orbits of the Local Group members, Peebles [19] concluded that our Galaxy and Andromeda are approaching each other. Given their present separation ~ 0.75 Mpc and relative radial velocity ~ 120 km/s, the two galaxies will approach to within 20–400 kpc in another 6 Gyr. Even if this first approach does not result in a direct collision, dynamical friction will sooner or later bring them closer to each other, and, in $10^{11}–10^{12}$ yrs, all the Local Group members will be merged into one giant stellar system. Its mass will not, however, differ appreciably from the mass of the Milky Way, since our Galaxy and M31 dominate in the Local Group.

In our study, we analyze the history of the Galaxy from its early phase ($\sim 10^8$ yrs) to the completion of "traditional" stellar evolution (~ 10 trillion yrs), neglecting its interaction with other members of the Local Group. We use for extrapolation our model of galactic evolu-

tion, which we describe in Section 2. The resulting variations in the basic parameters of the Galaxy are presented in Section 3, and we present our conclusions in Section 4.

2. EVOLUTIONARY MODEL OF THE GALAXY

In our study, we calculate variations in the basic parameters of the Galaxy—its luminosity, SFR, chemical composition, and stellar mass function—up to the completion of ordinary stellar evolution; i.e., to an age of 10^{13} yrs. Since the model of the Galaxy used was described in detail by Wiebe *et al.* [20], we present here only its main characteristics.

The model is based on the formalism of Tinsley [21], extended by Firmani and Tutukov [22]. The evolution of the system is described by differential equations expressing the conservation laws for the gas mass, the masses of individual chemical elements, and energy. The last equation makes it possible to take into account in a self-consistent way variations in the thickness of the gaseous disk of the Galaxy, assuming that energy balance is maintained via the input of energy from supernova explosions and its dissipation in collisions of clouds.

To calculate the star formation rate, we applied a Schmidt law

$$\psi = a\rho^n, \qquad (3)$$

where ρ is the average gas density, $a = 5 \times 10^7$ cm^3 g^{-1} s^{-1} [22], and $n = 2$. When calculating the evolution of the chemical composition, we used the results of Maeder [23] for helium and oxygen (type II and Ib supernovae), Thielemann *et al.* [24] for iron (type II and Ib supernovae), and Tsujimoto *et al.* [25] for the synthesis of iron in type Ia supernovae. We used a Salpeter law for the initial mass function, with limits of 0.1 and $100 M_\odot$. We have demonstrated the applicability of the model to studies of the prior evolution of disk-type galaxies in [26–28].

Here, we considered the following two processes affecting the evolution of the Galaxy over large time scales as additional parameters: the accretion of matter from the intergalactic medium (IGM) and the enrichment of the Galactic gas with heavy elements as a result of type Ia supernova explosions. The Table presents the versions of the models that we used to investigate the evolution of the Galaxy.

In the "standard" model of the Galaxy, the long-term accretion of intergalactic matter is not taken into account, and the rate for type I supernovae is determined by the star formation rate, assuming the lifetime for the supernova progenitors is ~300 Myr [29]. We selected the model parameters so that the model was consistent with the present values of the parameters of the Galaxy and with available data on its past evolution. In addition to the standard model, we calculated three other models, whose parameters are given in the table.

Parameters for calculated models

Model	Accretion IGM	Ia supernova explosion rate
CD (standard)	None	SFR with delay
AD	Yes	SFR with delay
CI	None	Inversely proportional to time
AI	Yes	Inversely proportional to time

These models differ from the standard model in their account for the accretion of intergalactic gas and the choice of technique for estimating the type-I supernova rate.

We took the accretion of surrounding matter to occur in two stages. The first is the initial "assembly" of the Galaxy, and is included in all models. By analogy with star-formation models, we assumed that the initial collapse of the Galaxy occurs in a regime with constant accretion rate $100 M_\odot$/yr (the constant accretion rate is due to the constant temperature of the accreted protogalactic matter), and lasts for about 2 Gyr. Thus, the mass of the Galaxy after completion of the initial accretion phase is $2 \times 10^{11} M_\odot$. Further, in models AD and AI, a phase of long-term accretion of external (relative to the already formed Galaxy) intergalactic gas begins. In this phase, the accretion rate decreases exponentially with a characteristic timescale of 10^{11} yrs. The rate constant was chosen so that the mass of the Galaxy doubles over 10^{11} yrs. The gradual decrease of the rate is due to the fact that the density of the intergalactic gas in clusters decreases as the gas is exhausted by accreting galaxies. This pattern for the development of accretion, based on theoretical concepts, appears to be qualitatively consistent with the accumulated information on the Galaxy as a member of the Local Group. It is not ruled out that the Local Group itself is a peripheral member of a larger cluster of galaxies in Virgo. Although the Galaxy is far from the gas-rich central regions of large galactic clusters (where gas accretion can play a very important role), it can still accrete some amount of intergalactic matter. The accretion rate at the present epoch is estimated to be ~$1 M_\odot$/yr (see, for example, [30]). On the other hand, "switching off" accretion makes it possible to distinguish the main evolutionary stages of the Galaxy more distinctly. Therefore, in the standard model, we took the Galaxy to be closed to inflowing matter.

The second important factor affecting the long-term evolution in the model is the way in which we calculate the type-I supernova (SN I) rate. From this point on, we will have in mind type-Ia supernovae when using this term. In [20], we included in the galactic energy balance only the contribution of type-II supernovae (SN II), whose rate is linked to the star formation rate, since, as we showed in [20], the contribution of SN I is negligible on time scales of the order of 10^{10} yrs. SN II currently dominate in the Galaxy, primarily due to the fact

that SN II explosions are much more frequent. The current SN II rate in the Galaxy is usually taken to be one per 30–100 yrs, whereas one SN I explosion occurs roughly every 1000 yrs [31]. There is, however, an important distinction between these two classes of objects. The massive stars that are the progenitors of type-II supernovae do not live longer than several tens of million of years, so that the SN II rate essentially tracks the SFR. On the contrary, the SN I rate "tracks" variations in the SFR with a substantial delay.

The lifetimes of SN I progenitors can reach values comparable to the time intervals considered here (10^{12} yrs). This means that the SN I rate at any time depends appreciably on the entire previous star-formation history. At late stages of Galactic evolution, when the SFR is already very low, the SN II rate is also reduced to zero. However, the SN I rate can remain relatively high, since some of these objects still "remember" times of more active star formation.

The mechanism for SN I explosions remains unclear; however, in essentially all hypotheses that have been put forth, these events occur in close binary systems in which one or both components are white dwarfs, i.e., they are essentially eternal objects. If a supernova explosion is associated with mass transfer onto a white dwarf from the secondary in a binary system, that is, a normal star that fills its Roche lobe, the interval between the formation of the system and the explosion is determined by the time required for the primary (more massive) component to become a white dwarf and for the secondary to reach the stage of Roche-lobe filling. If the explosion is due to the merging of the components of a binary (white dwarfs that lose angular momentum and approach each other due to gravitational-wave radiation), the interval between the formation of the system and the supernova explosion is limited only by the largest possible separation of the components, and can exceed 10^{12} yrs. Thus, first, variations of the SN I rate lag behind the SFR, and, second, SN I explosions are fairly "spread out" in time, and still occur at epochs when processes directly related to the SFR have ceased to operate.

Here, we use two methods to describe the SN I rate. In models CD and AD, we assumed that all type I supernova progenitors have the same lifetime of 300 Myr. Therefore, in these models, the evolution of the SN I rate essentially reproduces the star formation history (with accuracy to within a multiplicative coefficient). In models CI and AI, we assumed that type-I supernova progenitors have a range of lifetimes t_I obeying the law t_I^{-1}. If we assume that SN I explosions occur during the merging of white dwarfs in binaries whose components approach due to the radiation of gravitational waves, this law follows naturally from the distribution of semimajor axes in binary systems [32].

3. EVOLUTION OF THE GALAXY

Figure 1 presents variations of the parameters of the Galaxy for the case when the SN I rate is tied to the SFR with a delay of 3×10^8 yrs. The solid curves refer to accretionless models, and the dash–dot curves to models with accretion.

Three stages (periods) of star formation can distinctly be seen (Fig. 1a). In the first stage, which lasts about 2 Gyr, an increase (or outburst) in the star formation rate can clearly be recognized. The SFR increases rapidly to its maximum, which exceeds $200 M_\odot$/yr, then decreases to the current value of about $1 M_\odot$/yr after 10 Gyr. In the second stage, which lasts from $t \sim 2 \times 10^9$ to $t \sim 10^{11}$ yrs, the SFR decreases more and more slowly, due to the constant inflow of gas from evolved stars with masses $M_* > 0.4 M_\odot$. At $t \sim 10^{11}$ yrs, we can see a break in the curves for the SFR and gas mass. From this time, the SFR decreases much more rapidly, since stars with masses $M_* > 0.4 M_\odot$ lose hardly any gas in the course of their evolution. The Galaxy gradually "dies". By the end of the calculations ($t = 10^{13}$ yrs), the mass of gas in the Galaxy has decreased to nearly $10^5 M_\odot$ (Fig. 1c). Formally, star formation does not terminate, even with such a low gas mass; in reality, however, tidal forces, magnetic fields, ionization, etc. bring into effect some lower limit of the gas density, below which its condensation into stars becomes impossible.

The evolution of the half-thickness H of the gaseous disk (Fig. 1b) also goes through three stages. In the accretionless model, the collapse (flattening) of the disk is followed by a stage when the disk thickness remains essentially constant, and is sustained by the fading star formation. When the ejection of gas from evolved stars ceases, the disk responds to the break in the SFR via an increase in its contraction rate. This correlation between the behavior of the disk and the SFR is due to the fact that, in our model, the disk thickness and SFR are interrelated. A high star-formation rate results in an increase in the number of supernova explosions, a large flow of energy to the disk, a thickening of the disk, and a decrease in the average gas density and SFR. This feedback is caused by the fact that, in this model, the supernova explosion rate tracks the variations in the SFR for both types of supernovae (with a small delay that is negligible over large time intervals). Figure 1e presents the evolution of the supernova rate. We can see that the curves for the two types of supernovae are essentially parallel, and reproduce the SFR history. On the whole, in this model, the role of SN I is insignificant compared to that of SN II over the entire time interval considered.

Taking accretion into account increases the duration of the second stage compared to closed models, while the third stage is, on the contrary, shortened. The constant inflow of matter to the Galaxy extends its active life, but accelerates the processes that use up the gas in it. Therefore, by $t = 10^{13}$ yrs, when the intergalactic gas
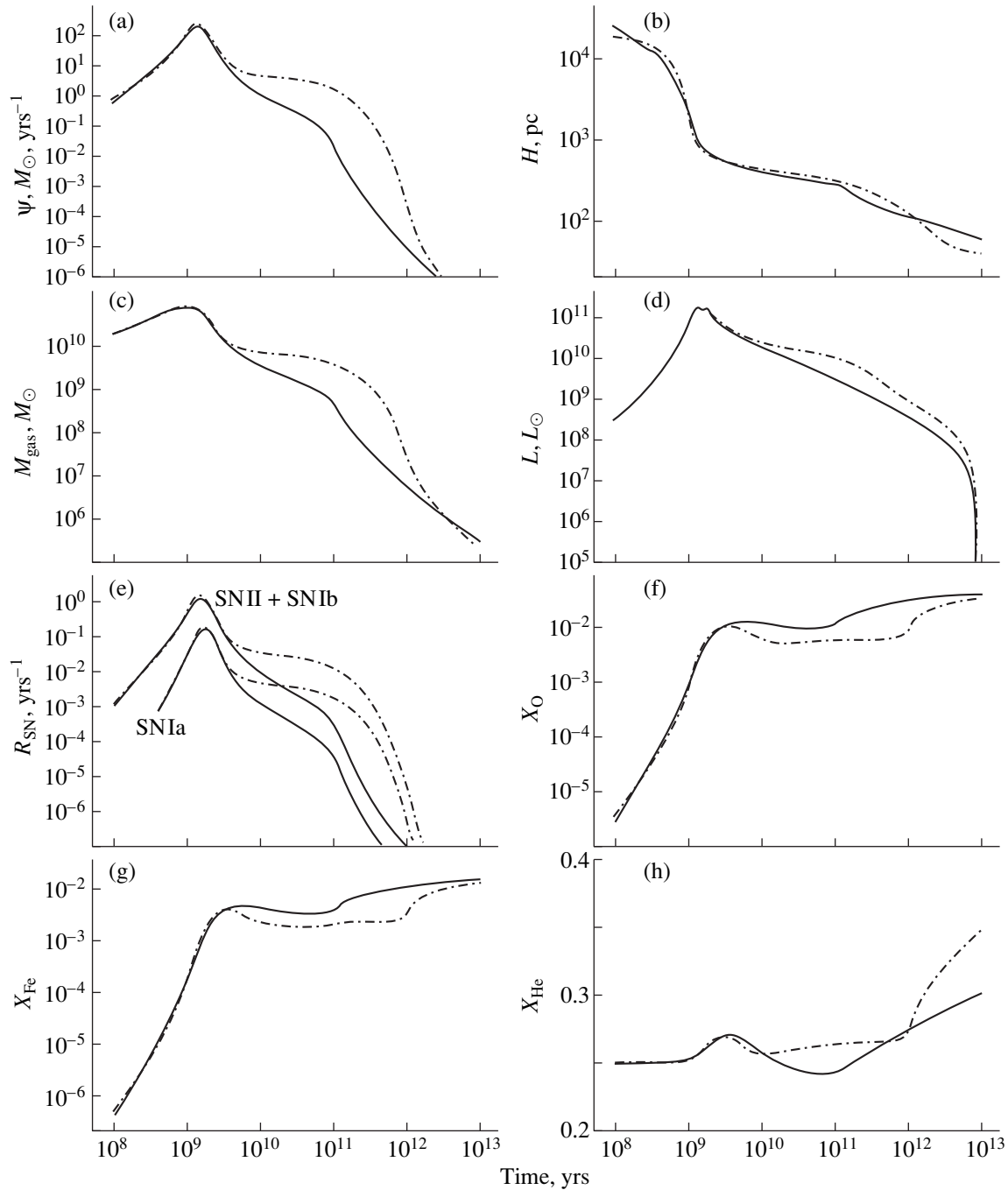
Fig. 1. Variations in the Galactic parameters in models CD and AD: (a) star formation rate, (b) half-width H of the gaseous disk, (c) gas mass, (d) luminosity, (e) supernova rate, (f)–(h) relative abundances of oxygen, iron, and helium.

has already been "scooped up" by the Galaxy, models with and without accretion do not differ substantially.

Variations in the luminosity of the Galaxy are presented in Fig. 1d. The transition from the second to the third stage is not pronounced here (unlike the SFR), since the effective termination of mass loss by stars is not related to their luminosity. By $t = 10^{13}$ yrs, the last, least-massive stars ($M \sim 0.1 M_\odot$) formed in the epoch of

large-scale star formation die, and the luminosity of the Galaxy rapidly decreases to nearly zero.

The shape of the evolutionary curves describing the chemical composition of the Galactic interstellar gas (Figs. 1f–1h) is also specified (in addition to the SFR) by the adopted forms of accretion and matter ejection from the Galaxy. The evolution of the oxygen and iron abundances follow the usual regular trends: a sharp increase at early stages of evolution, a decrease due to
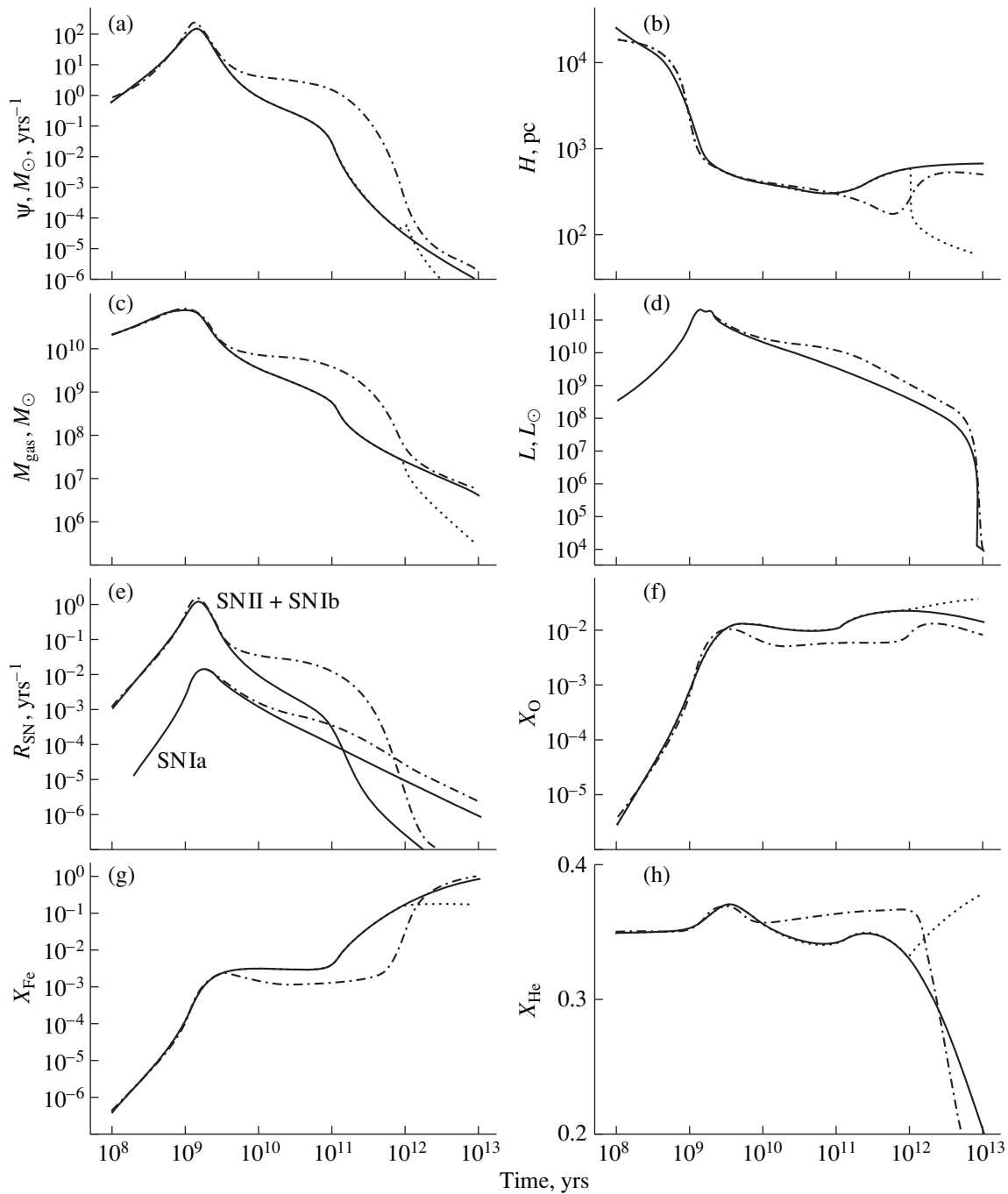
Fig. 2. Same as Fig. 1 for models CI and AI. The dashed curves indicate evolution of the parameters for a SN I switch-off after 10^{12} yrs.

a number of factors, discussed below, and then a small increase and gradual stabilization at a constant value. There is some decrease in the heavy-element abundance due to three factors: (a) return of gas by low-mass stars formed at early stages of evolution of the Galaxy from matter poor in heavy elements; (b) accretion of intergalactic gas (in models AD and AI), for which we assumed an initial chemical composition (75% hydrogen, 25% helium, 0% heavy elements); and (c) preferential ejection of heavy elements from the Gal-

axy. The two last assumptions are clearly oversimplified. First, it is known that the intergalactic gas contains a significant amount of heavy elements; i.e., its chemical composition differs from the initial composition. Second, although there is some reason to suppose that the metallicity of matter ejected from galaxies exceeds the solar value, this difference can hardly exceed a factor of a few [33]. However, as already mentioned above, we are interested in tendencies rather than accurate quantita-

tive predictions; for this reason, we decided to include "pure" accretion and "pure" ejection in the model.

Figure 1h presents evolutionary curves for the helium abundance. They differ from those for oxygen (Fig. 1f) and iron (Fig. 1g) due to the fact that interstellar helium is not "diluted" by accretion; on the contrary, its abundance is enhanced relative to the accretionless model. The constant helium abundance in model AD is due to the fact that, over a long time interval, accretion successfully competes with the ejection of helium from the Galaxy. In model CD, this ejection decreases the helium abundance between 10^{10} and 10^{11} yrs, even below its initial value. The increase of the helium abundance at late stages of evolution of the Galaxy is due to the fact that, unlike oxygen and iron, helium is also produced in low-mass stars.

In the standard model, we assumed that the lifetime of a type-I supernova progenitor is fixed and equal to 300 Myr. As noted above, in this case, there is little difference between the characteristic evolutions in the rates of the two types of supernovae, and the rate for SN I is always lower than that for SN II. This is also true for other delays, since they are small compared to the evolution timescale for the Galaxy considered.

A different situation arises in models with a smooth time dependence for the probability of SN I explosions. Figure 2 presents variations in the parameters of a Galaxy, in which the SN I rate is inversely proportional to the time. The notation is the same as in Fig. 1, but the evolution after the mechanism for making SN I is switched off is also displayed (dotted curves). In models CI and AI, we assumed that a SN I explosion can occur starting from 3×10^7 yrs after the formation of the precursor system and up to the limiting time $t_{\lim}$; as noted in Section 2, $t_{\lim}$ is close to 10^{12} yrs for models with merging white dwarfs. In this interval, the probability of a supernova explosion decreases as t^{-1}.

The most pronounced differences between these models and those with a constant delay can be seen in Figs. 2b, 2d, and 2e. At a certain time, the contribution from SN I exceeds that from SN II. Taking accretion into account moves this time into the future, without, however, making it less appreciable. Starting from this time ($\sim 10^{11}$ yrs) the gaseous disk becomes gradually thicker. The models displayed in Fig. 2 with the solid and dash–dot curves are calculated with $t_{\lim} = 10^{13}$ yrs, so that the thickening of the disk persists until the end of the calculation.

Taking into account the fact that SN I explosions are not instantaneous has little impact on the luminosity evolution, since the bulk of stars have already formed. On the other hand, at late stages of the evolution of the Galaxy, changes in the SN I "turn-on" regime can very appreciably affect the chemical composition of the interstellar gas. The "overproduction" of iron is especially noticeable. This is due to the fact that, in accordance with our current understanding, precisely SN I
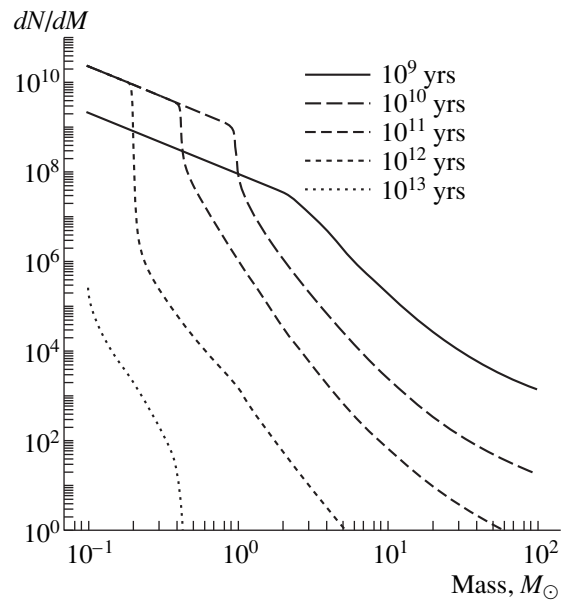

Fig. 3. Evolution of the stellar mass function in the Galaxy in the standard model.

are the main source of iron in our model. On the contrary, the relative abundances of oxygen and helium decrease, since the interstellar gas is continuously enriched. Thus, in our model, the final product of stellar evolution is a Galaxy whose disk is filled by iron gas.

Note that the chosen $t_{\lim} = 10^{13}$ yrs makes it possible to see more clearly trends in the evolution of the chemical composition of the Galaxy; this value is, however, unphysically large, and the value $X_{Fe} \sim 1$ will certainly not be reached in a real stellar system. A realistic value for $t_{\lim}$ should be smaller. The dashed curves in Fig. 2 are calculated with the parameters of model CI, but with $t_{\lim} = 10^{12}$ yrs. In this case, after the SN I "switch off," the Galaxy returns to its normal evolution.

Figure 3 presents variations in the stellar mass function in the Galaxy for the standard model. As expected, only stars with the lowest masses survive to the end of the stellar period of the Galaxy.

4. CONCLUSION

The extrapolated estimates for the basic parameters of the Galaxy considered here have enabled us to obtain the overall pattern of the evolution for our stellar system (and also for the main mass of radiant matter in the Universe as a whole) on time scales up to ten trillion years. We distinguish three periods (stages) of this evolution: a relaxation period featuring a vigorous burst of star formation (duration about 2 Gyr), a period of "normal" evolution (up to $\sim 10^{11}$ yrs), when the basic parameters of the Galaxy vary relatively little (excluding the gas mass and star formation rate), and a period of "dying", which finishes by $\sim 10^{13}$ yrs after the Gal-

axy's formation, when the evolution of the least massive stars ceases.

Note some aspects that we have not taken into account, but that could play a considerable role in the fate of the Galaxy. In the Introduction, we mentioned the possibility of a collision of the Galaxy with the nearest stellar systems. The observational manifestations of a "dying" galaxy could also, to some extent, be determined by stellar collisions. With the average density of the stars in the Galaxy, the rate of stellar collisions does not exceed 10^{-22} yr^{-1}; i.e., it is negligibly low. However, the collision rate could be many orders of magnitude higher in the dense central regions of the Galaxy, and outbursts resulting from stellar collisions could be a rare manifestation of activity in a dying Galaxy. Planetary systems possess much higher collisional cross sections, and, according to Massevich and Tutukov [32], their average collision rate can exceed 10^{-10} yr^{-1}. This substantial value implies that not only stars, but also single planets, may wander around the Galaxy over most of the stellar epoch of its evolution. In the central regions of the Galaxy, stellar collisions and accretion will certainly result in the formation of a massive black hole well before the end of the characteristic dynamical relaxation time.

Neither have we taken into account the role of dark matter in galactic evolution. We do not suppose this to appreciably affect the variation of the parameters of the Galaxy we have considered. Another important point is the insufficient theoretical elaboration of the evolutionary models for low-mass stars and stars with very high heavy-element abundances, as well as the uncertain information on the degree to which the interstellar medium is enriched with various elements from stars at various stages of their evolution. In spite of these simplifications, the model considered here could be quite relevant for studies of the possible development of physical and chemical processes in our Galactic stellar system.

ACKNOWLEDGMENTS

REFERENCES

1. F. C. Adams and G. Laughlin, Rev. Mod. Phys. **69**, 337 (1997).
2. L. M. Krauss and G. D. Starkman, astro-ph/9902189, 1999.
3. N. Trentham, Mon. Not. R. Astron. Soc. **294**, 193 (1998).
4. J. P. Ostriker and P. Steinhardt, Nature **377**, 600 (1995).
5. A. G. Riess, A. V. Filippenko, P. Challis, *et al.*, Astron. J. **116**, 1009 (1998).
6. N. C. Rana, Ann. Rev. Astron. Astrophys. **29**, 129 (1991).
7. M. C. Pardi and F. Ferrini, Astrophys. J. **421**, 491 (1994).
8. C. Chiappini, F. Matteucci, and R. Gratton, Astrophys. J. **477**, 765 (1997).
9. P. Modau, H. C. Ferguson, M. E. Dickinson, *et al.*, Mon. Not. R. Astron. Soc. **283**, 1388 (1996).
10. M. Pettini, M. Kellogg, C. C. Steidel, *et al.*, Astrophys. J. **508**, 539 (1998).
11. A. W. Blain, I. Smail, R. J. Ivison, and J.-P. Kneib, Mon. Not. R. Astron. Soc. **302**, 632 (1999).
12. A. J. Barger, L. L. Cowie, D. B. Sanders, *et al.*, Nature **394**, 248 (1998).
13. G. Laughlin, P. Bodenheimer, and F. C. Adams, Astrophys. J. **482**, 420 (1997).
14. A. Burrows and J. Liebert, Rev. Mod. Phys. **65**, 301 (1993).
15. S. Chandrasekhar, Astrophys. J. **97**, 255 (1943).
16. K. F. Ogorodnikov, *Dynamics of Stellar Systems* (Fizmatgiz, Moscow, 1958; Pergamon, Oxford, 1965).
17. L. Spitzer and M. Schwarzschild, Astrophys. J. **114**, 385 (1953).
18. V. L. Polyachenko and A. M. Fridman, *Equilibrium and Stability of Gravitating Systems* [in Russian] (Nauka, Moscow, 1976).
19. P. J. Peebles, Astrophys. J. **429**, 43 (1994).
20. D. S. Wiebe, B. M. Shustov, and A. V. Tutukov, Astron. Zh. **75**, 3 (1998) [Astron. Rep. **42**, 1 (1998)].
21. B. M. Tinsley, Fundam. Cosm. Phys. **5**, 287 (1980).
22. C. Firmani and A. V. Tutukov, Astron. Astrophys. **264**, 37 (1992).
23. A. Maeder, Astron. Astrophys. **264**, 105 (1992).
24. F.-K. Thielemann, K. Nomoto, and M. Hashimoto, Astrophys. J. **460**, 408 (1996).
25. T. Tsujimoto, K. Nomoto, Y. Yoshii, *et al.*, Mon. Not. R. Astron. Soc. **277**, 945 (1995).
26. B. M. Shustov, D. S. Wiebe, and A. V. Tutukov, Astron. Astrophys. **317**, 397 (1997).
27. D. S. Wiebe, B. M. Shustov, and A. V. Tutukov, Astron. Astrophys. **345**, 93 (1999).
28. D. S. Wiebe, B. M. Shustov, and A. V. Tutukov, in *Modern Problems of Stellar Evolution*, Ed. by D. S. Wiebe (Geos, Moscow, 1998), p. 43.
29. A. V. Tutukov and L. R. Yungelson, Mon. Not. R. Astron. Soc. **268**, 871 (1994).
30. J. R. Lepine and G. Duvert, Astron. Astrophys. **286**, 60 (1994).
31. E. Cappellaro, M. Turatto, D. Yu. Tsvetkov, *et al.*, Astron. Astrophys. **322**, 431 (1997).
32. A. G. Masevich and A. V. Tutukov, *Evolution of Stars: Theory and Observations* [in Russian] (Nauka, Moscow, 1988).
33. C. Heiles, Astrophys. J. **354**, 483 (1990).

Translated by K. Maslennikov

Radio Synchrotron Emission by Protons and Electrons in Pulsars and the Nuclei of Quasars

N. S. Kardashev

Astro Space Center, Lebedev Physical Institute, ul. Profsoyuznaya 84/32, Moscow, 117810 Russia

Received November 15, 1999

Abstract—Relations enabling estimation of the limiting brightness temperature of synchrotron radiation subject to self-absorption and inverse Compton scattering are presented for the case of relativistic electrons (positrons) and protons. Analogous expressions are presented for relativistic particles moving along curved magnetic lines of force (curvature radiation) and coherent radiation by relativistic particles. These relations can be used to determine the brightness temperatures expected for the central regions of active galactic nuclei, neutron stars, and other objects that produce relativistic particles. Radiation by relativistic protons yields higher intensities, and could be a source of the highest-energy cosmic rays. © *2000 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The most obvious sources of high-brightness-temperature radio emission are pulsars and the central regions of galaxies. Very high brightness temperatures may also be associated with radio sources observed immediately after γ-ray bursts in models invoking mergers of compact stars, radio sources in the early stages after supernova explosions, the neighborhoods of stellar-mass black holes, and certain other types of stars with strong radio emission. The brightness temperatures of pulsars probably reach $\sim 10^{25}$ K, as indicated by the intensity and duration of their pulses.

Studies of the structure of the central regions of extragalactic radio sources show that the brightness temperature rises with approach toward the central supermassive black hole. In a survey of 140 compact extragalactic radio sources at 22 GHz using the VLBA [1], only lower limits to the brightness temperature were found for 126 sources ($> 3 \times 10^{10}$ K); the brightness temperatures for 21 objects were $> 10^{12}$ K, and the sources DA 406 (1611+343) and OV 236 (1921–293) had lower limits of 5.7×10^{12} and 7.0×10^{12} K, respectively. Observations with the VSOP/VLBA space–ground interferometer at 5 GHz with projected baselines up to 30000 km indicate the brightness temperature of NRAO 530 (1730–130) to be $\geq 3 \times 10^{12}$ K [2].

Still higher brightness temperatures are implied by observations of very rapid (timescales less than a day) variability of the fluxes of some extragalactic sources. Brightness temperature estimates for 0405–385 [3], 0537–441 [4], 0716+714 [5], and 1308+328 [6] reach 10^{21} K.

The highest brightness temperatures are observed for pulsars (to 10^{25} K and higher). In this case, the radiation mechanism is probably associated with the motion of an inhomogeneous flux of relativistic particles along magnetic lines of force. The presence of such high brightness temperatures is probably connected to the specific properties of radiation by relativistic particles in the vicinity of a central engine—supermassive black hole or neutron star. The high intensity of this radiation makes it possible to study the structure of these regions with the extremely high angular resolution that can be achieved using space interferometers [7].

The main model recognized today suggests that the limiting brightness temperature for synchrotron radiation is $\sim 10^{12}$ K [8]. This limit is the result of rapid energy losses by relativistic electrons (positrons) due to inverse Compton scattering of their radiation. Higher brightness temperatures can be realized if the radiating particles are relativistic protons [9] or in the case of coherent radiation by clouds of relativistic particles.

Doppler boosting of relativistic synchrotron sources was first considered as a model for the one-sided jet of M87 by Shklovskiĭ [10], and was later widely discussed as a means to obtain high observed brightness temperatures due to the relativistic enhancement of radiation in the direction of motion of the radiating plasma [11–14]. It is possible to obtain very high observed brightness temperatures exceeding the Compton limit for non-stationary models; for example, for regions in a source where losses are compensated by particle acceleration [15]. In contrast to the physical conditions in the relativistic jets emanating from galactic nuclei, the conditions near supermassive black holes and in the envelopes of neutron stars could be non-stationary, without an equilibrium between the energies in particles and the magnetic field [16, 17].

We will derive basic relations determining limiting brightness temperatures below.

2. THE LIMITING INTENSITY
OF SYNCHROTRON RADIATION
BY ELECTRONS (POSITRONS) AND PROTONS

The threshold value corresponding to the limit imposed by multiple Compton scatterings yields an expression for the limiting intensity. This expression can be derived from the condition that the ratio of the inverse-Compton luminosity L_c and the synchrotron luminosity L_s be less than unity [8]:

$$L_c/L_s = \rho_s/(H^2/8\pi) \leq 1, \qquad (1)$$

where the energy density of the synchrotron radiation is

$$\rho_s = 4\pi \int_0^\infty I_\nu d\nu/c \approx 4\pi I_\nu \nu/c, \qquad (2)$$

$I_\nu = 2kT\nu^2/c^2$ is the intensity of the self-absorbed synchrotron radiation, T is the brightness temperature, ν is the peak frequency, and H is the magnetic field strength.

Let us consider a power-law particle spectrum [18, 19] $N(E) \propto E^{-\gamma}$ and a magnetic field strength $H_\perp$ that is the same for all particles of the projection and is oriented perpendicular to their direction of motion. In this case, the intensity of the radiation at low frequencies, where the optical depth to self-absorption is greater than unity, is

$$I_\nu = \frac{2kT}{c^2}\nu^2 = \frac{1}{\gamma+1}$$
$$\times \frac{\Gamma((3\gamma-1)/12)\Gamma((3\gamma+19)/12)}{\Gamma((3\gamma+2)/12)\Gamma((3\gamma+22)/12)}\left(\frac{8\pi m^3 c}{3cH_\perp}\right)^{0.5}\nu^{2.5}. \qquad (3)$$

Using (1)–(3), we find

$$T \leq a(\gamma)(mc^2/k)(m^2 c^4 e^{-3} H_\perp^{-1})^{1/7}, \qquad (4)$$

where

$$a(\gamma) = (\pi/6^3)^{1/7}$$
$$\times \left[\frac{1}{\gamma+1}\frac{\Gamma((3\gamma-1)/12)\Gamma((3\gamma+19)/12)}{\Gamma((3\gamma+2)/12)\Gamma((3\gamma+22)/12)}\right]^{6/7}. \qquad (5)$$

Here, Γ is the gamma function. For $\gamma = 1$, 2, and 3, respectively, we have $a(\gamma) = 0.85$, 0.28, and 0.19.

If we assume isotropic distributions for the directions of the magnetic field and particle velocities, (3) can be averaged: $\langle H_\perp^{-0.5}\rangle = \frac{\sqrt{\pi}}{2}\frac{\Gamma(0.75)}{\Gamma(1.25)}H^{-0.5}$. In this case, we must use in (4) the relation

$$a(\gamma) = (\pi^4 2^{-9} 3^{-3})^{1/7}$$
$$\times \left[\frac{\Gamma(0.75)}{\Gamma(1.25)}\frac{1}{\gamma+1}\frac{\Gamma((3\gamma-1)/12)\Gamma((3\gamma+19)/12)}{\Gamma((3\gamma+2)/12)\Gamma((3\gamma+22)/12)}\right]^{6/7}. \qquad (6)$$

For $\gamma = 1$, 2, and 3, respectively, we have $a(\gamma) = 0.69$, 0.30, and 0.19. We have for relativistic electrons (positrons) from Eq. (4)

$$T_e \leq a(\gamma) \times 1.07 \times 10^{12} H_\perp^{-1/7} \text{ K}. \qquad (7)$$

For synchrotron radiation by relativistic particles with a single energy E (a monoenergetic spectrum) [18, 19], we find in the self-absorbed region

$$I_\nu = (3e/(4\pi m^3 c^6))^2 E^5 H_\perp^2 S(x). \qquad (8)$$

Here, $x = \nu/\nu_c$,

$$\nu_c = 3eE^2 H_\perp/(4\pi m^3 c^5), \qquad (9)$$

$$S(x) = F(x)/K_{5/3}(x), \quad F(x) = x\int_x^\infty K_{5/3}(z)dz,$$

and $K_{5/3}$ is a Bessel function. For $x \ll 1$ $F(x) \approx 4\pi x^{1/3}/(\sqrt{3}\,\Gamma(1/3)2^{1/3})$, $S(x) \approx 1.5x^2$, and

$$I_\nu = 1.5E\nu^2/c^2; \qquad (10)$$

i.e.,

$$E = \frac{4}{3}kT. \qquad (11)$$

Here, the intensity does not depend on the magnetic field, and, in contrast to (3), the frequency dependence is the same as that for a Rayleigh–Jeans law.

We obtain from (1), (2), and (10) in Eq. (4) the quantity

$$a = 0.725. \qquad (12)$$

We then have for electrons (positrons) from (4), (9), and (12)

$$T_e \leq 0.77 \times 10^{12} H_\perp^{-1/7} \text{ K} \qquad (13)$$

at

$$\nu \ll \nu_c = 4eH_\perp(kT)^2/(3\pi m^3 c^5)$$
$$= 1.26 \times 10^{11} H_\perp^{5/7} \text{ Hz}. \qquad (14)$$

Taking into account the fact that $T \propto m^{9/7}$ in (4) and substituting the mass of the proton, we obtain the limiting brightness temperature for synchrotron radiation by protons

$$T_p \leq a(\gamma) \times 1.68 \times 10^{16} H_\perp^{-1/7} \text{ K}, \qquad (15)$$

$$T_p \leq 1.21 \times 10^{16} H_\perp^{-1/7} \text{ K}, \qquad (16)$$

at

$$\nu \ll 5.0 \times 10^9 H_\perp^{5/7} \text{ Hz} \qquad (17)$$

for power-law and monoenergetic spectra, respectively. It follows from (14) that the brightness temperature for an observing frequency ν is

$$T > (mc^2/k)(3\pi mc\nu/4eH_\perp)^{0.5} \qquad (18)$$

and this inequality must be satisfied simultaneously with (4). Then,

$$\nu < (4a^2/3\pi)(ce/m^3)^{1/7}H_\perp^{5/7}. \qquad (19)$$

For electrons (positrons), this corresponds to

$$\nu < 1.27 \times 10^{11}H_\perp^{5/7} \text{ Hz}, \qquad (20)$$

and for protons to

$$\nu < 5.06 \times 10^{9}H_\perp^{5/7} \text{ Hz}. \qquad (21)$$

It follows that the maximum source brightness temperatures in a given frequency range are expected for those objects with the largest magnetic fields. The peculiarities of synchrotron radiation by particles moving at small angles to the magnetic field have been considered in [20], and do not change this limiting brightness temperature.

The possibility that the radio emission from the central regions of quasars is associated with synchrotron radiation by protons was first discussed in [21], and an analogous mechanism for pulsars was considered in [9]. The main attraction of this model is its extremely low energy losses. For synchrotron radiation by particles with rest mass m and energy E in a magnetic field whose projection is $H_\perp$, the rate of energy loss is $-\dot{E} \propto H_\perp^2 E^2 m^{-4}$, which is a factor of 10^{13} lower for protons than for electrons (positrons). At the same time, the peak frequency $\nu_c \propto H_\perp E^2 m^{-3}$ is a factor of 10^{10} lower for protons than for electrons (positrons). The ratio of the energy-loss times for protons and electrons (positrons) with the same ν_c and $H_\perp$ values is $\tau_p/\tau_e = (m_p/m_e)^{2.5} = 1.4 \times 10^8$. This large difference suggests that synchrotron radiation by protons could play an appreciable role in regions with large magnetic fields (e.g. near black holes and neutron stars), where relativistic electrons are not accelerated as efficiently or rapidly lose their energy as a consequence of strong synchrotron losses and/or inverse Compton scattering. The spectrum of synchrotron radiation by a relativistic particle outside the region of self absorption is $P(\nu) = (\sqrt{3}\,e^3 H_\perp/mc^2)F(x)$. Note that, when $\nu \ll \nu_c$, the spectral density $P(\nu) \propto (H_\perp/E)^{2/3}\nu^{1/3}$, and does not depend on m; i.e., the spectral radiation densities for electrons (positrons) and protons with the same energy are the same.

The question of how realistic it is to observe synchrotron radiation by protons with very high brightness temperatures also depends on whether there are relativistic electrons (positrons) in the source that could absorb this radiation. However, as noted above, the density of relativistic electrons (positrons) should be very low due to the enormous energy losses expected in regions of strong magnetic fields.

3. THE LIMITING INTENSITY FOR RADIATION BY RELATIVISTIC PARTICLES MOVING ALONG CURVED MAGNETIC-FIELD LINES (CURVATURE RADIATION)

It is probable that this mechanism is predominant in pulsars, and it is also possible for the magnetospheres of supermassive black holes. Curvature radiation is described by formulas that are analogous to those for synchrotron radiation. The radius of curvature of the magnetic lines of force takes the place of the Larmor radius for the motion of a particle of the same energy in an equivalent magnetic field $\rho_H = E/eH_\perp$. The radius of curvature of the lines of force of a dipolar magnetic field is $\rho_c = 4R^2/3r$, where R and r are the distances from the center and the dipole axis, respectively. In this case, the equivalent magnetic field transverse to the particle motion is

$$H_\perp = 3rE/4R^2e. \qquad (22)$$

The magnitude of this projection could be appreciably smaller than that of the overall magnetic field (the radius of curvature of the magnetic force lines could be substantially smaller than the radius for Larmor motion). Therefore, the expected limiting intensity for curvature radiation could be appreciably higher than that for synchrotron radiation. For a monoenergetic particle spectrum, from (11) and (22),

$$H_\perp = kTrR^{-2}e^{-1}. \qquad (23)$$

In this case, we have from (4)

$$T \leq a^{7/8}(mc^2/k)(mc^2e^{-2}R^2r^{-1})^{1/8}. \qquad (24)$$

The characteristic frequency is

$$\nu_c = (3c/4\pi)(E/mc^2)^3/\rho_c. \qquad (25)$$

For a supermassive black hole with $M = 10^9 M_\odot$, $R = 10R_g = 3 \times 10^{15}$ cm $= r$, and $\rho_c = 4 \times 10^{15}$ cm, we find $T_e \leq 1.4 \times 10^{13}$ K and $T_p \leq 6.7 \times 10^{16}$ K. If we estimate the frequency range in which such high brightness temperatures can be realized by substituting the value for $H_\perp$ from (23) into (19), we obtain

$$\nu \ll \nu_c = (4/3\pi)(a^{21}m^3c^{14}/e^6)^{1/8}(R^2/r)^{-5/8}. \qquad (26)$$

For this example, this yields for electrons (positrons) $\nu < 6 \times 10^4$ Hz and for protons $\nu < 10^6$ Hz. A more acceptable frequency range could come about as a result of Doppler boosting, which we will consider in Section 5. For a

pulsar, adopting $R \sim 10^7$ cm and $r \sim 3 \times 10^6$ cm, we obtain $T_e < 1.4 \times 10^{14}$ K, $\nu_e < 5.5 \times 10^9$ Hz, $T_p < 6.7 \times 10^{17}$ K, and $\nu_p < 9 \times 10^{10}$ Hz. These values for T and ν could be enhanced in the same way, as discussed below.

4. THE SYNCHRO-COMPTON LIMIT WITH CONSTANT INJECTION OF RELATIVISTIC PARTICLES

Let us suppose that there is a constant injection of relativistic particles with energy E_0 into an emitting volume with a magnetic field. Let these particles leave the volume after a time t, during which they lose an appreciable fraction of their energy to inverse Compton scattering [15]. The final energy of the particles

$$E = (4\pi\nu_c^3\sigma_T m^{-2}c^{-6}t)^{-0.5} \tag{27}$$

does not depend on E_0 (here, $\sigma_T = 4\pi e^4/(3m^2c^4)$ and, from (11), the brightness temperature of their synchrotron radiation, taking into account self-absorption, is

$$T \le 0.75 m^2 c^5 (4\pi e^2 k)^{-1} \nu_c^{-1.5} (2t/3)^{-0.5} \text{ K}, \tag{28}$$

where ν_c is determined by Eq. (9) for ordinary synchrotron radiation or by (25) for curvature radiation. We obtain from (28) for electrons (positrons) and protons

$$T_e \le 3 \times 20^{13} (\nu_c/30 \text{ GHz})^{-1.5} (t/1 \text{ day})^{-0.5} \text{ K},$$
$$T_p \le 1 \times 10^{20} (\nu_c/30 \text{ GHz})^{-1.5} (t/1 \text{ day})^{-0.5} \text{ K}. \tag{29}$$

The time scale $t \sim 1$ day is typical of the observed variability of the synchrotron radiation of some active galactic nuclei. The characteristic particle energy from (11) and (29) for $\nu_c = 30$ GHz and $t = 1$ day is

$$E_e = 3.5 \times 10^9 \text{ eV}, \quad E_p = 1.2 \times 10^{16} \text{ eV}. \tag{30}$$

The characteristic magnetic field

$$H_\perp = (4\pi/3)e^{-1}m^3c^5\nu_c E^{-2}$$
$$= (4\pi)^3 3^{-2} e^3 m^{-1} c^{-5} \nu_c^4 t \tag{31}$$

and the radius of curvature

$$\rho_c = 3c(4\pi\nu_c)^{-1}(E/mc^2)^3$$
$$= 3^{2.5}(4\pi)^{-4}e^{-6}m^3c^{10}\nu_c^{-5.5}t^{-1.5} \tag{32}$$

are 1.5×10^{-4} G and 7.5×10^{10} cm for electrons (positrons) and 0.28 G and 4.6×10^{20} cm for protons. These could have other observed values, as shown in Section 5. For pulsars, assuming a characteristic working frequency $\nu_c \sim 1$ GHz and a characteristic time of

the order of a pulse duration $t \sim 0.01$ s, we have the following limits for the brightness temperature:

$$T_e \le 1.5 \times 10^{19}(\nu_c/1 \text{ GHz})^{-1.5}(t/0.01 \ c)^{-0.5} \text{ K},$$

$$T_p \le 4.9 \times 10^{25}(\nu_c/1 \text{ GHz})^{-1.5}(t/0.01 \ c)^{-0.5} \text{ K}, \tag{33}$$

$$E_e = 1.7 \times 10^{15} \text{ eV} \quad \text{and} \quad E_p = 5.7 \times 10^{21} \text{ eV}.$$

The magnitude of $H_\perp$ turns out to be unrealistically small, and ρ_c to be unrealistically large. However, taking into account possible coherence of the radiation, as discussed in Section 5, could yield more reasonable values.

Relations (29) and (30) predict very high brightness temperatures for the radio emission of quasars, accompanied by outflows of relativistic particles [electrons (positrons) or protons with energies to 10^{16} eV]. Based on the determining role of Compton losses [relation (27)], the power of the flow of relativistic particles should be larger than or equal to the radio power.

The above arguments are also relevant for an analogous description of the radio emission of pulsars using (33); here, the energy of the protons turns out to be close to the maximum energy for cosmic rays detected near the Earth ($\sim 10^{21}$ eV). It is also important to note that, in this model, synchrotron radiation by protons is not reduced by synchrotron absorption by electrons (positrons). Indeed, when losses to inverse Compton scattering dominate, the particle energy will be $E(t) = E_0/(1 + E_0\beta t) \longrightarrow (\beta t)^{-1}$, where $\beta = 32\pi e^4\rho_s/9m^4c^7$ and ρ_s is the energy density of the synchrotron radiation. The energy ratio is $E_p/E_e = (m_p/m_e)^4$; that is, if protons corresponding to the limiting synchrotron intensity are generated, most likely there will not be relativistic electrons (positrons). According to (9), if relativistic electrons (positrons) appear, the synchrotron frequency ratio is $\nu_p/\nu_c = (E_p/E_e)^2(m_e/m_p)^3 = (m_p/m_c)^5 = 2 \times 10^{16}$. In the same way, we have for curvature radiation from (25) $\nu_p/\nu_c = (E_p/E_e)^3(m_e/m_p)^3 = (m_p/m_e)^9 = 2 \times 10^{29}$. Thus, in this model, the range for the radiation and absorption of electrons (positrons) is many orders of magnitude lower than that for protons.

5. COHERENT RADIATION AND DOPPLER BOOSTING

Relation (4) can be used to estimate limiting brightness temperatures for synchrotron radiation by collections of N particles with size less than the wavelength, mass Nm, and charge Ne. In this case, the resulting brightness temperature grows by a factor of $N^{6/7}$ for synchrotron radiation or a factor of $N^{7/8}$ for curvature radiation. It is important to note that the power of the radiation here takes into account self-absorption. For an optically thin layer, the power of coherent radiation

from a cloud of N particles is a factor of N^2 higher than the corresponding power from N independent particles. In the case of coherent synchrotron radiation and a monoenergetic particle spectrum, we must have $\nu \ll \nu_c \propto H_\perp^{5/7} N^{-2/7}$ in relation (14); i.e., a large magnetic field is required for a given ν_c. Analogous to (26), we have for radiation in a curved magnetic field $\nu < \nu_c \propto \rho_c^{-5/8} N^{-3/8}$; i.e., a smaller curvature radius ν_c is required for a given ρ_c for coherent radiation.

In the case of a constant injection of particles into the emitting region, Eq. (28) does not depend on N; however, in (31), $H_\perp \propto N^2$ and, in (32), $\rho_c \propto N^{-3}$, making it possible to find consistency in the expected parameters over a wide range of values. For example, for the example described by (33), for $\rho_c = 10^9$ cm and $\nu_c = 1$ GHz, we would require $N = 10^7$ for electrons (positrons) and $N = 1.6 \times 10^{10}$ for protons.

Doppler boosting and deboosting of the brightness temperature in distant sources due to their redshift is described by the factor [22]

$$D = \delta/(1 + z), \qquad (34)$$

where $\delta = \gamma^{-1}(1 - \beta\cos\theta)^{-1}$, $\gamma = (1 - \beta^2)^{-0.5}$, $\beta = v/c$, v is the velocity of a source of radiation moving at an angle θ to the line of sight, z is the source's redshift. D is the ratio of the brightness temperature in the frame and at the observer's frequency to the temperature at the emission frequency in the frame near the source for which the synchrotron emission is isotropic. The intensity ratio in these systems is $I_\nu(\nu)/I_\nu(\nu/D) = D^3$. D is also the ratio of the observing and emission frequencies and the ratio of variability time scales in the same source frame and in the observer's frame.

As was already noted, according to variability observations of some extragalactic radio sources, the peak brightness temperature reaches 10^{21} K, i.e., a factor of 10^9 larger than 10^{12} K, the limit for synchrotron emission of electrons. In this case, to estimate the source size, we assume that $D \sim 1$. If $D \gg 1$, then the source's solid angle in the observer's frame for the model of a central explosion increases as $\propto D^2$, and 10^9 can be reached at $D = 10^3$, while $D = 3 \times 10^4$ for a one-dimensional ejection. Both estimates are unlikely.

Since many extragalactic sources show superluminal motions in their jets, we expect that $\delta \sim 10$, which yields for $z \ll 1$ and D an increase in the brightness temperature and a spectral shift of a factor of ten. An analogous effect could be important for pulsars.

6. CONCLUSION

We have presented relations determining the limiting brightness temperatures for radiation from an ensemble of relativistic particles. Our estimates show that the brightness temperature could be appreciably higher than the observed lower limits obtained from interferometric studies of the nuclei of extragalactic sources ($\sim 10^{12}$ K), which could explain the high temperatures expected based on observations of rapid variability of their fluxes. Similar estimates for pulsars yield brightness temperatures of 10^{25} K and higher. Important parameters that could bring about high brightness temperatures are the magnetic field strength and the radius of curvature of the magnetic lines of force; other possible factors include synchrotron radiation by relativistic protons, coherent radiation by groups of relativistic particles, and Doppler boosting of the radiation intensity for particles in a jet oriented close to the line of sight to the observer. The largest brightness temperatures are obtained for the non-steady-state case considered in Section 4. This model is quite plausible for both active galactic nuclei and pulsars. In this case, the brightness temperature could also be enhanced by Doppler boosting.

Synchrotron radiation by protons is a very probable mechanism for both active galactic nuclei and pulsars. The proton energy required for pulsars proves to be close to the limiting energy for near-Earth cosmic rays. The mechanism for acceleration of the protons (probably an induction mechanism) requires further study. As shown in [23], relativistic particles of even higher energies can be generated near supermassive black holes in galactic nuclei. The high brightness temperatures of the radio emission of such sources make it possible to study these interesting, very compact astronomical objects with very high angular resolution using space radio interferometers.

REFERENCES

1. G. A. Moellenbrock, K. Fujisawa, R. A. Preston, *et al.*, Astron. J. **111**, 2174 (1996).

2. G. C. Bower and D. C. Backer, Astrophys. J. Lett. **507**, L117 (1998).

3. L. Kedziora-Chudczer, D. L. Jauncey, M. H. Wieringa, *et al.*, Astrophys. J. Lett. **490**, L9 (1997).

4. G. E. Romero, G. Surpi, and H. Vucetich, Astron. Astrophys. **301**, 641 (1995).

5. S. J. Wagner, A. Witzel, J. Heidt, *et al.*, Astron. J. **111**, 2187 (1996).

6. J. Machalski and D. Engels, Mon. Not. R. Astron. Soc. **266**, L69 (1994).

7. N. S. Kardashev, Exp. Astron. **7**, 329 (1977).

8. K. I. Kellermann and I. I. Pauliny-Toth, Astrophys. J. Lett. **155**, L71 (1969).

9. F. Pacini and M. Rees, Nature **226**, 622 (1970).

10. I. S. Shklovsky, Astron. Zh. **40**, 972 (1964) [Sov. Astron. **7**, 748 (1964)].

11. L. Woltjer, Astrophys. J. **146**, 597 (1966).

12. M. Rees, Nature **211**, 468 (1966).

13. H. van der Laan, in *Les Noyaux of Galaxies*, Ed. by D. O'Connell (Pontificia Academia Scientiarum, Vatican City, 1971), p. 245.

14. R. D. Blandford and A. Königl, Astrophys. J. **232**, 34 (1979).

15. V. I. Slysh, Astrophys. J. **391**, 453 (1992).

16. A. Readhead, Astrophys. J. **426**, 51 (1994).

17. A. Güijosa and R. A. Daly, Astrophys. J. **461**, 600 (1996).

18. A. G. Pacholczyk, *Radio Astrophysics: Nonthermal Processes in Galactic and Extragalactic Sources* (Freeman, San Francisco, 1970; Mir, Moscow, 1973).

19. K. R. Lang, *Astrophysical Formulae: a Compendium for the Physicist and Astrophysicist* (Mir, Moscow, 1978; Springer, Berlin, 1999).

20. S. L. O'Dell and L. Sartori, Astrophys. J. Lett. **162**, L37 (1970).

21. J. D. Jukes, Nature **216**, 461 (1967).

22. K. Lind and R. Blandford, Astrophys. J. **295**, 358 (1985).

23. N. S. Kardashev, Mon. Not. R. Astron. Soc. **276**, 515 (1995).

Translated by D. Gabuzda

Determination of Molecular Gas Properties Using Methyl Cyanide Lines

S. V. Kalenskiĭ[1], V. G. Promislov[1], A. V. Alakoz[1], A. Winnberg[2], and L. E. B. Johansson[2]

[1]*Astro Space Center, Lebedev Physical Institute, ul. Profsoyuznaya 84/32, Moscow, 117810 Russia*

[2]*Onsala Space Observatory, Onsala, Sweden*

Received November 1, 1999

Abstract—A survey has been made of 27 Galactic star-forming regions in the (CH_3CN) 6_K–5_K, 5_K–4_K, and 8_K–7_K lines of methyl cyanide (CH_3CN) at 110, 92, and 147 GHz. Twenty-five sources were detected at 110 GHz, nineteen at 92 GHz, and three at 147 GHz. The strongest CH_3CN emission arises in hot cores in regions of massive star formation. The abundance of CH_3CN in these objects exceeds 10^{-9} as a consequence of grain mantle evaporation. Weaker CH_3CN lines were found in a number of sources. These can arise in either warm (30–50 K), dense ($>10^4$ cm^{-3}) clouds or in hot regions with cooler gas. © 2000 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

Cosmic methyl cyanide (CH_3CN) radio emission was first detected by Solomon *et al.* [1]. Like other symmetric-top molecules such as ammonia or methyl acetylene, methyl cyanide is a good probe of the interstellar gas temperature. Since radiative transitions between different K ladders are prohibited in symmetric-top molecules, the populations of different K ladders are determined by collisions, and the relative intensities of lines with different quantum numbers K are strongly linked to the kinetic temperature. CH_3CN has a large number of rotational transitions, which are grouped in series of closely spaced lines with the same J but different K quantum numbers. These lines can be observed simultaneously with a single receiver and spectrometer. Therefore, the kinetic temperature estimates are not affected by calibration and pointing errors [2, 3]. The spectral characteristics are well known [3, 4].

With its high dipole moment (3.91 *D*), CH_3CN is a tracer of dense gas. It is known that CH_3CN radio emission often arises in hot gas, in particular in hot cores—hot and dense neutral regions a few hundredths of a parsec in size. Interferometric observations of several strong sources have shown that hot-core emission dominates in their spectra, but is sometimes accompanied by much weaker halo emission ([5] and references therein). The reason for the strong emission in hot cores is grain-mantle evaporation, which leads to a substantial increase in the CH_3CN abundance. This can occur due to the evaporation of methyl cyanide itself or to a chain of gas-phase reactions starting from HCN or some other nitrogen-bearing molecule that appears in the gas phase due to mantle evaporation [6]. Apart from hot cores, CH_3CN has been detected toward warm clouds (30–50 K) in regions of massive star formation

[7] and in dark clouds [8]. In spite of the relatively large number of interferometric observations of hot cores in various lines, only a few surveys have been published, each devoted only to a small number of objects [7, 9, 10]. Therefore, following our previous studies of molecular clouds in various methanol lines [11, 12], we made a more extensive survey of Galactic star-forming regions in the 6_K–5_K and 5_K–4_K methyl cyanide lines near 110 and 92 GHz, respectively. Three sources were also observed in the 8_K–7_K lines near 147 GHz.

2. OBSERVATIONS

The observations at 92 and 110 GHz were carried out in April–May 1996 using the 20-m millimeter telescope of the Onsala Space Observatory. The pointing accuracy was checked using observations of SiO masers, and was found to be within 5″. The main beam's efficiency and half-power beamwidth (HPBW) at 92 GHz were 0.55″ and 39″, respectively. The observations were performed in a dual-beam switching mode with a switching frequency of 2 Hz and a beam separation of 11′. In no case did we detect absorption features indicating that the OFF beam intersected another source component. A cryogenically cooled low-noise SIS mixer was used at both frequencies. The single sideband receiver noise temperature was about 150 K. The system noise temperature corrected for atmospheric absorption, rearward spillover, and radome losses varied at both frequencies between 350 and 2000 K, depending on the weather conditions and source elevation. The data were calibrated using a standard chopper–wheel method. The backend consisted of two parallel filter spectrometers: a 256-channel spectrometer with 250 kHz frequency resolution and a 512-channel spectrometer with 1 MHz frequency resolution. The

Table 1. Frequencies and strengths of CH_3CN lines

Transition	Frequency[1], MHz	Line strength
$5_0–4_0$	91987.090	5.00
$5_1–4_1$	91985.317	4.80
$5_2–4_2$	91980.089	4.20
$5_3–4_3$	91971.310	3.20
$5_4–4_4$	91959.024	1.80
$6_0–5_0$	110383.522	6.00
$6_1–5_1$	110381.404	5.83
$6_2–5_2$	110375.052	5.33
$6_3–5_3$	110364.469	4.50
$6_4–5_4$	110349.659	3.33
$6_5–5_5$	110330.627	1.83
$8_0–7_0$	147174.592	8.00
$8_1–7_1$	147171.756	7.88
$8_2–7_2$	147163.248	7.50
$8_3–7_3$	147149.073	6.88
$8_4–7_4$	147129.236	6.00
$8_5–7_5$	147103.744	4.87

[1] Frequencies taken from the database http://physics.nist.gov/.

two objects NGC 1333/IRAS2 and L1157 were observed only with the 1 MHz resolution.

The observations at 147 GHz were carried out on October 4, 1998 using the 12-m telescope of the NRAO[1] at Kitt Peak in a remote-observing mode from the Astro Space Center in Moscow. The observations were performed in a position switching mode with a position separation of 5′. A cooled SIS-mixer dual-polarization channel receiver was used. Both polarizations were observed simultaneously. The data were calibrated using a standard vane method. The system noise temperature corrected for atmospheric absorption, rearward spillover, and ohmic losses varied between 300 and 500 K. The pointing accuracy was found to be within 5″. The main-beam efficiency and HPBW were 0.54″ and 42″, respectively. The backend consisted of two 512-channel filter bank spectrometers with 1 MHz frequency resolution connected to the different polarization channels.

The frequencies and strengths of the observed lines are presented in Table 1. Twenty-seven sources were observed at 110 GHz. Of these, 19 were observed at 92 GHz and three at 147 GHz. We reduced the spectra using the Grenoble CLASS package. We obtained Gaussian fits assuming that the different K components in each series have common LSR velocities and linewidths. The results are presented in Table 2.

The five objects Ori S6, G30.8–0.1, G34.26+0.15, W51E1/E2, and DR 21(OH) were mapped at 110 GHz. We made maps of the integrated intensity in the velocity range occupied by the blend of the $6_0–5_0$ and $6_1–5_1$ lines. The mapping data were obtained during excellent weather conditions, with a system noise temperature of about 350 K. The images were reconstructed using the maximum entropy technique using a method similar to that described by Wilczek and Drapatz [13], except for the computational algorithm to solve the optimization problem. The evolutionary algorithm used here is described in [14]. We estimated the effective HPBW of the reconstruction in a manner similar to that used in [15]. At each point of a map, we subtracted the contribution from the source, leaving only noise fluctuations. Further, for each point, we added the contribution from a "test" point source with flux equal to that of the real source convolved with the beam, located at the position of the brightness peak of the maximum entropy map. We applied a maximum entropy procedure to the maps produced in this manner to derive the brightness distribution, whose full width at half maximum FWHM was adopted as the effective HPBW of the reconstruction.

3. RESULTS

Emission at 110 GHz was detected toward 25 of the 27 objects observed (S252, W48, and DR 21 West were only marginally detected). Sixteen objects were detected at 92 GHz and three at 147 GHz. The spectra are presented in Figs. 1–3. The source coordinates and Gaussian parameters of the detected lines are given in Table 2.

Figure 4 shows the maximum entropy images. For a number of reasons (mainly the relatively large telescope beam and insufficient sampling and sensitivity), we could not resolve the sources even using the maximum entropy technique. On our maps, the FWHM of the image of each source proved to be very close to the corresponding effective HPBW of the reconstruction. Thus, the results of the mapping are consistent with compact source structures, supporting the indirect size estimates given in Table 4.

In addition to the lines of the main isotope of CH_3CN, we detected a weak blend of the $5_0–4_0$ and $5_1–4_1$ $CH_3^{13}CN$ lines at 91941.596 and 91939.834 MHz toward G34.26+0.15. The integrated intensities of the $K = 0$ and $K = 1$ lines are 0.27 and 0.31 K km s^{-1}. The $CH_3^{13}CN/CH_3^{12}CN$ line-intensity ratio, about 0.15, is much higher than the ratio of the abundances of ^{13}C and ^{12}C, strongly suggesting that the lines are optically thick. This was confirmed by our analysis of rotational diagrams and statistical equilibrium calculations (see below). Similar ratios between the $6_K–5_K$ $CH_3^{13}CN/CH_3^{12}CN$ line intensities in G34.26+0.15 were found in [19].

Table 2. Gaussian parameters for detected CH_3CN lines with 1σ[1] errors

Source	R.A.$_{B1950.0}$ DEC$_{B1950.0}$	$\int T_A^* dV$, K km s^{-1}						V_{LSR}, km s^{-1}	ΔV, km s^{-1}
		$K=0$	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$		
W3(OH)	$02^h23^m17\overset{s}{.}3$	0.97(0.06)	1.06(0.06)	0.80(0.01)	0.79(0.04)	0.24(0.04)	<0.12	−47.5(0.2)	8.4(0.2)
	61°38′58″	0.61(0.05)	0.47(0.05)	0.29(0.04)	0.27(0.04)	<0.09		−47.5(0.2)	6.9(0.3)
NGC 1333	03 25 55.0	<0.18							
	34 04 00	Not observed							
Ori S6	05 32 44.8	2.12(0.08)	2.00(0.08)	1.03(0.07)	0.83(0.07)	<0.21		6.9(0.1)	4.2(0.1)
	−05 26 00	0.97(0.07)	0.86(0.07)	0.53(0.07)	0.31(0.07)	<0.3		6.8(0.2)	4.7(0.2)
		0.77(0.07)	0.68(0.07)	0.36(0.07)	0.48(0.07)	<0.21		6.7(0.2)	4.4(0.2)
OMC 2	05 32 59.9	0.35(0.06)	0.44(0.07)	<0.18				11.2(0.1)	2.1(0.2)
	−05 11 29	<0.11							
S231	05 35 51.3	0.33(0.04)	0.31(0.04)	0.18(0.04)	0.23(0.04)	<0.12		−15.9(0.2)	3.0(0.0)
	35 44 16	0.27(0.06)	0.28(0.07)	0.14(0.07)	<0.09			−15.6(0.4)	3.8(0.5)
S235	05 37 31.8	0.31(0.06)	0.26(0.05)	0.22(0.05)	<0.15			−17.1(0.6)	4.7(0.4)
	35 40 18	0.22(0.04)	<0.1					−17.1(0.0)	3.7(0.0)
S252	06 05 53.7	0.24(0.08)[2]	0.18(0.08)[2]	<0.24				2.9(0.9)	5.0(0.0)
	21 39 09	<0.09							
S269	06 11 46.5	<0.06							
	13 50 39.0	Not observed							
NGC 2264	06 38 24.9	0.49(0.04)	0.41(0.04)	0.22(0.04)	0.17(0.04)	<0.12		8.5(0.2)	4.0(0.2)
IRS1	09 32 28	0.34(0.07)	0.26(0.07)	<0.2				7.8(0.5)	4.0(0.0)
G29.95−0.02	18 43 27.1	0.53(0.09)	0.73(0.09)	0.51(0.10)	0.39(0.09)	0.39(0.09)	<0.27	97.6(0.2)	5.5(0.4)
	−02 42 36	Not observed							
G30.8−0.1	18 45 11.0	3.09(0.11)	2.44(0.10)	1.74(0.08)	1.63(0.08)	0.58(0.07)	0.21(0.07)	98.9(0.1)	7.0(0.0)
	−01 57 57	2.18(0.16)	1.36(0.15)	1.18(0.08)	1.10(0.09)	0.32(0.08)		99.0(0.2)	7.9(0.3)
G34.26+0.15	18 50 46.1	3.72(0.15)	2.97(0.15)	2.20(0.14)	2.41(0.14)	1.31(0.13)	0.65(0.01)	58.7(0.1)	6.0(0.1)
	01 11 12	2.22(0.13)	2.06(0.12)	1.29(0.09)	1.28(0.10)	0.61(0.09)		58.5(0.2)	6.9(0.2)
G35.19−0.74	18 55 40.8	0.75(0.06)	0.88(0.07)	0.36(0.06)	<0.18			33.7(0.2)	4.6(0.2)
	01 36 30	Not observed							
W48	18 59 13.8	0.32(0.08)[2]	0.15(0.07)[2]	<0.2				43.9(0.6)	4.2(0.6)
	01 09 20.0	Not observed							
W49N	19 07 49.9	1.71(0.89)	1.69(0.98)	0.81(0.32)	1.59(0.33)	<0.93		9.4(1.6)	16.0(1.3)
	09 01 14	1.05(0.19)	0.90(0.15)	0.99(0.15)	1.16(0.15)	<0.45		9.2(1.0)	19.8(1.1)
W51E1/E2	19 21 26.2	4.89(0.16)	4.21(0.15)	3.34(0.10)	3.76(0.10)	1.87(0.10)	0.67(0.07)	56.8(0.1)	9.5(0.1)
	14 24 43	2.86(0.11)	2.13(0.11)	1.63(0.11)	1.87(0.08)	0.72(0.08)		57.3(0.1)	10.8(0.2)
W51 Met3	19 21 27.5	0.65(0.14)	0.59(0.13)	0.50(0.10)	0.42(0.11)	<0.27		55.5(0.6)	8.0(0.7)
	14 23 52	Not observed							
Onsala 1	20 08 09.9	0.56(0.05)	0.55(0.05)	0.24(0.05)	<0.12			11.9(0.2)	4.7(0.2)
	31 22 42	<0.18							
W75N	20 36 50.0	Not observed							
(0″, 0″)	42 26 58	Not observed							
		0.75(0.03)	0.67(0.03)	0.46(0.03)	0.47(0.03)	0.34(0.03)	0.09(0.03)	9.1(0.1)	5.6(0.1)
W75N	20 36 50.4	0.86(0.07)	0.84(0.07)	0.35(0.06)	0.36(0.06)	<0.18		9.4(0.2)	4.8(0.2)
(6″, 25″)	42 27 23	0.39(0.06)	0.30(0.06)	<0.18				9.6(0.3)	4.8(0.5)

Table 2. (Contd.)

Source	R.A.$_{\cdot B1950.0}$ DEC.$_{\cdot B1950.0}$	$\int T_{\mathrm{A}}^{*} dV$, K km s^{-1}						V_{LSR}, km s^{-1}	ΔV, km s^{-1}
		$K=0$	$K=1$	$K=2$	$K=3$	$K=4$	$K=5$		
DR 21 West	20 37 07.8	0.19(0.06)^2	0.16(0.08)^2	<0.21				−2.1(1.2)	5.0(0.0)
	42 08 44.0	Not observed							
DR 21	20 37 13.0	0.61(0.09)	0.47(0.08)	0.16(0.08)	0.23(0.09)	<0.21		−2.1(0.2)	2.7(0.0)
	42 08 50.0	0.63(0.05)	0.49(0.05)	0.20(0.05)	0.11(0.05)	<0.15		−2.2(0.1)	3.7(0.2)
DR 21(OH)	20 37 13.8	1.53(0.06)	1.43(0.06)	0.74(0.05)	0.64(0.05)	<0.15		−3.1(0.1)	4.7(0.1)
	42 12 13	0.97(0.07)	1.03(0.07)	0.30(0.06)	0.25(0.06)	<0.21		−3.0(0.2)	4.7(0.0)
L1157	20 38 39.6	<0.09							
(0″, 0″)	67 51 33	Not observed							
L1157	20 38 41.0	0.23(0.05)	0.21(0.05)	<0.12				2.1(0.5)	4.9(0.6)
(20″, −60″)	67 50 33	Not observed							
S140	22 17 41.2	0.39(0.04)	0.30(0.04)	0.26(0.05)	<0.12			−6.8(0.1)	3.1(0.2)
	63 03 43	0.34(0.12)	0.14(0.09)	0.15(0.09)	<0.36			−7.0(0.6)	3.8(1.1)
Cep A	22 54 19.2	0.30(0.04)	0.31(0.04)	0.08(0.03)	0.17(0.04)	<0.12		−9.9(0.2)	3.6(0.3)
	61 45 47	0.24(0.05)	0.22(0.05)	<0.15				−10.3(0.4)	3.9(0.5)
NGC 7538	23 11 36.6	0.45(0.04)	0.42(0.04)	0.22(0.04)	0.25(0.04)	<0.12		−57.7(0.2)	4.0(0.2)
IRS1	61 11 50	0.25(0.05)	0.15(0.05)	<0.12				−57.3(0.5)	4.0(0.0)
NGC 7538S	23 11 36.1	1.06(0.05)	0.84(0.05)	0.57(0.05)	0.34(0.04)	<0.12		−55.4(0.1)	4.5(0.1)
	61 10 30	0.74(0.09)	0.41(0.09)	0.22(0.08)	<0.24			−55.6(0.4)	4.5(0.4)
		0.40(0.02)	0.33(0.02)	0.18(0.02)	0.16(0.02)	<0.05		−55.7(0.1)	5.1(0.2)

[1] For each source, the 110 GHz parameters are presented in the upper line and the 92 GHz parameters in the lower line. The 147 GHz parameters for Ori S6, W75N(0″, 0″), and NGC 7538S are presented in the third line.
[2] Detection at the sensitivity limit.

We detected the 1–0 ^{13}CO line in all the observed sources. In addition, we found HNCO, HCOOCH$_3$, and the deuturated species NH$_2$D and CH$_3$OD in several objects. These data will be discussed in other papers.

4. ROTATIONAL DIAGRAMS

We derived the molecular gas properties in two ways. First, a rotational diagram analysis, as described by Turner [20], was carried out. The corresponding Boltzmann plots are shown in Fig. 5. Since the sources are typically smaller than the beams and the beam filling factors at 110, 92, and 147 GHz are different, we analyzed different groups of lines separately. The same main features are present in the 92 and 110 GHz rotational diagrams.

Figure 5 shows that most of the plots cannot be fitted by straight lines. First of all, in some objects, the points corresponding to the $K = 3$ levels are much lower than the fitted lines. Examination of previously published data (e.g. the line intensities presented in [7] or [10])

shows the same $K = 3$ depression for a number of sources, which proved to be a rather common feature and must be explained. Symmetry requirements separate the molecule into two independent species, A and E, with the $K = 3n$ rotational levels belonging to the A species and the $K = 3n \pm 1$ levels belonging to the E species. The $K = 3$ levels belong to the A species, suggesting that the A species may be underpopulated relative to E; but, in this case, the $K = 0$ levels should be underpopulated as well, in contradiction with Fig. 5.

Presumably the depressed $K = 3$ values are due to high optical depth. We constructed rotational diagrams following [20]; i.e., we plotted the upper level population N_u divided by the level statistical weight g_u as a function of E_u/k, where E_u is the upper-level energy. The upper level statistical weight g_u is the product of the rotational degeneracy $g_J = (2J + 1)$, the K-level degeneracy g_K, and the reduced nuclear spin statistical weight g_I. In symmetric top molecules, the product $g_I g_K$ for the $K = 3n$, $n \neq 0$ levels is twice that for the
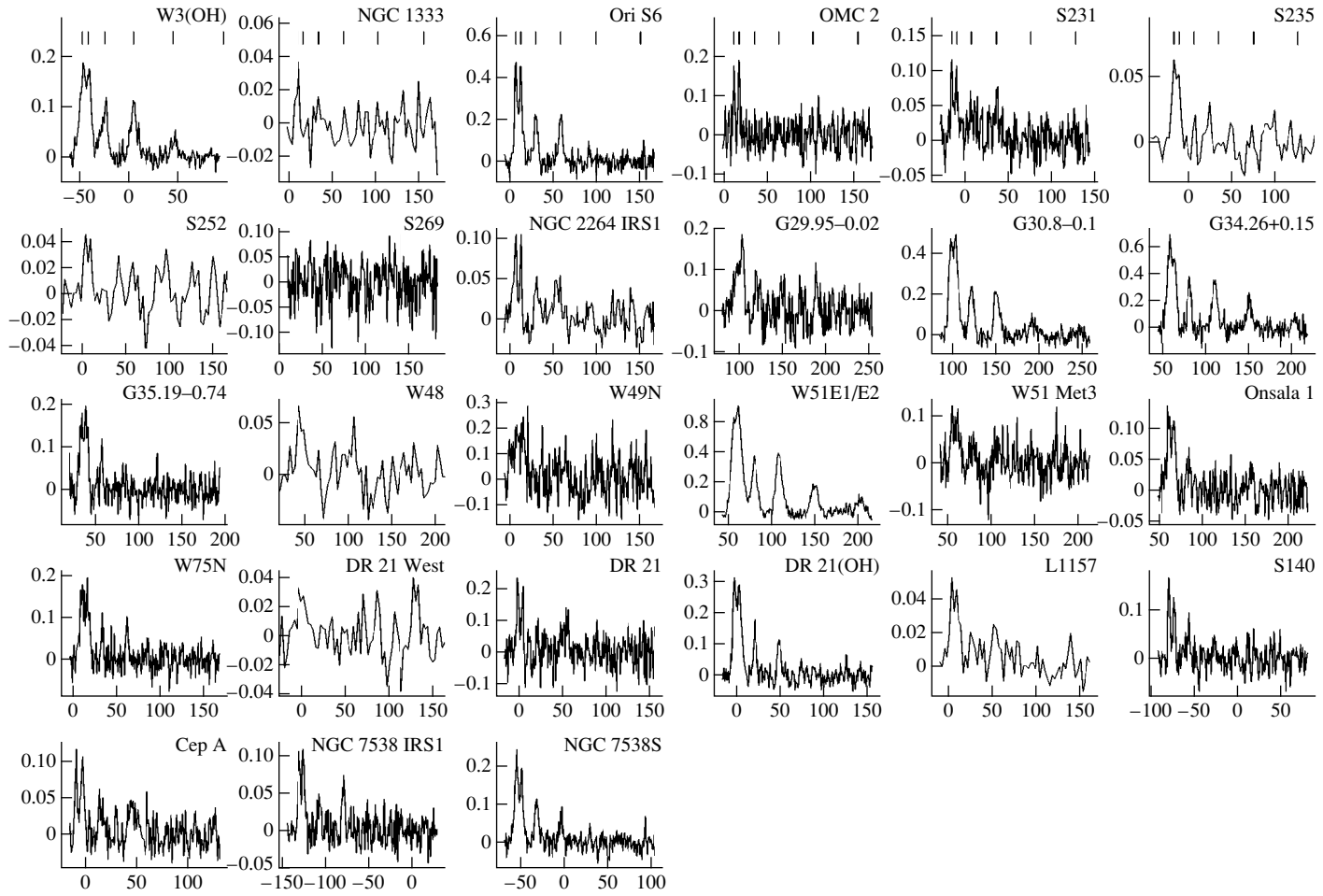
Fig. 1. Spectra of the sources observed at 110 GHz. X-axis: LSR velocity of the 6_0–5_0 line in km/s; Y-axis: antenna temperature in Kelvin. The vertical lines in the upper row indicate the positions of 0 to 5 K components, with the K values increasing to the right.
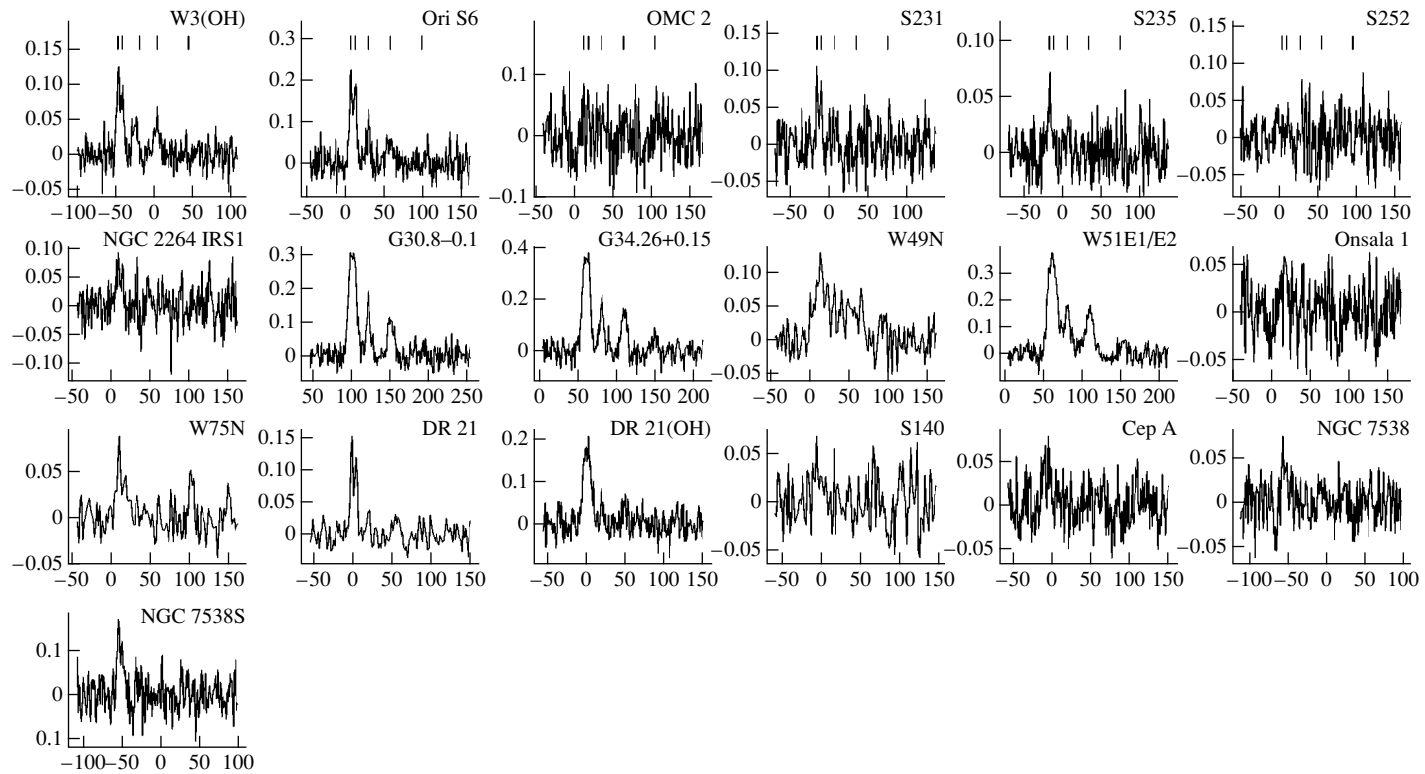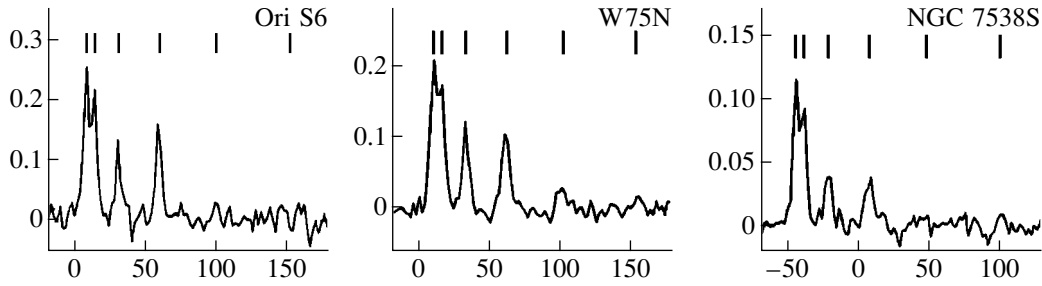
Fig. 2. Spectra of the sources observed at 92 GHz. X-axis: LSR velocity of the 5_0–4_0 line in km/s; Y-axis: antenna temperature in Kelvin. The vertical lines in the upper row indicate the positions of 0 to 4 K components, with the K values increasing to the right.

Fig. 3. Spectra of the sources observed at 147 GHz. X-axis: LSR velocity of the 8_0-7_0 line in km/s; Y-axis: antenna temperature in Kelvin. The vertical lines indicate the positions of 0 to 5 K components, with the K values increasing to the right.
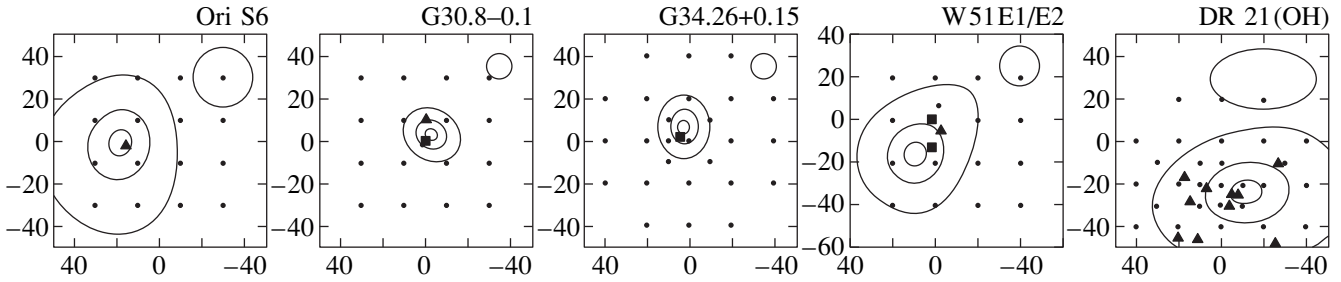
Fig. 4. Maximum entropy maps of five sources. The X and Y axes are offsets in right ascension and declination (arcsec) from the coordinates given in Table 2. The observed positions are marked by dots. The contours correspond to 10, 50, and 90% of the deconvolved integrated intensity peaks, which are 21.9 K km s^{-1} for Ori S6, 57.7 K km s^{-1} for G30.8–0.1, 77.6 K km s^{-1} for G34.26+0.15, 25.6 K km s^{-1} for W51E1/E2, and 7.3 K km s^{-1} for DR 21(OH). The effective HPBW of the reconstructions are represented by ellipses in the upper right corners. For DR 21(OH), the resolutions in the two coordinates are different due to the cosine effect. The filled squares show the positions of H_2O masers, and the filled triangles those of thermal methanol peaks taken from [16] (Ori S6), [17] (G30.8–0.1, G34.26+0.15, W51E1/E2), and [18] (DR 21(OH)).

other levels [20]. The level populations for rotational diagrams are calculated assuming optically thin lines:

$$\frac{N_u}{g_u} = \frac{3k \int T_{br} dV}{8\pi^3 \mu^2 \nu S_{ul} g_I g_K}. \qquad (1)$$

Here, μ is the permanent dipole moment, ν the line frequency, S_{ul} the line strength, k Boltzmann's constant, and T_{br} the brightness temperature. We used the main-beam brightness temperature T_{mb} instead of the unknown T_{br}. In the case of optically thick emission, Eq. (1) underestimates N_u/g_u; the higher the optical depth (proportional to the statistical weight of the upper level), the stronger the underestimation. Thus, the population of the more degenerate $K = 3$ levels is underestimated to a larger degree than is the population of adjacent levels.

The effect of optical depth is demonstrated in Fig. 5d, which shows model rotational diagrams for three parameter sets. We can see the same $K = 3$ depression in the diagram for high density and optically thick lines. Thus, like the marginal detection of the $^{13}CH_3CN$ lines, the rotational diagrams suggest that the lines are optically thick.

Figure 5d shows that there is no $K = 3$ depression for a density of 3×10^3 cm^{-3}, even in the case of optically thick lines. For low density and optically thick emission, the excitation of the 5_K-4_K and 6_K-5_K lines is determined by internal radiation rather than by collisions. The higher optical depth of the $K = 3$ lines leads to higher excitation and brightness temperatures for these transitions, and to the absence of the $K = 3$ depression. Examination of a large number of models showed that a significant $K = 3$ depression appears if the density is of the order of 10^5 cm^{-3} or larger. Hence, the absence of the $K = 3$ depression shows that either the lines are optically thin or the gas density is low.

The rotational temperature derived from optically thick lines is larger than the gas kinetic temperature. For example, the rotational temperatures for models 2 and 3 are 246 and 341 K, while the gas kinetic temperature is 100 K.

Table 3 presents the parameters derived from the rotational diagrams. Based on the discussion above, we divided the sources into two groups. The sources in group I are those that show significant $K = 3$ depressions. We conclude that the 110 and 92 GHz lines are optically thick, and the gas density is about 10^5 cm^{-3} or larger. Since rotational diagrams give overestimated temperatures for optically thick lines CH_3CN, the kinetic temperatures of the group I sources in Table 3 are upper limits. On the other hand, high optical depth,
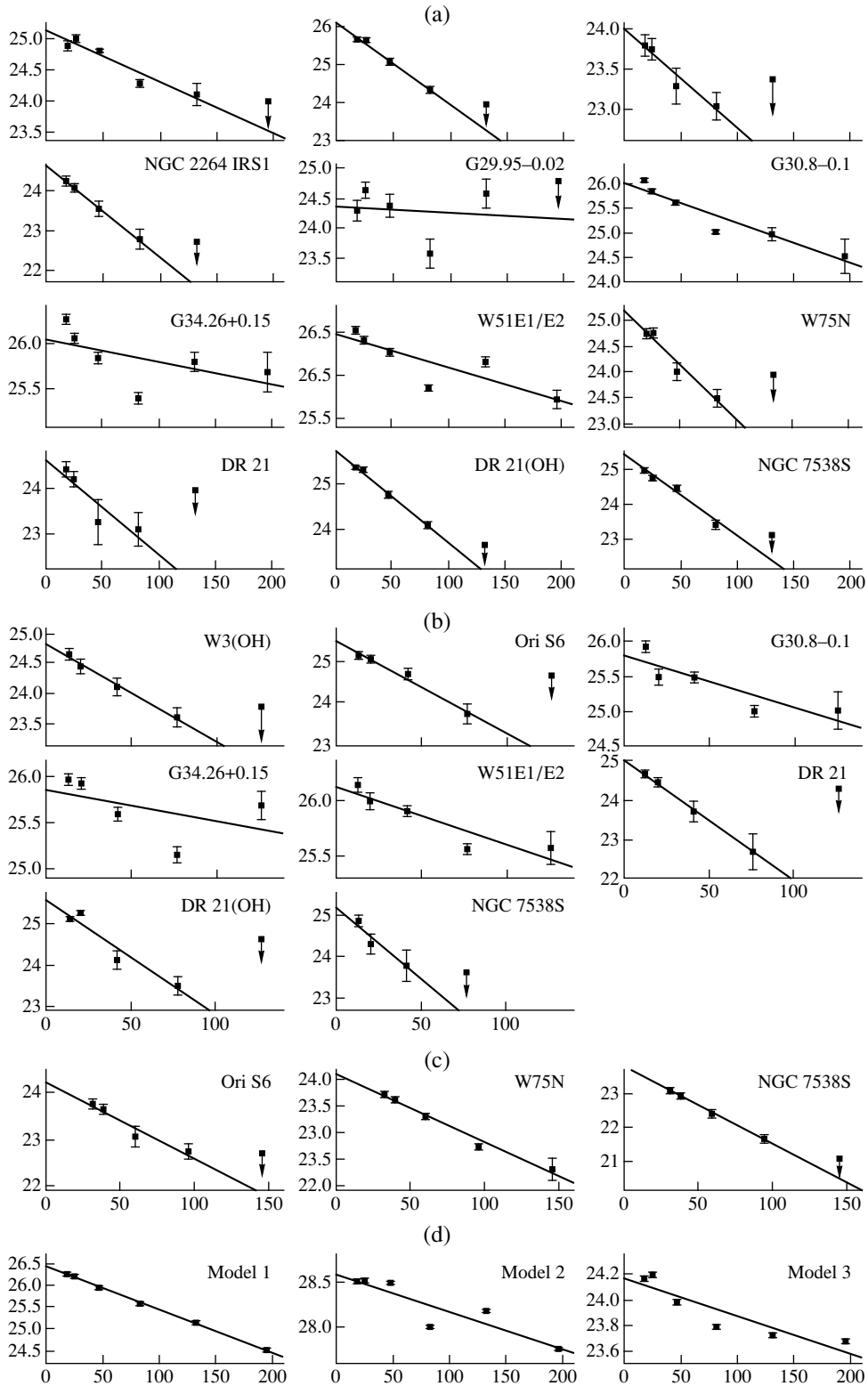
Fig. 5. Rotational diagrams constructed for the data at (a) 110, (b) 92, and (c) 147 GHz, and (d) for model rotational diagrams for the 110 GHz lines. $\ln(N_u/g_u)$ is plotted as a function of E_u/k (K). The arrows denote upper limits at the 3σ level. The parameters for Model 1 are $n_{H_2} = 10^6$ cm^{-3}, $T_{kin} = 100$ K, and optical depth of the 6_3–5_3 line $\tau_3 = 0.1$; for Model 2, $n_{H_2} = 10^6$ cm^{-3}, $T_{kin} = 100$ K, and $\tau_3 = 1.8$; and for Model 3, $n_{H_2} = 3 \times 10^3$ cm^{-3}, $T_{kin} = 100$ K, and $\tau_3 = 3.5$.

as well as beam dilution, leads to underestimation of the column density (note that both the maximum entropy images and the indirect size estimates in Table 4 demonstrate that the group I sources are very compact). Hence, the column densities of the group I sources in Table 3 are lower limits. The sources in group II are those where the $K = 3$ depression is either absent or cannot be detected because the $K = 4$ lines adjacent to the $K = 3$ lines were not measured. The reliability of the rotational temperatures and column densities of the group II sources is discussed in Section 7.

In G34.26+0.15 and, probably, G30.8–0.1 and W51E1/E2, the low excitation $K = 0$ and $K = 1$ levels have excess populations. The same effect can be noted in the model rotational diagram for low density (Fig. 5c). However, all three objects are group I sources, which we believe to be dense. Therefore, this excess indicates that they have complex structure. For simplicity, we will call them core–halo objects. Here, "halo" refers to any source that is cooler or less dense than the "core" sources. Core–halo emission structure has already been detected in some objects [5].

5. STATISTICAL EQUILIBRIUM CALCULATIONS

To derive the source properties via statistical equilibrium calculations, we used an approach similar to that applied in [7] and [9]. The LVG code was kindly made available by C.M. Walmsley and R. Cesaroni. The free parameters of the model are the gas kinetic temperature T_{kin}, molecular hydrogen number density n_{H_2}, and methyl cyanide density divided by the velocity gradient $n_{CH_3CN}/(dV/dR)$.

To obtain statistical equilibrium parameters, it is necessary to perform statistical equilibrium calculations for a number of parameter sets and choose those sets that are in agreement with the observational data. The agreement between our observations and the statistical equilibrium models can be evaluated by comparing the model line intensity ratios within the 110, 92, and 147 GHz series with the observed ratios. We used the ratios $R_i = T_{mb}(J_K - (J - 1)_K)/T_{mb}(J_2 - (J - 1)_2)$ for each line series.

We modeled the brightness temperatures of CH$_3$CN lines for kinetic temperatures 10–500 K, molecular hydrogen number densities 3×10^3–10^8 cm^{-3}, and densities divided by the velocity gradient 10^{-7}–10 cm^{-3}/(km s^{-1} pc^{-1}). Except for the microwave background, external radiation was neglected. The best-fit models are those that minimize the χ^2

$$\chi^2 = \sum_i \left(\frac{R_i^{obs} - R_i^{mod}}{\sigma_i^{obs}}\right)^2, \qquad (2)$$

Table 3. Parameters from rotational diagrams, rotational temperatures (from the 110, 92, and 147 GHz lines), and CH$_3$CN column densities (from the 110 GHz lines)

Source	Group	T_{rot}, K (110)	(92)	(147)	N_{CH_3CN}, 10^{12} cm^{-2}
W3(OH)	I[1]	120	63		
Ori S6	II	46	46	62	33
S231	II	80			9
NGC 2264	II	43			7
G29.95–0.02	I	1013			
G30.8–0.1	I	125	142		
G34.26+0.15	I	418	295		
G35.19–0.74	II	40			11
W49N	II	117			76
W51E1/E2	I	265	349		
W51 Met3	II	77			18
Onsala 1	II	36			7
W75N(0″, 0″)	II			78	
W75N(6″, 25″)	II	47			13
DR 21	II	23	32		5
DR 21(OH)	II	48	36		24
S140	II	42			18
Cep A	II	57			5
NGC 7538	II	62			9
NGC 7538S	II	42	30	44	15

[1] W3(OH) is a group I source, since it shows a $K = 3$ depression at 110 GHz, and the nature of this source (a hot core) is the same as that of other group I sources.

where σ_i^{obs} are the rms errors of the observed ratios. We calculated χ^2 for all models and then found the minimum values. The 1σ confidence levels were calculated as described in [21].

The minimum χ^2 values for G34.26+0.15, G30.8–0.1, and W51E1/E2 are much larger than expected based on the number of degrees of freedom of the model, confirming that these sources are inhomogeneous. Bearing in mind our suggestion that they have core–halo structure and arbitrarily assuming that the core contribution dominates in the $K \geq 2$ emission, we determined the core parameters using only the $K \geq 2$ lines (see below). This procedure virtually did not change the temperatures derived from all the observed lines. An interferometric spectrum of the hot core of G34.26+0.15 at 110 GHz [19] is well approximated by our core model, showing that this approach is reasonable.

Table 4. Source parameters from statistical equilibrium calculations: kinetic temperature, CH_3CN abundance, adopted distance, angular diameter, and density

Source	T_{kin}, K	X_{CH_3CN}, 10^{-10}	d, kpc	θ, $''$	n_{H_2}, 10^4 cm^{-3}
W3(OH)	70(50–95)	10–1000	2.2	2–4[1]	≥10
Ori S6	45(40–50)	0.3–30	0.5	≥10	10–100
S231	55(35–160)		2	≥3	
NGC 2264	45(25–65)		0.8	≥3	
G30.8–0.1[2]	60(45–75)	≥100	8	2–7	
G34.26+0.15[2]	140(65–220)	≥300	3.8	1–6[3]	
W51E1/E2[a]	160(120–180)	100–1000	8	1–3	≥30
W75N(6$''$, 25$''$)	40(35–60)		3	≥5	≥1
DR 21	30(20–55)		3	≥5	
DR 21(OH)	45(40–55)	1–100	3	≥8	≥3
NGC 7538	50(35–100)		3.5	≥3	
NGC 7538S	40(30–45)	10–100	3.5	5–9	10–100

[1] Less than 1$''$ according to [22].
[2] Parameters of the core.
[3] 3.9$'' \times$ 3.3$''$ according to [19].

Finally, for each model appropriate for a given source, we estimated the source angular size θ_S assuming a Gaussian brightness distribution:

$$\frac{\theta_S^2}{\theta_S^2 + \theta_{mb}^2} = \frac{T_{mb}}{T_{mod}}, \quad (3)$$

where T_{mod} is the model brightness temperature and θ_{mb} is the antenna beam FWHM. Using published distances for the sources, we estimated their linear sizes D, average velocity gradients $\Delta V/D$, and CH_3CN abundances

$$X_{CH_3CN} = n_m \Delta V/D n_{H_2}, \quad (4)$$

where n_m is the model $n_{CH_3CN}/(dV/dR)$ value.

Like previous workers in the field, we found that only the kinetic temperature can be determined fairly accurately using this method. For most sources, especially those in group II, we could not derive any densities and, hence, abundance estimates. To constrain these parameters, we left for further analysis only those models that agreed with the observed ratios between the 92, 110, and 147 GHz line intensities to within 20%.

Our statistical equilibrium calculations confirm that the $K = 3$ depression in the rotational diagrams appears as a result of the line opacities. For sources that were thought to be optically thick based on the $K = 3$ depression, only optically thick models are in good agreement with the observational data. The model gas temperatures were lower than the rotational temperatures, as expected for optically thick lines. For other objects, the kinetic temperatures proved to be close to the rotational temperatures, and either only optically thin or both

moderately optically thick and optically thin models are in agreement with the observational data.

The statistical equilibrium modeling for group I objects shows that only models for very compact sources are in agreement with the observed line intensity ratios. This conclusion is supported by the maximum entropy maps. The results of the statistical equilibrium analysis are summarized in Table 4.

6. COMMENTS ON SELECTED SOURCES

W3(OH). Wink *et al.* [22] published a map of this source in the 5_3–4_3 line obtained with the Plateau de Bure interferometer. This map revealed two compact objects—a strong clump located toward the water maser cluster and an HCN peak, known as the TW object (Turner and Welch [23])—and a much weaker object toward the ultracompact HII region. Our temperature estimate is in good agreement with that of [22]; however, we find a source size exceeding 2$''$, in contradiction with their upper limit of 1$''$ for the stronger clump. The lower limit for the core models was less than 1$''$. We suggest that our single-dish CH_3CN line ratios are affected by the contribution of the weaker clump and possible halo emission that is too weak or extended to be detected with the Plateau de Bure interferometer.

G34.26+0.15. An interferometer map of this well-known hot core [24] in the 6_K–5_K lines has been obtained by Akeson and Carlstrom [19]. The ratios of all line intensities for this source observed by us could not be fitted by homogeneous source models, suggesting a core–halo structure. The core model that agrees best with our data reproduces the interferometric spec-

trum of [19] fairly well. The inferred core temperature is about 140 K, although with a large uncertainty. This value is much lower than the 250 K estimate of [19]; the discrepancy is probably due to the different treatments of optical depth effects. We could not estimate the density from our data alone; however, the brightness temperature implied by the source size obtained by Akeson and Carlstrom [19] requires a density larger than 10^5 cm^{-3}.

W51E1/E2. The CH$_3$CN source was detected toward the radio continuum objects E1 and E2, which are associated with the H$_2$O and OH masers W51 Main/South, ammonia peaks NH$_3$–1 and NH$_3$–2 detected using the VLA [25], and emission by other molecules typical of hot cores [24]. The peak of the emission is coincident within the pointing error of 5″ with the NH$_3$–1 ammonia clump; however, the accuracy of the peak position is not high, since it is close to the edge of the mapped region. Without further observations, we cannot state with certainty that the peak is coincident with any specific object. The observed lines cannot be fitted using homogeneous source models. The core models give a kinetic temperature of about 160 K, a density exceeding 3×10^5 cm^{-3}, and a core diameter between 2.5″ and 4.9″.

7. DISCUSSION

Our statistical equilibrium analysis shows that the group I sources are very different from the others. These are the strongest objects in our sample, and have kinetic temperatures higher than the group II sources. All of them except G30.8–0.1 are well-known hot cores [22, 24]. The CH$_3$CN abundances in these objects are about equal to or larger than 10^{-9}, and are larger than those in the group II objects. Abundances of the order of 10^{-8} have been reported in hot cores, e.g. in Orion [26] and G10.47+0.03 [7]. It is now well established that the CH$_3$CN abundances in hot cores are enhanced due to the evaporation of molecular material from the icy mantles of interstellar grains. CH$_3$CN can be either a parent molecule or a product of a chain of gas-phase reactions, starting from HCN or some other nitrogen-bearing molecule that appears in the gas phase due to mantle evaporation (see, e.g., [6] and references therein). This scenario is in agreement with the current observational data on hot cores, and has further support from observations of comets. Both HCN and CH$_3$CN, as well as some other nitrogen-bearing species, have been identified as parent molecules in cometary comas [27, 28]. Since cometary ices may consist of interstellar material, this suggests that grain mantle evaporation can be a source of CH$_3$CN.

Most hot cores observed in lines are sites of massive star formation. Other regions where the gas-phase chemistry is strongly affected by grain mantle evaporation are bow shocks at the edges of young bipolar outflows driven by low-to-intermediate mass protostars

[29]. The temperatures in these objects are high enough to evaporate grain mantles, and the abundances of ammonia and HCN are strongly enhanced there, as well as those of some other molecules abundant in hot cores. Thus, it is natural to expect high CH$_3$CN abundances in these regions. We observed two such objects— NGC 1333 IRAS2 and L1157. In L1157, no lines were found toward the central source (coordinates 0″, 0″), but weak emission was detected toward the blue wing of the outflow (coordinates 20″, –60″), where the abundance of HCN, NH$_3$, CH$_3$OH and some other molecules is enhanced by up to several orders of magnitude ([30] and references therein). We estimated the CH$_3$CN abundance assuming that the emission arises in the same small region as the NH$_3$ emission (B1 in the notation of [31]). Assuming the size of B1 to be about 10″–15″ and its kinetic temperature to be 80 K [29], we obtained a column density of about $\sim 2 \times 10^{13}$ cm^{-2}. An H$_2$ column density of 2.4×10^{21} cm^{-2} follows from the C^{18}O data of [30]. Hence, the CH$_3$CN abundance is about 10^{-8}, as high as in hot cores. Thus, our data suggest that the CH$_3$CN abundance may be enhanced toward the blue wing of this bipolar outflow. Sensitive observations with high angular resolution are desirable to test this possibility.

We did not detect any emission toward the red wing of the NGC 1333 IRAS2 E–W outflow, where CS and CH$_3$OH enhancements were found by Sandell *et al.* [32], except for a weak spike at about the 2.5σ level, which can be attributed to the 6_0–5_0 line. The CS column density at this position derived in [33] is an order of magnitude lower than the CS column density toward L1157 B1 presented in [30], showing that either the gas column density or the molecular enhancement or both are lower in NGC 1333 IRAS E–W, leading to our non-detection of CH$_3$CN emission.

The nature of the group II sources is unclear. On the one hand, their rotational diagrams do not demonstrate significant deviations from the fitted lines. Therefore, the conditions necessary for a rotational-diagram analysis could be satisfied; in particular, the sources could be uniform and the observed lines optically thin. In this case, these objects are warm clouds (50 K or lower). On the other hand, the "clean" appearance of the rotational diagrams does not *prove* that these conditions are satisfied. The sources may be inhomogeneous or/and the lines may be optically thick. In particular, since the intensities of the $K = 4$ lines are unknown, we cannot distinguish between an opaque core–halo source and an optically thin uniform cloud that is cooler than the core. This can be explained as follows. Let us consider the rotational diagram of an optically thick core–halo object. A high optical depth leads to an apparent decrease of N_3/g_3 and to a $K = 3$ depression, while the contribution of the halo leads to an increase of N_0/g_0 and N_1/g_1. Therefore, the slope of the line fitted through the $K = 0$–3 points is larger than that corresponding to

the core, and yields a lower temperature. Since the signal-to-noise ratios in our spectra of group II sources are typically fairly low, neither rotational diagrams nor a χ^2 analysis can be used to distinguish between the two source models. Therefore, it is possible that at least some of the group II objects are similar to group I objects; i.e., they consist of compact, opaque hot cores and cooler halos. More sensitive observations are required to study the structure of these objects. Note that the presence of hot gas in most of the observed group II sources is indicated by water vapor maser emission, for example. Nevertheless, we assume below that the group II sources are homogeneous; all the results obtained from the rotational diagrams and statistical equilibrium analysis are valid only if this is the case.

The temperatures of the group II sources were $\lesssim 50$ K or lower, and their CH_3CN abundances were either below 10^{-9} or could not be estimated from our data. We have no size estimates for these sources, apart from very uncertain lower limits. Density estimates, when available, show that the densities are at least larger than 10^4 cm^{-3}. It is possible that the CH_3CN emission arises in the warm clouds indicated by observations of CS, NH_3, and other molecules. One of the best studied warm clouds is NGC 2264 IRS1. Our temperature estimate agrees with the value of 55 K obtained in [34] based on a careful study of this source in a large number of lines of various molecules. Thus, the CH_3CN emission in the group II sources may arise in warm clouds, but we cannot exclude the possibility that emission from hot, compact objects with enhanced CH_3CN abundances are present in at least some of these sources (see Section 7).

8. CONCLUSION

We have carried out a survey of 27 Galactic star-forming regions in the CH_3CN 6_K–5_K, 5_K–4_K, and 8_K–7_K lines at 110, 92, and 147 GHz. Twenty-five sources were detected at 110 GHz, nineteen at 92 GHz, and three at 147 GHz.

We derived a number of source parameters using a rotational diagram analysis and statistical equilibrium calculations. As expected, the strongest emission was found toward hot cores. The kinetic temperatures of these objects are $\gtrsim 100$ K, their CH_3CN abundances are higher than 10^{-9}, and the CH_3CN lines are optically thick. Hot cores are usually associated with cooler/less dense gas. The CH_3CN abundance in the hot gas associated with the young bipolar outflow L1157 is probably about 10^{-8}.

The nature of weaker objects is unclear. The CH_3CN emission may arise in warm (30–50 K), dense ($>10^4$ cm^{-3}) clouds. However, hot, compact objects with enhanced CH_3CN abundances may be present in at least some of these sources.

REFERENCES

1. P. M. Solomon, K. B. Jefferts, A. A. Penzias, and R. W. Wilson, Astrophys. J. **168**, L107 (1971).

2. J. M. Hollis, L. E. Snyder, D. H. Blake, *et al.*, Astrophys. J. **251**, 541 (1981).

3. R. B. Loren and L. G. Mundy, Astrophys. J. **286**, 232 (1984).

4. D. Boucher, J. Burie, A. Bauer, *et al.,* J. Phys. Chem. Ref. Data **9**, 659 (1980).

5. L. Olmi, R. Cesaroni, R. Neri, and C. M. Walmsley, Astron. Astrophys. **315**, 565 (1996).

6. T. J. Millar, in *Molecules in Astrophysics: Probes and Processes*, Ed. by E. F. van Dishoeck (Kluwer, Dordrecht, 1996), p. 75.

7. L. Olmi, R. Cesaroni, and C. M. Walmsley, Astron. Astrophys. **276**, 489 (1993).

8. W. M. Irvine, P. F. Goldsmith, and Å. Hjalmarson, in *Interstellar Processes*, Ed. by D. J. Hollenbach and H. A. Thronson, Jr. (Reidel, Dordrecht, 1987), p. 561.

9. P. Bergman and Å. Hjalmarson, in *The Physics and Chemistry of Interstellar Molecular Clouds*, Ed. by G. Winnewisser and J. T. Armstrong (Springer, Berlin, 1989), p. 124.

10. E. Churchwell, C. M. Walmsley, and D. O. Wood, Astron. Astrophys. **253**, 541 (1992).

11. S. V. Kalenskiĭ, A. M. Dzura, R. S. Booth, *et al.*, Astron. Astrophys. **321**, 311 (1997).

12. V. I. Slysh, S. V. Kalenskiĭ, I. E. Val'tts, and V. V. Golubev, Astrophys. J., Suppl. Ser. **123**, 515 (1999).

13. R. Wilczek and S. Drapatz, Astron. Astrophys. **142**, 9 (1985).

14. V. G. Promislov, in *Advances in Soft Computing: Engineering Design and Manufacturing*, Ed. by R. Roy, T. Furuhashi, and P. K. Chawdhry (Springer, London, 1999), p. 421.

15. G. H. Moriarty-Schieven, R. L. Snell, S. E. Strom, *et al.*, Astrophys. J. **319**, 742 (1987).

16. K. M. Menten, C. M. Walmsley, C. Henkel, and T. L. Wilson, Astron. Astrophys. **198**, 253 (1988).

17. S. Liechti and T. L. Wilson, Astron. Astrophys. **314**, 615 (1996).

18. S. Liechti and C. M. Walmsley, Astron. Astrophys. **321**, 625 (1997).

19. R. L. Akeson and J. E. Carlstrom, Astrophys. J. **470**, 528 (1996).

20. B. E. Turner, Astrophys. J., Suppl. Ser. **76**, 617 (1991).

21. M. Lampton, B. Margon, and S. Bowyer, Astrophys. J. **208**, 177 (1976).

22. J. E. Wink, J. Duvert, S. Guilloteau, *et al.*, Astron. Astrophys. **281**, 505 (1994).

23. B. E. Turner and W. J. Welch, Astrophys. J. **287**, L81 (1984).

24. M. Ohishi, in *Molecules in Astrophysics: Probes and Processes*, Ed. by E. F. van Dishoeck (Kluwer, Dordrecht, 1996), p. 61.

25. H. T. Ho, R. Genzel, and A. Das, Astrophys. J. **266**, 596 (1983).

26. G. A. Blake, E. C. Sutton, C. R. Masson, and T. G. Phillips, Astophys. J. **315**, 621 (1987).

27. D. Bockelée-Morvan, in *Molecules in Astrophysics: Probes and Processes*, Ed. by E. F. van Dishoeck (Kluwer, Dordrecht, 1996), p. 219.

28. N. Biver, D. Bockelée-Morvan, P. Colom, *et al.*, in *Proceedings of the First International Conference on Comet Hale-Boop, 1998*, Ed. by R. West and M. A'Hearn, Earth, Moon, Planets (in press).

29. R. Bachiller, in *Molecules in Astrophysics: Probes and Processes*, Ed. by E. F. van Dishoeck (Kluwer, Dordrecht, 1996), p. 103.

30. R. Bachiller and M. P. Gutiérrez, Astrophys. J. **487**, L93 (1997).

31. M. Tafalla and R. Bachiller, Astrophys. J. **443**, L37 (1995).

32. G. Sandell, L. B. G. Knee, C. Aspin, *et al.*, Astron. Astrophys. **285**, L1 (1994).

33. W. D. Langer, A. Castets, and B. Lefloch, Astrophys. J. **471**, L111 (1996).

34. K. Schreyer, F. P. Helmich, E. F. van Dishoeck, and Th. Henning, Astron. Astrophys. **326**, 347 (1997).

Translated by S. Kalenskiĭ

Discovery of Rapid Spectral Variability in the Ap Star χ Psc

S. G. Aliev and N. Z. Ismailov

Astrophysical Observatory, Shemakha, Azerbaijan

Received November 1, 1999

Abstract—Analysis of spectrograms with high spectral and temporal resolution taken during four nights in 1998 has revealed spectral-line variability of the peculiar star HD 220825 with a quasi-period of $0^{d}.0.12 \pm 0^{d}.0005$. The observed variations have the form of wavelike oscillations whose amplitude decreases or increases during the observations. The CrII, SiII, and TiII lines exhibit the strongest variability. The scale lengths of peculiar regions are estimated to be from 5000 to 15 000 km based on data for different lines. The results provide evidence for the presence of real wavelike dynamical processes at the stellar surface that change the physical conditions in the stellar atmosphere. © 2000 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

Systematic studies of short-period variability in peculiar stars with strong magnetic fields are of considerable interest for pulsation theory. Breger [1] pointed out that short-term photometric variability in magnetic stars could be associated with pulsations, as is the case in many main-sequence stars. According to the data of Percy [2], the peculiar star 21 Com exhibits photometric variability with a period of 30 min. Wood [3] also observed short-period photometric variations in peculiar stars. A number of authors have pointed out rapid pulsed variability in Ap stars (see, e.g., [4–6]).

Obviously, the short periods observed thus far in some peculiar stars cannot be explained by spots appearing on the visible side of the stellar disk as a result of axial rotation. Therefore, investigation of such short-period variations makes it possible to study the physics of local regions associated with spots. The aim of the present paper is to analyze rapid variability in the spectrum of χ Psc. HD 220825 (χ Psc) is a typical peculiar Cr, Sr star that is known to exhibit periodic photometric and spectral variations ($P = 0^{d}.5853$) [7]. The photometric observations of Rakos [8] and Genderen [7] yield similar photometric periods ($P = 0^{d}.5805$ and $0^{d}.5853$, respectively). Aliev [9] showed that the photometric period coincides with the period of the spectral variations. The star has a variable magnetic field [10].

2. OBSERVATIONS

We observed χ Psc with an échelle spectrometer attached to the Coudé focus of the 2-m telescope of the Shemakha Astrophysical Observatory of the Academy of Sciences of Azerbaijan. The spectrometer was equipped with a 530×580 CCD. Table 1 gives the main

characteristics of the Coudé focus, spectrometer, and CCD.

The optical design of the spectrometer was developed by Musaev [11] based on the Coudé spectrograph of the 2-m telescope of the Shemakha Astrophysical Observatory. The complex was assembled and was subject to preliminary testing at the 1-m telescope of the Special Astrophysical Observatory of the Russian Academy of Sciences [11]. Full-scale tests were performed at the Shemakha Astrophysical Observatory.

The entire spectral interval covered by the CCD is subdivided into 45 and 40 échelle orders, red and blue, respectively. The instrument can simultaneously cover the spectral regions near H_{α} in the 72nd order and near λ4350 Å in the 110th order, with reciprocal linear dispersions of 4.8 and 3.1 Å/mm, respectively. For our program observations, we used the spectrum from the 72nd through the 112nd orders.

Given the optical parameters of the spectrometer complex, the resulting two-pixel spectral resolution was $R = 30\,000$. For focal distances of the collimator and camera equal to $F_c = 5400$ mm and $F_k = 350$ mm, respectively, and the Coudé focal-plane scale of 2.86 arcsec mm^{-1}, we derived a normal width of $d = 2.01$ arcsec (0.73 mm) for the spectrometer slit.

We used the solar spectrum in the spectral interval considered to construct the dispersion curve. The standard deviation of individual reference lines from the mean dispersion curve did not exceed ±0.005 Å. To check for possible shifts during the observations, we obtained an Ne-Fe comparison spectra before and after each exposure. We reduced the spectrograms using standard techniques using a MS DOS program developed at the Special Astrophysical Observatory [12]. This program computes equivalent widths, halfwidths, radial velocities, and other spectral-line parameters.

We observed the two standard stars HR 8506 and HR 0265 to measure their heliocentric radial velocities. Table 2 gives the results for the 72nd, 74th, and 98th orders. The first row of the table gives the numbers of the spectral orders, the boundaries of the spectral intervals covered by the CCD window, and the mean dispersions for the orders. The radial velocities V_r were obtained by averaging four to six values in each order. As can be seen, for fairly narrow spectral lines and signal-to-noise ratios of no less than S/N ≪ 80–100, V_r can be measured with an rms error of ±0.6–1.0 km/s. Low signal-to-noise ratios increase the radial-velocity error to 5–6 km/s. The resulting radial velocities for the standard stars agree well with the data from [13], within the quoted errors.

Our measurements showed that the error in the equivalent width W_λ is determined mainly by the estimated continuum level and the signal-to-noise ratio. In our measurements, the rms error is about 5–10% of the equivalent width. The error in the halfwidth is about $\Delta\lambda_{1/2}$ ±0.02–0.03 Å.

Figure 1a shows a normalized section of the spectrum of the standard star HR 0265 near the H_α line for data obtained on several nights. It is clear that all significant features agree well for six spectrograms obtained on different nights, confirming the high stability of our equipment. Figure 1b shows the dispersion of the relative intensity of the same spectral section as a function of wavelength. We can see that the dispersion of the relative profile intensity of individual spectral sections does not exceed 0.5%.

The exposure times for χ Psc were 3–10 min, depending on the seeing. The signal-to-noise ratio was about 70–100 near the H_α line and about half this value in the blue region of the spectrum. We obtained a total of 34 spectrograms, with 15 spectrograms the first night, 13 the second, and 3 more during each of the remaining two nights. The first two observing runs lasted more than 1.5 hours each.

3. RESULTS

We computed the equivalent widths, halfwidths, depths, and radial velocities of spectral lines. The measurements of individual lines revealed complex, multicomponent structure in the CrII, SiII, TiII, FeII, and MgII lines. We obtained a mean halfwidth of 1.068 ± 0.01 Å for the MgII 4481 Å line, which corresponds to a rotational velocity of $v\sin i = 57.2 \pm 0.7$ km/s. If, as implied by its spectral type, the star has a radius of $R = 3R_\odot$, the true rotational velocity at the equator is $V_{eq} \sim 260$ km/s and the inclination of the rotation axis is $i \sim 12°$. Consequently, the star is observed almost pole-on.

Figure 2 shows the measured W_λ and V_r for H_α and H_β for two nights. This figure shows that both the equivalent widths and the radial velocities of the hydro-

Table 1. General parameters of the telescope-spectrograph-detector system

Parameter	Value
Aperture	2000 mm
Equivalent focus	72 000 mm
Collimator diameter	150 mm
Collimator focal length	540 mm
Camera focal length	350 mm
Camera diameter	352 mm
Échelle grids:	
line frequency	37.5 and 75 lines/mm
blaze angle	63.5°
size	200 × 300 mm
Working orders:	
at 37.5 lines/mm	70–140
at 75 lines/mm	35–70
Spectral interval	λλ 3600–8000 Å
Length of spectrum at λ6563 Å	57 Å
Dispersion at λ 6563 Å	0.0984 Å/px
Number of orders on CCD:	
in the red	46
in the blue	40
Optimum slit width (at H_α)	2″0
Spectral resolution (at H_α)	≈30 000
CCD size	530 × 580 px
Pixel size	24 μm
Linear CCD size	12.72 × 13.92 mm
Operating temperature	120 K
Accumulated signal:	
1-hour exposure	S/N = 10 (λ6563 Å, $V = 9^m4$)
10-min exposure	S/N = 60 (λ6563 Å, $V = 5^m0$)
Limiting magnitude	10^m (S/N = 10, 2-hour exposure)
Prism reflection angle	45°
Prism dispersion (at H_α)	150 Å/mm

gen lines exhibit oscillations. The radial-velocity variations of the H_β line are especially prominent, with an amplitude that exceeds that of the H_α-line variations.

Other measured lines also exhibit similar variability, including MgII 4481 Å, CrII 4836, 4848 Å, TiII 4571 Å, SiII 5041, 5056 Å, FeII 5316, 4991 Å, etc. By way of example, we show in Fig. 3 the time dependence of the radial velocities for the CrII 4836 Å and SiII 5041 Å lines for two different nights. As is evident from this figure, the radial velocities of these lines exhibit wave-like variations. During the first night (JD 2451057), the radial-velocity amplitude decreased (Fig. 3a), whereas
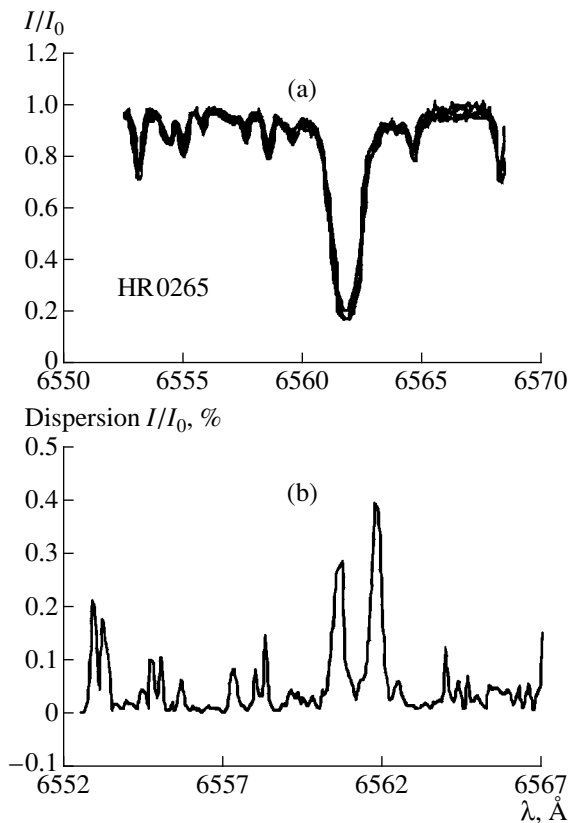
Fig. 1. (a) Section of the spectrum of HR 0265 (G2III) near H_α on various observing dates and (b) dispersion of the relative intensity.

it increased during the second night (JD 2451058). A similar effect can be seen in the H_β measurements in Fig. 2.

Figure 4 shows the time dependences of the equivalent widths of these same two lines for the same nights. The variations of W_λ virtually repeat the radial-velocity curve in Fig. 3. This provides evidence that we are observing the effect of real short-term dynamical processes that change the physical conditions in the stellar atmosphere.

Figure 5 presents the radial velocities of the CrII 4558 and 4836 Å lines folded with the photometric period derived by van Genderen [7] ($P = 0^{\text{d}}.5853$). It is clear that our measurements, which do not cover the entire photometric period, nevertheless show variations with this period. Our data show radial-velocity maxima and a very unusual enhancement of the line equivalent widths and halfwidths near phases 0.3–0.5. The radial velocity V_r rises sharply from minimum to maximum near phase 0.0. Consequently, the radial-velocity curve is asymmetric with respect to its extrema. This provides further support for the hypothesis that the angle i between the stellar rotational axis and the line of sight is small. The period in question could, indeed, plausibly be a manifestation of the star's axial rotation [9].

Thus, our program-star observations covering 1.5–2 h and consisting of 3–5 min exposures enabled variability of the star with a period of $0^{\text{d}}.012 \pm 0^{\text{d}}.0005$ to be
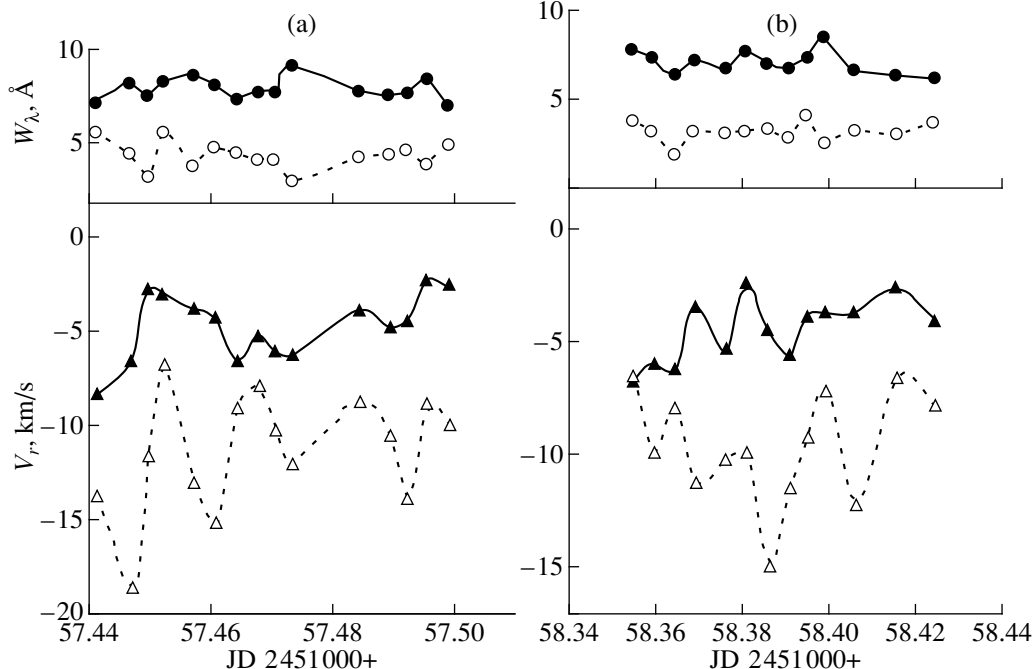

Fig. 2. Variations of equivalent widths (circles) and radial velocities (triangles) of the H_α (filled circles and triangles) and H_β (open circles and triangles) lines in the spectrum of HD 220825 on two different nights.
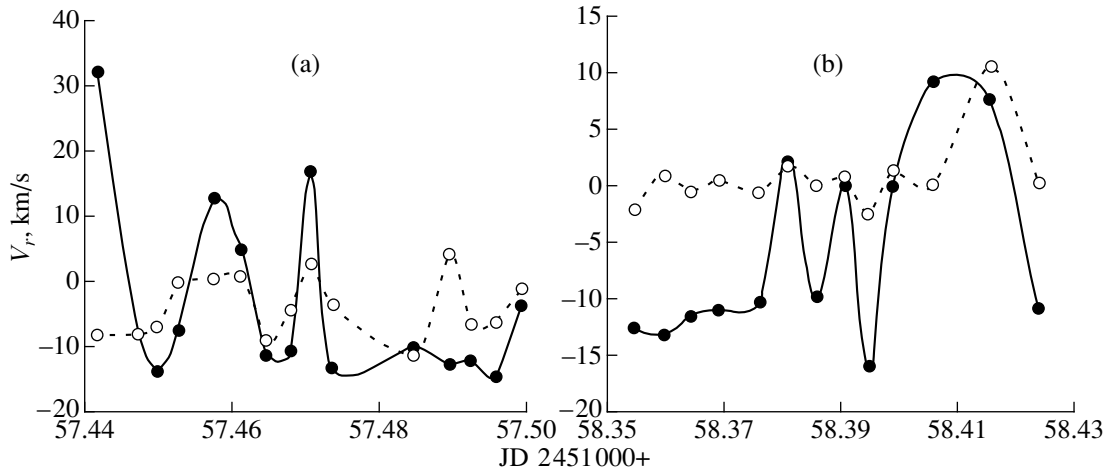
Fig. 3. Radial-velocity variations of the CrII 4836 Å (filled circles) and SiII 5041 Å (open circles) lines in the spectrum of HD 220825 on two nights.
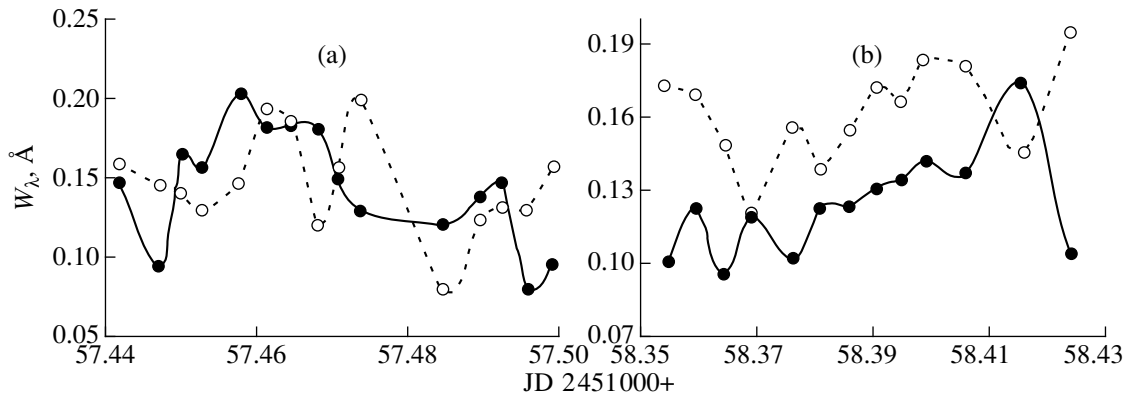

Fig. 4. Same as Fig. 3 for the line equivalent widths.

confidently detected. This period was observed rather stably on both nights (Figs. 2–4).

Figure 6 shows the radial-velocity curves for the MgII 4481 Å and CrII 4836, 4558 Å lines folded with a period of $P = 0^{d}.012$. The plot includes all data obtained during the four nights. It is clear that these lines exhibit periodic variations with $P = 0^{d}.012$ over 20 days, i.e., throughout the observing season. Figure 7 shows a similar graph with data for the H_α, H_β, FeI 5383 Å, SiII 5041 Å, and TiII 4571 Å lines in addition to the lines shown in Fig. 6. Overall, these lines also show periodic variations with $P = 0^{d}.012$. Since the radial velocities of the different lines are shifted somewhat relative to each other, we see a shift in the radial-velocity curves in Fig. 7 from minimum to maximum values (see also Fig. 6). On the whole, it is clear from Fig. 7 that the shifts for various elements are 10–12 km/s. The lower envelope in Fig. 7 is better defined than the upper envelope. The minimum radial velocities are observed for the most stable lines, such as FeII 5326 Å and FeI 5382 Å. The largest scatter is observed for the SiII, CrII, and TiII lines. We can also see from Fig. 7 that the total range of the radial-velocity variations for all the elements is ±30 km/s. The maximum amplitude of the V_r variations for individual elements varies from 5 to 15 km/s (see, e.g., Figs. 2 and 6). On the whole, the intervals for the radial-velocity variations for different elements overlap. As noted above, the most stable lines have the smallest radial velocities, and the more peculiar the element, the higher the amplitude of its V_r variations.

Figure 8 shows the phase dependences of the equivalent widths of the MgII 4481 Å, CrII 4558 Å, TiII 4571 Å, CrII 4848 Å, FeII 4991 Å, FeII 5316 Å, and FeI 5383 Å lines for a period of $0^{d}.012$. For two lines, trends represented by fourth-order polynomials are given. We can see that the equivalent widths of the MgII 4481 Å line show periodic variations. The equivalent widths of other lines also vary with phase to various degrees, although, for example, the FeII 5316 Å line shows nearly no equivalent-width variations.

Table 2. Radial-velocity measurements for the two standard stars

Date, UT		72nd order 6542–6600 Å 0.099 Å/px	74th order 6323–6367 Å 0.076 Å/px	98th order 4823–4866 Å 0.073 Å/px
		V_r, km/s		
HR 8506				
30.07.98	00^h40^m	-7.6 ± 6	-7.1 ± 3.1	-6 ± 4
27.08.98	23 59	-5.1 ± 2.1	-6.9 ± 1.0	-14.6 ± 1.8
28.08.98	21 41	-9.65 ± 1.0	-10.7 ± 1.4	-12.8 ± 3.3
29.08.98	21 50	-6.4 ± 1.4	-7.8 ± 0.6	-8.7 ± 2.6
01.09.98	22 30	-8 ± 1.8	-18 ± 0.9	-12 ± 1.2
HR 0265				
30.07.98	02 05	-36 ± 4.4	50 ± 0.6	-49 ± 3.1
28.08.98	00 35	-40 ± 1.3	-46.2 ± 0.8	-50.1 ± 2.4
28.08.98	23 15	-46.6 ± 1.1	-50.3 ± 0.6	-53.8 ± 1.8
29.08.98	22 10	-51 ± 2.1	-51.1 ± 1.5	-54.1 ± 2.4
01.09.98	01 30	49.9 ± 1.8	-51.9 ± 0.6	-53.1 ± 3.5

Since the amplitude of the photometric variability of the star in the V band is $0^m.01$ [7], the line equivalent widths can be expected to vary by up to 10%. However, the observed variations of W_λ exceed 50–100%. Especially strong equivalent-width variations are observed for the lines of the peculiar elements CrII, SiII, TiII, and for the Balmer H_β line.

4. DISCUSSION AND CONCLUSIONS

Our observations of the Ap star χ Psc made with high spectral and temporal resolution indicate rapid spectral variability. The variability time scale is $0^d.012$.
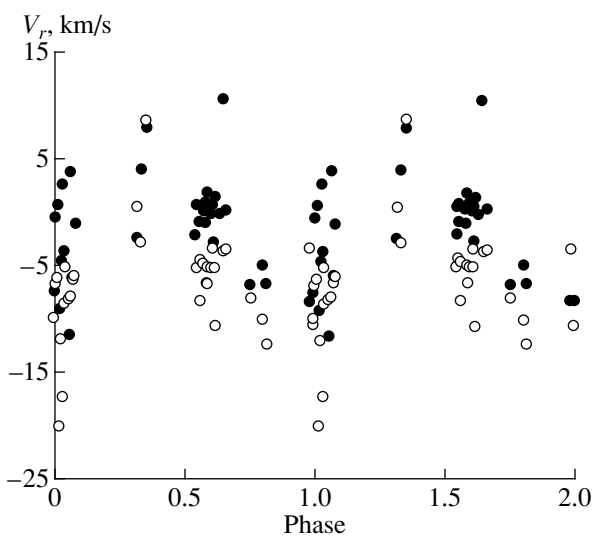

Fig. 5. Radial velocities of the CrII 4836 Å (filled circles) and 4558 Å (open circles) lines folded with the photometric period of $P = 0^d.5853$ [7].

It is apparently not always possible to reliably detect such variations in individual spectral lines due to the peculiar processes occurring on the star. Furthermore, such short-term variations in the stellar spectrum might be detectable only at certain axial-rotation phases and under certain physical conditions.

Our results agree fairly well with the photometric period derived earlier [7]. At the same time, the stellar spectrum exhibits pulsational variations with a period of 17 minutes, which remain stable over 20 days. We observed simultaneous variations in both the equivalent widths and halfwidths of spectral lines and in their radial velocities, providing evidence for the action of real short-term dynamical processes at the stellar surface.

The full amplitude of the Doppler-shift variations differs from one spectral line to another and varies from 5 to 50 km/s. Given the variation time scale, these amplitudes imply that the sizes of active regions in the atmosphere of χ Psc range from 5000 to 50000 km. Note that these are typical sizes for sunspots.

Figures 2–4 demonstrate wavelike variations in V_r and W_λ for the peculiar SiII and CrII lines and H_β. The time dependence of the variability amplitude also suggests the presence of wavelike impulsive processes acting at the star's surface. Judging from these figures, the characteristic times for the development and decay of these oscillations are from one to two hours. Rotation alone is too slow a process to explain such rapid, transient oscillations. The observed effect could be associated with a resonance between certain oscillation processes. For example, similar oscillations on the Sun have speeds of the order of the sound speed (10–30 km/s) [14].

Borra and Landstreet [15] showed that, as a rule, there is little evidence for irregular magnetic-field oscillations, except for oscillations with amplitudes of the order of 200 Oe observed in some Ap stars. There is
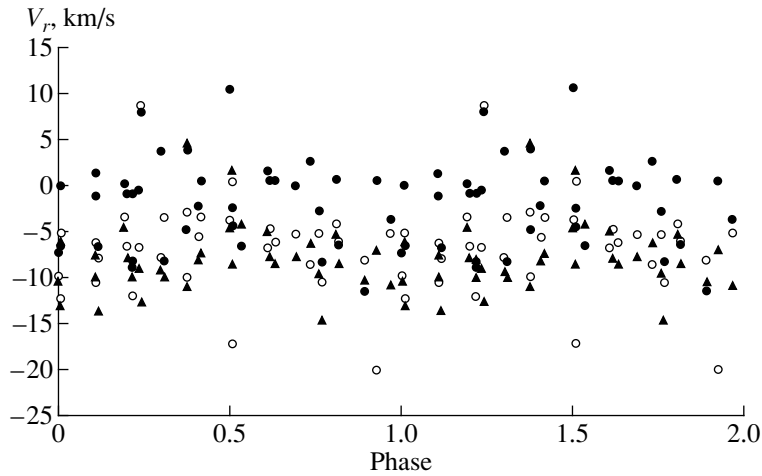
Fig. 6. Radial velocities of the MgII 4481 Å (filled triangles), CrII 4836 Å (filled circles), 4558 Å (open circles) lines folded with the period $P = 0.^d012$.
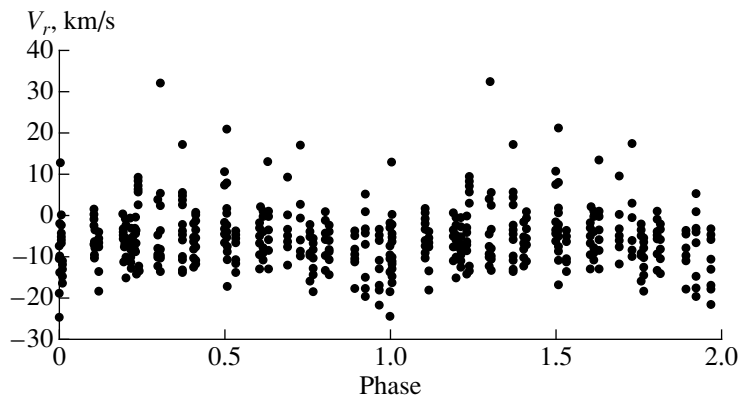

Fig. 7. Radial velocities of the H_α, H_β, MgII 4481 Å, CrII 4836, 4558 Å, FeI 5383 Å, SiII 5041 Å, and TiII 4571 Å lines folded with the period $P = 0.^d012$.
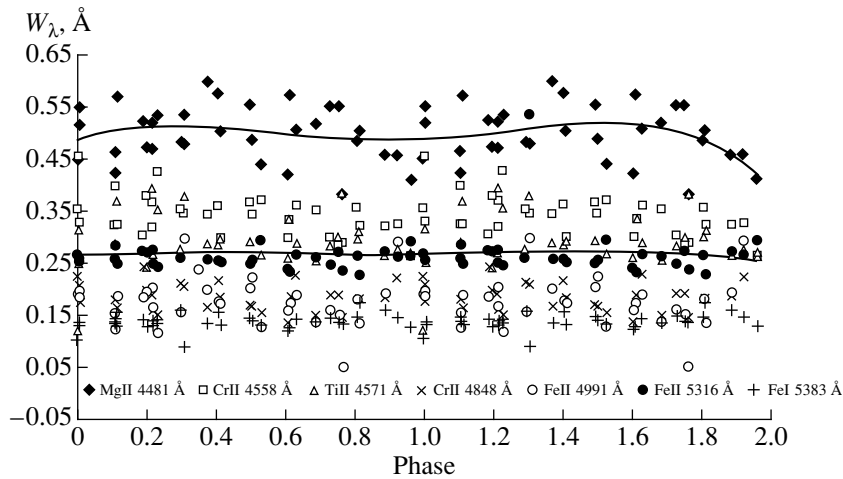

Fig. 8. Equivalent widths of lines in the spectrum of χ Psc folded with the period $P = 0.^d012$. Fourth-order polynomial trends are drawn for the MgII 4481 Å and FeII 5316 Å lines.

little doubt that, for most of these stars, atmospheric oscillations on long time scales can be explained using inclined rotator models. However, our results have shown the presence of shorter-term atmospheric variations in the atmospheres of Ap stars.

Our observational data lead us to the following conclusions.

(1) We have detected rapid variability of the spectrum of χ Psc with a quasi-period of $0.^{d}012 \pm 0.^{d}0005$. This variability is observed both in the line radial velocities and in other line parameters.

(2) The observed variations with a quasi-period of $P = 0.^{d}012$ have the form of damped or amplified oscillations. The amplitudes of these oscillations are comparable to those of acoustic waves on the Sun.

(3) On the whole, the variations of the spectral parameters can be described by the photometrically determined axial-rotation period. However, the radial-velocity variations are significant whether they are folded with the photometric period or with the period we have inferred in this paper.

(4) Based on the observed variations of the spectral-line parameters, we estimate the scale lengths of active formations at the stellar surface to be 5 to 50 thousand km.

REFERENCES

1. M. Breger, Astrophys. J. **176**, 373 (1972).
2. J. R. Percy, Astron. Astrophys. **22**, 381 (1973).
3. H. J. Wood, in *Magnetic and Related Stars*, Ed. by R. C. Cameron (Mono Book Comp., Baltimore, 1967), p. 485.
4. H. Wood, Publ. Astron. Soc. Pac. **76**, 158 (1964).
5. D. W. Kurtz, Mon. Not. R. Astron. Soc. **200**, 807 (1982).
6. J. M. Matthws, Publ. Astron. Soc. Pac. **103**, 3 (1991).
7. A. M. van Genderen, Astron. Astrophys. **14**, 48 (1971).
8. K. D. Rakos, Lowell Obs. Bull. **5** (117), 12 (1962).
9. S. G. Aliev, in *Magnetic Ap Stars* [in Russian] (Elm, Baku, 1975), p. 80.
10. H. W. Babcock, Astrophys. J., Suppl. Ser. **3**, 30 (1958).
11. F. A. Musaev, Pis'ma Astron. Zh. **19**, 776 (1993) [Astron. Lett. **19**, 315 (1993)].
12. G. A. Galazutdinov, Preprint No. 92 (SAO, 1992).
13. T. G. Barnes, III, T. J. Moffet, and M. H. Slovak, Publ. Astron. Soc. Pac. **98**, 223 (1986).
14. A. B. Severnyĭ, in *Some Problems of Solar Physics* [in Russian] (Nauka, Moscow, 1988), p. 102.
15. E. F. Borra and J. D. Landstreet, Astrophys. J., Suppl. Ser. **42**, 421 (1980).

Translated by A. Dambis

Reconstruction of the Brightness Distribution Across a Stellar Disk from Gravitational Lensing Caustic Passage Observations

M. B. Bogdanov[1] and A. M. Cherepashchuk[2]

[1]*Chernyshevskiĭ University, Saratov, Russia*
[2]*Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119899 Russia*
Received November 2, 1999

Abstract—This paper describes a method for analyzing observations of the passage of a star through a gravitational lensing caustic that enables the derivation of information about the brightness distribution across the stellar disk without knowledge of the lens parameters. The ill-posed inverse problem of reconstructing a one-dimensional strip brightness distribution across the stellar disk is solved using a regularization method, taking into account the non-negative nature of this function. Assuming the source is circularly symmetric, the search for the radial brightness distribution is carried out over compact sets of upward-convex, non-negative functions or functions that do not grow with distance from the stellar center. The method is used to analyze the gravitational microlensing event MACHO 98-SMC-1, of a star in the Small Magellanic Cloud, obtained by the PLANET international group. The resulting brightness distribution is compared with computations of limb darkening for model stellar atmospheres. © *2000 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Since the first detection of gravitational microlensing of stars by massive compact objects in the halo of the Galaxy (MACHOs) [1], possibilities for using this effect to study stars with high angular resolution have been investigated. These studies have shown that such cosmic lenses provide record resolving capability at visible and infrared wavelengths, and that, under certain conditions, the brightness distribution across the disk of a lensed star can be reconstructed. Unfortunately, in the case of single gravitational lenses, the sensitivity of the lensing curve to features of the brightness distribution is rather low [2–4]. One exception is the case when the gravitational lens is projected onto the stellar disk. This rare case occurred during the lensing of a Galactic bulge red giant known as MACHO Alert 95-30, which was observed by the MACHO and GMAN international projects in the summer of 1995 [5]. It proved possible to use the data for this event to derive information about the stellar brightness distribution, both by fitting models for plane-parallel stellar atmospheres with a non-linear limb-darkening law [5] and by reconstruction of the brightness distribution [6].

In the case of binary gravitational lenses, the nature of the potential leads to the appearance of caustics in the source plane. Intersection with these caustics can give rise to the appearance or disappearance of additional images of the star, accompanied by abrupt variations in the amplification coefficient [7, 8]. On a caustic itself, the amplification coefficient for the flux from a point source formally tends to infinity, sharply increasing the sensitivity of the lensing curve to the stellar brightness distribution [4, 9].

To date, several passages of stars through gravitational lensing caustics have been observed. In particular, the international group PLANET has observed two such events: gravitational microlensing of a star in the Small Magellanic Cloud, MACHO 98-SMC-1, and of a star in the Galactic bulge, MACHO 97-BLG-28. In the former case, model fitting was used to estimate the angular diameter of the star, assuming a uniformly bright disk [10]. In the latter case, the parameters for a non-linear limb-darkening law were determined [11].

The main difficulty encountered in the analysis of such data is the large number of free parameters in the models used. Even the simplest case of a static binary lens consisting of two material point sources and a source with a uniformly bright disk required nine model parameters [10]. When considering a non-linear limb-darkening law and taking into account the reduction coefficients for three independent series of data in two filters, the number of model parameters had to be increased to 19 [11]. Together with computational difficulties, this sharply exacerbates the problem of mutual interdependences of the parameters. In addition, the fact that this problem, like most inverse problems in astrophysics, is ill-posed begins to manifest itself [12]. The influence of noise in the input data can lead to instability and ambiguity in the estimates of the parameter values. For example, in [10], two different sets of model parameters were obtained that essentially

described the observed microlensing curve equally well.

It stands to reason that these circumstances do not reduce the importance of model fitting in the analysis of microlensing observations. Only via such fitting can the mass ratios, separations, and position angles of the components of binary lenses be estimated. However, if we are interested only in information about the brightness distribution across a lensed source, the problem must be drastically simplified, so that free parameters are either completely excluded or decreased in number.

In the current paper, we describe a method for analyzing observations of the passage of a star through the caustic of a gravitational lens. We also present the results of applying this method to the reconstruction of the brightness distribution across the disk of the star lensed in the MACHO 98-SMC-1 event.

2. MAIN INTEGRAL EQUATIONS

For an observer on the Earth, a gravitational lens achieves a smooth mapping of the source plane. In the case of a binary lens, singularities arise, which, as for any smooth mapping, are of two types: folds and gathers [13]. The corresponding system of caustics in the source plane usually has the form of a curvilinear polygon located between the projections of the binary components of the lens onto this plane [7, 8]. The sides of the polygon correspond to fold singularities, while the vertices—which are called cusps in gravitational-lensing terminology—correspond to gathers. At some points during the relative motion between the observer, lens, and lensed star, the star intersects this polygon. The first intersection of a caustic, accompanied by a jump-like growth and subsequent slower decrease in the flux, is usually noted only some time after it has been initiated, so that it is rarely observed with complete time coverage. However, once such a caustic passage has been detected, an alert warning system is activated, and various groups of observers try to detect the second intersection. Therefore, most observations usually occur during the second caustic intersection, which is characterized by a growth in the flux followed by a sharp decrease. It is also obvious that it is more probable to encounter a fold-type "singularity." According to the estimates of [4], the probability of a star passing close to a vertex of a caustic polygon during lensing by a binary lens does not exceed 20%.

Due to the small angular sizes of Galactic bulge stars and even more so stars of the Magellanic Clouds, the curvature of the caustic can be neglected, and we can take it to be linear. In this case, the character of an observed lensing curve will depend only on the one-dimensional projection $B(x)$ of the two-dimensional brightness distribution across the source $b(x, y)$ onto the x axis, perpendicular to the caustic. In astronomy, the projection $B(x)$ is usually called a strip distribution, and

is related to the two-dimensional brightness distribution $b(x, y)$ by the equation

$$B(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} b(\xi, y)\delta(\xi - x)d\xi dy, \qquad (1)$$

where $\delta(x)$ is the Dirac function. The integral equation (1) describes the scanning of the two-dimensional source brightness distribution by a knife beam oriented perpendicular to the x axis. Thus, only information about the strip brightness distribution can be obtained from observations of the intersection of a gravitational-lens caustic.

To elucidate the form of a two-dimensional brightness distribution, it is necessary to consider a set of strip distributions for various scanning directions. The task of reconstructing the structure of an object from such a set of one-dimensional projections is called "computer tomography," and this method and its solution have been well developed [14]. However, in the circularly symmetric case that is relevant for the vast majority of stars, the radial brightness distribution $b(r)$ is unambiguously related to $B(x)$ by the Abel integral equation:

$$B(x) = \int_{x}^{\infty} 2b(r)rdr/\sqrt{r^2 - x^2}. \qquad (2)$$

Let us consider for the sake of definiteness the second intersection of a caustic. We denote ξ to be the angular distance of a point in the strip distribution $B(\xi)$ from its center, and x to be the angular distance of the source center from the caustic. Then, we can write for the amplification coefficient for the flux from a source with an infinitely small size along the x axis near the time for intersection of the caustic [7, 8]

$$A(x, \xi) = A_0 + kH(\xi - x)/\sqrt{\xi - x}, \qquad (3)$$

where A_0 is the amplification coefficient associated with the lensed source images, which do not disappear during intersection of the caustic; k is a coefficient that depends on the parameters of the gravitational lens; and $H(\xi - x)$ is a Heavyside step function, which is equal to zero and unity for negative and positive arguments, respectively. Near the time of intersection of the caustic, A_0 and k can be taken to be constant. Then, the observed variations in the flux $I(x)$, usually called the lensing curve, can be written

$$I(x) = \int_{-\infty}^{\infty} A(x, \xi)B(\xi)d\xi. \qquad (4)$$

Expression (4) is an integral Fredholm equation of the first kind and, as is known, the inverse problem of deriving the strip brightness distribution $B(x)$ from the observed lensing curve $I(x)$ is ill-posed [12]. This fact was first pointed out in connection with studies of grav-

itational microlensing in [9], where the regularization method of Tikhonov was used. However, we cannot recommend this method for analyses of observations of lensed stars for two reasons. First, the regularization method uses minimum *a priori* information about the function $B(x)$, which lowers the stability of the solution to the influence of random noise. When the lensed sources are stars, we can impose a large number of restrictions on the solution, which have a qualitative nature and a clear connection to the physics of the problem (for example, the non-negativity of $B(x)$, the circular symmetry of the stellar disk, the requirement that $b(r)$ not grow with distance from the center, etc.). Second, the core of Eq. (4), $A(x, \xi)$, has a singularity at the point $\xi = x$, which leads to difficulties when searching for numerical solutions. Below, we describe a method for the analysis of lensing observations that is free from these drawbacks.

3. RECONSTRUCTION OF THE STRIP DISTRIBUTION AND RADIAL BRIGHTNESS DISTRIBUTION ACROSS A STELLAR DISK

When reconstructing $B(x)$ in [9], it was assumed that the amplification coefficient for a caustic intersection was so large that the term A_0 can be neglected in expression (3). This approximation is rather crude, and is not satisfied in practice. However, due to the linearity of Eq. (4), we can modify the lensing curve by subtracting the flux level to which the curve falls immediately after the caustic intersection in the region where the second term of (3) is zero. After this subtraction of the background level, the problem becomes degenerate in the sense that, if the lens parameters are unknown, the strip brightness distribution can be derived from (4) only with accuracy to within an arbitrary multiplicative factor.

Further, we will consider the operation of estimating and subtracting the background level to have been carried out, and, accordingly, will take $A_0 = 0$ and $k = 1$ in (3); as before, we will call the lensing curve modified in this way $I(x)$. The angular distance $x = V(t - t_0)$, where V is the angular velocity of the source toward the caustic and t_0 is the time of intersection of the caustic by the source center. V depends on the relative spatial velocity and distance to the source, and also on the angle at which it intersects the caustic. In our case, the velocity V is not known, which means that the angular size of $B(x)$ is likewise determined from Eq. (4) only to within a multiplicative factor.

Thus, without having at our disposal a model for the gravitational lens, we can derive from observations of the passage of a source through a caustic only the shape of the brightness distribution. However, it is precisely knowledge of this shape that is important for many goals. In particular, comparison of such data with theoretical calculations can test the correctness of model stellar atmospheres, independent of spectral data.

We can eliminate the influence of the singularity of the core $A(x, \xi)$ at the point $\xi = x$ as follows. Let $B(\xi)$ differ from zero in some symmetric measurement interval ξ, whose width can be chosen to be clearly larger than the expected source size. We introduce in this interval a uniform grid ξ_i with step Δ, such that $\xi_i = \Delta(i - 1/2)$, where $i = 0, \pm1, \pm2, \ldots, \pm N$. Equation (4) can then be written as the finite sum of integrals

$$I(x) = \sum_{i=-N}^{N} \int_{\xi_i - \Delta/2}^{\xi_i + \Delta/2} A(x, \xi)B(\xi)d\xi. \tag{5}$$

When Δ is sufficiently small, based on the mean theorem, we can bring $B(\xi_i)$ out from the integral and integrate expression (3) in intervals where $A(x, \xi)$ differs from zero, so that we can write (5) in the form

$$I(x) = \sum_{i=-N}^{N} \tilde{A}(x, \xi_i)B(\xi_i), \tag{6}$$

where

$$\tilde{A}(x, \xi_i) = 2(\sqrt{\xi_i + \Delta/2 - x} - \sqrt{\xi_i - \Delta/2 - x}) \tag{7}$$

for non-negative radicands. Otherwise, $\tilde{A}(x, \xi_i)$ is assumed to be equal to zero.

Expression (6) corresponds to a Fredholm integral equation of the first kind with a core without singularities. As already noted above, the ill-posed inverse problem for this equation can be solved using a regularization method. In this case, it is convenient to minimize the functional of Tikhonov via projection of the conjugate gradients onto the set of vectors with non-negative components [15], taking into account the *a priori* information that $B(\xi_i)$ is non-negative. This approach can be applied for sources with arbitrary strip brightness distributions. However, when analyzing observations of lensed stars, the volume of *a priori* information used can be appreciably increased.

Let the lensed source be circularly symmetrical, so that the brightness $b(r)$ depends only on the distance r from the center of its visible disk. It is obvious that, in this case, the function $B(\xi)$ is symmetric with respect to the y axis. Allowing for this symmetry, we can write Eq. (4) in the form

$$I(x) = \int_{0}^{\infty} B(\xi)[A(x, \xi) + A(x, -\xi)]d\xi. \tag{8}$$

Using the uniform grid ξ_i considered above and transforming Eq. (8) in the same was as was done in the derivation of (6), we obtain

$$I(x) = \sum_{i=1}^{N} K(x, \xi_i)B(\xi_i), \tag{9}$$

where the values of the new core $K(x, \xi_i)$ are either zero or

$$K(x, \xi_i) = \tilde{A}(x, \xi_i) + \tilde{A}(x, -\xi_i), \tag{10}$$

and $\tilde{A}(x, \xi_i)$ is determined from expression (7).

We can also use a regularization method when solving the inverse problem of reconstructing $B(x)$ from (9), taking into account the fact that the strip distribution is non-negative. However, it is important that, in the general case, even considering extended stellar atmospheres, $B(x)$ never grows with distance from the center. As is known, the set of monotonic and bounded functions is compact, and the inverse problem for Eq. (9) on this set is stable to the influence of random noise [12]. To search on compact sets for solutions for $B(x)$ that are in the best agreement with the observations, we can perform a minimization of a residual via projection of the conjugate gradients [12, 15]—a well established method.

If we assume circular symmetry, knowledge of the strip distribution enables us to find the radial brightness distribution across the stellar disk. It is known that the solution of the Abel integral Eq. (2) has the form

$$b(r) = -\int_r^\infty B'(x)dx/\pi\sqrt{x^2 - r^2}. \tag{11}$$

However, it is difficult to use Eq. (11) in practice, since it includes the derivative of the strip distribution $B'(x)$, and numerical estimation of this quantity represents another ill-posed problem [12]. On the other hand, we can consider the reconstruction of $b(r)$ to be an inverse ill-posed problem for the integral equation (2). In both cases, the problem is that $B(x)$, in turn, is obtained via the solution of an ill-posed problem, and, as a rule, precise values for its errors are not known. We can find an exit from this situation by formulating an equation relating $b(r)$ directly with $I(x)$.

Let us introduce a uniform grid in the radius r_k with step Δ, such that $r_k = \Delta k$, $k = 1, 2, \dots, M$. Then, using (2), the values for the strip brightness distribution on the grid considered earlier ξ_i can be written

$$B(\xi_i) = \sum_{k=1}^{M-1} \int_{r_k - \Delta/2}^{r_k + \Delta/2} 2b(r)rdr/\sqrt{r^2 - \xi_i^2}. \tag{12}$$

The integrals computed in the elementary intervals with width Δ in expression (12) are taken to be zero when $r_k < \xi_i$. The mean theorem again lets us write

$$B(\xi_i) = \sum_{k=1}^{M-1} F(\xi_i, r_k)b(r_k), \tag{13}$$

where the core $F(\xi_i, r_k) = 2\,(\sqrt{\xi_{k+1}^2 - \xi_i^2} - \sqrt{\xi_k^2 - \xi_i^2}\,)$ when $k \geq i$ and $F(\xi_i, r_k) = 0$ when $k < i$. Substituting

expression (13) into Eq. (9) and exchanging the order of the summation, we obtain

$$I(x) = \sum_{k=1}^{M-1} S(x, r_k)b(r_k), \tag{14}$$

where

$$S(x, r_k) = \sum_{i=1}^{N} K(x, \xi_i)F(\xi_i, r_k). \tag{15}$$

Expression (14) is a Fredholm integral equation of the first kind relating the lensing curve $I(x)$ and the radial brightness distribution $b(r)$, whose core, determined by Eq. (15), has no singularities. For stars with possibly extended atmospheres, the solution of this equation can be obtained on the compact set of monotonic, non-negative functions. For normal stars, it is convenient to search for solutions to (14) on the compact set of convex-upward, non-negative functions [12, 14]. Thus, the restrictions imposed on the possible form of $b(r)$ both have a direct connection to the physics of the problem at hand and make the solution stable with respect to the influence of random noise.

4. ANALYSIS OF THE MACHO 98-SMC-1 STELLAR MICROLENSING EVENT

The event MACHO 98-SMC-1, which corresponds to microlensing of a star in the Small Magellanic Cloud, was observed by the international PLANET group in June, 1998. We retrieved data files containing the Julian dates for the observations, stellar magnitudes in the I band, and the corresponding rms errors via the Internet, from the PLANET web server (http://www.astro.rug.nl/~planet/). In our analysis, we used 55 measurements near the brightness maximum associated with the second caustic intersection. These were all obtained using a CCD array on the 1-m telescope of the South African Astronomical Observatory (SAAO), which removes the problem of reducing the data to a single photometric system.

The stellar magnitudes and their errors were translated into fluxes, in flux units corresponding to a magnitude $I = 17^{m}.0$. Before deriving the microlensing curve $I_0(t)$ from the flux measurements, we subtracted the background level, which was taken to be the rms average of nine measurements near the jump-like drop in brightness and was equal to $I = 19^{m}.623 \pm 0^{m}.023$. The Julian dates, corresponding measurements $I_0(t_i)$, and their rms errors σ_i are presented in Table 1. Adopting the angular velocity $V = 1$, we will determine the values $x_i = (t_i - t_0)$ in fractions of a day. The time t_0 is not known *a priori*, and, generally speaking, must be treated as a free parameter. Examination of the lensing curve suggested a preliminary estimate for t_0 of JD 2450982.6. Below, we will demonstrate how this

value can be refined. The lensing curve measurements $I_0(x_i)$ are shown by circles in Fig. 1.

In the first stage of our analysis, we reconstructed the strip brightness distribution $B(x)$ using the regularization method, taking into account the non-negative nature of the solution [15]. As is known, this method finds the smoothest solution whose lensing-curve residual $I_c(x_i)$ agrees with the value expected based on the specified errors. It is convenient to adopt

$$\chi_N^2 = \sum_{i=1}^{N} [(I_0(x_i) - I_c(x_i))/\sigma_i]^2, \qquad (16)$$

as the residual, where N is the number of lensing-curve measurements and the σ_i are estimates of the errors in these measurements. The expected value of χ_N^2 is the number of degrees of freedom N. Unfortunately, the formal flux-measurement errors provided by modern programs for the reduction of CCD photometric data in fields with high stellar densities tend to be fairly severely underestimated. This is indicated, in particular, by the fact that the minimum χ_N^2 obtained when microlensing data are fit usually exceeds the number of degrees of freedom by a factor of three to five [5, 10, 11]. Thus, the expected residual cannot be accurately estimated. This essentially makes it impossible to apply regularization methods, since the resulting solutions depend directly on the residual. Nevertheless, we have used such a method to reconstruct $B(x)$ after doubling the *a priori* error estimates.

We carried out a search for solutions of Eq. (6) on a uniform grid of 160 points in the interval [$-0^{\text{d}}.275$, $+0^{\text{d}}.275$]. The resulting strip brightness distribution is mainly of interest because it allows us to correct the origin of the x axis. It follows from (4) or (6) that the shift in the time t_0 will be equal to the shift in the center of gravity of the reconstructed profile $B(x)$. We expect that our imprecise knowledge of the residual will not appreciably affect the position of this center of gravity. The time t_0 corrected in this way is JD 2450982.623, which is the value we used when constructing the plots in Fig. 1. The strip brightness distribution reconstructed using this corrected value is shown in Fig. 2. We can see that, in spite of the overestimated residual, the solution of (6) is "not sufficiently regularized." This is indicated by the noise oscillations superposed on the smooth $B(x)$ profile. Their noise-like nature is confirmed by the fact that the corresponding lensing curve $I_c(x_i)$, shown by the dashed curve in Fig. 1, begins to track clearly random deviations in the observed curve. Nonetheless, analysis of Fig. 2 enables us to estimate the characteristic radius of the star and to confirm the coincidence of the center of gravity of the $B(x)$ profile and the time origin. The next stage in reduction of the observations requires incorporation of a large amount of *a priori* information.
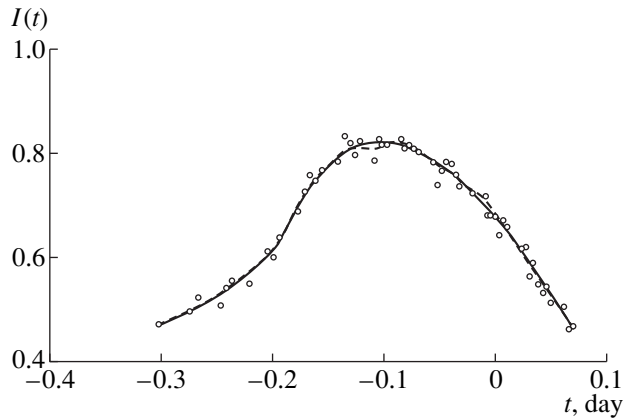
Fig. 1. Readout values of the MACHO 98-SMC-1 lensing curve, observed in the *I* band near the brightness maximum during the second intersection of a caustic (circles). The dashed curve shows the lensing curve corresponding to the reconstructed strip brightness distribution, and the solid curve shows the radial brightness distribution across the stellar disk.

Near the maximum brightness for MACHO 98-SMC-1, a spectrum of the star with a resolution of about 7 Å was obtained on the 1.9-m SAAO telescope. Analysis of this spectrum indicated that the spectral type of the lensed star was A6, and that its effective temperature was $T_{\text{eff}} = 8000$ K [10]. According to our current understanding, such a star has a rather thin atmosphere, and the radial brightness distribution across its disk $b(r)$ should be a convex-upward function of r. Note that the radius of the star R, which determines the overall width of the interval over which the search for $b(r)$ is carried out, is not known *a priori*, although a somewhat crude estimate can be made based on Fig. 2. When reconstructing $B(x)$, and even when searching for $b(r)$ on the class of monotonic and convex-upward functions, it was good enough to take an interval for variation of the argument that clearly (but not by too much!) exceeded the size of the source; however, when finding a convex-upward solution, we must use an accurate value for R. Thus, the radius of the star becomes a free parameter. In order to preserve complete independence from our inexact knowledge of the flux errors, we have also introduced a second free parameter—a correction to the time origin Δt.

We minimized the residual (16) on the class of convex-upward, non-negative functions for various values of R and Δt using a modified version of the PTISR program [12, 15], written in FORTRAN. To eliminate the effect of rounding errors, all REAL variables in PTISR and its associated subroutines were declared as DOUBLE PRECISION variables; this corresponds to the presence of 16 significant digits in the mantissa of a floating-point number. The search for $b(r)$ was carried out on a uniform grid in radius with step $\Delta = 0^{\text{d}}.001727$. It is evident that, in this case, we do not need to have pre-

Table 1. Readouts of $I_0(t_i)$ at times t_i and their rms errors σ_i, readouts of $I_c(t_i)$ corresponding to the reconstructed radial brightness distribution and O–C

i	JD t_i-2 450 000	$I_0(t_i)$	σ_i	$I_c(t_i)$	O–C	i	JD t_i-2 450 000	$I_0(t_i)$	σ_i	$I_c(t_i)$	O–C
1	982.315	0.4710	0.0088	0.4685	0.0025	29	982.548	0.8053	0.0089	0.8031	0.0022
2	982.342	0.4952	0.0064	0.4949	0.0004	30	982.561	0.7857	0.0124	0.7872	−0.0015
3	982.350	0.5217	0.0082	0.5037	0.0179	31	982.565	0.7402	0.0145	0.7814	−0.0413
4	982.370	0.5066	0.0070	0.5287	−0.0220	32	982.569	0.7674	0.0100	0.7754	−0.0081
5	982.375	0.5405	0.0065	0.5357	0.0048	33	982.573	0.7857	0.0095	0.7689	0.0168
6	982.380	0.5534	0.0072	0.5430	0.0104	34	982.578	0.7817	0.0087	0.7604	0.0213
7	982.396	0.5487	0.0066	0.5695	−0.0208	35	982.582	0.7603	0.0099	0.7532	0.0071
8	982.412	0.6115	0.0068	0.6025	0.0090	36	982.585	0.7387	0.0096	0.7475	−0.0088
9	982.417	0.6000	0.0067	0.6150	−0.0150	37	982.597	0.7243	0.0087	0.7231	0.0012
10	982.422	0.6392	0.0047	0.6287	0.0104	38	982.609	0.7168	0.0120	0.6959	0.0209
11	982.439	0.6884	0.0064	0.6959	−0.0075	39	982.611	0.6813	0.0108	0.6911	−0.0098
12	982.445	0.7280	0.0061	0.7178	0.0102	40	982.614	0.6813	0.0088	0.6838	−0.0025
13	982.450	0.7595	0.0056	0.7344	0.0251	41	982.618	0.6791	0.0075	0.6739	0.0053
14	982.455	0.7486	0.0055	0.7492	−0.0006	42	982.622	0.6432	0.0077	0.6635	−0.0203
15	982.460	0.7697	0.0064	0.7625	0.0072	43	982.625	0.6721	0.0075	0.6557	0.0164
16	982.475	0.7873	0.0095	0.7932	−0.0059	44	982.629	0.6603	0.0067	0.6448	0.0154
17	982.481	0.8354	0.0101	0.8022	0.0332	45	982.641	0.6167	0.0074	0.6111	0.0056
18	982.486	0.8219	0.0068	0.8082	0.0136	46	982.645	0.6193	0.0069	0.5993	0.0200
19	982.491	0.7995	0.0156	0.8131	−0.0136	47	982.649	0.5629	0.0073	0.5873	−0.0244
20	982.495	0.8252	0.0076	0.8162	0.0090	48	982.652	0.5893	0.0071	0.5782	0.0111
21	982.508	0.7881	0.0110	0.8219	−0.0338	49	982.657	0.5481	0.0071	0.5626	−0.0145
22	982.512	0.8295	0.0100	0.8223	0.0071	50	982.661	0.5307	0.0074	0.5498	−0.0190
23	982.515	0.8194	0.0099	0.8223	−0.0029	51	982.664	0.5434	0.0071	0.5401	0.0033
24	982.519	0.8185	0.0106	0.8218	−0.0032	52	982.668	0.5122	0.0076	0.5269	−0.0147
25	982.532	0.8295	0.0108	0.8164	0.0130	53	982.680	0.5039	0.0075	0.4863	0.0176
26	982.536	0.8119	0.0090	0.8137	−0.0019	54	982.684	0.4613	0.0073	0.4722	−0.0110
27	982.540	0.8177	0.0076	0.8106	0.0071	55	982.688	0.4664	0.0087	0.4582	0.0082
28	982.544	0.8110	0.0075	0.8070	0.0040						

cise values of the flux errors. It is sufficient to find a global minimum for the residual. Table 2 presents the residuals for various values of the free parameters. We can see that the minimum residual $\chi_N^2 = 268.8$ for $N = 55$ degrees of freedom is obtained when $R = 113$ $\Delta = 0^d.1951$ and $\Delta t = -0^d.002$. The small value for Δt reflects the good accuracy of our estimate of the shift in the time origin derived from our analysis of the position for the center of gravity for the $B(x)$ profile.

The radial brightness distribution corresponding to the global minimum in the residual is presented in Table 3 and shown by the solid curve in Fig. 3. The normalization was determined by the conditions $R = 1$, $b(0) = 1$. Values for the corresponding lensing curve $I_c(x_i)$ and O–C values are given in Table 1. This lensing curve is also shown by a solid curve in Fig. 1, and is in fairly good agreement with the observational data.

It is of interest to compare the resulting brightness distribution $b(r)$ with computations for model stellar atmospheres. The most detailed grid of models for plane-parallel stellar atmospheres and their corresponding brightness distributions for various wavelengths has been computed by Kurucz [16]. A more compact limb-darkening fit is achieved using two-parameter, non-linear laws; for example,

$$b_m(\mu) = b(1) [1 - a_1 (1 - \mu) - a_2(1 - \mu)^2], \qquad (17)$$

where μ is the cosine of the angle between the line of sight and the normal to the stellar surface and a_1 and a_2 are parameters. The values of these parameters for stars with various T_{eff} and free-fall accelerations g for photometric infrared bands can be found in [17]. Claret *et al.* [17] by the nonlinear law (17) fit the model brightness distributions [16] with errors of no more than 1%. If, in accordance with [18], we adopt a spectral class of A6
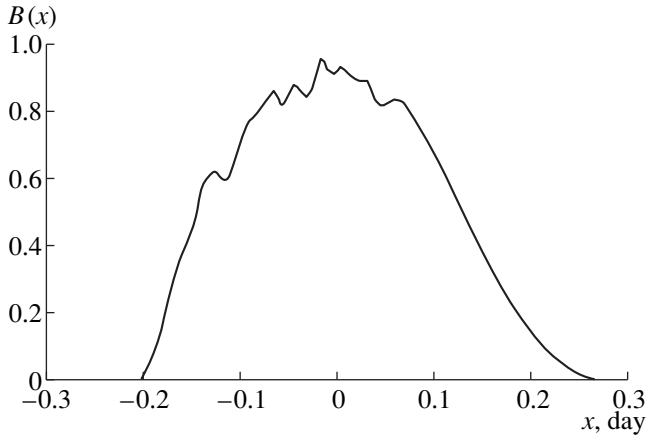
Fig. 2. One-dimensional strip brightness distribution across the disk of the lensed star reconstructed using a regularization method with account for the fact that the solution is non-negative.
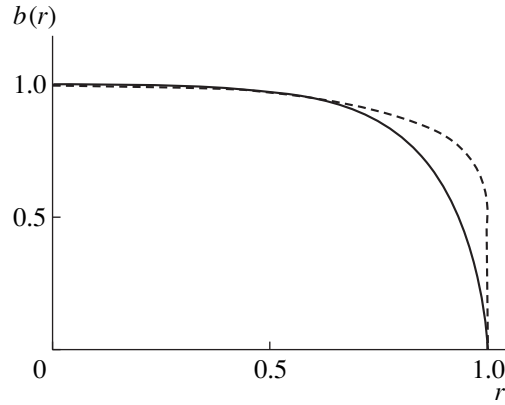

Fig. 3. Radial brightness distribution across the stellar disk found on the set of convex-upward, non-negative functions (solid curve). The dashed curve shows the distribution for a non-linear limb-darkening law for the A6 star.

and $\log g$ [cm s^{-2}] = 4.2 for the lensed star, then, if T_{eff} = 8000 K, we will have a_1 = 0.219 and a_2 = 0.234 in the I band [17]. When calculating the brightness distribution using (17), we took into account the fact that the readout values μ_i corresponding to the uniform grid of values r_i are related to them by the expressions $\mu_i = \sqrt{R^2 - r_i^2}/r_i$.

The resulting model brightness distribution $b_m(r_i)$ is presented in Table 3 and shown by the dashed curve in Fig. 3. We can see that the limb darkening for the reconstructed brightness distribution is substantially stronger than for $b_m(r)$. There is little doubt that model brightness distributions for such simple objects as A stars differ from the real distribution by more than several percent. It is most likely that this discrepancy is due to errors in the reconstruction of $b(r)$. Unfortunately, as noted above, we do not have accurate estimates of the flux errors, so we cannot judge the significance of discrepancies between the curves in Fig. 3.

For comparison, we can consider an analysis of observations of the near-caustic passage of a Galactic-bulge K giant MACHO 97-BLG-28, which, as mentioned above, were obtained by fitting models with non-linear limb-darkening laws [11]. Errors in the estimates of the limb-darkening parameters were not determined in [11]. However, the deviations of the fit limb-darkening law from the theoretical law were appreciably smaller than the discrepancies in Fig. 3. Note, however, that the flux from the star at maximum brightness was nearly two orders of magnitude higher for the MACHO 97-BLG-28 event than in our case, so that the errors in the lensing curve should have been appreciably smaller.

We can also obtain some idea of the accuracy in $b(r)$ from the solution of another problem with a similar character—analysis of observations of eclipsing vari-

able stars. Estimation of the degree of limb darkening from curves created by mutually eclipsing components proves to be a rather difficult problem, which requires high-accuracy observational data [19]. In this case it is usual to restrict the solutions to those for linear limb-darkening laws, which are obtained from (17) when a_2 = 0. The general conclusion that can be drawn for various analysis methods is that estimation of the limb-darkening coefficient with an accuracy of 0.1 requires knowledge of the light curve with a relative error of no more than 1%. In the presence of flux errors of 2–3%, deviations of the disk brightness distribution curves from their exact values comparable to those in Fig. 3 are quite possible.

Nevertheless, our analysis of this lensing curve using various methods making use of *a priori* information leads to a good agreement between the various results, and the constructed radial brightness distribution is in qualitative agreement with the model-atmosphere computations. This confirms the capabilities of our method in obtaining stable estimates of brightness distributions, even in the presence of comparatively

Table 2. Dependence of χ_N^2 on the stellar radius R (in units of the grid size Δ) and corrections to the time origin Δt in fractions of days ($N = 55$)

$R/\Delta t$	−0.006	−0.004	−0.002	0.000	0.002	0.004
110	353.6	353.6	353.6	332.6	332.6	332.6
111	382.7	382.7	382.7	296.2	296.2	296.2
112	423.8	423.8	275.6	275.6	275.6	464.1
113	475.2	268.8	268.8	268.8	387.0	387.0
114	274.0	274.0	274.0	331.9	331.9	331.9
115	289.4	289.4	289.4	296.8	296.8	296.8
116	305.6	305.6	280.3	280.3	280.3	555.0

Table 3. Reconstructed distribution $b(r_i)$ for the MACHO 98-SMC-1 event and the non-linear limb-darkening model distribution $b_m(r_i)$ for $T_{\text{eff}} = 8000$ K and $\log g$ [cm s^{-2}] = 4.2

i	$b(r_i)$	$b_m(r_i)$	i	$b(r_i)$	$b_m(r_i)$	i	$b(r_i)$	$b_m(r_i)$	i	$b(r_i)$	$b_m(r_i)$
1	1.0000	1.0000	30	0.9943	0.9918	58	0.9664	0.9642	87	0.8344	0.8901
2	1.0000	1.0000	31	0.9939	0.9913	59	0.9645	0.9627	88	0.8248	0.8859
3	1.0000	0.9999	32	0.9934	0.9906	60	0.9624	0.9611	89	0.8145	0.8815
4	0.9999	0.9999	33	0.9929	0.9900	61	0.9603	0.9595	90	0.8036	0.8769
5	0.9999	0.9998	34	0.9923	0.9893	62	0.9580	0.9578	91	0.7920	0.8720
6	0.9998	0.9997	35	0.9918	0.9887	63	0.9557	0.9561	92	0.7797	0.8670
7	0.9998	0.9996	36	0.9912	0.9880	64	0.9532	0.9543	93	0.7665	0.8617
8	0.9997	0.9994	37	0.9905	0.9872	65	0.9505	0.9524	94	0.7524	0.8562
9	0.9996	0.9993	38	0.9899	0.9865	66	0.9477	0.9505	95	0.7373	0.8504
10	0.9995	0.9991	39	0.9892	0.9857	67	0.9448	0.9485	96	0.7212	0.8443
11	0.9994	0.9990	40	0.9885	0.9848	68	0.9417	0.9464	97	0.7039	0.8378
12	0.9993	0.9988	41	0.9877	0.9840	69	0.9384	0.9443	98	0.6854	0.8310
13	0.9991	0.9985	42	0.9869	0.9831	70	0.9350	0.9421	99	0.6655	0.8238
14	0.9990	0.9983	43	0.9860	0.9822	71	0.9314	0.9398	100	0.6441	0.8162
15	0.9988	0.9980	44	0.9851	0.9813	72	0.9275	0.9375	101	0.6210	0.8080
16	0.9986	0.9978	45	0.9842	0.9803	73	0.9235	0.9351	102	0.5959	0.7993
17	0.9984	0.9975	46	0.9832	0.9793	74	0.9192	0.9325	103	0.5688	0.7900
18	0.9982	0.9972	47	0.9822	0.9782	75	0.9147	0.9299	104	0.5396	0.7799
19	0.9980	0.9968	48	0.9811	0.9772	76	0.9099	0.9273	105	0.5071	0.7689
20	0.9977	0.9965	49	0.9799	0.9761	77	0.9048	0.9245	106	0.4718	0.7569
21	0.9975	0.9961	50	0.9787	0.9749	78	0.8995	0.9216	107	0.4328	0.7437
22	0.9972	0.9957	51	0.9774	0.9737	79	0.8938	0.9186	108	0.3894	0.7288
23	0.9969	0.9953	52	0.9761	0.9725	80	0.8878	0.9155	109	0.3406	0.7119
24	0.9966	0.9949	53	0.9747	0.9712	81	0.8814	0.9123	110	0.2850	0.6920
25	0.9963	0.9944	54	0.9732	0.9699	82	0.8747	0.9089	111	0.2199	0.6675
26	0.9959	0.9940	55	0.9716	0.9686	83	0.8676	0.9054	112	0.1403	0.6341
27	0.9956	0.9935	56	0.9700	0.9672	84	0.8600	0.9018	113	0.0310	0.5470
28	0.9952	0.9929	57	0.9682	0.9657	85	0.8520	0.8981			
29	0.9948	0.9924				86	0.8434	0.8942			

high and imprecisely known noise levels. Note also that, for our chosen grid in ξ, a similarity in the quadrature formulas (5)–(6) is demonstrated in the integral equation (4) (see, for example, [20]).

5. CONCLUSION

Our method for the analysis of lensing curves observed during the passage of a star through a gravitational-lensing caustic can be used to reconstruct the shape of the brightness distribution across the stellar disk without knowledge of the characteristics of the lens. In the case of accurately known errors in the measured fluxes, it is not necessary to introduce any additional free parameters in the solution. Even in the complete absence of information about the accuracy of the measurements, only one free parameter is required to reconstruct the radial brightness distribution on a compact set of monotonic and bounded functions; for reconstruction on the set of convex-upward, non-negative functions, two free parameters are sufficient.

Application of our methods for reconstructing the strip and radial brightness distributions across a stellar disk to observations of the MACHO 98-SMC-1 microlensing event yields results that are in good agreement with each other and in qualitative agreement with data of model-atmosphere computations. The largest difficulty encountered in the analysis of observed microlensing caustics, as in the case of single gravitational lenses [6], remains our inexact knowledge of the flux-measurement errors. Elimination of this uncertainty for fields with high stellar densities will require a large effort in perfecting algorithms and computer programs for the reduction of CCD-photometry data.

Recently, observational programs to search for so-called high-amplitude events in the multiple images of gravitationally lensed quasars have begun. These are manifest as a sharp intensification of the image flux, associated with the passage of the quasar core through a caustic that has arisen due to microlensing by stars in the lensing galaxy. In spite of the large number of papers dedicated to the nature of the cores of these objects, there are currently no models that adequately describe their expected brightness distributions. We expect that application of our methods to the analysis of observations of such high-amplitude phenomena in quasar images could make it possible to derive important information about the brightness distribution in and around quasar accretion disks.

ACKNOWLEDGMENTS

REFERENCES

1. C. Alcock, C. W. Akerlof, R. A. Allsman, *et al.*, Nature **365**, 621 (1993).
2. M. B. Bogdanov and A. M. Cherepashchuk, Astron. Zh. **73**, 783 (1996) [Astron. Rep. **40**, 713 (1996)].
3. M. A. Hendry, I. J. Coleman, N. Gray, *et al.*, New Astron. Rev. **42**, 125 (1998).
4. S. B. Gaudi and A. Gould, Astrophys. J. **513**, 619 (1999).
5. C. Alcock, W. H. Allen, R. A. Allsman, *et al.*, Astrophys. J. **491**, 436 (1997).
6. M. B. Bogdanov and A. M. Cherepashchuk, Astron. Zh. **76**, 688 (1999) [Astron. Rep. **43**, 601 (1999)].
7. P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (Springer, Berlin, 1992).
8. A. F. Zakharov, *Gravitational Lenses and Microlenses* [in Russian] (Yanus, Moscow, 1997).
9. B. Grieger, R. Kayser, and T. Schramm, Astron. Astrophys. **252**, 508 (1991).
10. M. D. Albrow, J.-P. Beaulieu, J. A. Caldwell, *et al.*, Astrophys. J. **512**, 672 (1999).
11. M. D. Albrow, J.-P. Beaulieu, J. A. Caldwell, *et al.*, Astrophys. J. **522**, 1011 (1999).
12. A. V. Goncharskiĭ, A. M. Cherepashchuk, and A. G. Yagola, *Ill-Posed Problems of Astrophysics* [in Russian] (Nauka, Moscow, 1985).
13. V. I. Arnold, *Catastrophe Theory* (Springer, Berlin, 1986; Nauka, Moscow, 1990).
14. F. Natterer, *The Mathematics of Computerised Tomography* (Wiley, Chichester, 1986; Mir, Moscow, 1990).
15. A. N. Tikhonov, A. V. Goncharskiĭ, V. V. Stepanov, and A. G. Yagola, *Regularizing Algorithms and a Priori Information* [in Russian] (Nauka, Moscow, 1983).
16. R. L. Kurucz, Preprint No. 3348 (Harvard, 1991).
17. A. Claret, J. Diaz-Cordoves, and A. Gimenez, Astron. Astrophys., Suppl. Ser. **114**, 247 (1995).
18. C. W. Allen, *Astrophysical Quantities* (Athlone Press, London, 1973; Mir, Moscow, 1977).
19. *Eclipsing Variable Stars*, Ed. by V. P. Tsesevich [in Russian] (Nauka, Moscow, 1971).
20. S. M. Belotserkovskiĭ and I. K. Lifanov, in *Numerical Methods in Singular Integral Equitions, Their Application for Aerodynamics, Theory of Elasticity, and Electrodynamics* [in Russian] (Nauka, Moscow, 1985).

Translated by D. Gabuzda

Are Anomalous X-ray Pulsars
and Soft Gamma-ray Repeaters Magnetars?

Ya. I. Istomin[1] and B. V. Komberg[2]

[1]Lebedev Institute of Physics, Leninskiĭ pr. 53, Moscow, 117924 Russia
[2]Astro Space Center, Lebedev Physical Institute, ul. Profsoyuznaya 84/52, Moscow, 117810 Russia
Received November 15, 1999

Abstract—The magnetic fields of soft gamma-ray repeaters and anomalous X-ray pulsars have been estimated, taking into account the appreciable increase in the deceleration of the neutron star if it is embedded in a dense interstellar medium. These estimates yield the usual values of $B \simeq 10^{12}$ G. © *2000 MAIK "Nauka/Interperiodica".*

Only four soft gamma-ray repeaters (SGR) are currently known, listed in Table 1. The first—SGR 0526–66—was detected as early as March 5, 1979, and identified with the supernova remnant (SNR) N 49 in the Large Magellanic Cloud [1]. In Table 1 and the subsequent text, P is the period and $\dot{P}$ the time derivative of the period. Other SGRs are also located in young SNRs near the Galactic plane [2–5]. Their X-ray spectra are softer than those of ordinary gamma-ray bursts (mean energy 30 keV), and their energy releases are in the range from 10^{35}–10^{36} erg/s.

Many properties of SGRs (e.g., P, $\dot{P}$, and the energy release) are similar to those for anomalous X-ray pulsars (AXPs), of which fewer than ten are currently known [6]. A list of AXPs is presented in Table 2. AXPs are also located in SNRs near the Galactic plane. They are slightly weaker than ordinary accretion pulsars [7, 8], although AXPs do not exhibit features of binarity, in contrast to accretion pulsars (see, for example, [9]). Similarities in the properties of SGRs and AXPs have also been noted in [10–12].

In standard models in which the neutron star's rotational energy loss is assumed to be due to magnetic-dipole emission, both SGRs and AXPs must have large magnetic-field intensities at their surfaces:

$$B = \left(\frac{3c^3 I}{8\pi^2 R^6} P\dot{P} \right)^{1/2} = 3.2 \times 10^{19}(P\dot{P})^{1/2} > 10^{14} \text{ G. (1)}$$

Here, I is the moment of inertia of the neutron star and R is its radius.

Based on such estimates, such objects were classified as so-called "magnetars" (as distinct from normal pulsars, whose magnetic fields do not exceed the critical value $B_{cr} = 4.4 \times 10^{13}$ G). Thomson and Duncan [13] propose that such strong fields could be generated by a dynamo mechanism driven by convective turbulence. A number of papers have presented arguments that magnetars, indeed, exist. For example, Gotthelf *et al.* [8] have

noted that the SNR Kes 73 ($P = 12$ s) bears a similarity to accretion pulsars; the shape of its X-ray pulse profile and the linear character of the $\dot{P}(t)$ dependence have been maintained over ten years. In addition, the authors of [4, 13, 14] suggested the presence of magnetic fields exceeding B_{cr} to explain the emission of X-ray luminosities of SGRs and AXPs at levels of $\sim 10^{35}$ erg/s. This conclusion was based on the assumption that the energy release associated with the loss of rotational energy, $\dot{E} = I\Omega\dot{\Omega}$ (where $\dot{\Omega} = -k\Omega^n$, Ω is the angular velocity, n and k are constants, and $n = \Omega\ddot{\Omega}/\dot{\Omega}^2$ is the deceleration index), would be insufficient to explain their observed X-ray luminosities. Nevertheless, we note that the energy emitted by a rotating neutron star whose

Table 1. Soft gamma repeaters

SGR	P, s	$\dot{P}$, s/s
0526–66	8	
1806–20	7.47	8×10^{-11}
1900+14	5.16	10^{-10}
1627–41	6.7(?)	

Table 2. Anomalous X-ray pulsars

AXP	P, s	$\dot{P}$, s/s
PSR J1844–02.58	6.97	
1E 1841–04.5	11.76	6×10^{-11}
1E 2259+58.6	6.98	8×10^{-13}
4U 0142+61.5	8.69	2×10^{-12}
1E 1048–59.37	6.44	10^{-11}
RXI 1708.49–400.90	11.00	
Pulsar in SNR Kes 73	12.00	

surface magnetic field is B need not be $I\Omega\dot{\Omega}$. In fact, the energy loss of a neutron star could be associated not only with the amplitude of B (as for magnetars), but also with the field dissipation rate, since $E \sim \dfrac{d}{dt} \int \dfrac{B^2}{8\pi} dV$. In particular (see, for example, [13]), if superfluid and superconducting vortices are associated with each other, an increase in the "output" rate of superfluid vortices to the neutron-star crust (i.e., an increase in $\dot{\Omega}$) will result in a corresponding increase in the output of magnetic flux.

As proposed in [4], SGRs ($\tau \simeq 10^4$ yrs) could evolve into AXPs ($\tau \approx 10^5$ yrs), with a subsequent transition to solitary, cold neutron stars characterized by very low luminosities. A slightly different evolutionary scheme was put forward in [15]: from magnetars to accretion pulsars with very long rotational periods (up to $P = 2^h8$ for 2S 0114+650). In this type of picture, the evolution of a neutron star includes three successive stages (see also, for example, [16]).

(1) *The radio pulsar stage.* A rapidly rotating neutron star (with a period of the order of a millisecond) with a normal magnetic field forms after the explosion of a massive star (type II supernova). The deceleration of its rotation is governed by magnetic-dipole energy losses; the deceleration index is $n = 3$, and the characteristic deceleration time is $\tau = -\Omega/2\dot{\Omega} = P/(n-1)\dot{P}$. This phase continues until the pressure of the ambient medium becomes (in the course of the pulsar's deceleration) greater than the pulsar-wind pressure within the gravitational capture radius $R_{gr} = 2GM_{ns}/v_{ns}^2$, where M_{ns} is the mass of the neutron star and v_{ns} is the velocity of the neutron star with respect to the interstellar medium.

(2) *The "propeller" stage [17].* As the ambient matter penetrates to the pulsar magnetosphere, it accumulates inside the corotation radius $R_c = [GM(P/2\pi)^2]^{1/3}$, which gradually approaches the Alfvenic radius R_A. As a result, the rotational period of the neutron star tends toward the equilibrium value

$$P_{eq} = 20 B_{12}^{6/7} \dot{M}_{15\,accr}^{-3/7} R_6^{18/7} M_{1.4M_\odot}^{-5/7} \text{ s.} \qquad (2)$$

Here, $\dot{M}_{15\,accr}$ is the time variation of the mass in units of 10^{15} g/s, B_{12} is the magnetic field in units of 10^{12} G, and $M_{1.4M_\odot}$ is the mass in units of $1.4M_\odot$ for the time

$$\tau = 2.5 \times 10^8 B_{13}^{-1} \dot{M}_{11\,accr}^{-1/2} v_{10^8\,cm/s}^{-1} \text{ yrs},$$

where $\dot{M}_{11\,accr}$ is in units of 10^{11} g/s, B_{13} is in units of 10^{13} G, and $v_{10^8\,cm/s}$ is in units of 10^8 cm/s. The presence of strong magnetic fields is required in order for objects with $P = 10$ s to be formed within $\tau < 10^7$ yrs. This is precisely the case in which a magnetar can

evolve into an accretion pulsar over the time during which the flow of matter from a massive companion can maintain a sufficient accretion rate.

(3) *The stage of decay of the strong field and transformation of the magnetar into a slowly-rotating neutron star in a close binary system (i.e., into an accretion pulsar).*

We should note, however, that the evolutionary scheme considered in [15] assumes that both SGRs and AXPs are components of binary systems. Unfortunately, the available observations have not yet confirmed this supposition.

In general, the idea that there are fields exceeding B_{cr} in magnetars has recently encountered problems. For example, according to [18], magnetars should not possess appreciable radio emission. On the other hand, pulsed radio emission was detected for SGR 1900+14, at a frequency of 111 MHz, with $P = 5.16$ s and $\dot{P} = 1.2 \times 10^{-10}$ s/s [19]. Thus, this object has turned out to be similar to the radio pulsar J1907+0919. Moreover, according to the data of [20], there was an appreciable change in the value of $\dot{P}$ between two periods of activity (~1976 and 1998) in SGR 1900+14. This is difficult to understand in the framework of models with large field intensities. This behavior may be associated with the loss of momentum of the neutron star due to its relativistic wind.

In addition, the hypothesis of large magnetic fields in SGRs and AXPs encounters a number of other difficulties (see, for example, [21, 22]). In particular, it results in a considerable disagreement between the ages of SNRs and of the active objects in them (it is always the case that $\tau_{SNR} \gg \tau_{SGR,\,AXP}$). In addition, as a rule, the deceleration index n differs strongly from $n = 3$ to smaller values, characteristic of pure magnetic-dipole losses. Note that the same is observed for ordinary radio pulsars [23].

All these facts led some investigators to conclude that the rapid deceleration of SGRs and AXPs might be associated with interaction with an ambient plasma (which was partially studied for the "propeller" mechanism), rather than with the presence of very strong fields. Let us consider this possibility in more detail.

In agreement with [24], the deceleration moment of a magnetic dipole rotating in vacuum is

$$\mathbf{R} = -\frac{2}{3}\frac{\Omega^2}{c^3}[[\mathbf{M\Omega}]\mathbf{M}],$$

where $\mathbf{M}$ is the magnetic moment. On the other hand, if a magnetic dipole rotates in a medium with dielectric permittivity $\dot{P} \neq 1$, its deceleration moment increases to

$$\mathbf{R}_* = \varepsilon^{3/2}\mathbf{R},$$

and, consequently, $\dot{P} \sim \varepsilon^{3/2}$. As a result, the characteristic deceleration time τ becomes less than the kinematic time:

$$\tau_* = \tau_{\text{kin}}/\varepsilon^{3/2}.$$

The dielectric permittivity can be estimated using the formula

$$\varepsilon \approx 1 + \frac{4\pi m_p n_p^{(L)} c^2}{B_L B_0} \gamma_p, \qquad (3)$$

where $n_p^{(L)}$ is the density of relativistic protons with Lorentz factor $\gamma_p = [1 - (v/c)^2]^{-1/2}$ that penetrate the region of the light cylinder, B_L is the magnetic field near the light cylinder, B_0 is the magnetic field in the ambient plasma, and m_p is the proton mass. Substituting $B_L = 10^{12} B_{12} (R\Omega/c)^3$ and $n_p^{(L)} = n_p(B_L/B_0)$ into (3), we obtain

$$\varepsilon = 1 + 2.4 \frac{P^3}{B_{12}} \left(\frac{n_p}{1 \text{ cm}^{-3}}\right) \left(\frac{B_0}{10^{-3} \text{ G}}\right)^{-1} \gamma_p. \qquad (4)$$

In fact, this value for ε is only an upper limit, since not all protons of the surrounding medium can penetrate the shock wave formed by the interaction of the pulsar wind with the ambient plasma. It is also evident that n_p will decrease as the angle between $\boldsymbol{\Omega}$ and $\mathbf{v}_{\text{ns}}$ increases.

The results of [24] can be used to estimate the parameters of the surrounding medium required for substantial deceleration of the neutron star's rotation without assuming the presence of very strong magnetic fields. For example, let us consider the case of SGR 1900+14 ($P \simeq 5$ s, $\dot{P} \simeq 10^{-10}$ s/s). Since $\dot{P} = \varepsilon^{3/2}(2\pi)^2 M^2 i/c^3 PI$ (where $i \approx 1$ is the dimensionless current and $I = 10^{45}$ g cm^2 is the moment of inertia), we have

$$\dot{P} = 1.5 \times 10^{-12} \left(\frac{P}{5 \text{ s}}\right)^{7/2} B_{12}^{1/2} \left(\frac{n_p}{1 \text{ cm}^{-3}}\right)^{3/2}$$
$$\times \left(\frac{B_0}{10^{-3} \text{ G}}\right)^{-3/2} \gamma_p^{3/2} \left(\frac{R}{10 \text{ km}}\right)^6 \left(\frac{I}{10^{45} \text{ g cm}^2}\right)^{-1} \text{ s/s} \qquad (5)$$

or

$$B_{12} = 0.44 \left(\frac{\dot{P}}{10^{-12}}\right)^2 \left(\frac{P}{5 \text{ s}}\right)^{-7} \left(\frac{n_p}{1 \text{ cm}^{-3}}\right)^{-3} \left(\frac{B_0}{10^{-3} \text{ G}}\right)^3$$
$$\times \gamma_p^{-3} \left(\frac{R}{10 \text{ km}}\right)^{-12} \left(\frac{I}{10^{45} \text{ g cm}^2}\right)^2. \qquad (6)$$

Taking $P = 5$ s, $\dot{P} = 10^{-10}$ s/s, $\gamma_p \approx 1$, $B_0 = 10^{-4}$, $n_p = 1$, and other values standard for a neutron star, we find

$$B_{12} = 4.4 < B_{12\text{cr}}.$$

Thus, substantial deceleration of the neutron star's rotation can be obtained without assuming the presence of very strong magnetic fields. In this case, if $\varepsilon \longrightarrow 3$, then the deceleration index $n = \dfrac{3(3 - \varepsilon)}{2\varepsilon} \longrightarrow 0$, which is observed in SGR 1900+14. Appreciable deceleration of neutron stars embedded in surrounding media can also resolve the disagreement between the ages of SNRs and their corresponding neutron stars, since the true ages of the objects derived from $\dot{P}$ will increase ($\dot{P}$ is large only during the passage of the object through a dense cloud). In addition, recall that, in a medium, $\tau \approx \tau_{\text{kin}}/\varepsilon^{3/2} \leq \tau_{\text{kin}}$ (since $\varepsilon > 1$).

This approach was applied in [25] to PSR 1509–58, which is associated with SNR G320.4–01.2. The age of this radio pulsar ($P = 0.15$ s, $\dot{P} = 1.5 \times 10^{-12}$ s/s) estimated using the formula for magnetic-dipole losses turns out to be < 2000 years, whereas the age of the SNR is estimated to be $\sim 10^4$ years. Attempts to remove this disagreement have usually focused on trying to decrease the age of the SNR. In this case, one must assume that $E_{\text{SN}} = 10^{53}$ erg and $n_{\text{ext}} \simeq 0.1$–$0.01$ cm^{-3}. Based on the results presented in [24], Gvaramadze [25] proposed another approach. He assumed that $v_{\text{PSR}} \simeq 100$ km/s and $\varepsilon \approx 3$–5 and evaluated the resulting B and n_p. Indeed, since

$$\dot{\Omega} = -\varepsilon^{3/2} \frac{2M^2\Omega^3}{3c^2 I}, \quad M = \frac{BR^3}{2},$$
$$\varepsilon = 1 + 2.5 \times 10^{-6} \left(\frac{B}{10^{13} \text{ G}}\right)^{-2} \left(\frac{v_{\text{ns}}}{100 \text{ km/s}}\right)^{-3} n_p, \qquad (7)$$

these values of ε and v_{PSR} lead to

$$B \approx 10^{13} \text{ G}, \quad n_p \approx 10^6 \text{ cm}^{-3}, \text{ and } \tau_{\text{PSR}} = 10^3 \tau_{\text{kin}}.$$

The disagreement between the ages of PSR 1757–24.4 and SNR G5.27–0.9 was resolved in the same way in [26]. As a result, the velocity $v_{\text{PSR}} = 2300$ km/s (!) obtained in [27] can be reduced to the more reasonable value $v_{\text{PSR}} = 460$ km/s (for $\gamma_p = 1$, $n_p = 1$ cm^{-3}, and $B_0 = 10^{-6}$ G).

Indirect evidence confirming the role of deceleration of the rotation of neutron stars in dense media is the fact that objects with large $\dot{P}$ are located at low Galactic latitudes. (Along with SGRs and AXPs, this is true of the pulsars PSR 1757–24 and PSR 1509–58.) It is interesting that, for 15 pulsars associated with dense HII regions listed in [28], we find $\bar{\dot{P}} = 7 \times 10^{-12}$ s/s, which is considerably larger than the estimate $\bar{\dot{P}} = 10^{-13}$ s/s obtained for all radio pulsars. If we plot the data for 400 radio pulsars with $P > 0.1$ s in a $\dot{P} - |b|$ diagram (where b is Galactic latitude), we can see (Table 3

and the figure) that objects with $\dot{P} > 3 \times 10^{-13}$ s/s are located at $|b| < 10°$. Moreover, while there is no $\dot{P}(b)$ dependence at $|b| > 10°$, we find $\dot{P} \propto |b|^{-0.7}$ at $|b| < 10°$. Since the average time τ_{kin} increases with $|b|$ for radio pulsars (namely, $\tau_{kin} \propto |b|^{+0.7}$), and standard pulsar theory predicts $\dot{P} \propto \tau^{-1/2}$, we should observe $\dot{P} \propto |b|^{-0.35}$.

Another piece of indirect evidence in favor of deceleration of neutron stars in surrounding media would be a relation between the pulse widths of radio pulsars and $\dot{P}$ (or n_{ext}). The pressure of the ambient medium decreases the light-cylinder radius R_L, which, in turn, increases the polar-cap size $r_{pc} \approx R_0(R_0/R_L)^{1/2}$. As a result, some dependence between the mean widths of the radio pulses and the values of $|b|$ for pulsars could form (Table 4). Indeed, an increase in the widths of radio pulses is observed with approach toward the Galactic plane.

Entry of a normal radio pulsar moving with a speed $v_{PSR} = 300$ km/s into a dense cloud of ionized gas with $M_{cl} \simeq 10^{-2}M_{\odot}$, $D_{cl} = 0.3$ pc, and $n_p \simeq 10$ cm^{-3} could lead to a change in its period by

$$\Delta t = \dot{P}\Delta t \simeq 10^{-10} \times 3 \times 10^{10} = 3 \text{ s}$$

for $\Delta t \simeq 1000$ yrs and $\dot{P} \simeq 10^{-10}$ s/s. When the pulsar leaves the cloud, its $\dot{P}$ should sharply decrease, and the neutron star may enter the "death region" for radio pulsars in the $\dot{P}$–P plane. Thus, neutron stars embedded in surrounding media have accelerated rates of evolution compared to normal radio pulsars in the $\dot{P}$–P diagram.

We should also note that, as follows from formula (7), the most efficient deceleration in a medium should be experienced by radio pulsars with small values of B and v_{ns}. The lowest B are typical of millisecond pulsars, whose rapid rotation is thought to be the result of accretion of matter from a companion in a close binary system [29, 30]. (The pair could be disrupted after the explosion of the companion.) After its entry into a dense cloud, the $\dot{P}$ of a millisecond pulsar should sharply change, since $\dot{P} \propto \varepsilon^{3/2}P^{-1}$. Consequently, the millisecond pulsar could enter for some time Δt the region of magnetars in the $\dot{P}$–P diagram, and be manifest observationally as an SGR (but without an ambient SNR). After leaving the dense cloud, the millisecond pulsar (whose $\dot{P}$ decreased but period P increased during its residence in the cloud for a time Δt) can again enter the region of normal radio pulsars.

In the framework of the hypothesis under consideration, we suggest that SGRs could, in fact, be normal radio pulsars whose $\dot{P}$ have sharply increased during their passage through dense magnetized clouds of inter-

$\bar{P}$, 10^{-13} s/s
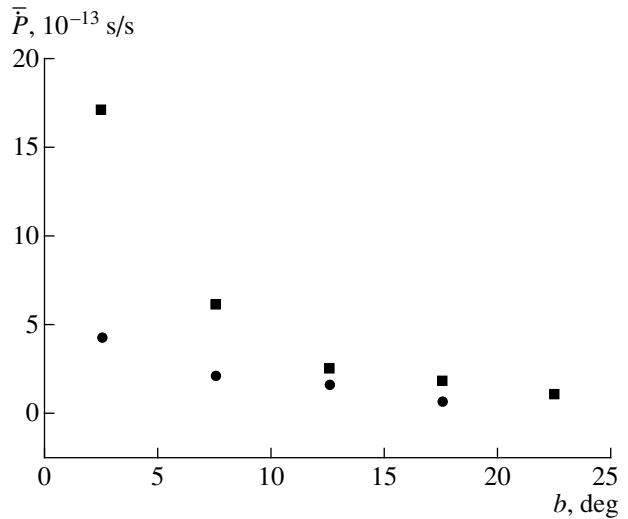
Dependence of the average rate of change in the period P on Galactic latitude for 400 radio pulsars with $P > 0.1$ s. The squares denote mean values, and the circles median values.

stellar gas. It is clear that the large values of $\dot{P}$ should bring about frequent reconstructions of the neutron-star crust under the influence of the variable moment of rotation. Therefore, period glitches should be frequent, which could plausibly result in gamma-ray bursts (see, for example, [31–33]). Thus, in the picture considered here, SGRs should not be numerous, since they appear

Table 3. Dependence of the rate of change of the period on Galactic latitude from data for 400 radio pulsars

Interval of b, deg	$\dot{P}$, 10^{-13} s/s		Dispersion	Mean deviation	Standard deviation
	median	mean			
0–5	4.25	17.06	2484.77	20.94	49.84
5–10	2.13	6.07	356.44	6.88	18.88
10–15	1.61	2.47	11.64	1.95	3.41
15–20	0.67	1.79	4.99	1.68	2.23
20–25	0.95	1.11	0.89	0.76	0.94

Table 4. Dependence of the effective pulse width on Galactic latitude from data for 400 radio pulsars

Interval of b, deg	Effective pulse width (percent of the period)		Standard deviation
	median	mean	
0–5	37.5	53.67	51.99
5–10	30.65	34.47	18.27
10–15	28.60	43.48	44.69
15–20	23.45	32.18	26.01
20–25	19.40	22.03	9.99

only for a time $\Delta t \simeq 10^3$ yrs, which is $\sim 10^{-5}$ of the lifetime of an old neutron star. Note that the number of such neutron stars is quite sufficient to explain the observed number of gamma-ray bursts if a considerable fraction are Galactic in nature, and each neutron star produces at least one gamma-ray burst every 1000 years (see, for example, [34]).

In conclusion, we wish to discuss another interesting fact: the spatial proximity of the two peculiar objects SGR 1900+14 and SS 433 [35, 36]. Both are located at approximately the same distance from us (~ 5.5 kpc) and are separated by an angular distance of $\sim 4°$, which corresponds to ~ 300 pc. Since the age of SNR W50, in which the binary SS 433 lies ($P_1 = 13^d$, $P_2 = 165^d$) is estimated to be 10^5 yrs, we suggest that an object that later evolved to become SGR 1900+14 was ejected 10^5 yrs ago from a triple system, which was a predecessor of SS 433 [37]. The corresponding velocity of ejection would be only ~ 300 km/s, which is in reasonable agreement with the velocity acquired by a component during a supernova explosion in a multiple system. It would be interesting to determine the direction of the velocity of SGR 1900+14: if this object was ejected from the SS 433 system, the proper velocity of SGR 1900+14 should be directed away from SS 433. (It is noted in the recent paper [37] that the 226-ms pulsar J1907+0918 is also within $\sim 5°$ of SGR 1900+14, and could have been born in a single massive binary system together with SGR 1900+14.)

ACKNOWLEDGMENTS

REFERENCES

1. E. P. Mazets, S. V. Golenetski, and Y. A. Gur'yan, Pis'ma Astron. Zh. **5**, 641 (1979) [Sov. Astron. Lett. **5**, 343 (1979)].
2. C. Kouvelioton, T. Strochmayer, K. Hurley, *et al.*, Astrophys. J. Lett. **510**, L115 (1999).
3. E. P. Mazets, R. L. Aptekar, P. S. Butterworth, *et al.*, astro-ph/9902292.
4. C. Kouvelioton, Proc. Natl. Acad. Sci. USA **96**, 5351 (1999).
5. J. S. Heyl and L. Hernquist, Mon. Not. R. Astron. Soc. **304**, L37 (1999).
6. S. Mereghetti and L. Stella, Astrophys. J. Lett. **442**, L17 (1995).
7. P. Reig and P. Roche, Mon. Not. R. Astron. Soc. **306**, 100 (1999).
8. E. V. Gotthelf, G. Vasisht, and T. Datani, astro-ph/9906122.
9. C. A. Wilson, S. Dieters, M. H. Finger, *et al.*, Astrophys. J. **513**, 464 (1999).
10. A. Melatos, Astrophys. J. Lett. **519**, L77 (1999).
11. C. Kouvelioton, S. Dieters, T. Strochmayer, *et al.*, Nature **393**, 235 (1998).
12. S. Mereghetti, G. L. Israel, L. Stella, *et al.*, Mon. Not. R. Astron. Soc. **296**, 689 (1998).
13. C. Thomson and R. Duncan, Astrophys. J. Lett. **392**, L9 (1992).
14. H. C. Spruit, Astron. Astrophys. **341**, L1 (1998).
15. X.-D. Li and E. P. van den Heuvel, Astrophys. J. Lett. **513**, L45 (1999).
16. V. M. Lipunov and S. B. Popov, astro-ph/9906012.
17. A. F. Illarionov and R. A. Sunyaev, Astron. Astrophys. **39**, 185 (1975).
18. M. G. Baring and A. K. Hurding, Astrophys. J. Lett. **507**, L55 (1998).
19. Yu. P. Shitov, IAU Coll. No. 7110 (1999).
20. D. Marsden, R. E. Rothschied, and E. Lingenfelter, astro-ph/9904244.
21. L. Stella, S. Mereghetti, and G. L. Israel, astro-ph/9605058.
22. S. Mereghetti, L. Stella, and G. L. Israel, astro-ph/9712254.
23. A. Baykal, M. A. Alpar, P. E. Boyton, *et al.*, Mon. Not. R. Astron. Soc. **306**, 207 (1999).
24. Ya. N. Istomin, Astron. Astrophys. **283**, 85 (1994).
25. V. Gvaramadze, Astron. Astrophys., 2000 (in press).
26. G. M. Dubner, M. Holdaway, W. N. Goss, and I. F. Mirabel, Astron. J. **116**, 1842 (1998).
27. D. A. Frail and S. R. Kulkarni, Nature **352**, 785 (1991).
28. A. V. Pynzar' and V. I. Shishov, Astron. Zh. **76**, 16 (1999) [Astron. Rep. **43**, 436 (1999)].
29. G. S. Bisnovatyĭ-Kogan and B. V. Komberg, Pis'ma Astron. Zh. **2**, 338 (1976) [Sov. Astron. Lett. **2**, 130 (1976)].
30. M. Toscano, J. S. Sandhu, M. Bailes, *et al.*, Mon. Not. R. Astron. Soc. **307**, 925 (1999).
31. G. S. Bisnovatyĭ-Kogan, V. M. Chechetkin, V. S. Imshennik, and D. K. Nadezhin, Astrophys. Space Sci. **35**, 23 (1975).
32. A. I. Tsygan, Astron. Astrophys. **44**, 21 (1975).
33. K. Chen and M. Ruderman, Astrophys. J. **402**, 264 (1993).
34. B. V. Komberg and D. A. Kompaneets, Astron. Zh. **74**, 690 (1997) [Astron. Rep. **41**, 611 (1997)].
35. R. C. Vermeulen, Preprint No. 35, OVRO (1995).
36. G. Pugliese, H. Falcke, and P. L. Biermann, Astron. Astrophys. **344**, L37 (1999).
37. K. Xilouris and D. R. Lorimer, astro-ph/9911027.

Translated by Yu. Dumin

The Large-Scale Solar Magnetic Field and 11-Year Activity Cycles

V. I. Makarov and A. G. Tlatov

Kislovodsk Mountain Astronomical Station, Main Astronomical Observatory, Russian Academy of Sciences,
P.O. Box 1, Kislovodsk, 357741 Russia

Received May 14, 1995

Abstract—Magnetic Hα synoptic maps of the Sun for 1915–1999 are analyzed and the intensities of spherical harmonics of the large-scale solar magnetic field computed. The possibility of using these Hα maps as a database for investigations of long-term variations of solar activity is demonstrated. As an example, the magnetic-field polarity distribution for the Hα maps and the analogous polarity distribution for the magnetographic maps of the Stanford observatory for 1975–1999 are compared. An activity index $A(t)$ is introduced for the large-scale magnetic field, which is the sum of the magnetic-moment intensities for the dipole and octupole components. The 11-year cycle of the large-scale solar magnetic field leads the 11-year sunspot cycle by, on average, 5.5 years. It is concluded that the observed weak large-scale solar magnetic field is not the product of the decay of strong active-region fields. Based on the new data, the level of the current (23rd) solar-activity cycle and some aspects of solar-cycle theory are discussed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The origin of the large-scale solar magnetic field, its role in the organization of the overall magnetic field of the Sun, and its relation to sunspot activity are likely key points for our understanding of the 22-year magnetic cycle. One remarkable feature of solar activity is the 11-year cycle for the appearance of local magnetic fields in the solar atmosphere, with induction values of up to 5 kG, which have different polarities in the northern and southern hemispheres of the Sun and change their sign at the activity minimum. The large-scale solar field also changes its polarity, but at the activity maximum of each 11-year cycle. The characteristics of polarity reversals of the solar magnetic field have been studied in detail for more than 120 years [1]. At latitudes higher than 50°, the polar-activity cycle begins after the polarity reversal of the polar magnetic field; it also lasts about 11 years, but proceeds in antiphase with the sunspot cycle. These cycles are related, and manifest a single process of global solar activity at all latitudes, from the poles to the equator [2]. This structure of the global magnetic cycle of the Sun has been discussed on the basis of both observations [3, 4] and theory [5, 6].

Over the past two decades, Makarov and Sivaraman [7] constructed magnetic Hα maps of the polarity distribution for the large-scale solar magnetic field from observations carried out at the Kodaikanal Observatory (India) from 1910 to 1975. These data were later augmented with Hα maps from the Kislovodsk Mountain Astronomical Station of the Main Astronomical Observatory of the Russian Academy of Sciences. Based on the magnetic Hα maps for 1910–1999, the topology of

the large-scale solar magnetic field and its role in the development of solar activity were studied in [8–13]. Makarenko *et al.* [14] found qualitative evidence that the weak large-scale solar magnetic field may be primary with respect to the strong local magnetic fields of sunspots. However, no quantitative characterization of this relationship was found. On the other hand, early theories of the solar cycle took the weak large-scale magnetic field to result from the decay of strong active-region fields and their poleward drift [15, 16]; i.e., to be a secondary product of the activity of local magnetic fields. In some studies, meridional circulation was invoked to account for the poleward drift [17–20]; however, the question of the durations of the solar cycle and the field polarity reversals remains open.

On the other hand, studies of long-term, poleward magnetic-field drifts [1, 3, 4], equatorward field drifts or torsional oscillations [11–13, 21], high-latitude and polar-facula magnetic activity [2, 3, 22], and coronal intensity variations over the course of the global activity cycle [11, 12] have shown that the magnetic activity of the Sun is not manifest only in local magnetic fields. According to global solar-cycle models, a new activity cycle starts at high latitudes several years before the emergence of the earliest new active regions of the new cycle, whereas its pattern-forming activity appears only later [8].

Here, we discuss the quantitative relationship between the 11-year cycle of the large-scale magnetic field and the 11-year cycle of local magnetic fields for the past eight solar-activity cycles, from 1915 to 1999. We use new data to discuss the activity level of the current (23rd) cycle, as well as some aspects of solar-cycle theory.

2. OBSERVATIONAL DATA AND PROCESSING TECHNIQUES

Information about the intensities of modes for the large-scale magnetic field and poleward and equatorward drifts of the fields were derived from Hα synoptic magnetic maps for 1915–1999 constructed using the technique described in [7]. Data on Wolf numbers, numbers of groups, and sunspot areas for this period were obtained in accordance with [2, 23, 24]. To compare the magnetic Hα maps with the magnetographic observations of the Stanford observatory, we expanded synoptic maps taken to have magnetic-field polarities of +1 or –1 G in spherical harmonics.

The surface magnetic field of the Sun can be represented as a function of latitude θ and longitude φ using the spherical-harmonic expansions

$$B_r = \sum_l \sum_m P_l^m (g_l^m \cos(m\phi) + h_l^m \sin(m\phi)),$$

where P_l^m are the associated Legendre functions. The expansion coefficients g_l^m and h_l^m can be found via a surface integration:

$$g_l^m = \frac{(2l+1)}{2\pi}\frac{(l-m)!}{(l+m)!}\int_0^{2\pi} d\phi \cos(m\phi)$$

$$\times \int_0^\pi B_r(\theta, \phi) P_l^m(\cos(\theta)) \sin(\theta) d\theta,$$

$$h_l^m = \frac{(2l+1)}{2\pi}\frac{(l-m)!}{(l+m)!}\int_0^{2\pi} d\phi \sin(m\phi)$$

$$\times \int_0^\pi B_r(\theta, \phi) P_l^m(\cos(\theta)) \sin(\theta) d\theta.$$

Here, $B_r(\theta, \phi)$ is the surface magnetic field. In our case, we used only the sign of the magnetic field for the Hα synoptic maps: +1 or –1 G for positive and negative magnetic fields, respectively. If the expansion coefficients g_l^m and h_l^m are known, the synoptic map of the magnetic field can be reconstructed and its distribution analyzed.

When performing the expansion, we assumed that the number of harmonics was $l = 9$. We carried out comparisons for the period of the observations at the Stanford Observatory, for Carrington rotations 1642–1950 (i.e., for the years 1975–1999).

3. RESULTS

Figure 1 displays latitude–time diagrams for the magnetic-field polarity distribution obtained from the Kislovodsk Hα synoptic maps and the Stanford magne-

tographic observations. A comparison reveals fairly good agreement in these distributions for the period under study. The observed small differences may be due to inaccuracy in constructing the Hα maps and/or possible displacements of the zero point in the magnetographic observations. The similarity of the patterns of the magnetic-field poleward drift for 1975–1999 obtained using the two different sources testifies to the uniformity and reliability of the Hα-map series for the Kislovodsk Mountain Astronomical Station data. The general pattern of the poleward drift and sign reversals of the magnetic field for 1915–1999 derived from the spherical-function expansions is shown in Fig. 1.

The time variations in the coefficients of the spherical-function expansion for the magnetic field reflect changes in the topology of the large-scale magnetic field. Obviously, the presence of bipolar active sunspot regions during the period of activity maximum increases the contribution of high-l harmonics. Conversely, the simplification of the distribution of large-scale fields during the period of activity decline and minimum increases the contribution of lower harmonics. Thus, in general, periods of intensity maxima for the harmonics should be displaced with respect to maxima in the sunspot distribution; i.e., they should have a phase shift that depends on l.

Let us consider the time behavior of low-l harmonics. We represent the magnetic dipole ($l = 1$) and octupole ($l = 3$) moments in the form

$$\mu_1 = \left(\sum_{m, l=1} (g_l^m g_l^m + h_l^m h_l^m) \right)^{1/2},$$

$$\mu_3 = \left(\sum_{m, l=3} (g_l^m g_l^m + h_l^m h_l^m) \right)^{1/2},$$

where g_l^m and h_l^m are the coefficients of the spherical-function expansion. We introduce a parameter characterizing the intensity of the dipole and octupole magnetic moments:

$$A(t) = (\mu_1^2 + \mu_2^2/3)^2.$$

The top panel of Fig. 2 shows the distribution of $A(t)$ for 1915–1999. The 11-year cycle of the large-scale solar magnetic field can clearly be seen. We applied a sliding-averaging procedure with a 2-year window to reduce the noise. The bottom panel of Fig. 2 presents the 11-year solar-activity cycle in the Wolf number W. Both $A(t)$ and $W(t)$ exhibit cyclic behavior with a period of about 11 years, with the phase shift of $A(t)$ relative to the Wolf number being about 5.5 years (Fig. 3). The cycle of the large-scale magnetic field leads the sunspot cycle. The relative variations in the amplitudes of $A(t)$ are similar to the cycle-to-cycle variations in the amplitudes of $W(t)$, but are phase shifted. As for the Wolf numbers, $A(t)$ increases from 1920 to 1956, then
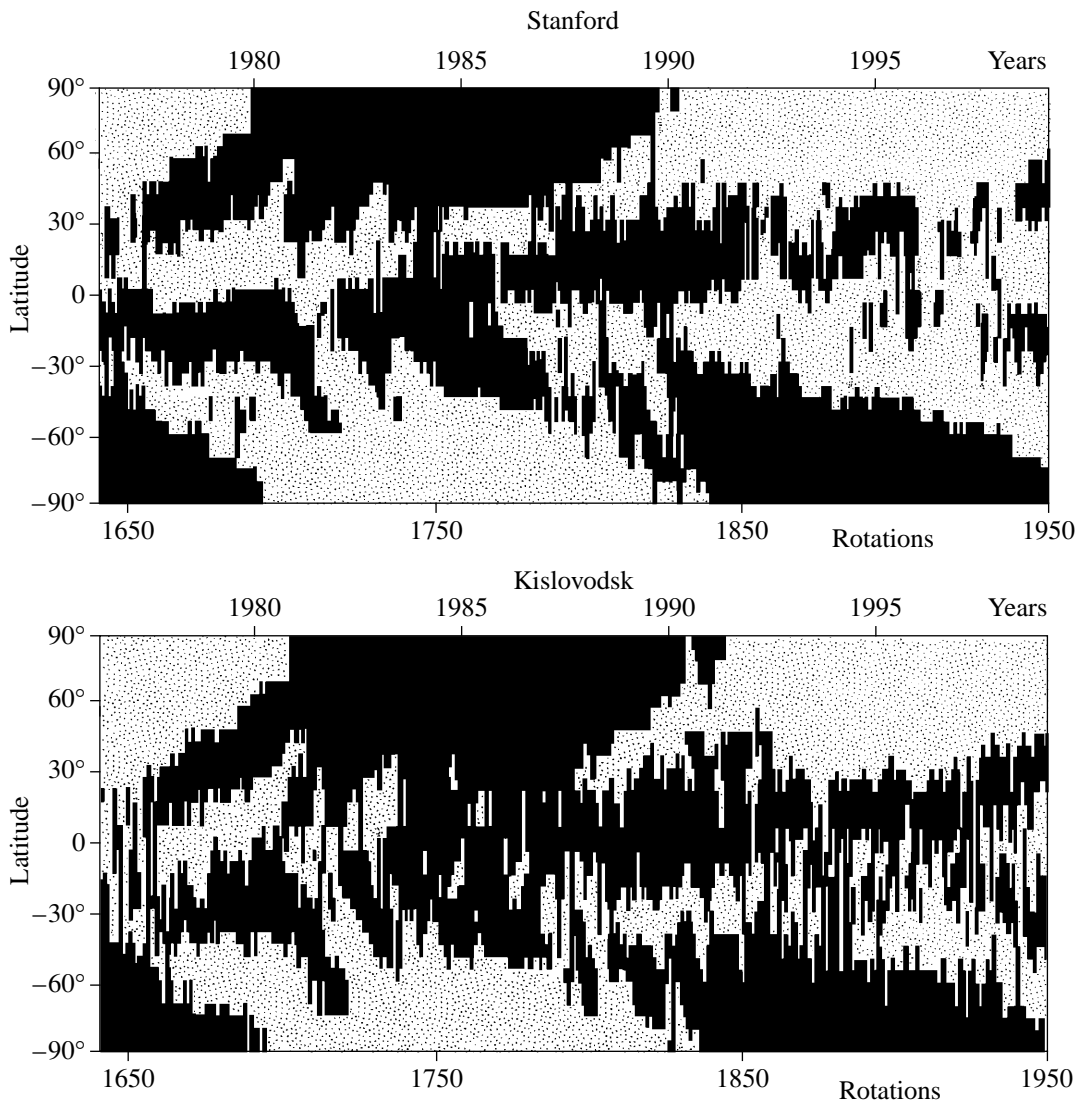
Stanford

Kislovodsk

Fig. 1. Latitude–time distribution of polarities of the large-scale solar magnetic field derived from Stanford magnetographic observations (top) and Hα synoptic magnetic maps obtained at the Mountain Astronomical Station of the Main Astronomical Observatory (bottom) for 1975–1999. The light and dark regions correspond to + and – magnetic-field polarities, respectively. The zonal structure of the solar magnetic field, latitude drift, and reversal of the polar magnetic field can be seen.

sharply falls during the period of the minimum preceding cycle 20. Later, the values at the minima before cycles 21 and 22 are roughly equal, and relatively low values are observed at the minimum preceding activity cycle 23.

The parameter $A(t)$ describing cyclic variations in the large-scale magnetic field includes the intensities of the magnetic moments of only the dipole and octupole components. Higher odd harmonics, e.g., $l = 5, 7, 9, \ldots$, can also be taken into account; however, as can be seen from the calculations, they enhance only the noise component. The intensity of the even modes $l = 2, 4, 6, \ldots$ is low, and their contribution is significant only in analysis of the quasi-biennial cycle (Fig. 4).

Our results confirm that the large-scale magnetic field plays a leading role in the organization and

dynamics of the solar cycle. The amplitude of the total power of the dipole and octupole components ($l = 1, 3$) depends on the field topology at middle and high latitudes. Therefore, the field of a new cycle, which has the sign of the leading spots of the next activity cycle, determines the polarity topology—primarily for middle and high latitudes at the phases of activity decline and minimum of the current cycle. Obviously, the stronger the high-latitude field, the less likely the penetration of the opposite-polarity field to the poles (associated with meridional circulation), and the higher the smoothed value of the index $A(t)$. Comparisons between $A(t)$ and the Wolf-number index $W(t)$ demonstrate possibilities for solar-activity forecasting. For instance, the current (23rd) activity cycle is expected to be weaker than cycle 22, since $A(23)$ is about 11.4,
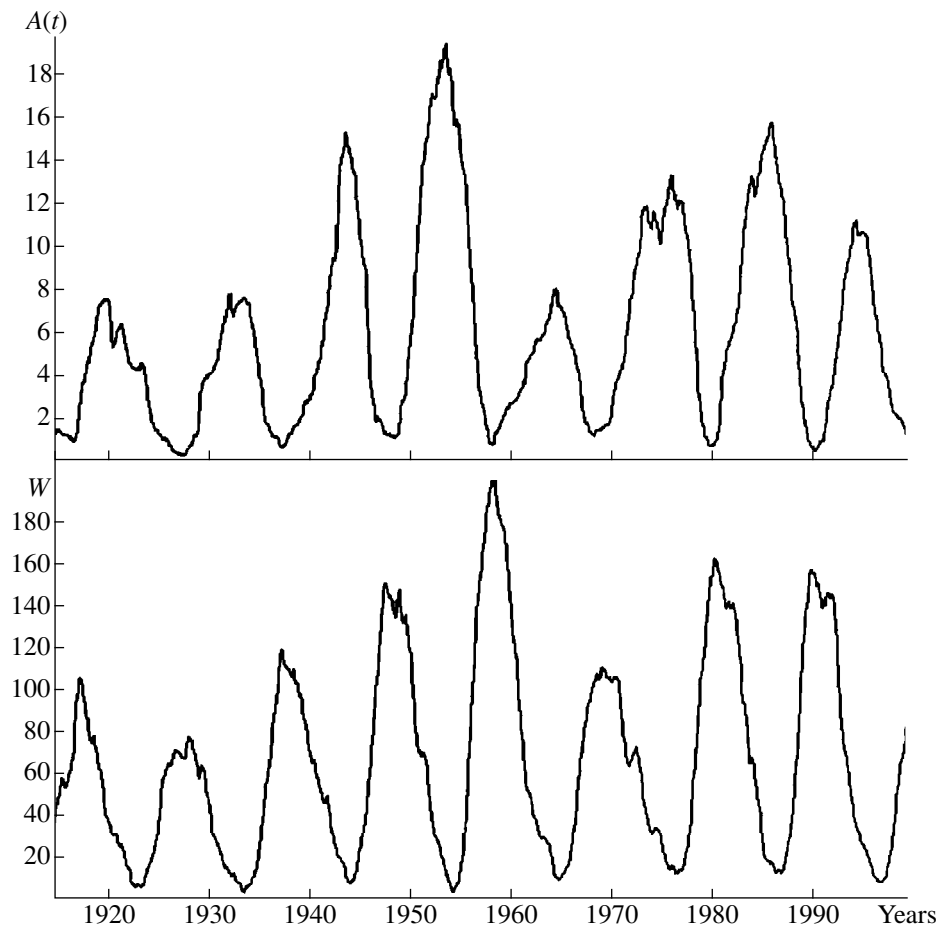
Fig. 2. Top: 11-year cycle of the large-scale solar magnetic field represented by the parameter $A(t)$. $A(t)$ is calculated from Hα synoptic magnetic maps, for 1915–1999, where the magnitude and polarity of the field are taken to be either +1 or –1 G. Bottom: 11-year solar-activity cycle represented by the Wolf number, for 1915–1999. The activity cycles of the large-scale solar magnetic field lead the corresponding cycles of sunspot activity by 5.5 years, and are in antiphase with the latter.
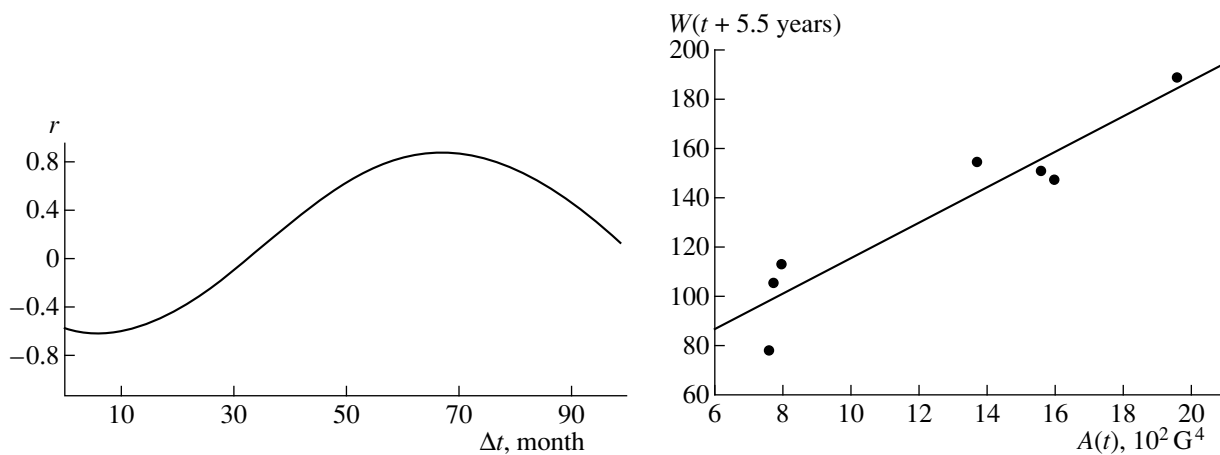
Fig. 3. Top: mutual-correlation coefficient as a function of the shift Δt between $A(t)$ and $W(t)$ for 1915–1999. The mutual-correlation coefficient is maximum for a shift of 65 months (i.e., 5.5 years), with $A(t)$ leading $W(t)$. Bottom: relationship between the maximum magnitude $A(t)$ of the large-scale magnetic field and the maximum Wolf number in a cycle, taken 5.5 years later. For cycle 23, $A(t)$ is 11.4; equivalently, $W(23) = 130 \pm 10$.
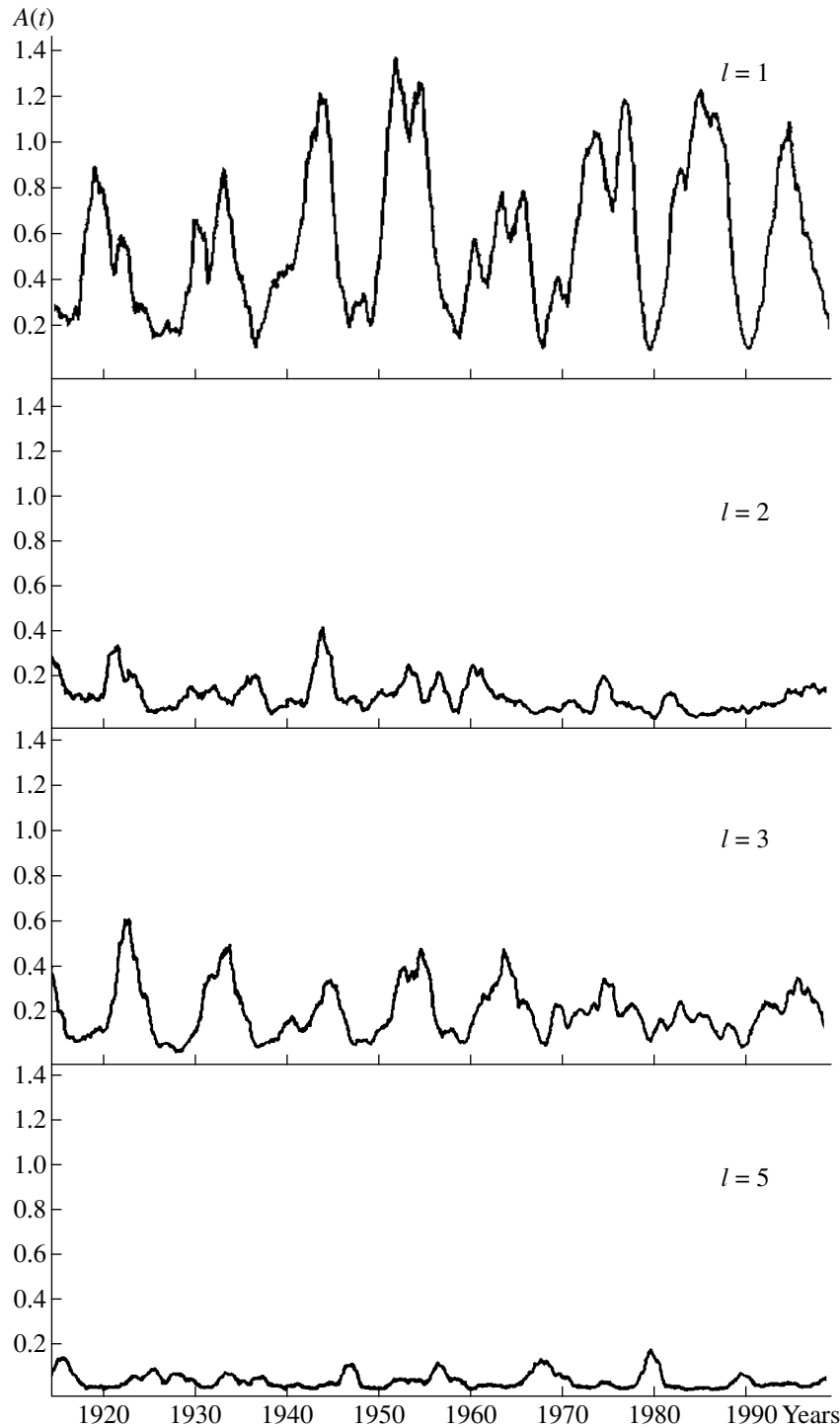
Fig. 4. Relative intensity of the harmonics $A^*(t)$ calculated from the Hα maps according to the formula $A^*(t) = \sum_{m,l}(g_l^m g_l^m + h_l^m h_l^m)$, for $l = 1, 2, 3, 5$.

whereas $A(22)$ was equal to 15.9 for cycle 22 (Fig. 2). In view of this, the maximum value of $W(23)$ should reach about 130.

The solar magnetic cycle can be schematically represented as periodic waves of propagation of a new magnetic field excited at high latitudes [11]. The new magnetic field is probably excited close to the lower boundary of the convection zone. The wave of field generation travels with the meridional flow from the poles to the equator over 15–18 years. This wave is

traced by a slow wave of torsional oscillations. The reason for the observed reversal of the polar magnetic field of the Sun remains unclear. In the context of our data, we can consider the transfer of the magnetic field from the surface to the base of the convection zone, which accompanies the sinking of material near the poles [11]. Sources of the large-scale magnetic field emerge at all latitudes, from the poles to the equator. As this takes place, sources belonging to two successive cycles can be superposed and form a zonal polarity distribution, with a neutral line at middle latitudes, between the sources of opposite polarities, which can persist for several years [13]. The sunspot cycle is the final dynamical stage of the magnetic cycle.

ACKNOWLEDGMENTS

REFERENCES

1. V. I. Makarov and K. R. Sivaraman, Sol. Phys. **119**, 35 (1989).
2. V. I. Makarov and V. V. Makarova, Sol. Phys. **163**, 267 (1996).
3. V. I. Makarov, Sol. Phys. **150**, 359 (1994).
4. V. N. Obridko and B. D. Shelting, Sol. Phys. **184**, 187 (1999).
5. V. I. Makarov, A. A. Ruzmaikin, and S. V. Starchenko, Sol. Phys. **111**, 267 (1987).
6. G. Belvedere, K. Kuzanyan, L. Paterno, and D. Sokoloff, in *Proceedings of the 9th European Meeting on Solar Physics, Magnetic Fields and Solar Processes,* Ed. by A. Wilson, 1999, p. 59.
7. V. I. Makarov and K. R. Sivaraman, Kodaikanal Obs. Bull. **7**, 2 (1986).
8. V. I. Makarov, Doctoral Dissertation, 1989.
9. V. I. Makarov and K. R. Sivaraman, in *The Solar Photosphere: Structure, Convection, and Magnetic Fields,* Ed. by J. O. Stenflo (1990), p. 281.
10. V. I. Makarov, A. G. Tlatov, and D. K. Callebaut, Sol. Phys. **170**, 373 (1997).
11. V. I. Makarov and A. G. Tlatov, Astron. Zh. **72**, 749 (1995) [Astron. Rep. **39**, 668 (1995)].
12. A. G. Tlatov, Astron. Zh. **74**, 621 (1997) [Astron. Rep. **41**, 548 (1997)].
13. V. I. Makarov and A. G. Tlatov, Izv. Akad. Nauk, Ser. Fiz. **62** (9), 1853 (1998).
14. N. G. Makarenko, L. M. Karimova, V. I. Makarov, and K. S. Tavastsherna, in *Proceedings of the Conference "Topical Problems of Solar Activity," Pulkovo* [in Russian], 1998, p. 139.
15. H. W. Babcock, Astrophys. J. **133**, 572 (1961).
16. R. B. Leighton, Astrophys. J. **140**, 1547 (1964).
17. C. R. DeVore, N. R. Sheeley, Jr., and J. P. Boris, Sol. Phys. **92**, 1 (1984).
18. Y.-M. Wang, A. G. Nash, and N. R. Sheeley, Jr., Science **245**, 681 (1989).
19. B. R. Durney, Sol. Phys. **169**, 1 (1996).
20. A. R. Choundhuri, M. Schlussler, and M. Dipkati, Astron. Astrophys. **303**, L29 (1995).
21. B. J. LaBonte and R. Howard, Sol. Phys. **75**, 161 (1982).
22. T. Homann, F. Kneer, and V. I. Makarov, Sol. Phys. **175**, 81 (1997).
23. H. S. Jones, *Sunspot and Geomagnetic-Storm Data* (London, 1995).
24. D. V. Hoyt and K. H. Schatten, Sol. Phys. **157**, 340 (1998).

Translated by A. Getling

Variations of the Solar-Wind Stream Structure in the Region of Subsonic Flow during the 11-Year Solar Cycle

N. A. Lotova[1], V. N. Obridko[1], and K. V. Vladimirskiĭ[2]

[1]Institute of Terrestrial Magnetism, Ionosphere, and Radio Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142092 Russia
[2]Lebedev Institute of Physics, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117942 Russia
Received November 19, 1999

Abstract—The large-scale stream structure of the solar wind near the Sun and its evolution during the 11-year solar activity cycle are investigated. The study is based on observations of scattering of the radiation from compact natural radio sources at radial distances $R \leq 14R_s$ (R_s is the solar radius). Regular observations were conducted in 1981–1998 on the RT-22 and DKR-1000 radio telescopes of the Russian Academy of Sciences at Pushchino, at $\lambda = 1.35$ cm and 2.7 m, respectively. The radial dependences of the interplanetary scintillations $m(R)$ and the scattering angle $2\theta(R)$ are considered together with the structure of large-scale magnetic fields in the solar corona at $R = 2.5R_s$. The entire range of variations in the level of scattering and the associated helio-latitude flow structures in the subsonic solar wind forms over the 11-year solar cycle, as a direct result of the large-scale structure of the evolving magnetic fields at the source of the solar-wind streamlines. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In this paper, we study variations of the spatially nonuniform stream structure of the solar wind during the 11-year solar activity cycle near the Sun, in the region of subsonic flow. Basic data on the structure of the solar wind flows and mechanisms for their generation can be obtained in experiments on the transmission of radio waves emitted by natural or artificial sources through the near-solar plasma [1–9]. Such experiments established the existence of considerable nonuniformity in the spatial stream structure, which is always present, irrespective of the solar activity phase [8, 10–12]. It is also known that the stream structure of the flow is a key factor in solar-wind acceleration [13–15].

We undertook a detailed investigation of the large-scale stream structure of the solar wind at radial distances $R \leq 14R_s$ and its correlation with the large-scale magnetic field structure of the solar corona at $R = 2.5R_s$. We used radio astronomical data on the scattering of radiation from compact natural radio sources, data on the structure of coronal magnetic fields derived from photospheric optical observations [16–19], and optical observations of the white corona.

2. EVOLUTION OF THE RADIAL DEPENDENCE OF RADIO SCATTERING OVER THE 11-YEAR SOLAR CYCLE

This study is based on yearly observations of scintillations of the water maser source IRC-20431 conducted in 1981–1988. In 1984 to 1987, these observations were carried out as part of a coordinated program at the Institute of Terrestrial Magnetism, Ionosphere, and Radio Propagation (Troitsk, Russia) and the Radio Astronomy Institute of Bonn University (Bonn, Germany) [20]. In 1985–1998, we also obtained scattering observations of the radio source 3C 144. The maser source IRC-20431 probes the northern and 3C 144 the southern hemisphere of the heliosphere. The experiments with IRC-20431 at $\lambda = 1.35$ cm enabled us to obtain 13 curves for the radial dependence of the level of the scintillations $m(R)$. Experiments with 3C 144 yielded at $\lambda = 2.7$ m 13 radial curves for the scattering angle. Significant variations of $m(R)$ and $2\theta(R)$ were detected during these experiments, at small distances from the Sun ($R \leq 14R_s$). These radial distances were identified earlier as a zone of subsonic solar-wind flow [21].

Characteristic examples of the radial dependence of the scintillation level $m(R)$ are given in Fig. 1. The scattering angle as a function of radial distance, $2\theta(R)$, shows a similar behavior. The data show that the variety of radial dependences for the radio scattering can be divided into two basic types. The first (Fig. 1a), typical of the solar maximum phase, is characterized by a power-law dependence at the considered radial distances:

$$m(R) \sim 2\theta(R) \sim R^{-1.6}. \tag{1}$$

The second type (Fig. 1b) differs in its gently sloping radial dependence and reduced scattering level at small distances from the Sun.

When discussing the results of radio astronomical observations at different phases of the solar cycle, we must take into account the particular nature of observations of natural sources passing near the Sun. When the
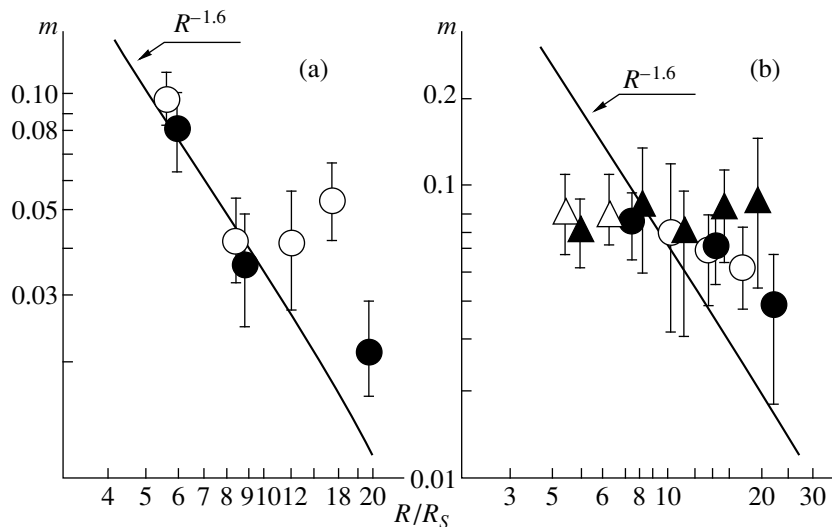
Fig. 1. Radial dependence of the scintillation level $m(R)$ from observations of IRC-20 431 in (a) December 1992 and (b) December 1986. Data from observations made at Pushchino (Russia) and Effelsberg (Germany) are indicated by circles and triangles, respectively. Open circles correspond to approach of the source toward the Sun (eastern hemisphere of the heliosphere), while filled circles correspond to recession from the Sun (western hemisphere).

radio telescope beam is at large distances from the Sun, day-to-day variations in the heliolatitude are negligible, and it is primarily the radial distance from the Sun that changes. However, as the source approaches the Sun, substantial variations in the heliolatitude arise. The corresponding variations in the level of scattering are small only near solar activity maximum, when the circumsolar plasma flow is closest to being spherically symmetric. This explains the domination of the first type of radial dependence at the phase of maximum. In years of decreased solar activity, the solar corona is quite anisotropic, and the approach of the telescope beam to the polar region of the heliosphere, with its high-velocity plasma flows [5–7, 10], leads to significant variations in the radial dependence of the scattering. In this case, the main factor determining the scattering characteristics are the heliolatitude variations that accompany the decrease of the radial distance.

3. LARGE-SCALE CORONAL MAGNETIC FIELDS

The characteristic variations of radio scattering with heliolatitude are not the only subject of studies of the subsonic solar-wind flow. The circumsolar plasma flow is generally quite nonuniform, beginning from the smallest distances from the Sun. In connection with this, we also considered data on the magnetic field structure in the solar corona in our study of the region of subsonic solar-wind flow. For this, we used optical measurements of the photospheric magnetic field available via the Internet, obtained at the John Wilcox Observatory (Stanford, Ca, USA). We then calculated the magnetic fields between the solar surface and a sphere of radius $R = 2.5R_s$ using previously developed

methods [16–19]. These calculations were performed in a potential approximation, ignoring the effect of the circumsolar medium in this region. The initial conditions for the magnetic-field line computations were specified on a uniform grid of points at the surface of a sphere with $R = R_s$ (Fig. 2). Calculations were made for the times and spatial coordinates for the scattering observations. The results enabled us to directly compare structural features of the circumsolar magnetic fields with the radio scattering observations for the same region of circumsolar space. Figure 2 presents examples of these calculations that are consistent with the radio data presented in Fig. 1. It follows that radial dependences of type (1), without kinks, correspond to closed magnetic field structures, which do not contain field lines extending out into space. The reduced scattering at small distances (Fig. 1b) corresponds to open magnetic structures and powerful polar coronal holes with magnetic field structures that are close to dipolar (Fig. 2b).

4. DISCUSSION

The data obtained in 25 series of radio scattering observations are collected in Tables 1 and 2. The results for phases when the source approached the Sun and receded from it are, in many cases, presented separately, providing a total of 40 essentially independent experiments. As a parameter defining the form of the radial dependence for the scattering, we give in the tables the coordinates of the "deviation point"—the first point in an experimental radial scattering curve [$m(R)$ or $2\theta(R)$] that departs from a power-law dependence of the form (1) at small distances from the Sun. Columns 2–6 in the tables contain the characteristics of
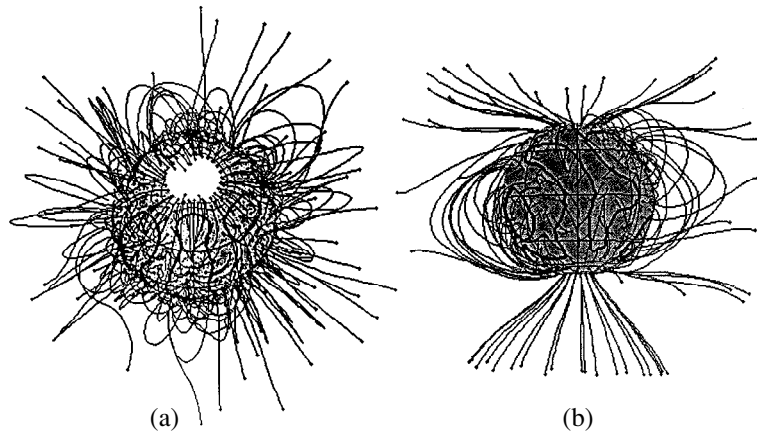
Fig. 2. Calculated magnetic-field structure in the solar corona on (a) December 21, 1992 (E, $\varphi_0 = 58°$, high-latitude streamer; W, $\varphi_0 = 41°$, powerful streamer) and (b) December 23, 1986 (E, $\varphi_0 = 46°$, transition from streamer to polar coronal hole; W, $\varphi_0 = 64°$, polar coronal hole).

this deviation point: the date of detection; its coordinates (eastern or western hemisphere, radial distance R_0, and heliolatitude φ_0), and the type of radial dependence (a power-law of type (1) or a deviant curve). Further, we give characteristics of the circumsolar plasma corresponding within the probable accuracies to the same spatial and temporal coordinates: the magnetic-field structure and morphology of the white corona. In cases when the radial scattering dependence exhibits no kinks [type (1)], the tables contain the coordinates of the point in the radial scattering curve that is nearest to the Sun. Comparisons with the magnetic-field and white-corona data are also given for these points, whose coordinates in the tables are marked by asterisks.

The data in Tables 1 and 2 show that radial scattering dependences that do not deviate from a power law of type (1) are due, in all cases, to closed magnetic field structures at the source of the streamlines; streamers are observed here in the white corona. The second type of radial dependence, with decreased scattering at small distances from the Sun, is associated with open coronal magnetic-field structures. The corresponding structures in the white corona are very diverse; polar and local coronal holes dominate.

The coordinates of the deviation points in Tables 1 and 2 have obvious physical meaning as points of a boundary separating regions with different levels of turbulence of the subsonic solar-wind flows in heliolatitude. To investigate the long-term variations of these effects, we present in Fig. 3a the heliolatitudes of the deviation points as a function of date for measurements in the northern hemisphere of the heliosphere (IRC-20431). Figure 3b presents the same data for the southern hemisphere (3C 144). The open circles correspond to approach of the source toward the Sun, and filled circles to recession from the Sun. When a deviation point was recorded at low heliolatitude, we have also plotted the position of the point in the radial scattering curve following the deviation point. In many experiments in

years of solar activity maximum, we did not observe deviations from a power-law radial dependence; in these cases, we have plotted the points of the radial curve nearest to the Sun as a lower limit for the deviation-point heliolatitude.
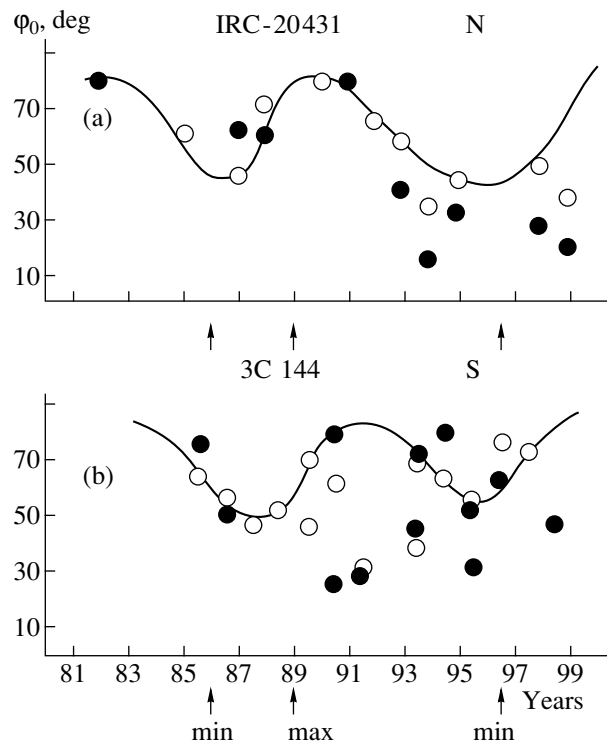

Fig. 3. Heliolatitude of deviation points φ_0 as a function of date, determined from measurements of (a) the scintillation level $m(R)$ for IRC-20431 in the northern hemisphere of the heliosphere and (b) the scattering angle $2\theta(R)$ for 3C 144 in the southern hemisphere of the heliosphere. Open circles correspond to approach of the source toward the Sun, and filled circles to recession from the Sun.

Table 1. Radial dependence of $m(R)$ near the Sun from observations of IRC-20431

N	Date	E/W	R_0/R_s	φ_0, deg	$m(R)$ type	Field structure	White corona morphology	Solar cycle phase
	2	3	4	5	6	7	8	9
1	23.12.1981	W	4.5*	80	Type (1)	Closed	High-latitude streamer	After maximum
2	22.12.1984	E	5.3	61	Deviation from (1)	Open	Polar coronal hole	Inclined
3	22.12.1986	E	6.5	46	Deviation from (1)	Open	Transition from streamer to polar coronal hole	Minimum
4	24.12.1986	W	5.0	64	Deviation from (1)	Open	Polar coronal hole	Minimum
5	23.12.1987	E	4.8	72	Deviation from (1)	Open	Polar coronal hole	After minimum
6	24.12.1987	W	4.7	62	Deviation from (1)	Open	Transition from streamer to polar coronal hole	After minimum
7	23.12.1989	E	4.5*	80	Type (1)	Closed	High-latitude streamer	Maximum
8	23.12.1990	W	5.0*	80	Type (1)	Closed	Transition from streamer to polar coronal hole	After maximum
9	23.12.1991	E	5.3	65	Deviation from (1)	Open	Transition from streamer to polar coronal hole	Inclined
10	21.12.1992	E	5.7*	58	Type (1)	Closed	Streamer	Inclined
11	23.12.1992	W	6.0*	41	Type (1)	Closed	Streamer	Inclined
12	21.12.1993	E	9.6	35	Deviation from (1)	Open	Local coronal hole	Inclined
13	26.12.1993	W	11.5	16	Deviation from (1)	Open	Transition from streamer to local coronal hole	Inclined
14	22.12.1994	E	7.2	45	Deviation from (1)	Open	Transition from streamer to polar coronal hole	Inclined
15	25.12.1994	W	7.3	31	Deviation from (1)	Open	Local coronal hole	Inclined
16	22.12.1997	E	6.5	50	Deviation from (1)	Open	Polar coronal hole	Inclined
17	25.12.1997	W	8.0	27	Deviation from (1)	Open	Region between two streamers	Inclined
18	24.12.1998	E	7.4	39	Deviation from (1)	Open	Local coronal holes between two streamers	Inclined
19	25.12.1998	W	14.0	20	Deviation from (1)	Open	Local coronal hole	Inclined

Note: See text for explanation of the asterisks in column 4.

The field of points in Fig. 3 shows that open magnetic-field structures—and, accordingly, decreased levels of plasma turbulence and scattering at small distances from the Sun—can arise over a broad range of heliolatitudes, depending on the structure of the magnetic fields and plasma flows at the source of the solar-wind streamlines. Nevertheless, some regularity in the phase dependence of the solar activity is evident. At the activity maximum, when loop magnetic-field structures dominate, deviations from a dependence of the form (1) are rare. Near solar-activity maximum, when the sizes of coronal holes are minimum, large values of φ_0 are observed. As the activity minimum approaches, the sizes of polar coronal holes, with their characteristic quasi-dipolar structure, increase. Large-scale formations with loop magnetic-field structures move toward lower heliolatitudes. Accordingly, decreased φ_0 values are observed.

In addition to the quasi-periodicity of heliospheric processes, Fig. 3 also reflects the asymmetry of processes on the Sun. Comparison of the positions of the filled and open circles reveals a significant east–west asymmetry of the solar-wind flow. Comparison of Figs. 3a and 3b shows that large-scale variations in the state of

Table 2. Variations in the scattering angle $2\theta(R)$ near the Sun from observations of 3C 144

N	Date	E/W	R_0/R_s	φ_0, deg	$2\theta(R)$ type	Field structure	White corona morphology	Solar cycle phase
	2	3	4	5	6	7	8	9
1	14.06.1985	E	5.4*	−64	Type (1)	Closed	High-latitude streamer	Before minimum
2	15.06.1985	W	5.0*	−76	Type (1)	Closed	High-latitude streamer	Before minimum
3	14.06.1986	E	5.9*	−56	Type (1)	Closed	Boundary of high-latitude streamer; polar coronal holes	Minimum
4	16.06.1986	W	6.0*	−51	Type (1)	Closed	Boundary of high-latitude streamer; polar coronal holes	Minimum
5	13.06.1987	E	9.0	−32	Deviation from (1)	Open	Transition from streamer to local coronal hole	After minimum
6	13.06.1988	E	7.0	−51	Deviation from (1)	Open	Local coronal holes, streamer	Before maximum
7	13.06.1989	E	7.7	−46	Deviation from (1)	Open	Local coronal holes, streamer	Maximum
8	14.06.1990	E	6.0	−62	Deviation from (1)	Open	Local coronal holes, streamer	After maximum
9	17.06.1990	W	8.8	−26	Deviation from (1)	Open	Local coronal holes	Inclined
10	12.06.1991	E	12.4	−30	Deviation from (1)	Open	Local coronal holes	Inclined
11	17.06.1991	W	8.0	−30	Deviation from (1)	Open	Local coronal holes	Inclined
12	13.06.1993	E	7.8	−45	Deviation from (1)	Open	Large coronal mass ejection	Inclined
13	16.06.1993	W	6.6	−40	Deviation from (1)	Open	Large coronal mass ejection	Inclined
14	14.06.1994	E	6.0*	−61	Type (1)	Closed	High-latitude streamer	Near minimum
15	15.06.1994	W	4.8*	−81	type (1)	Closed	High-latitude streamer	Near minimum
16	14.06.1995	E	6.5	−54	Deviation from (1)	Open	Polar coronal hole	Near minimum
17	17.06.1995	W	7.9	−31	Deviation from (1)	Open	Local coronal holes	Near minimum
18	14.06.1996	E	5.1*	−78	Type (1)	Closed	High-latitude streamer	Before minimum
19	15.06.1996	W	5.2*	−61	Type (1)	Closed	High-latitude streamer	Before minimum
20	15.06.1997	E	5.0	−73	Deviation from (1)	Open	Local coronal holes	Minimum
21	16.06.1998	W	6.0	−47	Deviation from (1)	Open	Large local coronal hole	After minimum

Note: See text for explanation of the asterisks in column 4.

the heliosphere occur at different times in the northern and southern hemispheres, in accordance with our current understanding of the evolution of large-scale solar magnetic fields.

5. CONCLUSION

A detailed comparison of the observed variations of the deviation points in heliolatitude with the corresponding predicted structures of the coronal magnetic field suggest that the quasi-periodic variations in the position of the deviation points at high heliolatitudes, $40° \leq |\varphi_0| \leq 80°$, are due to evolution of the sizes of polar coronal holes during the 11-year solar activity cycle. The appearance of deviation points at low heliolatitudes ($15° \leq |\varphi_0| \leq 35°$) is associated with local open coronal magnetic fields: local coronal holes, side lobes of streamers, etc. Deviation points at high heliolatitudes are related to the appearance of streamers at high heliolatitudes. By and large, the entire variety of radial variations of the level of radio scattering and related heliolatitude structures of the subsonic solar-wind flow forms directly as a result of the evolution of large-scale coronal magnetic fields at the source of the solar-wind streamlines.

ACKNOWLEDGMENTS

REFERENCES

1. M. K. Bird, H. Volland, M. Patzold, *et al.*, Astrophys. J. **426**, 373 (1994).

2. A. I. Efimov, Space Sci. Rev. **70**, 397 (1994).

3. N. A. Lotova, in *Proceedings of the 3rd COSPAR Colloquium, Solar Wind 7*, Ed. by E. Marsch and R. Schwenn (Pergamon, New York, 1992), p. 217.

4. P. K. Manoharan, Sol. Phys. **148**, 153 (1993).

5. N. R. Sheeley, Y.-M. Wang, Jr., S. H. Hawley, *et al.*, Astrophys. J. **484**, 472 (1997).

6. R. R. Grall, W. A. Coles, M. T. Klinglesmith, *et al.*, Nature **379**, 429 (1996).

7. M. Patzold, J. Karl, and M. K. Bird, Astron. Astrophys. **316**, 449 (1996).

8. N. A. Lotova, Sol. Phys. **117**, 399 (1988).

9. S. Ananthakrishnan, V. Balasubramanian, and P. Janardhan, Space Sci. Rev. **72**, 229 (1995).

10. P. Janardhan, M. K. Bird, P. Edenhofer, *et al.*, Sol. Phys. **184**, 157 (1999).

11. M. Kojima, M. Tokumaru, H. Watanabe, and A. Yokobe, J. Geophys. Res. **103**, 1980 (1998).

12. N. A. Lotova, K. V. Vladimirskiĭ, and O. A. Korelov, Sol. Phys. **172**, 225 (1997).

13. K. V. Vladimirskiĭ, N. A. Lotova, and O. A. Korelov, Kratk. Soobshch. Fiz., Nos. 1–2, 12 (1996).

14. F. Malara, Astrophys. J. **396**, 297 (1992).

15. F. Malara, L. Primavera, and P. Veltri, in *Proceedings of the COSPAR Colloquium, Solar Wind 8, Dana Point, USA, 1995,* Abstract 0-1-04.

16. J. T. Hoeksema, J. M. Wilcox, and P. H. Scherrer, J. Geophys. Res. **87**, 10331 (1982).

17. J. T. Hoeksema, J. M. Wilcox, and P. H. Scherrer, J. Geophys. Res. **88**, 9910 (1983).

18. V. N. Obridko, A. F. Kharshiladze, and B. D. Shel'ting, in *Solar Magnetic Field and Helioseismology*, Ed. by V. A. Dergachev (Fizikotekhn. Inst., St. Petersburg, 1994), p. 71.

19. V. N. Obridko and B. D. Shelting, Sol. Phys. **184**, 187 (1999).

20. N. A. Lotova, I. Yu. Yurovskaya, Ya. V. Pisarenko, *et al.*, Astron. Zh. **69**, 173 (1992) [Sov. Astron. **36**, 88 (1992)].

21. N. A. Lotova, K. V. Vladimirskiĭ, I. Yu. Yurovskaya, and O. A. Korelov, Astron. Zh. **72**, 757 (1995) [Astron. Rep. **39**, 675 (1995)].

Translated by A. Kozlenkov

The Solar Dynamo and Integrated Irradiance Variations in the Course of the 11-Year Cycle

V. V. Pipin and L. L. Kichatinov

Institute of Solar–Terrestrial Physics, Siberian Division, Russian Academy of Sciences, Irkutsk, 664033 Russia

Received November 18, 1999

Abstract—The effect of the large-scale magnetic fields generated by the solar dynamo on the radiation flux issuing from the convection zone is studied. A governing equation describing convective heat transfer is obtained in the framework of mean-field magnetohydrodynamics, with account for the influence of magnetic fields and differential rotation on the energy budget of the convection zone. The principal effects are illustrated using a one-dimensional numerical model. Calculations indicate that the influence of large-scale magnetic fields can modulate the solar irradiance with a relative amplitude of ~0.07%. © 2000 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

Measurements of the integrated solar irradiance indicate that the brightness of the Sun varies over the 11-year activity cycle. In their analysis of satellite observations, Wilson [1] and Wilson and Hadson [2] found that the radiation flux varies in phase with the magnetic cycle, with an amplitude of about 0.1%. This correlation between the irradiance and the magnetic cycle suggests that the influence of magnetic fields on the heat flux in the convection zone and at the photospheric level is probably the main source of long-term luminosity variations. Modeling of the photospheric effects of magnetic fields on the luminosity is complicated by many factors. It is now generally thought that excess emission dominates on the scale of the cycle [2, 3]. Foukal and Lean [3] and Lean *et al.* [4] assert that an important contribution to the variability of the Sun is given by two components—the deficient emission of sunspots and the enhanced emission of faculae. The redistribution of the energy blocked by spots and pores on short time intervals and emitted on longer time intervals by decaying active regions, faculae, and network fields is frequently discussed [5, 6].

Note that all solar radiation—both the emission blocked in spots and the excess emission of faculae and network elements—comes from the convection zone. For this reason, it is presumably the magnetic fields generated by the solar dynamo in the convection zone that modulate the radiation flux. Discussions of various mechanisms for modulation of the solar irradiance by the magnetic fields in the convection zone and at the photospheric surface can be found in the review by Spruit [5] and in Parker's paper [6]. The estimates obtained by Spruit [5, 7] and Balmforth *et al.* [8] for the contribution of *local* subphotospheric fields to luminosity variations show that the effect of local fields situated at depths greater than 10^4 km does not exceed 0.01% of the solar luminosity. The modulation of luminosity by large-scale, dynamo-generated magnetic fields has not yet been examined.

We consider here the influence of the large-scale magnetic fields generated by the solar dynamo on the radiation flux from the convection zone. We will study the effect of the *averaged* magnetic fields on the *mean* heat flux carried by convection. From the point of view of mean fields, the overall energy-flux balance in the convection zone can be represented as follows. The heat flux coming from the radiative zone is converted into a convective flux. A small fraction of the convective energy together with rotation energy are converted into differential-rotation energy (e.g., via the Λ effect [9]); other convective energy plus some differential-rotation energy is converted into energy of large-scale fields (e.g., via the $\alpha\Omega$ dynamo mechanism). The maintenance and dissipation of differential rotation, as well as the generation and dissipation of large-scale fields in the dynamo cycle, provide additional energy sinks and sources for the convection zone; at the surface, all energy is emitted into space. In addition to these processes, we must take into account the decrease of the efficiency of convective heat transfer in the presence of a magnetic field.

All these effects can act as sources of luminosity variations in the course of the solar cycle. Let us consider them in greater detail. We will first write a basic equation describing the variation of the mean thermal energy in the convection zone and estimate the amplitudes of the effects. Further, we describe a one-dimensional numerical model for variations in the energy flux from a convective layer where a dynamo operates, present the calculation results, and discuss them.

2. GOVERNING EQUATIONS AND ESTIMATES FOR THE AMPLITUDES OF VARIOUS EFFECTS

The heat-transfer equation for a rotating turbulent medium can be obtained using the corresponding conservation laws of magnetohydrodynamics. In our case, these laws must be adapted for a turbulent medium. As is usually done in the theory of turbulence, all physical quantities are decomposed into their mean and random components; averaging over an ensemble of fluctuations then yields the mean-field equations. We defer the derivation of the heat-transfer equation to the Appendix. The result has the form

$$\bar{\rho}\bar{T}\partial_t\bar{s} = -\bar{\rho}\bar{T}V_i\partial_i\bar{s} - \partial_i(F_i^{con} + F_i^r)$$

$$+ \bar{\rho}(-Q_{ij}^\Lambda + N_{ijkl}\partial_k V_l)\partial_j V_i \quad (1)$$

$$-\frac{1}{4\pi}(\alpha_{ij}B_j(\nabla \times B)_i - \eta_T(\nabla \times B)^2),$$

where $\bar{\rho}$ $\bar{T}$ and $\bar{s}$ are the mean density, temperature, and entropy, respectively, and **V** and **B** are the large-scale components of the velocity field and magnetic field, separated into poloidal and toroidal components. $\mathbf{F}^{con}$ is the convective heat flux and $\mathbf{F}^r$ the radiative heat flux (including molecular processes). The tensors Q_{ij}^Λ and α_{ij} describe convective sources providing the generation of differential rotation and magnetic field (the so-called Λ and α effects), and the tensor N_{ijkl} and the scalar η_T represent the eddy viscosities of large-scale flows and magnetic fields. The expressions for radiative heat transfer and the turbulent effects of the magnetic field and differential rotation will be specified below when we describe the specific model to be considered. Equation (1) takes into account the fact that the large-scale fields are axisymmetric, and that the eddy viscosities are much greater than the molecular viscosities. Note also that the production rate for the thermal energy per unit volume written above has the meaning of the power output of the turbulent flows and magnetic fields in a given volume. The physical interpretation of (1) is as follows. Variation of the mean thermal energy in some elementary volume of the convection zone is the result of heat transfer due to meridional circulation, convection, heat conduction, and radiation; energy consumption due to the generation of differential rotation and magnetic fields; and the release of heat due to the viscous dissipation of flows and Joule dissipation of magnetic fields. The large-scale fields in (1) are governed by the dynamo and differential-rotation equations, which will be specified below in the context of our one-dimensional model. If we neglect turbulent effects, (1) becomes the classic equation for the production of entropy, which can be found, e.g., in the textbook of Landau and Lifshitz [10]. We will assume that the large-scale velocity field and magnetic field in (1) have their origins in turbulence. If we neglect turbulent effects, (1) should reduce to the classical equation for heat conduction. In this transition, we must set not only the intensity of the turbulence, but also all terms to which it gives rise—differential rotation, meridional circulation, the α effect, and the large-scale field—equal to zero.

Let us assume, as usual, that the integrated radiation flux from a star can be approximated as pure black-body emission: $F_r = \sigma T_{eff}^4$, where T_{eff} is the surface temperature. In this case, a variation δF in the radiation flux is related to the temperature variation δT by the expression $\delta F/F_r = 4\delta T/T_{eff}$. We will consider radiation-flux variations over times of the order of the cycle duration; i.e., over intervals much longer than the time to establish hydrostatic equilibrium. In this case,

$$\delta F = 4F_r\frac{T_e}{T_{eff}}\frac{\delta\bar{s}}{c_p}, \quad (2)$$

where T_e is the temperature at a certain level above the surface. We will assume that the dynamo is confined to the region of hypothetical giant-cell convection; i.e., it acts below the base of the supergranular cells. We also assume that the transport of energy via supergranulation and granulation renders heat transfer so efficient that the radiation flux at the boundary of the dynamo layer can be matched to the flux of black-body radiation at the photospheric surface [11]. In this case, in view of (2), the boundary condition for the radiation flux issuing from the dynamo region can be written

$$F_r\big|_{r=r_e} = \frac{L_\odot}{4\pi r_e^2}\left(1 + 4\frac{T_e}{T_{eff}}\frac{\delta\bar{s}}{c_p}\right), \quad (3)$$

where r_e is the outer boundary of the dynamo region and T_e and $\delta\bar{s}/c_p$ are the temperature and thermal perturbation at the boundary of the dynamo region, respectively. Thus, the amplitude of variations in the energy flux at the surface is controlled by the amplitude of entropy variations, which can be estimated using (1).

Order-of-magnitude energy estimates for thermal variations due to the influence of magnetic fields and differential rotation on convection were previously obtained by Spruit [5] and Parker [6]. They considered the effect of a large-scale magnetic field localized in tubes under the convection zone on the surface radiation flux. Because of the very high efficiency of heat transfer deep in the convection zone, local thermal non-uniformities arising under the influence of a magnetic tube are dispersed on diffusion scales comparable to the tube diameter (see, e.g., [8]); for this reason, their contribution to variations in the radiation flux at the photospheric surface is extremely small. If the large-scale field is distributed over the entire convection zone, the estimates below suggest that the contribution of thermal perturbations in the convection zone to the surface flux may be significant.

Let us consider the contributions expected from each term in (1) separately. The $\bar{\rho}\bar{T}V_i\partial_i\bar{s}$ term represents the transfer of heat by meridional circulation. Let

us assume that, in some steady state of the system without magnetic fields but with meridional-circulation and rotation effects, some distribution of entropy is established. The effect of magnetic activity is to weakly modulate this distribution over the course of the cycle. In this case, entropy variations due to variations in the meridional circulation will be higher order effects compared with the direct influence of the magnetic field on heat transfer. Therefore, meridional circulation can be neglected at this stage.

Changes in the efficiency of convection due to the presence of a magnetic field are known to be among the origins of convective-flux modulation by the magnetic field. This effect was calculated for mean magnetic fields by Kitchatinov et al.[12]. Using the known relationships of mixing-length theory, we can write

$$F_i^{\mathrm{con}} = -\bar\rho \bar T \chi_{ij}(B)\partial_j \bar s, \tag{4}$$

where

$$\chi_{ij}(B) = \langle u_i' u_j'\rangle = \langle u'^2\rangle\left(\frac{1}{3}\Psi(\beta)\delta_{ij} + \frac{B_i B_j}{B^2}\Psi_1(\beta)\right),$$

$$\langle u'^2\rangle = -\frac{l_{\mathrm{cor}}^2 g}{4c_p}\frac{\partial \bar s}{\partial r} \tag{5}$$

are the tensor of effective heat conductivity and the turbulence intensity determined for the actual sources of turbulence (in the case of the Sun, the superadiabatic gradient $\partial \bar s/\partial r$) in the absence of magnetic fields and rotation. g is the free-fall acceleration, l_{cor} the turbulence correlation length, and β the magnetic-field strength normalized to the kinetic-energy density for the convective motion. The functions Ψ and Ψ_1 take into account the suppression of turbulence by the magnetic field.

The characteristic lifetime of a thermal fluctuation due to the effects of large-scale fields is the dynamo-cycle time $\delta t \sim 3.5 \times 10^8$ s. Let us assume that, in the middle of the convection zone, β varies from 0–1 over the course of the cycle, so that $|\chi_{ij}|$ varies by a factor of two. The superadiabatic gradient normalized to the pressure scale height, $\Delta\nabla = \dfrac{1}{c_p}\dfrac{\partial \bar s}{\partial lg\bar p}$, varies from 10^{-7}–10^{-3}. We adopt the value $\Delta\nabla \sim 10^{-6}$, typical of the middle region of the convection zone [13]. Let $\nu_T \sim 10^{13}$ cm^2/s; then, the amplitude of the thermal variations can be estimated as

$$\left.\frac{\delta \bar s'}{c_p}\right|_{\mathrm{con}} \sim \nu_T\delta t_{\mathrm{cyc}}\frac{\partial}{\partial r}\frac{1}{c_p}\frac{\partial \bar s}{\partial r} \sim \frac{\nu_T\delta t_{\mathrm{cyc}}}{R_\odot H_p}\Delta\nabla \sim 2.5\times10^{-5},$$

where $H_p \sim 7 \times 10^9$ cm is the pressure scale height. Since $4T_e/T_{\mathrm{eff}} \sim 130$, we find that the relative variation in the surface solar-energy flux due to the decrease of the efficiency of convective heat transfer in the large-scale magnetic field is about 3×10^{-3}, and this variation will be in antiphase with the magnetic cycle.

Energy consumed by the generation of differential rotation and thermal heating due to viscous dissipation are important components of the thermal budget of the solar convection zone. In a steady state (in the absence of a magnetic field), the sources of differential rotation and viscous losses are balanced both locally and integrally. The large-scale magnetic field modifies this balance, since some of the differential-rotation energy goes into generation of the field. At the same time, the feedback of the large-scale fields on the turbulence produces additional paths for angular-momentum transfer. Furthermore, the energy expended by the large-scale fields to counteract the tension due to differential rotation must be taken into account. We see that the net energy balance is a complex combination of the interacting turbulence, differential rotation, and magnetic field. To simplify our estimates, note that, in an $\alpha\Omega$ dynamo such as that operating in the Sun, the energy of the large-scale field toroidal component is much greater than the energy of the poloidal component. Let us assume that the differential-rotation energy that goes into field generation is nearly completely compensated by heating due to dissipation of the large-scale field. Then, the remainder of the energy imbalance is equal to the energy of torsional oscillation. Let us estimate separately the contributions from dissipation of the large-scale field and torsional oscillation.

Adopting $V_{\mathrm{tors}} \approx 10$ m/s as a typical value for the torsional-oscillation velocity, we obtain the characteristic magnitude of thermal fluctuations due to torsional oscillations

$$\left.\frac{\delta \bar s'}{c_p}\right|_{\mathrm{tors}} \sim \frac{V_{\mathrm{tors}}^2}{c_p T_e} \sim \frac{10^6}{3.5\times10^8\times2.2\times10^5} \sim 1.7\times10^{-8}.$$

Therefore, the torsional-oscillation energy per unit volume is a negligibly small fraction of the heat content $c_p T$. The corresponding relative variations in the energy flux will be about 2×10^{-6}. We thus conclude that the effect of the thermal fluctuations that arise in the course of the cycle due to magnetic modulation of the generative and dissipative sources of differential rotation is small. On the other hand, it is known that modulation of the differential rotation in secular cycles can reach about 100 m/s, which results in a relative energy-flux variation of 2×10^{-4}. Such a variation in the solar irradiance can be detected observationally.

We now estimate the thermal variation due to sinks and the dissipation of large-scale fields in the dynamo cycle.

By analogy with the preceding estimates, we find

$$\left.\frac{\delta \bar s'}{c_p}\right|_{\mathrm{J.heat}} \sim \frac{\eta_T\delta t_{\mathrm{cyc}}(\Delta B_\phi)^2}{d^2 c_p T_e \, 4\pi\rho_e} \sim \frac{\eta_T\delta t_{\mathrm{cyc}}}{d^2}\frac{\langle u'^2\rangle}{c_p T_e} \sim 1.1\times10^{-5},$$
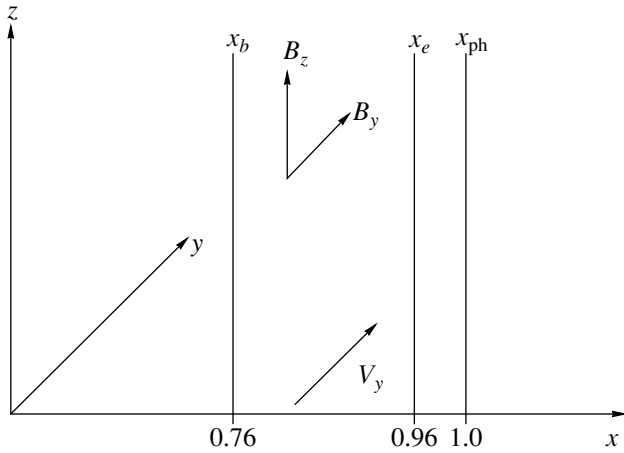
Fig. 1. General schematic of the model. The radial direction corresponds to the x coordinate, latitude to the z coordinate, and longitude to the y coordinate. The dynamo region is situated between x_b and x_e; x_{ph} corresponds to the photosphere. In the dynamo region, there is a large-scale velocity field, which has a shear such that the velocity increases from the pole to the equator (i.e., in the negative z direction).

where $d \sim 10^{10}$ cm is the half-thickness of the convection zone and $\langle u'^2 \rangle = \dfrac{(\Delta B_\phi)^2}{4\pi\rho_e\beta^2} \sim 100$ m/s for $\beta \sim 1$, in accordance with the model of Stix [13]. The corresponding luminosity variation comprises about 10^{-3}.

Based on these estimates for the main effects of the mean fields on heat transfer in the convection zone, we have ascertained that modulation of heat transfer by variations of the large-scale field in the activity cycle can result in modulation of the solar luminosity with a relative amplitude in the range 10^{-4}–10^{-3}. Moreover, we have demonstrated the important effect of long-term luminosity variations associated with secular variations in differential rotation. To illustrate these effects, we describe below a simple one-dimensional model for a dynamo in a convective layer.

3. NUMERICAL MODEL, RESULTS, AND DISCUSSION

We choose a Cartesian coordinate system, as shown in Fig. 1. The x axis is directed along the radius, the y axis in the direction of longitude, and the z axis in the direction of latitude. Distance along the x axis is measured in fractions of the solar radius: $r = xR$. Let us imagine a convective layer extended in the z direction. Let the boundary of the outer layer $x_{ph} = 1$ emit as a black body at $T_{eff} = 5870$ K. Let the upper boundary of the dynamo region be below the outer boundary, at $x_e = 0.96$. We specify a large-scale flow V_y in the dynamo region, with a shear $\partial V_y/\partial z = -W$ (the rotation speed is larger at the equator). We will assume that all physical

quantities vary only along the x axis. Then, the large-scale magnetic field has the form $\mathbf{B} = (0, T(x, t), P(x, t))$, where T and P are the toroidal and poloidal components, respectively. The corresponding dimensionless set of dynamo equations has been studied in detail by Kitchatinov and Mazur [14]:

$$\frac{\partial T}{\partial t} = -DP + \frac{\partial}{\partial x}\left(\hat{\eta}\frac{\partial T}{\partial x}\right),$$

$$\frac{\partial P}{\partial t} = \frac{\partial(\hat{\alpha}T)}{\partial x} + \frac{\partial}{\partial x}\left(\hat{\eta}\frac{\partial P}{\partial x}\right),$$

(6)

where $D = W\alpha_0 R^3/\eta_0^2$ is the dynamo number for the given model, and α_0 and η_0 are typical values of the coefficient α and the diffusivity, and $\hat{\alpha}$ and $\hat{\eta}$ are dimensionless functions of order unity. We measure the toroidal field in units of the pressure $\sqrt{4\pi p_e}$ (where p_e is the pressure at the outer boundary of the dynamo layer) and the poloidal field in units of $\alpha R\sqrt{4\pi p_e}/\eta$.

Equations (6) must be supplemented with the vacuum boundary conditions at the outer boundary of the dynamo region, $\mathbf{B}|_{x=x_e} = 0$, and the superconductor conditions at the bottom of the convective layer,

$$\left(\frac{\partial T}{\partial x}, \hat{\alpha}T + \hat{\eta}\frac{\partial P}{\partial x}\right)\Bigg|_{x=x_b} = 0.$$

We introduce the dimensionless quantity $s' = \bar{s}/c_p$ characterizing the distribution of entropy. We will take the heat capacity to be constant with depth. According to the model of [13], this is true for most of the solar convection zone. The superadiabatic temperature gradient is small over most of the convection zone; therefore, we will take the mean distribution of thermodynamic parameters in the convective layer to be adiabatic. Let us also assume that the condition of hydrostatic equilibrium is satisfied and that the free-fall acceleration increases inward according to the law $g = g_e/x^2$, where $g|_{x=x_e} = g_e$ (this is equivalent to the assumption that most of the solar material resides beneath the convective layer). Integrating the equation of hydrostatic equilibrium, we obtain the following distributions for the thermodynamic parameters:

$$\frac{p}{p_e} = f(x) = \left(\frac{x_e - x}{xx_e}c_f + 1\right)^{\frac{\gamma}{\gamma-1}};$$

(7)

$$T = T_e f(x)^{\frac{\gamma-1}{\gamma}}; \quad \rho = \rho_e f(x)^{\frac{1}{\gamma}},$$

with $c_f = \dfrac{2}{5}\dfrac{g_e\rho_e R_e}{p_e}$ and $\gamma = 5/3$; here, we have adopted the values $g_e = 2.53 \times 10^4$, $\rho_e = 5.1020 \times 10^{-3}$, $T_e = 2.2286 \times 10^5$ K, and $p_e = 1.42 \times 10^{11}$ in accordance with

the model of [13], expressed in the CGS system. To determine the radiative heat flux

$$F^{\text{rad}} = -c_p \rho \chi_D \frac{1}{R} \frac{\partial T}{\partial x}, \quad \chi_D = \frac{16}{3} \frac{\sigma T^3}{\kappa \rho^2 c_p}$$

we used a Kramers opacity law $\kappa = c_\kappa \rho T^{-7/2}$ [15] ($c_\kappa = 2.04 \times 10^{24}$ CGS). Thus, in view of (7), the expression for the radiative flux can be rewritten in the form

$$F^{\text{rad}} = \frac{5}{2} \frac{\chi_D^{(0)}}{R} \frac{p_e c_f f(x)^{1.4}}{x^2}, \quad \chi_D^{(0)} = \frac{16}{3} \frac{\sigma T_e^{6.5}}{\rho_e^2 c_\kappa c_p}.$$

The convective flux is determined from (4), (5), and (7) to be

$$F^{\text{con}} = -\frac{\gamma - 1}{\gamma R} \eta_0 \hat{\chi}_T \Psi(\beta) f(x) p_e \frac{\partial s'}{\partial x},$$

where

$$\hat{\chi}_T = \frac{\tau_{\text{cor}} \langle u'^2 \rangle}{3 \eta_0}.$$

The function $\Psi(\beta)$ is given in [12]. The parameters of the turbulence are given by the relations

$$\langle u'^2 \rangle = -\frac{\alpha_{MLT}^2}{4} \frac{H_p^2}{x^2 \tau_{ff}^2} \frac{\partial s'}{\partial x},$$

$$\tau_{\text{cor}} = \left(\frac{8}{15} \frac{\alpha_{MLT}^2 x^6 f(x)^{1.8} R p_e \tau_{ff}^2}{c_f^2 T_{\text{eff}}^4} \right)^{1/3}$$

$$\times \left(1 - \frac{\gamma}{\gamma - 1} \frac{\chi_D^{(0)} c_f p_e f(x)^{1.4}}{R \sigma T_{\text{eff}}^4} \right)^{-1/3},$$

where α_{MLT} is the mixing-length constant, $H_p = R f'(x)$ the pressure scale height, $\tau_{ff} = \sqrt{R/g_e}$ the gravitational-collapse time, and τ_{cor} is estimated without taking into account the effect of the large-scale fields on turbulence—by equating the total (convective plus radiative) energy flux to the radiation flux from the surface of a black body. As noted above, we assume that the differential rotation (velocity-field shear) is given; i.e., we do not take into account the turbulent energy that goes into the generation of differential rotation. For this reason, the turbulent energy involved in the formation of the toroidal large-scale field via differential rotation remains unaccounted for. In order to recover the energy balance, we introduce *a priori* a term describing the differential-rotation energy spent to generate the toroidal field in the thermal-balance equation. Later, with the development of spherical dynamo models that take into account all sources and sinks of turbulent energy in the convection zone, the need for such a procedure will be removed.

It is clear from the first equation of the dynamo system (6) that the power expended by the differential rotation to generate the large-scale field is DPT. Using the above expressions, the dimensionless heat transfer equation assumes the form

$$\frac{\partial s'}{\partial t} = -\frac{1}{f(x)} \frac{\partial}{\partial x} \left[\frac{\gamma - 1}{\gamma} \hat{\chi}_T \Psi(\beta) f(x) \left(-\frac{\partial s'}{\partial x} \right) \right.$$

$$\left. + \frac{\chi_D^{(0)}}{\eta_0} \frac{f(x)^{1.4}}{x^2} c_f \right] + \frac{\gamma - 1}{\gamma} \left\{ \frac{\hat{\alpha} T}{f(x)} \frac{\partial P}{\partial x} + \frac{\hat{\eta}}{f(x)} \right. \quad (8)$$

$$\times \left[\left(\frac{\partial P}{\partial x} \right)^2 + \left(\frac{\partial T}{\partial x} \right)^2 \right] + \frac{D}{f(x)} PT \right\},$$

where $\eta_0 = 1.1 \times 10^{13}$ cm²/s is the characteristic diffusivity for our model. All parameters of the convective layer are specified once two constants—c_κ and α_{MLT}—are given. For the Sun, $c_\kappa = 2.04 \times 10^{24}$ (in CGS units). The bottom of the convection zone is defined by the equality between the radiative flux and the total flux of radiation. In a planar geometry, the total flux of radiation does not vary with depth, whereas the convective flux grows. Therefore, we expect a thinner convective envelope in a planar geometry. For the above-specified parameters, the bottom of the convective layer turns out to be at $x_b = 0.76$. The mixing-length theory constant is taken to be in the range $\alpha_{MLT} = 1$–2. For equation (8), the boundary condition at $x = x_e$ is relation (3); the convective flux is assumed to vanish at the bottom of the convective layer.

Equations (6) and (8) were solved numerically, in time steps, using a finite-difference method. Prior to these computations, we calculated the initial entropy distribution from the time-independent heat-transfer equation (without magnetic fields) $F_x^{\text{con}} + F_x^{\text{rad}} = \sigma T_{\text{eff}}^4$. We then specified a weak magnetic field and followed the nonlinear evolution of the magnetic field and entropy distribution until the system reached a steady state. Below, we consider only such steady-state behavior of the fields.

Figure 2 presents the calculation results for $\alpha_{MLT} = 1$. We assumed that the coefficients of the magnetic-field generation and diffusion were independent of the magnetic-field strength: $\hat{\alpha} = f_\alpha(\beta)$ and $\hat{\eta} = f_\eta(\beta)$. The functions $f_\alpha(\beta)$ and $f_\eta(\beta)$ are described in [14].

Contours of the toroidal magnetic field and heat increment in the convective layer are shown in the top diagram of Fig. 2. The magnetic-field contours are separated by 2 kG, with a field-amplitude maximum of 10 kG in the middle of the layer; the maximum value of β (the ratio of the magnetic energy to the convective-flow energy) is 1.3, so that the dynamo under consideration is weakly nonlinear.
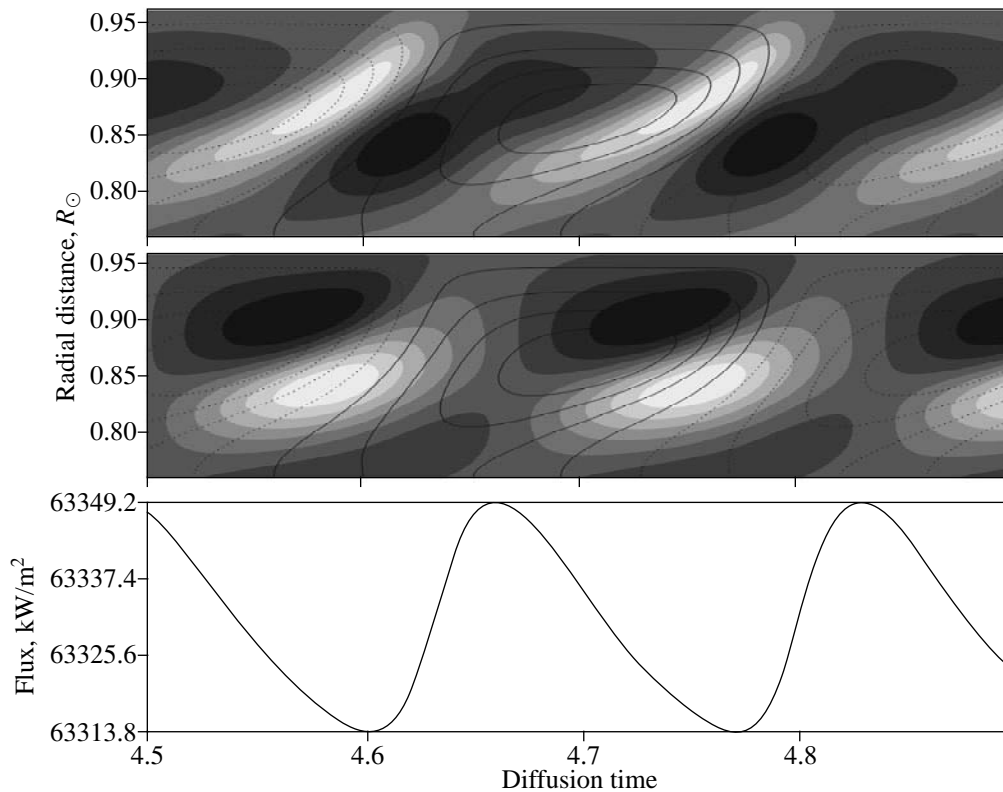
Fig. 2. Top and middle: contours of the toroidal magnetic field in radial distance–time coordinates. Superposed are gray scales representing the current (instantaneous) increments of thermal energy (top diagram) and variations in the mean amount of heat (middle diagram) obtained by integrating the heat increments shown in the top diagram over time. Black regions correspond to negative values. Bottom: variations in the radiation flux at the photospheric level.

The current (instantaneous) increments of the thermal energy in time ($\Delta Q = \bar{\rho}\,\bar{T}\,\Delta\bar{s}$) are shown by a gray scale; the dark and light regions correspond to negative and positive increments, respectively. By and large, negative increments coincide with regions of growing magnetic fields and positive increments with regions of decaying fields; therefore, they can be interpreted as energy sinks associated with the generation of large-scale fields and with dissipative heating, respectively. The integral amount of heat in the layer averaged over a cycle remains constant. To demonstrate this more convincingly, we have integrated the distribution of heat increments over time. The result is shown in the middle diagram of Fig. 2, along with contours of the toroidal field and variations in the averaged heat (gray scale; dark regions correspond to cooling and light regions to heating). As expected, the pattern of departures of the mean thermal energy is displaced with respect to the pattern of increments by $\pi/2$.

We can clearly seen from the diagram of departures of the mean thermal energy that the general decrease in the surface energy flux is associated with the blocking of this flux by the magnetic field, whose strength is peaked in the middle of the layer. We can see that

increase in the surface energy flux is associated, first and foremost, with the release of energy stored during the preceding cycle and the dissipation of large-scale magnetic fields.

The bottom diagram of Fig. 2 shows the variation in the surface radiation flux. The amplitude of the variation is about 38 kW/m^2, or 0.07% of the total radiation flux, which, in principle, agrees with the observed variations to order of magnitude. The maximum brightness corresponds approximately to the field maximum in the layer, i.e., to the cycle maximum. If we assume that magnetic activity at the surface is determined by the field in the lower half of the convection zone, the brightness will be maximum at the decay phase of these fields and minimum at the growth phase. This seems to agree with the results of analyses based on series of the integrated solar irradiance reported in various studies [2, 3]. Most authors incline toward the view that the general increase in luminosity during the magnetic cycle is associated with an overcompensation of the deficiency of emission from spots by the excess emission in decaying active regions (faculae and network magnetic elements). Recently, this was also demonstrated by Mordvinov and Wilson [16] based on a wavelet analysis of series of observations of the inte-

grated solar irradiance and comparisons with the development of magnetic activity at the solar surface.

When α_{MLT} is increased to 2 with fixed dynamo number, the amplitude of variations in the surface flux is reduced to 20 kW/m^2. This can be accounted for by the fact that, at larger α_{MLT}, the rms convection velocities are larger, so that the ratio of the magnetic energy to the convective-flow energy (β) is smaller; accordingly, the amplitude of modulations of the effects of thermal and magnetic diffusion by the magnetic field will be smaller.

In this theory, the observed variability of the Sun over the 11-year activity cycle can be interpreted as the result of modulation of the radiation flux by the large-scale field generated by the solar dynamo. The asymmetric shape of the brightness curve is associated with the nonlinear nature of the dynamo. The phase of magnetic-field growth turns out to be shorter than the decline phase due to the nonlinearity of the magnetic diffusivity with respect to the strength of the large-scale field.

The results of our study indicate that the large-scale field generated by the solar dynamo can substantially affect the solar luminosity. The principal factors in this process are changes in the efficiency of convection in the presence of a magnetic field, as well as energy expenditures and releases due to the generation and dissipation of large-scale magnetic fields.

ACKNOWLEDGMENTS

APPENDIX

The heat-transfer equation for a rotating turbulent medium with a magnetic field can be obtained using the corresponding energy-conservation law for magnetohydrodynamics. In accordance with Landau and Lifshitz [10], we write this in the form

$$\rho \frac{d}{dt}\left(\frac{v^2}{2} + \epsilon\right) + \frac{\partial}{\partial t}\frac{H^2}{8\pi} \qquad (9)$$
$$= (\mathbf{v} \cdot \mathbf{f}_e) - \mathbf{div}(\mathbf{S} + \mathbf{F} + \mathbf{v} \circ \hat{\sigma}),$$

where ρ is the density, v the flow velocity, ϵ the internal energy, H the magnetic field, $\mathbf{S}$ the Umov–Poynting vector of electromagnetic energy flux, $\mathbf{F}$ the vector of heat flux due to radiation and molecular heat conduction, and $\sigma_{ij} = \sigma'_{ij} - p\delta_{ij}$ (where p is the pressure and $\hat{\sigma}'$ is the viscosity tensor), f_e are the external forces. We

assume that the medium is a perfect gas and that the inelastic-flow condition $\mathrm{div}\,\rho\mathbf{v} = 0$ is satisfied. In this case, equation (9) can be rewritten using the well-known thermodynamic relationships in a form more convenient for further manipulation:

$$\rho T\partial_t s + \mathrm{div}\,c_p\rho\mathbf{v}T + \rho\frac{d}{dt}\frac{v^2}{2} + \frac{\partial}{\partial t}\frac{H^2}{8\pi} \qquad (10)$$
$$= (\mathbf{v} \cdot \mathbf{f}_e) - \mathbf{div}(\mathbf{S} + \mathbf{F} + \mathbf{v} \circ \hat{\sigma}),$$

where T and c_p are the temperature and specific heat, respectively.

In accordance with the usual practice in the theory of turbulence, we decompose all quantities into their mean and random components ($v_j = V_j + u'_j$, etc.) and average over an ensemble of fluctuations. We neglect the contributions of second-order fluctuations of thermodynamic quantities, but retain the products of these fluctuations with the velocity field. This corresponds to an inelastic approximation for the solar convection [17]. Since the convective mass flux is small, we omit terms of the form $\langle \rho' u'_j \rangle$ (where $\langle \dots \rangle$ denotes averaging). In view of the aforesaid, averaging the first two terms in (10) yields

$$\langle \rho T\partial_t s + \mathrm{div}\,c_p\rho\mathbf{v}T \rangle = \bar{\rho}\bar{T}(\partial_t\bar{s} + V_i\partial_i\bar{s}) + \partial_i F_i^{\mathrm{con}}, \qquad (11)$$

where $F_i^{\mathrm{con}} = c_p\bar{\rho}\langle u'_i T' \rangle$ is the convective heat flux. Convective turbulence is a source of small-scale fluctuations of the velocity and magnetic field.

In a mixing-length-theory approximation, the intensity of convection is described completely by the distribution of the mean entropy. Therefore, in the approximation used, we must take the entropy production rate to include the power of the velocity and magnetic-field fluctuation fields. Thus, we assume that $\bar{\rho}\bar{T}\partial_t\bar{s}$ contains derivatives of the turbulent energy $\bar{\rho}\frac{d}{dt}\frac{\langle u'^2 \rangle}{2} + \frac{\partial}{\partial t}\frac{\langle h^2 \rangle}{8\pi}$, where h is the small-scale magnetic field. In this case, we take into account the effect of small-scale fields on the evolution of the mean fields and do not consider the evolution of the small-scale fluctuations themselves.

This theory will correctly describe variations in heat transfer on the time scale of the mean-field variations, i.e., on the dynamo-cycle time scale. If we wish to take into account shorter period variations, we must reject mixing-length theory, and use more accurate tech-

niques of turbulence theory to describe the small-scale fields. With account of the aforesaid, averaging equation (10) yields

$$\bar{\rho}\bar{T}(\partial_t \bar{s} + V_i \partial_i \bar{s}) + \rho \frac{d}{dt} \frac{V^2}{2} + \frac{\partial}{\partial t} \frac{B^2}{8\pi}$$

$$= -\partial_i(F_i^{\text{con}} + F_i^r + V_j \hat{T}_{ij} + \hat{S}_i) + (V \cdot f_e). \quad (12)$$

Here, we take into account the fact that heat transfer F_i^r, by radiation and molecular heat conduction can play an important role in stellar atmospheres. **V** and **B** are the large-scale components of the velocity and magnetic field, respectively. We also introduce the turbulent stress tensor

$$\hat{T}_{ij} = \rho \langle u_i' u_j' \rangle + \frac{1}{4\pi}\left(\langle h_i h_j \rangle - \frac{\delta_{ij}}{2} \langle h^2 \rangle\right)$$

$$= -Q_{ij}^\Lambda + N_{ijkl} \partial_k V_l,$$

where the tensor Q_{ij}^Λ describes the so-called Λ effect for the generation of differential rotation [9] and the viscosity tensor N_{ijkl} describes the turbulent dissipation of shear flows (calculations of N_{ijkl} are presented in [12]). $\hat{S} = \frac{1}{4\pi}(((V \times B) \times B) - \bar{\varepsilon} \times B)$ is the energy flux of the large-scale field, where $\bar{\varepsilon}_i = \langle u' \times h \rangle_i = \alpha_{ij} B_j - \eta_T (\nabla \times B)_i$ is the mean emf of small-scale fields expressed in a simple form, with α_{ij} describing the α effect for the generation of large-scale field and η_T describing turbulent magnetic diffusion. These effects were calculated with allowance for the feedback of the magnetic fields on the turbulence in [12, 18].

To complete the derivation, we need equations for the energies of the large-scale field and mean flow, which can be obtained from the mean-field equations using a standard procedure. The equation for the energy of the large-scale field turns out to be

$$\frac{\partial}{\partial t} \frac{B^2}{8\pi} + \partial_i \hat{S}_i = \frac{1}{4\pi}(V \cdot (B \times (\nabla \times B))$$

$$+ \alpha_{ij} B_j (\nabla \times B)_i - \eta_T (\nabla \times B)^2). \quad (13)$$

The first term on the right-hand side of (13) describes the work of the magnetic force during deformation of the large-scale field by the velocity field. The product of the kinetic and magnetic helicities describes the power of the turbulent sources generating the large-scale field. Finally, the last term represents the energy sink due to the turbulent dissipation of magnetic fields.

We write the energy equation for the large-scale velocity field in the form

$$\bar{\rho}\frac{d}{dt} \frac{V^2}{2} = V_j \partial_i(\hat{T}_{ij} - \delta_{ij}\bar{P})$$

$$- \frac{1}{4\pi}V \cdot (B \times (\nabla \times B)) + (V \cdot f_e). \quad (14)$$

By subtracting (13) and (14) from (12), we obtain an equation describing the rate of variation of the mean entropy in a turbulent medium under the influence of large-scale velocity and magnetic fields:

$$\bar{\rho}\bar{T}\partial_t \bar{s} = -\bar{\rho}\bar{T} V_i \partial_i \bar{s} - \partial_i(F_i^{\text{con}} + F_i^{\text{rad}})$$

$$+ (-Q_{ij}^\Lambda + N_{ijkl}\partial_k V_l)\partial_j V_i \quad (15)$$

$$- \frac{1}{4\pi}(\alpha_{ij} B_j (\nabla \times B)_i - \eta_T (\nabla \times B)^2).$$

The pressure term in (15) can be eliminated using the incompressibility condition for the mean flows.

The resulting equation admits a simple physical interpretation. Let us assume that the large-scale fields are axisymmetric; then, the first term on the right-hand side represents heat transfer via meridional circulation, the following two terms describe turbulent convection and radiative transfer, the next two describe energy spent on the generation of differential rotation and energy supplied by dissipation of the flows, and the last two terms describe energy consumed in the generation of the large-scale field by kinetic helicity and thermal energy supplied by magnetic-field dissipation. The resulting equation describes the balance of thermal energy in the convection zone, with account for its transformation into large-scale magnetic fields and flows. Since the physical situation is repeated in each magnetic cycle, the total change in the integrated thermal energy over a cycle must be zero.

REFERENCES

1. R. C. Wilson, Space Sci. Rev. **38**, 203 (1984).

2. R. C. Wilson and H. S. Hadson, Nature **332**, 810 (1988).

3. P. V. Foukal and J. Lean, Astrophys. J. **328**, 347 (1988).

4. J. L. Lean, J. Cook, W. Marquette, and A. Johannesson, Astrophys. J. **492**, 390 (1998).

5. H. C. Spruit, in *The Sun in Time*, Ed. by C. Sonett, M. Giampapa, and M. S. Matthews (Univ. of Arizona Press, Tucson, 1991), p. 118.

6. E. Parker, Astrophys. J. **440**, 415 (1995).

7. H. C. Spruit, Astron. Astrophys. **108**, 348 (1982).

8. N. J. Balmforth, D. O. Gough, and W. J. Merryfield, Mon. Not. R. Astron. Soc. **278**, 437 (1996).

9. L. L. Kitchatinov and G. Rüdiger, Astron. Astrophys. **269**, 581 (1993).

10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics, Vol. 8: Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).

11. P. A. Gilman and G. A. Glatzmaier, Astrophys. J., Suppl. Ser. **45**, 335 (1981).

12. L. L. Kitchatinov, V. V. Pipin, and G. Rüdiger, Astron. Nachr. **315**, 157 (1994).

13. M. Stix, *The Sun: An Introduction* (Springer, Berlin, 1990).

14. L. L. Kitchatinov and M. V. Mazur, Pis'ma Astron. Zh. **25**, 549 (1999) [Astron. Lett. **25**, 471 (1999)].

15. R. Kippenhahn and A. Weigert, *Stellar Structure and Evolution* (Springer, Berlin, 1994).

16. A. V. Mordvinov and R. C. Wilson, Geophys. Res. Lett. **26**, 3613 (1999).

17. D. O. Gough, J. Atmos. Sci. **26**, 448 (1969).

18. L. L. Kitchatinov and G. Rüdiger, Astron. Astrophys. **276**, 96 (1993).

Translated by A. Getling