Astronomy Reports, Vol. 48, No. 5, 2004, pp. 353–366. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 387–402.
Original Russian Text Copyright © 2004 by Kasimova, Shchekinov.

Single-Zone Model for the Evolution of Damped Lyα Systems

E. R. Kasimova and Yu. A. Shchekinov

Department of Physics, Rostov State University, Rostov-on-Don, 344090 Russia

Received December 21, 2003

Abstract—We use a numerical model to investigate the chemical evolution of spiral galaxies. Unlike previous models, our scheme includes the effects of the condensation of heavy elements onto dust grains and the influence of dust on the colors of the galaxies and (both relative and absolute) heavy-element abundances in the gaseous phase. The calibration of the numerical model is based on the observed colors of spiral galaxies, oxygen abundances in HII regions, rates of type Ia and II supernovae, and the mass fraction of gas. The model is used to investigate the possibility that damped Lyα systems observed in the spectra of quasars are associated with spiral galaxies at appropriate evolutionary stages (at redshifts from $z = 5$ to $z = 0$). In the closed, single-zone model used, the damped Lyα systems can be interpreted as Sa–Sd galaxies at early stages of their evolution. In this case, however, the calculated relative abundances of elements such as Si and Fe are, on average, underestimated. We conclude that the incorporation of a dust component is essential for a correct description of the chemical evolution of damped Lyα systems and correct interpretation of the elemental abundances observed in them. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It has become clear over the past decade, owing primarily to advances in observational astronomy, that damped Lyα systems with hydrogen densities $N(\mathrm{HI}) > 2 \times 10^{20}$ cm^{-2} (DLA systems) are associated with the gaseous disks of galaxies at various stages of their evolution [1–4]. The nature of the galaxies themselves—in particular, their morphological type—is unclear, although the few observations of low-redshift DLA systems in emission suggest that these objects may represent a wide range of morphological types: spiral, dwarf, and low surface-brightness galaxies [5]. One characteristic feature of DLA systems is that they contain appreciable amounts of heavy elements (metals), with the corresponding abundances ranging from 10^{-2} to 0.3 of the solar value. From this viewpoint, DLA systems provide a tool for studies of galactic evolution. However, confident interpretation of the abundances observed in DLA systems is hindered by an important factor—the possible presence of dust. Heavy elements can condense onto dust grains, substantially decreasing the heavy-element abundance in the gaseous phase [6]. Pettini and Bowen [7] found direct evidence for a depletion of gaseous-phase heavy elements in DLA systems. They found that the metallicities in low-redshift DLA systems (i.e., in systems at nearly the same evolutionary stage) can vary over two orders of magnitude. In addition, extinction due to dust in DLA systems may decrease the number of observable systems [8].

In the current paper, we describe a model for the photometric and chemical evolution of galaxies and carry out a theoretical analysis of the evolutionary status of DLA systems. An important feature of this model is that it explicitly includes equations describing the formation and destruction of dust, which affects both the photometric and chemical properties of the modeled system, so that, in this sense, our model is self-consistent. We report here only our results for spiral galaxies, for which the model of the photometric and chemical evolution can be described in the simplest way. We will report our results for other types of galaxies and their agreement with DLA systems in a separate paper.

Section 2 gives a general description of the model and its components. Section 3 discusses the constraints on the model parameters imposed by observations of galaxies at the present epoch. Section 4 reports the results of our modeling of the chemical evolution of spiral galaxies as DLA systems. Finally, Section 5 summarizes our conclusions.

2. DESCRIPTION OF THE MODEL

We report here results obtained using a closed, single-zone model for the chemical and photometric evolution of a galaxy. The model obviously provides an oversimplified description of spiral galaxies, which are characterized by radial gradients of their chemical composition and are not closed systems, since they exchange gas with the intergalactic medium. However, although it is limited, this description provides a

1063-7729/04/4805-0353$26.00 © 2004 MAIK "Nauka/Interperiodica"

qualitative understanding of the role played by various processes in the galaxies and indicates the evolutionary factors that must be further refined in order to improve the modeling.

Our aim is to calculate in a self-consistent way the properties of the dust component of the interstellar medium (ISM) in a model for the evolution of a galaxy and to use this model to interpret the observed chemical compositions of damped Lyα systems. A refined version of the model will include the physical processes accompanying the exchange of gas and dust between the galaxy and intergalactic medium.

2.1. Chemical and Photometric Model

The standard method for investigating the spectrophotometric and chemical evolution of galaxies is to construct models for these processes. Our calculations are based on the PÉGASE.2 code [9, 10], which can be used to calculate the spectra, colors, and metallicities of galaxies at any stage of their evolution. PÉGASE.2 uses the stellar tracks [11–15] and isochrones [16] developed by the Padua University group and also the set of theoretical stellar spectra [17, 18], based on the stellar model atmospheres [19, 20]. This set includes stellar spectra for all the metallicities of the Padua isochrones, spanning a broad wavelength interval from 90 Å to 160 μm with a spectral resolution of 10–20 Å.

One of the advantages of the PÉGASE.2 code is that it enables a chemically self-consistent calculation of the metallicity of a galaxy in the process of its evolution. This is more correct than fixing the metallicities of newly born stars, since the ISM is gradually enriched in heavy elements as new generations of stars live and die. The PÉGASE.2 code also enables a self-consistent treatment of the internal extinction of the galaxies when calculating their spectrophotometric properties. In this case, the model for spiral galaxies assumes that the stars, gas, and dust are distributed uniformly in a plane-parallel layer whose optical depth is estimated as a function of the mass and metallicity of the gas [9]. The optical properties of silicate and graphite grains used in PÉGASE.2 are taken from [21, 22].

At the beginning of its evolution, the modeled galaxy forms a gaseous cloud with chemical composition Z_0 and mass M_0. Given the star-formation rate (SFR) $\Psi(t)$ and initial mass function (IMF) $\varphi(m)$, the evolution of the gaseous and stellar components of the galaxy can be described by the following set of equations:

$$\frac{dG(t)}{dt} = -\Psi(t) + E(t) + A(t) - W(t), \quad (1)$$

$$\frac{dS(t)}{dt} = \Psi(t) - E(t), \quad (2)$$

where $M_t(t)$ is the total mass of the galaxy; $M_g(t)$ is the mass of gas in the galaxy; $G(t) = M_g(t)/M_t(t)$ and $S(t) = M_s(t)/M_t(t)$ are the masses of gas and stars normalized to the total mass of the galaxy; respectively, $\Psi(t)$ is the SFR;

$$E(t) = \int_{m_{low}}^{m_{up}} \Psi(t-\tau_m)\, Q_m(t-\tau_m)\, \varphi(m)\, dm \quad (3)$$

is the rate at which stars return gas to the ISM; $A(t)$ [s^{-1}] is the accretion rate to the galaxy from the ambient medium; and $W(t)$ [s^{-1}] is the rate at which gas is lost by the galaxy due to winds and/or sporadic large-scale mass ejections. To analyze the photometric evolution of a galaxy, we calculate the monochromatic flux $S_M(\lambda, t, Z)$ produced by all stars born in a single burst of star formation and then the monochromatic flux of the entire galaxy:

$$S_{gal}(\lambda, t, Z) \quad (4)$$
$$= \int_0^t \int_{m_{low}}^{m_{up}} S_*(\lambda, m, t-\tau, Z) \cdot \varphi(m) \cdot \Psi(\tau) \cdot dm \cdot d\tau,$$

where $S_*(\lambda, \tau, Z)$ is the spectrum of a star of mass m, metallicity Z, and age τ. We finally calculate the model colors of the galaxies by convolving the model spectra with filter response curves (see [23] for details).

The chemical-evolution equations for the single-zone model have the form [24, 25]

$$dG(X,t)/dt = -\Psi(t)Z(X,t) \quad (5)$$
$$+ \int_{m_{low}}^{M_b^{\min}} \Psi(t-\tau_m) Q_m(X, t-\tau_m) \varphi(m) dm$$
$$+ \beta \int_{M_b^{\min}}^{M_b^{\max}} \varphi(M_b) \left[\int_{\mu_{\min}}^{0.5} f(\mu) \Psi(t-\tau_{m_2}) \right.$$
$$\left. \times Q_m(X, t-\tau_{m_2}) d\mu \right] dM_b + (1-\beta)$$
$$\times \int_{M_b^{\min}}^{M_b^{\max}} M_b^{\max} \Psi(t-\tau_m) Q_m(X, t-\tau_m) \varphi(m) dm$$
$$+ \int_{M_b^{\max}}^{m^{up}} \Psi(t-\tau_m) Q_m(X, t-\tau_m) \varphi(m) dm$$
$$+ Z_A(X) \cdot A(t) - Z_W(X) \cdot W(t),$$

where $\varphi(m)$ is the stellar IMF defined in the mass interval (m_{low}, m_{up}) and normalized to unity, $G(X,t) = M(X,t)/M_t(t)$ is the mass fraction of chemical element X, β is the fraction of binaries in the mass interval from $M_b^{\min}$ to $M_b^{\max}$ resulting in type Ia supernovae, $Z(X,t) = G(X,t)/G(t)$ is the mass density of element X in the gaseous phase, $\tau(m)$ is the main-sequence lifetime of a star of mass m, $Q_m(X, t-\tau_m) = (m - m_i)/m$ is the mass fraction of element X ejected by a star of mass m [26, 27], $Z_A(X)$ is the mass density of element X in the accreted gas, and $Z_W(X)$ is the mass density of element X in the galactic wind.

The terms on the right-hand side of (5) describe (i) the rate of the decrease of the amount of element X in the ISM as a result of star formation; (ii) the rate at which stars in the mass interval $(m_{low}, M_b^{\min})$ enrich the ISM in element X; (iii) the contribution of type Ia supernovae occurring as a result of the evolution of binaries in the mass interval $(M_b^{\min}, M_b^{\max})$; (iv) the evolution of single stars with masses in the interval $(M_b^{\min}, M_b^{\max})$; (v) the contribution to the evolution of element X of stars with masses greater than $M_b^{\max}$; and (vi), (vii) the rates of accretion and ejection of element X as the galaxy exchanges gas with the intergalactic medium.

2.2. Evolution of Dust

A realistic description of the galaxy's evolution must include a self-consistent calculation of the dust component of the ISM and its effect on the photometric and chemical properties of the system. A correct description of the evolution of the dust component can be found in [28, 29]. Following [29], our model describes the evolution of dust in galaxies using the following set of equations:

$$dD(X,t)/dt = -\Psi(t)Z^d(X,t) \tag{6}$$
$$+\int_{m_{low}}^{M_b^{\min}} \Psi(t-\tau_m)Q_m(X,t-\tau_m)\varphi(m)\delta(X)dm$$
$$+\beta\int_{M_b^{\min}}^{M_b^{\max}} \varphi(M_b)\left[\int_{\mu_{\min}}^{0.5} f(\mu)\psi(t-\tau_{m_2}) \times Q_m(X,t-\tau_{m_2})\delta^I(X)d\mu\right]dM_b + (1-\beta)$$
$$\times\int_{M_b^{\min}}^{M_b^{\max}} \psi(t-\tau_m)Q_m(X,t-\tau_m)\delta^{II}(X)\varphi(m)dm$$
$$+\int_{M_b^{\max}}^{m^{up}} \psi(t-\tau_m)Q_m(X,t-\tau_m)\delta^{II}(X)\varphi(m)dm$$
$$-D(X,t)/\tau_{snr}(X,t) + D(t) \times [1 - D(X,t)/G(X,t)]/\tau_{acc}(X,t)$$
$$-Z_w^d(X)\cdot W(t) - [dD(X,t)/dt]_{rad},$$

where $D(X,t)$ is the mass of element X contained in dust, $D(t) = \Sigma_X D(X,t)$ is the total mass of dust, $Z^d(X,t) = D(X,t)/G(t)$ is the mass fraction of element X in dust, $\delta(X)$ is the efficiency of the condensation of dust in ordinary stars, $\delta^I(X)$ is the efficiency of the condensation of dust in type Ia supernovae, $\delta^{II}(X)$ is the efficiency of the condensation of dust in type II supernovae, $\tau_{snr}(X)$ is the time scale over which element X returns from the condensed (dust) phase to the gaseous phase during the destruction of dust grains by shocks produced by supernovae, τ_{acc} is the time scale for condensation (accretion) of metals from the gaseous phase onto the surface of dust grains, $Z_w^d(X)\,W(t)$ is the rate at which particles of element X contained in dust are swept out by the galactic wind, and $[dD(X,t)/dt]_{rad}$ is the rate at which particles of element X contained in dust are swept out by stellar radiation pressure.

The terms on the right-hand side of (6) correspond to the following processes that determine the dust-mass balance:

(i) the decrease of the dust mass during condensation from the gaseous phase into stars;

(ii) the ejection of dust into the ISM by the stellar wind;

(iii) the condensation of dust in type Ia supernovae;

(iv) the condensation of dust in type II supernovae;

(v) the destruction of dust by shocks produced by supernovae, with the time scale for this process calculated as [28, 30]

$$\tau_{snr} = \frac{G}{\varepsilon M_{snr} R}, \tag{7}$$

where ε is the efficiency of the destruction of dust in the case of a fairly strong shock propagating from a supernova at $v_s > 50$ km/s; M_{snr} is the mass of the gaseous remnant of the supernova, i.e., of the interstellar gas set into motion by the shock (in units of the initial mass ofthe galaxy); and R is the supernova rate in the galaxy;

(vi) the increase of the mass of the dust component as a result of the condensation of gas onto dust grains in interstellar clouds [31, 32], with the time scale for condensation (accretion) being [28, 30–32]

$$\tau_{acc}(X) = \frac{m_{gr}}{\pi a^2}[\alpha(X)\mu n_c v_{tc}\delta_c], \tag{8}$$

where m_{gr} is the mass of the dust grain, a is its radius, $\alpha(X)$ is the sticking coefficient between the atom and the dust grain, μ is the mean mass of the particles of the ambient gas, n_c and v_{tc} are the number density and thermal velocity of these particles in the clouds, and δ_c is the fraction of volume occupied by the clouds.

We assume here that the rate of accretion of heavy elements onto dust grains is proportional to the total mass of elements in the condensed phase, i.e., $\propto D(t)$ (see (6)), in contrast to [28], where the accretion rate is $\propto D(X,t)$ (see (6) in [28]). This is due to the fact that the number of collisions of particles of a given element X with dust grains per unit volume per unit time is determined by the total number density of dust grains n_d and not by the partial fraction of the element X contained in them. We have $\tau_{acc} \sim 10^8/\alpha(X)$ yr for typical parameters of dense clouds in the Galaxy. Note that τ_{acc} depends on the cloud density as $\tau_{acc} \propto n_c^{-1}$. Although the observed molecular-hydrogen densities in DLA systems are extremely low [33], we can nevertheless assume that these systems contain a cloud phase. 21-cm absorption observations of some DLA systems indicate that they have a multiphase structure, with the minimum temperature ranging from $T = 103$ K [34] to $T = 180$ K [35]. Similar temperatures are implied by the observed populations of the rotational levels of H_2 molecules ($T = 70$ K [36]), although, on the whole, the cloud phase in DLA systems seems to be hotter [33, 36, 37]. Moreover, 21-cm emission observations of DLA systems show that the inhomogeneity of the HI distribution increases with the angular resolution of the observations, and its structure becomes more clumpy [37]. Under such conditions, as was shown by Liszt [38], the low molecular-hydrogen densities may be due to fairly high intensities of ultraviolet radiation and a relatively low (proportional to the metallicity) dust content in DLA systems.

We assumed that the time scale τ_{acc} for the accretion of gas onto dust grains is equal to the Galactic value noted above. In reality, however, as is evident from its definition, it is also a function of the state of the ISM and can vary strongly from galaxy to galaxy, depending on the fraction of gas contained in dense clouds, their spatial structure, which determines how they are affected by shocks, the temperature of dust grains in the clouds, which can influence the sticking coefficients of atomic particles, etc. We restrict our analysis to a simplified treatment of these processes. Table 1 gives the numerical values of the parameters characterizing the destruction of dust grains by shocks and their condensation in the ejecta of type Ia and II supernovae. Table 2 gives the numerical values of $\alpha(X)$ adopted for dust grains in the selective model.

In our Galaxy and its neighbors, the total mass of dust in the ISM is approximately equal to the mass of metals, although the ratio varies somewhat from galaxy to galaxy. The higher SFR (and the associated supernova rate) in DLA systems may result in the more efficient destruction of dust grains, so that their mass can be substantially smaller than the mass of metals produced by stars, as has been pointed out by Pettini *et al.* [30] and Molaro *et al.* [40]. When calculating the destruction of dust by shock waves in galaxies of various types, we normalized τ_{snr} to its value for the Milky Way and assumed it to be inversely proportional to the supernova rate, so that the rate of dust destruction is $\propto R$. Since the production of dust in cool stellar atmospheres and the ejecta of supernova are also proportional to the SFR, and, consequently, to R, the ratio of the amounts of dust and metals produced should remain constant up to some Ψ and R. This is made even more true by the fact that the rate of accretion onto dust grains in clouds can also increase with the SFR Ψ. The turbulent pressure of the interstellar gas varies as a function of the supernova rate per unit volume as $p \propto \nu^{3/4}$ [41], and if the supernova rate ν is not too high and interstellar clouds are not disrupted, the gas density in clouds should also vary as $n_c \propto \nu^{3/4}$. The exact ν at which the cloud phase is disrupted depends strongly on the physical conditions in the interstellar gas, in particular, on the sizes, density, and spatial distribution of interstellar gas clouds, their formation rate, and a number of other factors, many of which remain incompletely studied.

2.3. Accretion and Ejection

The accretion of external gas onto the galaxy and the ejection of matter from it are difficult to describe correctly in a single-zone approach. Each particular case (each type of galaxy) requires its own specific formulation of the problem, with corresponding qualifications. The rate of gas accretion from the surrounding space depends on the gravitating mass of the galaxy and, strictly speaking, on its geometry, the parameters of the ambient gas, and the relative motion of the galaxy and the intergalactic gas. The chemical composition of the accreted matter depends on whether the galaxy is isolated or a member of a cluster. Even when the galaxy is isolated, the ambient intergalactic matter may be enriched in heavy elements to some degree (see the discussion in [42–44]) and even in dust (see the discussion in [45–48]). Moreover, in the single-zone model, both accretion and ejection can also include mass exchange between different regions of the same galaxy. The description of ejection [the last term in (5) and (6)] is equally uncertain. The ejection of matter by shocks produced by

Table 1. Numerical values of the parameters in the model [29]

Parameter	S	Si	Cr	Fe	Ni	Mn	Al	O
$\tau_{SNR}(X)$, Gyr	0.40	0.40	0.50	0.50	0.50	0.50	0.40	0.40
$\tau_{acc}(X)$, Gyr	0.20	0.20	0.25	0.25	0.25	0.25	0.20	0.20
δ^{Ia}_{cond}	0.8	0.8	0.8	0.8	0.8	0.8	0.8	–*
δ^{II}_{cond}	0.8	0.8	0.8	0.8	0.8	0.8	0.8	–*

* Condensation of oxygen atoms into dust is calculated as $M_{dust}(O,m) = 16\Sigma_X \delta(X) M_{ej}(X,m)/\mu(X)$ [29].

supernovae depends very strongly on the distribution of the gas density in the galaxy (especially its scale height) and, consequently, on the distribution and total mass of the gravitating matter, the luminosity function of OB associations [49], and the presence and magnitude of magnetic fields [50]. Moreover, it is also important that dust grains (and, consequently, metals) can be ejected from the galaxy by stellar radiation pressure [45, 51] virtually without affecting the matter in the gaseous phase. The formulation of an adequate phenomenological model for the exchange terms in (5) and (6), $A(t)$ and $W(t)$, requires a separate analysis, and we restrict our present investigation to the single-zone model.

In their recent analysis of the effect of dust on elemental abundances in DLA systems, Calura *et al.* [52] scaled the degree of depletion of elements due to condensation onto dust grains to the dust content and total heavy-element abundance in the galaxy found in [53, 54]. An important aspect of this approach is that the dust content in the galaxy must be assumed to be known. This assumption is justified to some extent, but only within certain limits, which depend appreciably on the characteristic properties of star formation, the structure of the ISM, and mass exchange between the galaxy and its environment. Moreover, the time-independent scaling transformation itself assumes that the condensation of heavy elements into dust grains has reached a steady-state asymptotic regime. Although our calculations show that a steady-state element-depletion regime does become established in spiral galaxies, such an approach may be inappropriate when there is the possibility of short ($\sim 10^8$ yr) bursts of star formation in the galaxy with durations comparable to the time scales for heavy-element depletion.

2.4. Parameters of the Model

The chemical and photometric evolution of galaxies in the single-zone model is mostly determined by the SFR $\Psi(t)$ and the stellar IMF $\phi(m)$. The processes that determine these functions are not completely understood; however, *bona fide* empirical relations are used when describing the evolution of galaxies. Another important parameter of the model that influences the chemical evolution of galaxies is the fraction β of binaries that end as type Ia supernovae.

2.4.1. Star formation rate. Here, it is standard to assume the Schmidt law [55], which was first proposed in the form of the dependence of Ψ on the gas volume density:

$$\Psi = A^{-1}\rho_g^n, \tag{9}$$

where ρ_g is the gas density and n is the index of the power law (Schmidt adopted $n = 2$). More recently, the Schmidt law has usually been written as the functional dependence on the surface gas density Σ_g:

$$\Psi = A^{-1}\Sigma_g^n, \tag{10}$$

where the dimensions of these quantities are $[\Psi] = [M_\odot\ \mathrm{yr}^{-1}\ \mathrm{pc}^{-2}]$ and $[\Sigma_g] = [M_\odot\ \mathrm{pc}^{-2}]$. The index n is usually taken to be close to unity, although the n values show a fairly large scatter. Kennicutt [56] derived $n = 1.3 \pm 0.3$ based on observations of 15 spiral galaxies. Subsequent observations of a larger sample including 61 spiral galaxies spanning a wide range of gas densities and SFRs [57] led to the refined estimate $n = 1.4 \pm 0.14$. The most recent estimates of the parameters of the Schmidt law were reported for a sample of seven spiral galaxies by Wong and Blitz [58], who found that n can vary over the fairly wide range from 1.1 to 1.7. In view of the large uncertainty in n, we performed our modeling for $n = 1$, 1.5, and 2.

2.4.2. Initial mass function. The IMF is defined as the distribution of masses of stars at the time of

Table 2. Parameters of elements in the selective model

Element	Zn	S	Mn	Si	Cr	Fe	Ni	O
$\alpha(A)\sqrt{A}$	0.01	0.20	0.95	0.95	0.99	0.99	0.996	0.40

their birth and is modeled by the power-law relation

$$\frac{dN}{dm} = \varphi(m) = Am^{-\alpha}, \qquad (11)$$

where N is the number of stars born in the mass interval $[m, m + dm]$. The widely used Salpeter mass function was determined in [59] as a power-law function with $\alpha = 2.35$ for solar-neighborhood stars with masses in the interval 0.4–10 $M_\odot$. Subsequent studies showed that the slope α of the IMF decreases at small masses and increases at high masses compared to the Salpeter value [60–62]. Kroupa [63] analyzed the effect on the slope of the IMF of unresolved multiple stars and statistical errors due to the analysis of small samples of stars. He found that the slope of the IMF corrected for the above effects has the form

$$\begin{aligned} \alpha_1 = 1.80 \pm 0.5, \quad 0.08 \le m/M_\odot < 0.50, \qquad (12)\\ \alpha_2 = 2.70 \pm 0.3, \quad 0.50 \le m/M_\odot < 1.00,\\ \alpha_3 = 2.30 \pm 0.7, \quad 1.00 \le m/M_\odot. \end{aligned}$$

We report here calculations obtained using the piecewise-continuous IMF of Kroupa [63], which we extended to the stellar-mass interval $m/M_\odot < 120$. The slope of the IMF for massive stars is very uncertain, and we accordingly adopted the following modified piecewise-continuous approximation:

$$\begin{aligned} \alpha_1 = 1.80, \quad 0.08 \le m/M_\odot < 0.5, \qquad (13)\\ \alpha_2 = 2.70, \quad 0.5 \le m/M_\odot < 1,\\ \alpha_3 = 2.35, \quad 1 \le m/M_\odot < 3,\\ \alpha_4 = 2.60, \quad 3 \le m/M_\odot < 120. \end{aligned}$$

This choice was motivated primarily by the relatively small amount of observational data for Galactic field stars with masses exceeding the solar mass [64]. The lifetimes of massive stars are much shorter than the age of the Galaxy, and these stars can be observed primarily in young star clusters and OB associations. The IMF of massive stars in such stellar groups is unfortunately difficult to analyze due to the influence of numerous selection effects. In addition, the effect of stellar multiplicity on analyses of the IMF of massive stars has not been studied in detail, although it is known that massive stars can be members of binary and multiple systems [65, 66]. The corresponding corrections for unresolved binaries and multiple stars could lead to an IMF slope that is much higher than the Salpeter value for stars with masses $m > 3M_\odot$. The IMF of single stars with such masses could have a slope of $\alpha = 2.7$ or even higher. A discussion of this problem can be found, for example, in [63]. Our choice of the IMF slope for stars with $m > 3\ M_\odot$ is close to the value $\alpha = 2.7$ determined for the IMFs considered in [60, 62]. These IMFs were analyzed using many models for the evolution of the Galaxy [67, 68], and good agreement was achieved between the model calculations and observations.

2.4.3. Binary frequency. The parameter β determines the fraction of binaries in the mass interval $M_b^{\min} < M_b < M_b^{\max}$ whose evolution ends in type Ia supernovae. This parameter was first determined by Matteucci and Greggio [25], who analyzed a single-zone model for the chemical evolution in the solar neighborhood. They derived the value $\beta = 0.1$ by requiring that the calculated type Ia and II supernova rates agree with the observed values. Although β is a free parameter, it is nearly always fixed at $\beta = 0.1$ in chemical-evolution models, by analogy with [25]. We dropped this simplifying assumption and derived β by requiring agreement between the calculated and observed type Ia and II supernova rates for various types of spiral galaxies [69]. The calculated results agreed well with observations for $\beta = 0.05$–0.1, depending on the type of the spiral galaxy. Note that a more complex multizone model for the chemical evolution of the Milky Way agrees well with observations for the same values, $\beta = 0.05$–0.09 [70], depending on the choice of the IMF.

3. OBSERVATIONAL CONSTRAINTS OF THE MODEL

Constructing an evolutionary track for a galaxy based on its chemical characteristics is a typical inverse problem. It is important to bear in mind that all the structural parameters of and dynamical processes in a galaxy can affect the total abundances of metals and the relative abundances of various elements to varying degrees. This means that modeling the chemical and photometric evolution of a galaxy is a multiparameter inverse problem. Under these conditions, to reduce the time required to find the domain of reasonable parameter values and arbitrariness in the choice of parameters, we must identify the observational parameters of galaxies that are most sensitive to significant model parameters, thereby making it possible to constrain the domain of possible parameter values. To this end, we use four reliably measurable observational parameters of spiral galaxies: the oxygen abundance in HII regions, supernova rate, color parameters, and the gas fraction in the galaxy. When modeling the depletion of metals due to condensation onto dust grains, we used the parameters in system (6) that yielded the mean observed distribution of elemental depletions for our Galaxy.

3.1. Oxygen Abundance in HII Regions of Spiral Galaxies

To quantify the radial gradient of the heavy-element abundance [71] in the single-zone model,

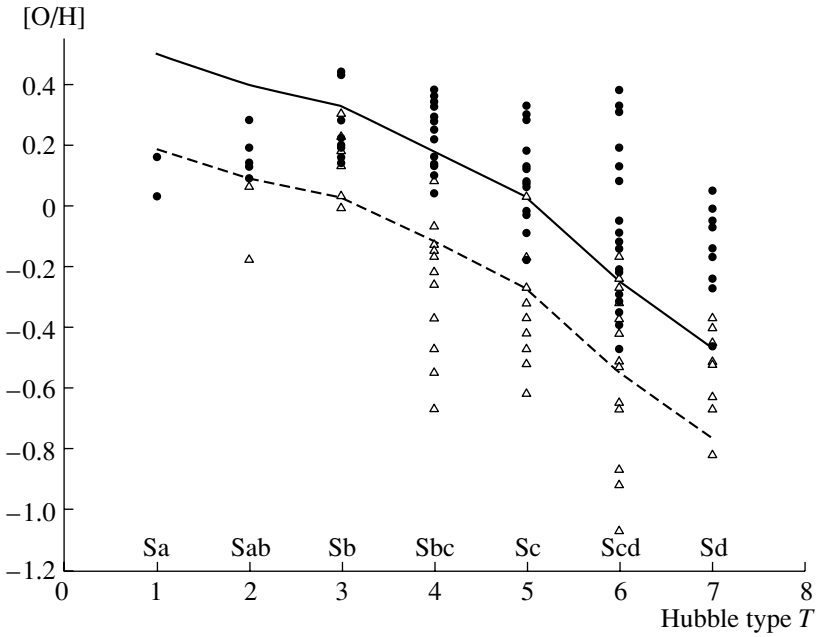

Fig. 1. Observed oxygen abundance in HII regions of spiral galaxies in the gaseous phase at two characteristic radii (circles and triangles) compared to the calculated oxygen abundance in the gaseous phase (lower curve) and the total oxygen abundance in the gaseous and dust phase (upper curve).

we chose as observational constraints for our model the oxygen abundance in the gaseous phase at two characteristic galactocentric distances, $0.4\rho_0$ and ρ_0, where ρ_0 is the distance from the galactic center along the major axis of the galaxy where the B surface brightness reaches $\sigma_B = 25.0^m/(\mathrm{arcsec})^2$.

The gaseous-phase oxygen abundances in the HII regions of spiral galaxies have been obtained from numerous observations [72–77]. Figure 1 summarizes this information for 60 spiral galaxies of various morphological types ranging from Sa to Sd. We can see that spiral galaxies of the same morphological type show a large dispersion of observed oxygen abundances at the distances $0.4\rho_0$ and ρ_0. This could be due to both physical factors (possible azimuthal variations of the oxygen abundance [78], the effect of local conditions and large-scale gas motions) and errors in the methods used ($\sim$0.2dex). It is also important that determinations of the heavy-element abundances in HII regions are affected by strong observational selection effects that favor metal-poor and much brighter HII regions in late-type spiral galaxies [79].

Note that we are concerned with the oxygen abundances in the *gaseous* phase. A detailed analysis of the effect of the condensed (dust) phase on derived metallicities showed that the corresponding error in the oxygen abundances is about $\sim$0.2dex [80, 81], and, in any case, the oxygen abundances in the most metal-rich HII regions of Sa–Sb galaxies cannot be lower than twice the solar value [72]. It follows that the gaseous-phase oxygen abundances of Sa–Sb galaxies in Fig. 1 are probably severely underestimated [79]. The higher metallicities of these systems compared to late-type galaxies may also imply a higher dust content, in which case the decrease in the gaseous-phase oxygen abundance observed in these galaxies could indicate that a considerable amount of oxygen is in the form of dust [81].

3.2. Supernova Rates in Spiral Galaxies

The number of supernovae is difficult to determine because of numerous selection effects, such as internal extinction in the galaxy, the galaxy's inclination, and difficulties in identifying supernovae in central regions of the galaxy. The uncertainty in the estimated supernova rate can therefore be high, from 30 to 50%. We chose the type Ia and II supernova rates as our observational parameter, because, combined with the constraining parameters noted above, they can be used to substantially reduce the interval of possible values of the free parameters of the model. Available data on the evolution of supernovae of other types (Ib/c, IIb, etc.) and their role in heavy-element enrichment of the ISM are not yet sufficient for analyses of the effect of these stars on galactic evolution.

The "supernova rate" is measured in SNu, where $1\mathrm{SNu} = 1\mathrm{SN}(100\,\mathrm{yr})^{-1}(10^{10}L^B_\odot)^{-1}$, or one supernova per century per luminosity $10^{10}L^B_\odot$ in the B

Table 3. "Supernova rate" (in units of SNu/h^2) in spiral galaxies [68]*

Type of galaxy	SNIa	SNIb/c	SNII	All supernovae
S0a–Sb	0.18 ± 0.07	0.11 ± 0.06	0.42 ± 0.19	0.72 ± 0.21
Sbc–Sd	0.21 ± 0.08	0.14 ± 0.07	0.86 ± 0.35	1.21 ± 0.37

* The supernova rates were calculated for a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1}.

band. The normalization to the luminosity $L_\odot^B$ is usual and is convenient because the integrated B luminosity has been determined for a large number of galaxies and is correlated with the total mass of the galaxy. The most complete compilation of observed supernova rates is presented by Cappellaro [69]. Two patterns are immediately apparent from the data in Table 3, which we adopted from [69]: (i) the type Ia supernova rate is essentially the same for spiral galaxies of all types and (ii) the type II supernova rate increases appreciably toward late-type galaxies. The presented data correspond to a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1} and are corrected for inclination and extinction. The supernova rates listed in this table can easily be rescaled for other H_0 values by multiplying them by $h^2 = (H_0/75)^2$.

3.3. Integrated Color Indices

It is well known that the integrated color indices of spiral galaxies decrease along the Hubble sequence from early- to late-type galaxies (Fig. 2). Already in their early studies of galactic evolution, Tinsley [82] and Larson and Tinsley [83] interpreted this correlation between the color indices on two-color diagrams, first, as a tendency for a monotonic decrease in the SFR with age for all spiral galaxies and, second, as evidence for more active star formation in early-type galaxies in the past. The large color dispersion within a given morphological type can be explained by the fact that the mean SFR varies strongly from galaxy to galaxy [84].

Buta *et al.* [85] used the data of the RC3 catalog [86] to calculate the integrated $U-B$ and $B-V$ colors for a large number of spiral galaxies. Buta and Williams [87] then used the same catalog to estimate the $V-R$ and $V-I$ color indices corrected for internal extinction and inclination. The color indices reported in [85, 87] represent a homogeneous sample, since they were calculated using the same technique and the same catalog data. We use these integrated $U-B$, $B-V$, $V-R$, and $V-I$ color indices (shown in Fig. 2) as one of the main observational constraints for our model.

3.4. Gas Content in Spiral Galaxies

McGaugh and de Blok [88] developed a technique for estimating the gas content, $f_g = M_g/(M_g + M_*)$, and applied it to a large number of spiral galaxies. The correlation between f_g and the $B-V$ and $V-I$ color indices and a comparison of these results with two-color diagrams (Fig. 2) showed that the gas contents in spiral galaxies vary over a wide range, on average, from $f_g = 0.1$ for Sa to $f_g = 0.7$ for Sd galaxies.

4. MODELING RESULTS

Our model agrees well with the observational data for spiral galaxies described above if we adopt the IMF (13), the binary frequency $\beta = 0.05$–0.1, and the set of parameters listed in Table 4, which characterizes the SFR $\Psi(t)$.

Figure 1 shows the steady-state oxygen abundance plotted as a function of galaxy type. It is clear that the calculated and observed distributions agree fairly well for all galaxy types except early-type Sa and Sb galaxies, for which the calculated abundances are somewhat overestimated. However, this discrepancy between the calculated and observed abundances is not significant, given the uncertainties and possible underestimation of the oxygen abundances in HII regions in Sa and Sb galaxies noted above.

Our calculated extinction-free galaxy colors also agree well (within the errors) with the observed colors. The points connected by line segments in Fig. 2 show the colors calculated with allowance for reddening due to the dust formed by the present epoch, $z = 0$. We can see that, in essentially all cases, the calculated colors lie inside the domains constraining the observed colors. The calculated colors for Sa galaxies are close to the uncertainty region for the observed colors.

The distribution of metal depletion due to condensation onto dust grains is in good agreement with the observed distribution in the Milky Way for the elemental dependence $\alpha(X)$ listed in Table 2. The data on the mean depletion in the Milky Way were taken from [89] and are plotted in Fig. 3. In our scheme, the distribution of elements in the gaseous phase is determined by the rate of their accretion

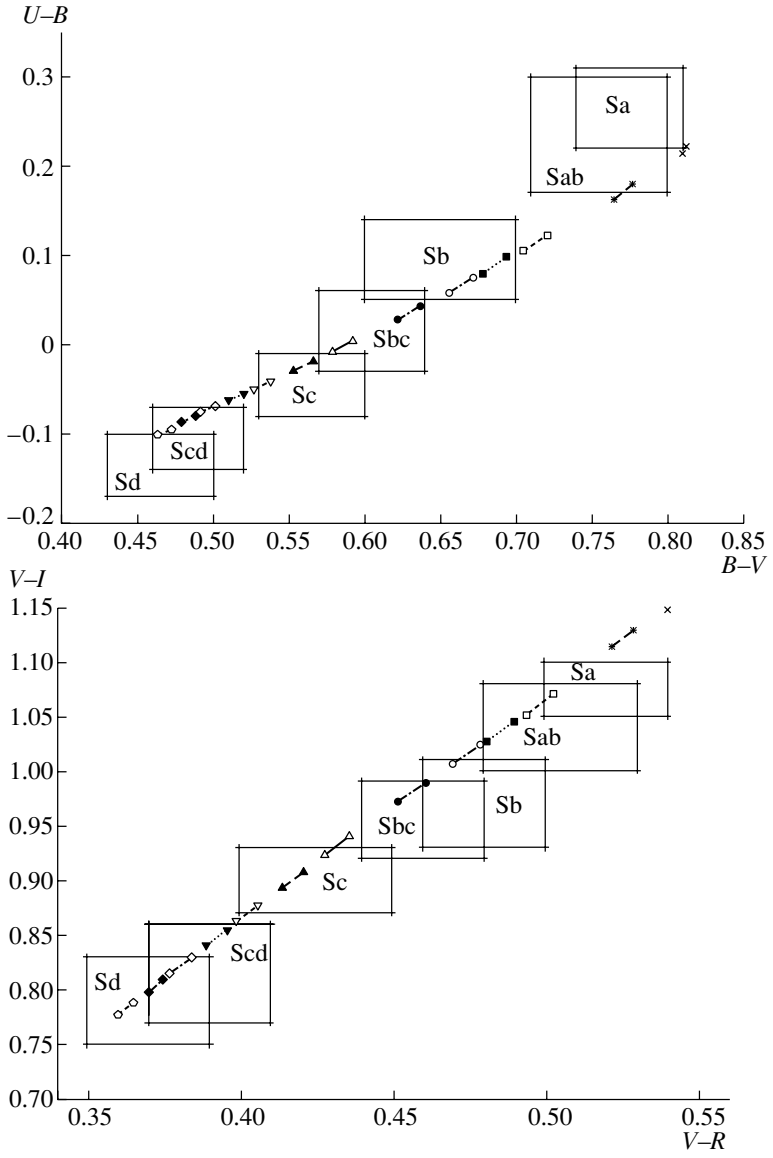

Fig. 2. Observed (a) $(B-V)-(U-B)$ and (b) $(V-R)-(V-I)$ two-color diagrams for spiral galaxies shown with the color indices calculated in the model with the parameters listed in Table 4.

onto the surfaces of dust grains in clouds. Note that accretion onto dust grains is also the determining factor in the model of Dwek [29], with the difference that the accretion rates for all elements are assumed to be equal to the rates of return via the destruction of dust by supernova-produced shocks; as a result, the steady-state depletions of all elements are equal to about half their total abundances.

4.1. Overall Distribution of Elements

Figure 4 shows the dependences of the distribution of several elements (Fe, Si, Zn) in spiral galaxies ranging from type Sa (solid curves) to Sd (dashed curves) as a function of redshift, without (thin curves) and with (thick curves) allowance for the depletion of elements due to condensation onto dust grains. On the whole, the observed metallicities in DLA systems lie within the band of metallicities for spiral galaxies calculated with allowance for dust formation. When

Table 4. Parameters characterizing the SFR for a Schmidt law $\Psi = A^{-1}\rho_g^n$, where ρ_g is the volume density of gas and $n = 1$ is the power-law index for 12 scenarios of spiral galaxies. $A[(M_\odot/10^9\,\mathrm{yr})^{-1}]$ is normalized to the total mass of model galaxy equal to $1M_\odot$.

N	1	2	3	4	5	6	7	8	9	10	11	12
A	1	2	3	4	5	7	9	12	15	20	25	35

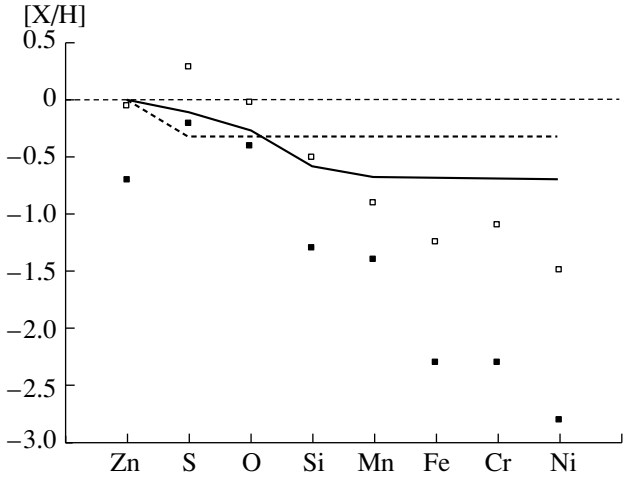

Fig. 3. Calculated distributions of the degree of depletion of metals for the model of Dwek [29] (polygonal dashed line) and our model with a selective condensation of metals (solid line) compared to the mean observed depletion of metals in the warm (open squares) and cold (filled squares) ISM. The thin dashed horizontal line indicates the solar metallicity.

dust formation and the associated depletion of metals in the gaseous phase are disregarded, the observed abundances are substantially lower than those predicted for spiral galaxies. At the same time, an excess metallicity is fairly evident in early-type galaxies (Sa–Sb), even for models with a depletion of metals in the gaseous phase. This may suggest that DLA systems are represented by the later types for some reason. In this connection, we note an interesting feature pointed out by Pettini *et al.* [90]: when measured using the abundance of zinc, which does not condense into a solid phase, the evolution of DLA systems is slower than the evolution of Population II objects in the Milky Way, resulting in a clear excess at $z \sim 2$ of metals in Population II compared to DLA systems. This may also provide evidence that early-type spirals are not represented among DLA systems at redshifts $z > 2$. However, no firm conclusions can be drawn without more complete modeling of the chemical evolution of spiral and other galaxies, including a sufficiently complete treatment of all the processes noted above.

4.2. Relative Abundances

Modern high-resolution ($R \approx 30\,000$) and high signal-to-noise ($S/N \approx 30$) observations [91] enable very accurate estimation of the relative abundances of metals in DLA systems (with errors of $\sim 10\%$). This makes it possible to separate fairly confidently the effects of nucleosynthesis and condensation on dust grains on the overall abundance patterns, making it possible, in turn, to draw certain conclusions about the history of star formation and the possible evolution of the IMF. Figure 5 compares the observational data of [92] with our calculated relative abundances [Fe/Zn] as a function of [Zn/H] and [Si/Fe] as a function of [Si/H].

The theoretical dependences are determined by the ratios of the coefficients for condensation onto dust grains for various elements, which are reflected in the pattern of the depletion of gaseous-phase elements shown in Fig. 3. This is due to the fact that the time scale for the accretion of metals onto dust grains determined by (6) depends only on the properties of the grains, and is equal to $\sim 10^8$ yr, independent of the metallicity of the gas. It follows that the distribution of the depletion of gaseous-phase metals shown in Fig. 3 is established quickly and depends only weakly on the total metallicity. We can see that allowance for selective condensation of iron on dust grains provides a qualitatively correct description of the observed "sag" in the dependence of the relative abundance [Fe/Zn] near $[\mathrm{Zn/H}] \simeq -0.75$, although the quantitative discrepancy is appreciable and reaches 0.3–0.5 dex in various [Zn/H] intervals. The dependence of [Fe/Zn] on [Zn/H] calculated without condensation onto dust grains falls within the interval of observed values, but its overall behavior—a monotonic increase with [Zn/H]—differs from the observed dependence.

The agreement between the calculated and observed dependences of [Si/Fe] on [Si/H] is appreciably poorer, both quantitatively and qualitatively, independent of whether the model includes condensation on dust grains. The observed relative abundances [Si/Fe] increase monotonically with metallicity; e.g., [Si/Fe] varies from ~ 0.2 to ~ 0.7 as [Si/H] varies from ~ -2.3 to ~ -0.5 [91]. The lower limit of [Si/Fe] is believed to correspond to the situation when nucleosynthesis is dominated by the contribution of type Ia supernovae, while the upper limit corresponds to the depletion of iron in the gaseous phase as a result of condensation onto dust grains. Our results suggest that this interpretation is not correct—on the whole, the calculated dependences decrease monotonically. However, inclusion of the effects of selective blowout of dust grains from galaxies by radiation pressure makes it possible to explain the observed increase of [Si/Fe] with metallicity as being due to the depletion of iron due to condensation onto dust grains, since the amount of Si in the gaseous phase then increases with the total metallicity, whereas the amount of iron decreases steadily as it is removed from the galaxy. Wiebe *et al.* [93] analyzed this mechanism in connection with the enrichment of gas in clusters of galaxies.

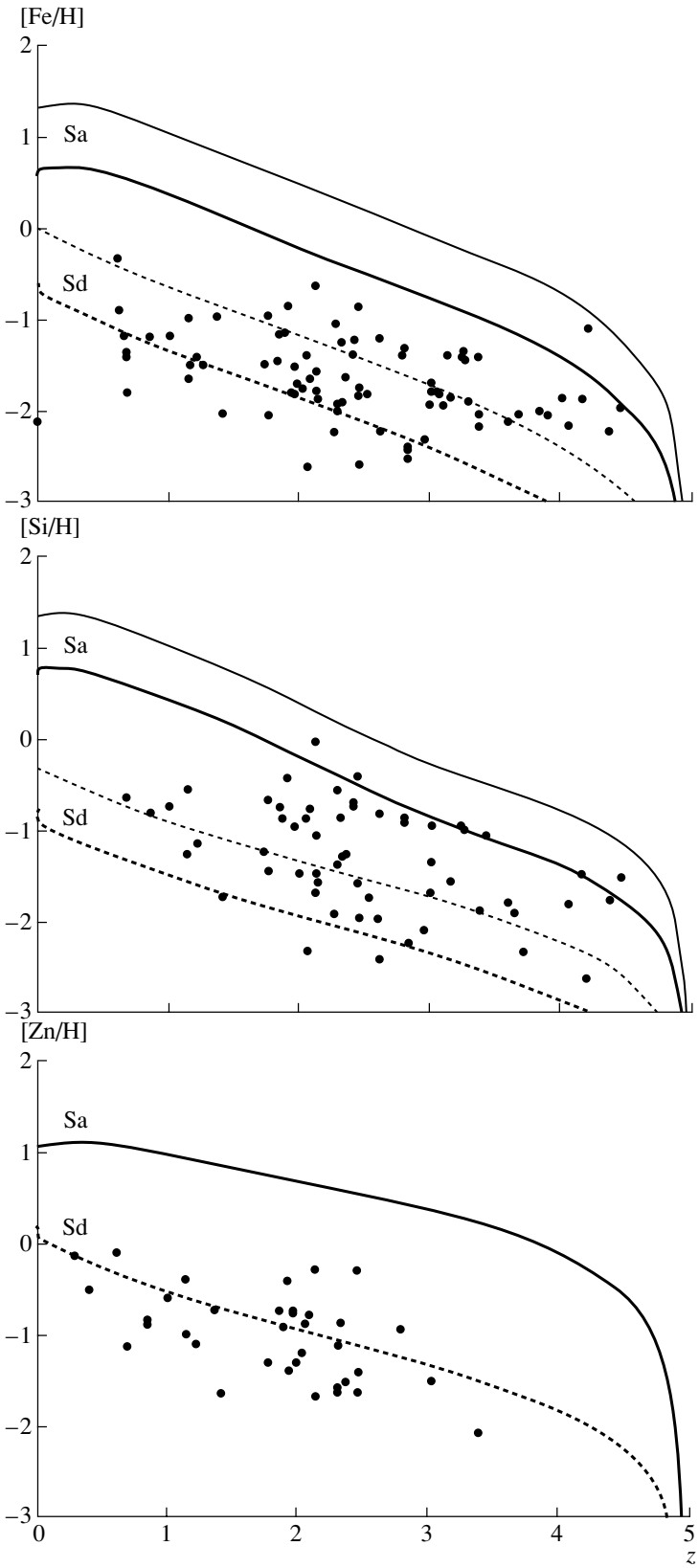

Fig. 4. Calculated distributions of the Fe, Si, and Zn abundances in Sa (solid) and Sd (dashed) spiral galaxies as a function of redshift. The thin and thick lines show the distributions calculated without and with allowance for the depletion of metals due to condensation onto dust grains, respectively.

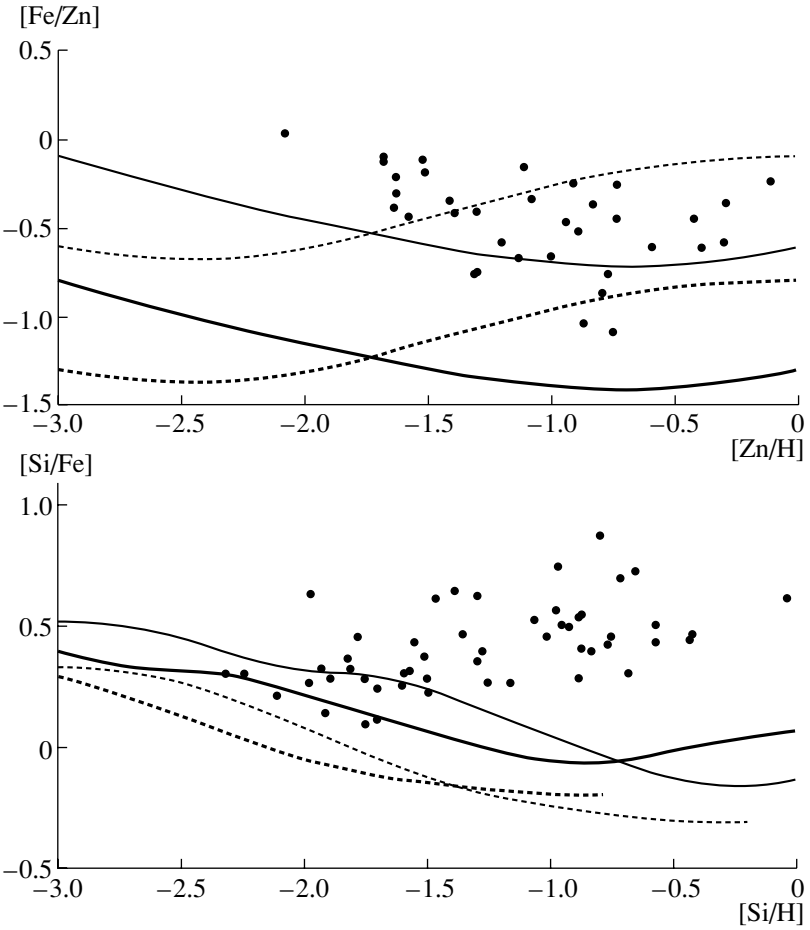

Fig. 5. Comparison of the calculated relative abundances [Fe/Zn]–[Zn/H] and [Si/Fe]–[Si/H]. Same notation as in Fig. 4.

We will perform a detailed analysis of this possibility for DLA systems in a separate paper.

In this connection, we should also mention the so-called problem of "hidden" or "invisible" metals pointed out by Pettini [94], which is described and analyzed in [95]. The essence of this problem is that the total amount of metals produced in the Universe by $z \simeq 2.5$ for an SFR of ~ 0.2–$0.4\ M_\odot\ \mathrm{Mpc}^{-3}\ \mathrm{yr}^{-1}$, as is indicated by optical, ultraviolet, and infrared observations [95], is about an order of magnitude higher than is observed in DLA systems [94]. Such an amount of metals could, in principle, be scattered in the intergalactic medium [46, 47], but the production of these metals must be more efficient than is observed in DLA systems. The resolution of this problem may involve an excess production of metals (compared to the solar amount) in early-type spiral galaxies and the selective blowing out of metals condensed into the solid phase by radiation pressure. We will return to this issue in a future paper.

5. CONCLUSIONS

(1) We have constructed a model for the chemical and photometric evolution of galaxies that includes the formation of dust, depletion of heavy elements due to condensation onto the surface of dust grains, and variation of the photometric parameters due to extinction.

(2) We chose the model parameters to achieve agreement of the results with the observed oxygen abundances in HII regions in spiral galaxies, as well as the color parameters, rates of various types of supernovae, and the fraction of remaining gas of the galaxies.

(3) In our closed, single-zone model, the mass fraction of metals condensed into the solid phase (dust) is about 50% when the calculations use the condensation scheme described by Dwek and Scalo [28]. When allowance is made for the differential (element-dependent) condensation of metals onto dust grains in clouds with the accretion coefficients in Table 2 (which reproduce the mean behavior of the metal depletion in the local ISM), the calculated

gaseous-phase abundances are close to those observed in DLA systems, especially if the latter are represented by late-type spirals.

(4) The relative abundances of iron and zinc (which condenses only weakly onto dust grains) in the gaseous phase reproduces the qualitative behavior of the observed dependence, leading us to conclude that the relative overabundance of zinc in DLA systems is due, to a considerable extent, to the presence of dust in these objects. However, the observed substantial variations of both the absolute abundances and their ratios for different elements could also be due to characteristic features of the evolution of the protogalaxies making up the DLA system. The recent observation of an unusual distribution of metal depletion in the Small Magellanic Cloud [96], with both silicon and zinc being very strongly depleted, is noteworthy in this connection.

(5) The relative abundances [Si/Fe] calculated in our closed model are lower than the observed abundances and do not increase monotonically with the metallicity. This suggests that the observed monotonic increase of [Si/Fe] as a function of [Si/H] cannot be unambiguously attributed to the condensation of Fe onto dust. One possible solution to this discrepancy is that we must consider open models with the selective blowout of dust by radiation pressure.

The results reported here demonstrate that allowance for the presence of dust in damped Lyα systems is clearly important for the correct interpretation of the heavy-element abundances observed in these systems.

ACKNOWLEDGMENTS

We are grateful to L.S. Pilyugin, A.V. Tutukov, and B.M. Shustov for discussing the results and critical comments. This work was supported by the Russian Foundation for Basic Research (project code 00-02-17689).

REFERENCES

1. J. X. Prochaska and A. M. Wolfe, Astrophys. J. **507**, 113 (1998).
2. R. Jimenez, D. V. Bowen, and F. Matteucci, Astrophys. J. **514**, 83 (1999).
3. U. Lindner, U. Fritze-von Alvensleben, and K. J. Fricke, Astron. Astrophys. **341**, 709 (1999).
4. J. L. Hou, S. Boissier, and N. Prantzos, Astron. Astrophys. **370**, 23 (2001).
5. V. Le Brun and J. Bergeron, Astron. Astrophys. **332**, 814 (1998).
6. B. T. Draine, Astrophys. Space Sci. **233**, 111 (1995).
7. M. Pettini and D. V. Bowen, Astron. Astrophys. **327**, 22 (1997).
8. Y. C. Pei and S. M. Fall, Astrophys. J. **454**, 69 (1995).
9. M. Fioc and B. Rocca-Volmerange, Astron. Astrophys. **326**, 950 (1997).
10. B. Rocca-Volmerange and M. Fioc, Astrophys. Space Sci. **269**, 233 (1999).
11. A. Bressan, F. Fagotto, G. Bertelli, and C. Chiosi, Astron. Astrophys., Suppl. Ser. **100**, 647 (1993).
12. F. Fagotto, A. Bressan, G. Bertelli, and C. Chiosi, Astron. Astrophys., Suppl. Ser. **105**, 29 (1994).
13. F. Fagotto, A. Bressan, G. Bertelli, and C. Chiosi, Astron. Astrophys., Suppl. Ser. **105**, 39 (1994).
14. F. Fagotto, A. Bressan, G. Bertelli, and C. Chiosi, Astron. Astrophys., Suppl. Ser. **104**, 365 (1994).
15. L. Girardi, A. Bressan, C. Chiosi, G. Bertelli, and E. Nasi, Astron. Astrophys., Suppl. Ser. **117**, 113 (1996).
16. G. Bertelli, A. Bressan, C. Chiosi, F. Fagotto, and E. Nasi, Astron. Astrophys., Suppl. Ser. **106**, 275 (1994).
17. T. Lejeune, F. Cuisinier, and R. Buser, Astron. Astrophys., Suppl. Ser. **125**, 229 (1997).
18. T. Lejeune, F. Cuisinier, and R. Buser, Astron. Astrophys., Suppl. Ser. **130**, 65 (1998).
19. R. Kurucz, *IAU Symposium 149: The Stellar Populations of Galaxies* (Kluwer, Boston, 1992), p. 225.
20. R. Kurucz and B. Bell, *Atomic Line Data*, Kurucz CD-ROM, No. 23 (Mass. Smithsonian Astrophysical Observatory, Cambridge, 1995).
21. B. T. Draine and H. M. Lee, Astrophys. J. **285**, 89 (1984).
22. A. Laor and B. T. Draine, Astrophys. J. **402**, 441 (1993).
23. G. A. Bruzual, *Proc. XI Canary Islands Winter School of Astrophysics on Galaxies at High Redshift*; astro-ph/0011094.
24. C. Chiappini, F. Matteucci, and R. Gratton, Astrophys. J. **477**, 765 (1997).
25. F. Matteucci and L. Greggio, Astron. Astrophys. **154**, 279 (1986).
26. S. E. Woosley and T. A. Weaver, Astrophys. J., Suppl. Ser. **101**, 181 (1995).
27. K. Iwamoto *et al.*, Astrophys. J., Suppl. Ser. **125**, 439 (1999).
28. E. Dwek and J. M. Scalo, Astrophys. J. **239**, 193 (1980).
29. E. Dwek, Astrophys. J. **501**, 643 (1998).
30. C. F. McKee, in *IAU Symposium 135: Interstellar Dust* (Kluwer, Boston, 1989), p. 431.
31. I. G. Kolesnik, Astron. Zh. **55**, 191 (1978) [Sov. Astron. **22**, 109 (1978)].
32. B. V. Vainer, Astron. Zh. **67**, 209 (1990) [Sov. Astron. **34**, 107 (1990)].
33. S. A. Levshakov, M. Dessauges-Zavadsky, S. Dódorico, and P. Molaro, Astrophys. J. **565**, 696 (2002).
34. W. M. Lane and F. H. Briggs, Astrophys. J. **561**, 27 (2001).
35. N. Kanekar, T. Ghosh, and J. N. Chengalur, Astron. Astrophys. **373**, 394 (2001).
36. J. Ge and J. Bechtold, Astrophys. J. **477**, 73 (1997).
37. J. N. Chengalur and N. Kanekar, Astron. Astrophys. **388**, 383 (2002).

38. H. Liszt, Astron. Astrophys. **389**, 393 (2002).
39. M. Pettini, S. L. Ellison, C. C. Steidel, and D. V. Bowen, Astrophys. J. **510**, 576 (1999).
40. P. Molaro, S. A. Levshakov, S. Dódorico, P. Bonifacio, and M. Centurion, Astrophys. J. **549**, 90 (2001).
41. C. F. McKee, in *The Evolution of the Interstellar Medium* (Astron. Soc. Pac., San Francisco, 1990), p. 3.
42. A. Songaila, Astrophys. J. **561**, 153 (2001).
43. A. Ferrara, M. Pettini, and Y. A. Shchekinov, Mon. Not. R. Astron. Soc. **319**, 539 (2000).
44. Y. A. Shchekinov, Astron. Astrophys. Trans. **21**, 131 (2003).
45. B. M. Shustov and D. Z. Wiebe, Astron. Zh. **72**, 650 (1995) [Astron. Rep. **39**, 578 (1995)].
46. B. B. Nath, S. K. Sethi, and Y. Shchekinov, Mon. Not. R. Astron. Soc. **303**, 1 (1999).
47. A. Ferrara, B. B. Nath, S. K. Sethi, and Y. A. Shchekinov, Mon. Not. R. Astron. Soc. **303**, 301 (1999).
48. A. Aguirre and Z. Haiman, Astrophys. J. **532**, 28 (2000).
49. Y. A. Shchekinov, Astron. Astrophys. **314**, 927 (1996).
50. A. Steinacker and Y. A. Shchekinov, Mon. Not. R. Astron. Soc. **325**, 208 (2001).
51. A. Ferrara, F. Ferrini, J. Franco, and B. Barsella, Astrophys. J. **381**, 137 (1991).
52. F. Calura, F. Matteucci, and G. Vladilo, Mon. Not. R. Astron. Soc. **340**, 59 (2003).
53. G. Vladilo, Astron. Astrophys. **391**, 407 (2002).
54. G. Vladilo, Astrophys. J. **569**, 295 (2002).
55. M. Schmidt, Astrophys. J. **129**, 243 (1959).
56. R. C. Kennicutt, Jr., Astrophys. J. **344**, 685 (1989).
57. R. C. Kennicutt, Jr., Astrophys. J. **498**, 541 (1998).
58. T. Wong and L. Blitz, Astrophys. J. **569**, 157 (2002).
59. E. E. Salpeter, Astrophys. J. **121**, 161 (1955).
60. P. Kroupa, C. A. Tout, and G. Gilmore, Mon. Not. R. Astron. Soc. **262**, 545 (1993).
61. J. M. Scalo, Fundam. Cosm. Phys. **11**, 1 (1986).
62. J. M. Scalo, in *The Stellar Initial Mass Function*, Ed. by G. Gilmor and D. Howell, ASP Conf. Ser. **142**, 201 (Astron. Soc. Pac., San Francisco, 1998).
63. P. Kroupa, Mon. Not. R. Astron. Soc. **322**, 231 (2001).
64. P. Kroupa, in *IAU Symposium 200: The Formation of Binary Stars*, Ed. by H. Zinnecker and R. D. Mathieu (Potsdam, 2001), p. 199.
65. B. D. Mason, D. R. Gies, W. I. Hartkopf, *et al.*, Astron. J. **115**, 821 (1998).
66. G. Duchene, T. Simon, J. Eislöffel, and J. Bouvier, Astron. Astrophys. **379**, 147 (2001).
67. C. S. Möller, U. Fritzev. Alvensleben, and K. J. Fricke, Astron. Astrophys. **317**, 676 (1997).
68. S. Boissier and N. Prantzos, Mon. Not. R. Astron. Soc. **312**, 398 (2000).
69. E. Cappellaro, R. Evans, and M. Turatto, Astron. Astrophys. **351**, 459 (1999).
70. F. Matteucci, in *Proc. XII Canary Islands Winter School of Astrophysics*; astro-ph/0203340 (2002).
71. R. B. C. Henry and G. Worthey, Publ. Astron. Soc. Pac. **111**, 919 (1999).
72. M. S. Oey and R. C. Kennicutt, Jr., Astrophys. J. **411**, 137 (1993).
73. D. Zaritsky, R. C. Kennicutt, Jr., and J. P. Huchra, Astrophys. J. **420**, 87 (1994).
74. S. D. Ryder, Astrophys. J. **444**, 610 (1995).
75. A. M. N. Ferguson, J. S. Gallagher, and R. F. G. Wyse, Astron. J. **116**, 673 (1998).
76. L. van Zee, J. J. Salzer, M. P. Haynes, *et al.*, Astron. J. **116**, 2805 (1998).
77. L. S. Pilyugin, M. Mollá, F. Ferrini, and J. M. Vílchez, Astron. Astrophys. **383**, 14 (2002).
78. R. C. Kennicutt, Jr. and D. R. Garnett, Astrophys. J. **456**, 504 (1996).
79. R. C. Kennicutt, Jr., M. S. Oey, D. Zaritsky, and J. P. Huchra, Rev. Mex. Astron. Astrofis. **27**, 21 (1993).
80. R. B. C. Henry, Mon. Not. R. Astron. Soc. **261**, 306 (1993).
81. J. C. Shields and R. C. Kennicutt, Jr., Astrophys. J. **454**, 807 (1995).
82. B. Tinsley, Astron. Astrophys. **20**, 383 (1972).
83. R. Larson and B. Tinsley, Astrophys. J. **219**, 46 (1978).
84. R. C. Kennicutt, Jr., Astrophys. J. **498**, 541 (1998).
85. R. Buta, S. Mitra, de G. Vaucouleurs, and H. G. Corwin, Jr., Astron. J. **107**, 118 (1994).
86. G. de Vaucouleurs, A. de Vaucoulerurs, H. Corwin, *et al.*, *RC3*, *Third Reference Catalogue of Bright Galaxies* (Springer, New York, 1991).
87. R. Buta and K. L. Williams, Astron. J. **109**, 543 (1995).
88. S. S. McGaugh and W. J. G. de Blok, Astrophys. J. **481**, 689 (1997).
89. B. D. Savage and K. R. Sembach, Ann. Rev. Astron. Astrophys. **34**, 279 (1996).
90. W. Pettini, R. W. Hunstead, D. L. King, and L. J. Smith, in *QSO Absorption Lines*, Ed. by G. Meylan (1995), p. 71.
91. J. X. Prochaska, in *Hubble's Science Legacy: Future Optical-Ultraviolet Astronomy from Space,* ASP Conf. Ser. **291**, 164 (Astron. Soc. Pac., San Francisco, 2003).
92. J. X. Prochaska and A. M. Wolfe, Astrophys. J. **566**, 68 (2002).
93. D. S. Wiebe, B. M. Shustov, and A. V. Tutukov, Astron. Astrophys. **345**, 93 (1999).
94. M. Pettini, in *Chemical Evolution from Zero to High Redshift*, Ed. by J. R. Walsh and M. R. Rosa (Springer, Berlin, 1999), p. 233.
95. L. Dunne, S. A. Eals, and M. G. Edmunds, Mon. Not. R. Astron. Soc. **341**, 341 (2003).
96. D. E. Welty, J. T. Lauroesch, J. C. Blades, L. M. Hobbs, and D. G. York, Astrophys. J. **554**, 75 (2001).

Translated by A. Dambis

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 367–373. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 403–410.
Original Russian Text Copyright © 2004 by Moiseev, Tutukov, Wiebe, Shustov.

A Possible Origin of the High Metallicity in the Nuclei of Massive Galaxies and Quasars

A. V. Moiseev[1], A. V. Tutukov[2], D. S. Wiebe[2], and B. M. Shustov[2]

[1] *St. Petersburg State University, St. Petersburg, Russia*

[2] *Institute of Astronomy, Moscow, Russia*

Received November 10, 2003

Abstract—According to current observations, the relative abundance of gas-phase metals in distant quasars with ages of only $\sim 10^9$ yr ($z \sim 5$) can be appreciably higher than the solar abundance. We show that there are two main ways to explain the high metallicity of these galactic nuclei: a high gas density in the central regions, or an increase in the minimum masses of forming stars to several solar masses. The results of numerical modeling confirm this conclusion. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quasars, while being extremely interesting objects in their own right, also provide unique opportunities to investigate the physical and chemical states of the matter located along the line of sight between them and the observer. In particular, spectroscopic studies of quasars and active galactic nuclei can be used to determine the chemical composition of the gas in the immediate vicinity of these objects. One peculiarity of this gas is that it appears to have an appreciably enhanced abundance of heavy elements Z relative to the solar value ($Z_\odot \approx 0.02$). The ratio $Z/Z_\odot$ can reach three to five for distant quasars with ages of only $\sim 10^9$ yr [1, 2]. For a considerable fraction of the quasar's lifetime, corresponding to the redshift range $z \sim 2$–5, the heavy-element abundance remains nearly unchanged [3]. This means that most of the metals were produced at epochs earlier than $\sim 7 \times 10^8$ yr, which corresponds to the ages of the most distant quasars ($z \sim 5$–6).

Several hypotheses have been put forth to explain the origin of these high metallicities, both within the framework of conventional stellar nucleosynthesis and in models with a "nonstellar" production of heavy elements in the accretion disk around the central black hole [1]. The former scenario seems more attractive, since it requires no additional assumptions. However, even in this case, it is necessary to explain the very rapid and efficient enrichment of the circumstellar gas in the vicinity of the nuclei of massive galaxies with the products of supernova explosions.

Models of stellar evolution predict the mass of heavy elements synthesized in stars of various masses, $M_Z(M)$, and the mass of unprocessed material lost by these stars during their lifetime, $M_{\rm ej}(M)$. To illustrate the roles of stars of various masses in determining the chemical compositions of galaxies, we show in Fig. 1 the values of $M_Z(M)$ and $M_{\rm ej}(M)$ multiplied by a Salpeter initial mass function (IMF), $\phi(M) \propto M^{-2.35}$. The functions in Fig. 1 are normalized so that their integrals specify the fractions of ejected matter and synthesized heavy elements per unit mass of gas transformed into stars. It is evident that the lion's share of heavy elements are produced by stars with masses 20–30 $M_\odot$ and lifetimes $\sim 3 \times 10^6$–5×10^6 yr. However, matter is returned to the interstellar medium (ISM) mainly by low-mass stars with $M < 8\ M_\odot$. Their lifetimes are longer, and, for stars with masses of about one solar mass, are comparable to the lifetime of the Galaxy ($\sim 10^{10}$ yr). Immediately after the initial burst of star formation, there follows an epoch when the ISM in a young galaxy is enriched in the products of nucleosynthesis, but is not yet "diluted" by matter with low Z that has been ejected by low-mass stars. The length of this stage of "supermetallicity" and the maximum value of Z depend primarily on the height and duration of the star-formation peak: for sharper peaks, more heavy elements are ejected into the ISM before low-mass stars begin to return matter with low Z. In addition, the efficiency of the metal enrichment depends on the fraction of massive stars that produce heavy elements. It is currently believed that stars with masses greater than the solar mass dominated in the early Universe. The main arguments supporting this hypothesis are presented by Larson [4].

Both these possibilities have been considered in the literature. Cen and Ostriker [5] found that, in a model with a fixed production of heavy elements per unit mass of gas transformed into stars (i.e., essen-

1063-7729/04/4805-0367$26.00 © 2004 MAIK "Nauka/Interperiodica"

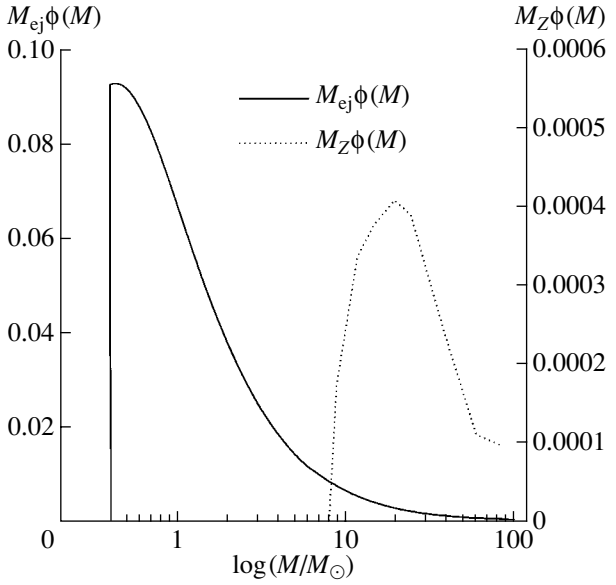

Fig. 1. Mass of unevolved matter returned to the ISM by planetary nebulae and supernovae, M_{ej}, and the mass of synthesized heavy elements, M_Z, multiplied by a Salpeter IMF for stars of various initial masses M.

tially with a fixed IMF), the metallicity is correlated with the mean density of matter in the star-formation zone. In this model, the metallicity must grow faster in zones of high density than in zones of low density. Cen and Ostriker [5] concluded that an enhanced density would be quite sufficient to explain the supermetallicity of quasars and its rapid growth.

The second possibility—that the IMF was initially dominated by massive stars—was considered by Romano *et al.* [6], who investigated IMFs less steep than the Salpeter IMF, i.e., with a higher fraction of massive stars. Romano *et al.* [6] concluded that the supermetallicities of quasars can be explained in a model in which the quasar host galaxies are identified with modern elliptical galaxies. In this model, the increasing activity of the quasar results in the termination of star formation in elliptical stellar systems.

A detailed analysis of the possibility of supermetallicities in the nuclei of disk galaxies has not been carried out despite observational data suggesting that such stellar systems are the hosts of many quasars [7] and the identification of such galaxies with distant Lyα galaxies (damped Ly α systems) [8]; the latter fact provides evidence that there may have been an abundance of such systems in the early stages of the evolution of the Universe. Later, some fraction of disk galaxies could be transformed into elliptical galaxies as a result of mergers [9].

In the present paper, we consider the origin of supermetallicities in a disk galaxy or its nucleus and the dependence of the evolution of galaxies with masses of the order of the mass of the Galaxy on the fraction of massive stars and the density of matter in circumnuclear regions. Section 2 describes the model for a disk galaxy used. Our results are presented in Section 3 and discussed in Section 4. Our main conclusions follow in Section 5.

2. MODEL OF A DISK GALAXY

In studies of the evolution of the chemical composition and stellar populations of galaxies, we can distinguish two main approaches to calculating the star-formation rate (SFR), ψ. The first, historically, usually assumed a constant SFR within a galaxy or described the SFR by an analytical function (e.g., assumed it was an exponential decreasing with time). The second approach takes into account the dependence of the SFR on the parameters of the galaxy (the gas density, number of massive stars, etc.). The assumption that, for given IMF, the SFR is determined by the degree of ionization of the interstellar gas leads to the expression [10]

$$\psi \propto \rho^2, \tag{1}$$

where ρ is the density of the interstellar gas. Firmani and Tutukov [10] were the first to take into account for a model disk galaxy variations of the thickness of the gaseous disk with time by assuming they were determined by the dissipation of the kinetic energy of supernova shells. Since, in the model we are considering, the supermetallicity of quasars is associated with an enhanced supernova rate (due to variation of the lower mass limit of the IMF and/or accelerated star formation), it is necessary to take this feedback into account.

Our single-zone model for a disk galaxy based on the formalism proposed by Firmani and Tutukov [10] is described in detail in [11, 12], and we note here only its main features. The evolution of the mass of interstellar gas is described by the equation

$$\frac{dM_g}{dt} = -\psi(t) \tag{2}$$
$$+ \int_{M_{min}}^{M_{max}} \psi(t-\tau_M) M_{ej} \phi(M) dM - M_g^{out} + M_g^{in},$$

where M_g is the mass of the gas in the galaxy, $\psi(t)$ is the SFR, $\phi(M)$ is the IMF, M_{min} and M_{max} are the limiting masses for forming stars, τ_M is the lifetime of a star with mass M, M_{ej} is the mass of the ejected envelope, and M_g^{out} and M_g^{in} are the rates of mass loss and accretion by the galaxy, respectively.

The production of heavy elements is described by the equation

$$\frac{d}{dt}(ZM_{\rm g}) = \int_{M_{\rm min}}^{M_{\rm max}} \psi(t-\tau_M) \times \left[Z(t-\tau_M)M_{\rm ej} + P_Z(M)\right]\phi(M)dM + P_Z^{\rm SNIa} - Z(t)\psi(t) - M_Z^{\rm out} + M_Z^{\rm in}. \quad (3)$$

Here, $Z(t)$ is the relative abundance of heavy elements at time t and $P_Z(M)$ is the mass of heavy elements synthesized in type II supernovae or asymptotic giant branch stars with mass M (Fig. 1). The term $P_Z^{\rm SNIa}$ represents the production of heavy elements in type Ia supernovae. We assumed that the rate of type Ia supernovae normalized to the current rate was proportional to the SFR, with the time delay $\sim 3\times 10^8$ yr [13]. The last two terms describe the exchange of heavy elements with the ISM. The rate of accretion was assumed to be zero in all models. We assumed that the fraction of heavy elements lost to the ISM, $f_{\rm esc}$, depends only on the mass of the galaxy $M_{\rm G}$ and is described by the expression [11]

$$f_{\rm esc} = \begin{cases} 0.2(12-\log M_{\rm G}), & M_{\rm G} < 10^{12} M_\odot, \\ 0, & M_{\rm G} \geq 10^{12} M_\odot. \end{cases} \quad (4)$$

Studies for stars in the solar neighborhood show that the IMF for stars with masses exceeding the solar mass is a power law with index -2.35 to -2.5 [14–16]. For lower masses, the absolute value of the index decreases, approaching -1.5 [17–19]. It is clear that the stars containing most of the baryonic matter of galaxies have masses near the break in the mass function; in this sense, this mass can be considered typical. The real lower limit of the IMF may be below $0.1\ M_\odot$ [20], but these stars contain a negligibly small fraction of the mass of the galaxy.

Below, we shall assume a Salpeter IMF with index -2.35 and a variable lower limit $M_{\rm min}$, which essentially marks the break of the IMF. There is observational evidence supporting the possibility that $M_{\rm min}$ can vary. For example, Smith and Gallagher [21] and Foster-Schreiber *et al.* [22] have shown that the typical mass of stars in young globular clusters in M82 is $2–3\ M_\odot$. This provides evidence that the value of $M_{\rm min}$ in regions of enhanced density may be higher than the average over the galaxy, justifying the inclusion of a variable $M_{\rm min}$ as one possible explanation for the accelerated enrichment of the nuclei of massive galaxies and quasars in heavy elements.

The assumption that the mass of young stars depends on the physical conditions in the region in which they form is related to the influence of the initial gas temperature T_0 on the rate of accretion by a gas–dust star in hydrostatic equilibrium [23]. For $T_0 \sim 10$ K, typical stellar masses are close to several tenths of a solar mass and proportional to T_0^3. The temperature of the interstellar gas, in turn, is determined by the local radiation density, $T_0 = L^{1/4}R^{-1/2}$, where L is the optical luminosity of the galaxy and R is its radius. It is obvious that, during the initial burst of star formation, the typical mass of stars should be higher than its current value. Moreover, the value of $M_{\rm min}$ in the dense, bright nucleus of the galaxy should be higher than the typical mass of the stars at the galaxy's periphery. In the first generations of stars, this effect will be enhanced by the low rate of cooling due to the absence of metals.

Another factor that should be mentioned is the metallicity dependence of the parameters of a star's evolution (the star's lifetime and luminosity, the production of heavy elements, the masses of remnants). We have used the results of the computations of Maeder [26], which cover metallicities from 0.004 to 0.02. In the galaxy models we consider here, the heavy-element abundance varies over a wider range, from 0 to $\sim$0.1. However, studies of the evolution of high-metallicity stars with initial masses from 0.1 to $3\ M_\odot$ [27] have shown that the luminosity and lifetime of stars depend only relatively weakly on the heavy-element abundance if this abundance does not exceed 0.1. This justifies the application of our model to $Z \sim 0–0.1$.

3. RESULTS

We considered several sets of models corresponding to different origins for the supermetallicities of quasars. Our "standard" model has a mass and radius similar to those of our Galaxy ($M = 2\times 10^{11}\ M_\odot$, $R = 20$ kpc) and $M_{\rm min} = 0.1\ M_\odot$; the model is described in detail in our previous papers (see, e.g., [12]). This model reproduces the evolution of the Galaxy well. At an evolutionary lifetime that is approximately equal to the age of the Galaxy 13 Gyr, the heavy-element abundance in the model approaches $Z_\odot$.

3.1. A Variable Lower Limit for the IMF

In the first part of our study, we attempted to reproduce the epoch of supermetallicity in a model that treats the galaxy as a single object. It is clear that an IMF with an enhanced fraction of massive stars will gradually be transformed into the current IMF, but we have no reliable information on the course of this process. We have chosen the following approximate algorithm for investigating the dependence of Z on time for a model with a variable typical mass for newborn stars. The initial $M_{\rm min}$ was assumed to be

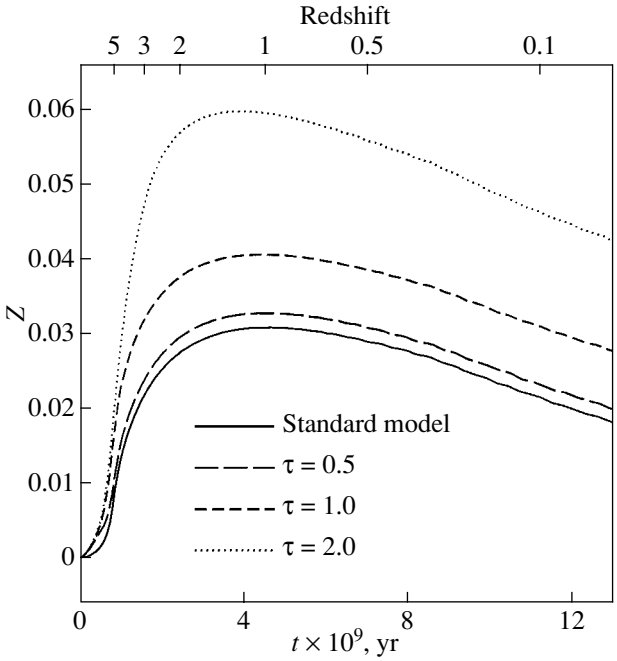

Fig. 2. Evolution of the heavy-element abundance in the case of a variable minimum mass $M_{\min}$ for various values of τ. The radius and mass of the model galaxy are $R =$ 20 kpc and $M_G = 2 \times 10^{11}\ M_\odot$. The upper scale shows the redshift, which is related to the age of the galaxy as $z = (13 \times 10^9/t)^{2/3} - 1$. The units of τ are 10^9 yr.

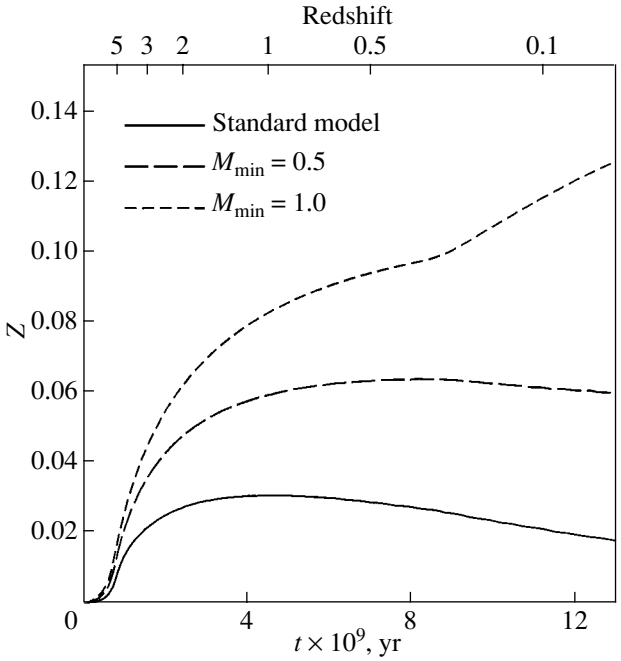

Fig. 3. Evolution of the heavy-element abundance for various values of the lower limit of the IMF. $M_{\min}$ is in $M_\odot$.

higher than the current value (i. e., $M_{\min} > 0.1\ M_\odot$). The value of $M_{\min}$ then linearly decreases with time to $0.1\ M_\odot$. The results of computations for a model galaxy with $M = 2 \times 10^{11}\ M_\odot$ and $R = 20$ kpc are shown in Fig. 2. The time parameter τ is expressed in units of 10^9 yr. The initial value of $M_{\min}$ was $2\ M_\odot$.

The increase in the minimum mass of newborn stars at early epochs is not sufficient to reproduce the high heavy-element abundance at high z. Even when $\tau = 2 \times 10^9$ yr, the maximum value of Z is achieved at $z \sim 1.5$, which is too late compared to the observational data. Moreover, the maximum metallicity itself ($\approx 3Z_\odot$) is lower than the Z values observed in some quasars. It is interesting that enhancing $M_{\min}$ in the early stages of the evolution of the model galaxy leaves the modern values of important parameters such as the disk thickness and the SFR virtually unaffected. This is undoubtedly related to the fact that the model is self-regulating and is able (like a real galaxy) to "adjust" to varying conditions. We conclude that the predominance of massive stars in the initial stages of the evolution does not lead to appreciable observational consequences for the galaxy as a whole.

The insufficient increase of the metallicity cannot be corrected, even if we assume that $M_{\min}$ remains high during the entire lifetime of the galaxy. Figure 3 shows the dependence of the heavy-element abundance on time and z for several values of $M_{\min}$ at $\tau = \infty$. It is obvious that the resulting heavy-element abundance grows with the initial mass of the forming stars. Increasing the lower limit of the IMF to $1\ M_\odot$ makes it possible to obtain Z values a factor of two higher than the solar value at $z \sim 3$. The maximum value $Z \sim 0.1$ in the ISM is attained at the age $\sim 7 \times 10^9$ yr ($z \sim 0.5$). However, this is not sufficient to explain the supermetallicities of quasars, since the heavy-element abundance must appreciably exceed the solar value at $z \sim 4$–6. Moreover, the other parameters of the model galaxy are in this case fairly different from the parameters for the standard model. For example, the current SFR in models with $M_{\min} \sim 1\ M_\odot$ exceeds $10\ M_\odot$/yr, and the half-thickness of the gaseous disk is 700–800 pc.

The value of Z grows slowly, because the initial peak of the SFR is lower in models with an increased lower limit for the IMF than in the standard model, although the burst of star formation lasts longer. The increase in $M_{\min}$ results in a relative increase in the supernova rate; as a result, the thickness of the disk decreases more slowly than in the standard model. The gas in the disk is less dense and the SFR decreases. If the values of $M_{\min}$ are higher than listed above, then, due to the very high supernova rate, the compression of the disk changes into expansion, leading to the disruption of the galaxy.

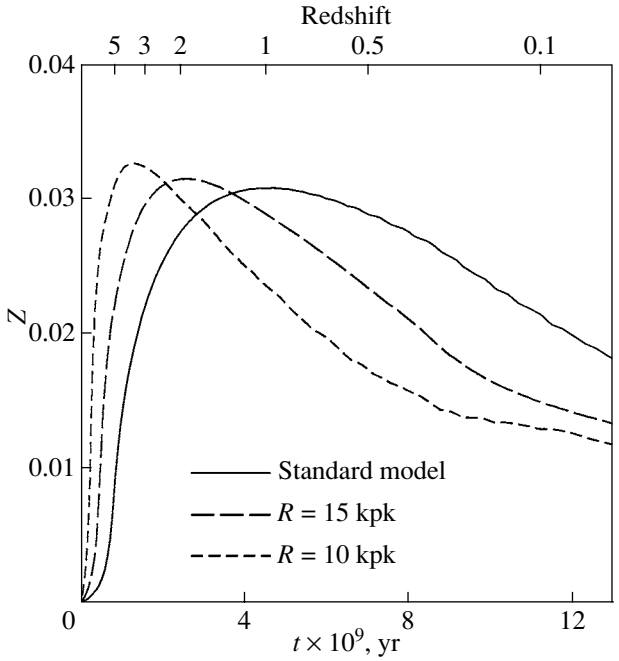

Fig. 4. The evolution of the heavy-element abundance for various initial densities in the star-formation region. R is given in kpc.

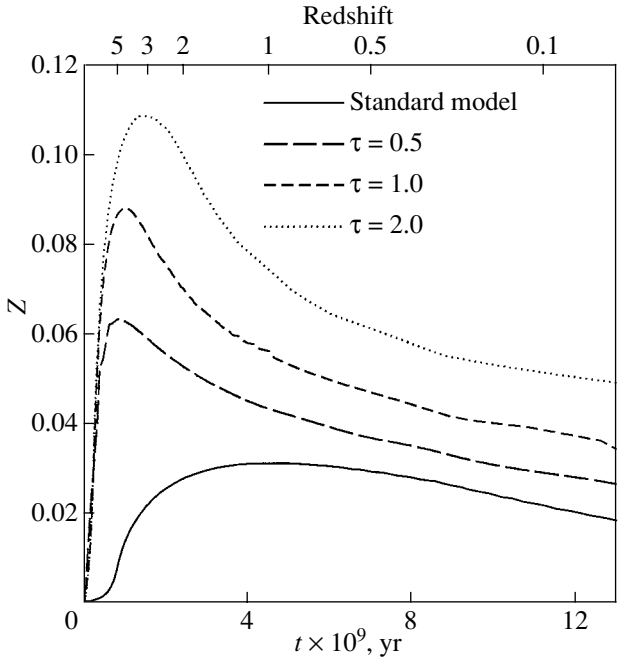

Fig. 5. Evolution of the heavy-element abundance in a subsystem of the galaxy with a high initial density and variable $M_{\min}$. The initial density in the star-formation region is determined by the radius and mass of the model galaxy, $R = 10$ kpc and $M_G = 2 \times 10^{11}\ M_\odot$. τ is in units of 10^9 yr.

Summarizing the results of this section, the seemingly natural assumption that massive stars dominated in the IMF does not lead to the desired results for the galaxy as a whole. This is largely due to feedback—the transfer of energy to the disk during supernovae. This means, in particular, that the supermetallicities of quasars cannot be considered to be characteristic of their host galaxies as a whole.

3.2. Enhanced Matter Density

One characteristic of our model of the galaxy is the self-consistent computation of the half-thickness of the disk. However, this limits its applicability in studies of different subsystems of the galaxy, in particular, the circumnuclear region. Therefore, to model the denser central region, we restricted our consideration to models with radii of 10 and 15 kpc but with the same mass for the galaxy ($2 \times 10^{11} M_\odot$). In all cases, the initial half-thickness of the disk was taken to be equal to the radius.

First, to exclude effects due to the influence of the lower limit of the IMF, we assumed that the minimum mass of forming stars was constant and equal to 0.1 $M_\odot$. Figure 4 shows the results of the computations for various disk radii. The SFR depends strongly on the gas density, resulting in an initially high SFR (up to 835 $M_\odot$/yr in the model with $R = 10$ kpc). This leads to an earlier enrichment of the ISM in heavy elements. The maximum value of Z is similar in all the models ($\sim 1.5\ Z_\odot$) but is attained at $t \sim 5 \times 10^9$ yr in the model with $R = 20$ kpc (the standard model; $z \sim 0.5$) and already at $t \sim 1.3 \times 10^9$ yr in the model with $R = 10$ kpc ($z \sim 4$). However, in this case as well, the supermetallicity is not as high as is observed. The enhanced heavy-element abundance produced by a powerful burst of star formation decreases to approximately $Z_\odot$ at the current epoch. For both radii, the current SFR in the models is close to 1 $M_\odot$/yr and the half-thickness of the disk is close to 200 pc.

A heavy-element abundance $Z > 3Z_\odot$ can be obtained if there is a high initial density *and* a variable lower limit for the IMF. Figure 5 shows the heavy-element abundance as a function of time in models with reduced radii and initially increased lower limits for the IMF. If the lower mass limit is sufficiently high and it decreases sufficiently slowly, the ISM can be considerably enriched in heavy elements. For $M_{\min} = 2.0\ M_\odot$ and $\tau = 2 \times 10^9$ yr, we obtain metallicities that are a factor of five higher than the solar value by $z \sim 5$. As the current epoch is approached, the metallicity drops to $2Z_\odot$—the value typical of the circumnuclear regions of our Galaxy. The current half-thickness of the disk is 200 pc, and the SFR is 1 $M_\odot$/yr. Due to the self-regulation in the models, these parameters are virtually the same in models with different τ.

4. DISCUSSION

The results presented above suggest that both the lower limit of the IMF and the initial density in the regions of star formation are equally important. Comparison of Figs. 4 and 5 shows that, as a result of an enhanced density and constant $M_{\min}$, the metallicity decreases relatively slowly after its rapid initial growth. Since the value of $M_{\min}$ could be much higher in the initial stages of the evolution of the galaxy than its modern value, it is possible to obtain both a high metallicity in the early evolutionary stages and a rapid decrease of Z in later stages.

The decrease of the heavy-element abundances in quasars observed in the transition from $z \sim 2$ to $z \sim 1$ [28] provides evidence in support of this hypothesis. However, we should bear in mind that this may be an observational selection effect and not a consequence of the evolution of the heavy-element abundance. Indeed, we can observe only the most luminous objects at large z, while objects with relatively low luminosities become observable at low z [29, 30]. Therefore, the relatively low metallicities of nearby quasars ($z \sim 1-2$)—only $(1.5-2)Z_\odot$—may be a manifestation of a correlation between the metallicities and luminosities of quasars. In another words, evolutionary variations of the heavy-element abundance may be superposed on the dependence of Z on the mass (density) of the galaxy. The separation of these two effects requires further studies of the observed dependence $Z(z)$ for quasars with fixed mass.

Note that, in contrast to the model of Romano *et al.* [6], our self-consistent model describes a system with *ongoing* star formation. This imposes much more stringent constraints on the parameters of the galactic evolution. This is especially true for the early stages of evolution, when the enhancement in heavy elements occurs. For example, a very large increase in $M_{\min}$ and in the initial density would lead to a very rapid synthesis of heavy elements, but the gaseous component of the galaxy would be destroyed by supernovae (the stability of the remaining stellar system would also be questionable). As we already noted above, the computation results reflect processes in a certain subsystem of the galaxy only. Bearing this in mind, the models in Figs. 2 and 3 reflect variations in the heavy-element abundance occurring at the periphery of a galaxy with a relatively low density. At the same time, the plots in Figs. 4 and 5 correspond to the evolution of the central regions of the galaxy, where the density is higher.

Of course, our single-zone model is methodically restricted. The minimum radius we have considered (10 kpc) is too large to describe the circumnuclear region. In our Galaxy, the Sun would be located within this radius, but the modern metallicity in this model is twice the solar value. However, we must take into account the existence of an additional factor influencing the evolution of the chemical composition of a disk galaxy: hot winds that carry some fraction of the synthesized heavy elements into intergalactic space. Wiebe *et al.* [31] noted that the efficiency of this ejection should increase with R, which may explain the observed radial gradient of the heavy-element abundance. Therefore, the high value of Z may be retained only at the very center of the galaxy. The heavy-element abundance will decrease with distance from the galactic center due not only to the return of matter by low-mass stars but also to the ejection of matter into intergalactic space. Detailed numerical studies of these processes require a multizone model.

5. CONCLUSION

We have investigated possible reasons for the enhanced abundance of heavy elements in the gaseous medium of quasars and active galactic nuclei. We conclude that solar and even supersolar heavy-element abundances can be achieved in the very early stages of the evolution of massive galaxies ($M \geq 10^{11} M_\odot$), at $z \sim 5$ ($t \sim 10^9$ yr). The early enrichment of massive galaxies in heavy elements is consistent with modern ideas concerning the onset of the formation of the first massive stars at $z \sim 10-20$. This means that the first stars were born $\sim 10^8$ yr after the Big Bang, which corresponds to the free-fall time for the current densities of galaxies.

ACKNOWLEDGMENTS

The authors thank A.A. Boyarchuk for discussions of methods for estimating the chemical composition of a highly ionized gas. This work was supported by the Federal Scientific and Technological Program "Astronomy" and the Russian Foundation for Basic Research (project no. 03-02-16254).

REFERENCES

1. F. Hamann and G. Ferland, Ann. Rev. Astron. Astrophys. **37**, 487 (1999).
2. F. Hamann, K. T. Korista, G. J. Ferland, *et al.*, Astrophys. J. **564**, 592 (2002).
3. L. Pentericci, X. Fan, H.-W. Rix, *et al.*, Astron. J. **123**, 2151 (2002).
4. R. Larson, Mon. Not. R. Astron. Soc. **301**, 569 (1998).
5. R. Cen and J. Ostriker, Astrophys. J. **519**, L109 (1999).
6. D. Romano, L. Silva, F. Matteucci, and L. Danese, Mon. Not. R. Astron. Soc. **334**, 444 (2002).
7. T. S. Hamilton, S. Casertano, and D. A. Turnshek, Astrophys. J. **576**, 61 (2002).

8. J. X. Prochaska and A. M. Wolfe, Astrophys. J. **487**, 73 (1997).
9. J. Silk and R. Bouwens, New Astron. Rev. **45**, 337 (2001).
10. C. Firmani and A. Tutukov, Astron. Astrophys. **264**, 37 (1992).
11. B. M. Shustov, D. S. Wiebe, and A. V. Tutukov, Astron. Astrophys. **317**, 397 (1997).
12. D. S. Wiebe, A. V. Tutukov, and B. M. Shustov, Astron. Zh. **75**, 3 (1998) [Astron. Rep. **42**, 1 (1998)].
13. A. V. Tutukov and L. R. Yungel'son, Astron. Zh. **79**, 738 (2002) [Astron. Rep. **46**, 667 (2002)].
14. E. Salpeter, Astrophys. J. **121**, 161 (1955).
15. S. Vereshchagin, Z. Kraicheva, E. Popova, *et al.*, Pis'ma Astron. Zh. **13**, 63 (1987) [Sov. Astron. Lett. **13**, 26 (1987)].
16. G. Chabrier, Publ. Astron. Soc. Pac. **115**, 763 (2003).
17. R. Kennicutt, P. Tamblyn, and C. Congolon, Astrophys. J. **435**, 22 (1994).
18. G. Chabrier, Astrophys. J. **567**, 304 (2002).
19. K. L. Luhman, G. H. Rieke, E. T. Young, *et al.*, Astrophys. J. **540**, 1016 (2000).
20. M. Bate, I. Bonnell, and V. Bromm, Mon. Not. R. Astron. Soc. **339**, 577 (2003).
21. L. Smith and J. Gallagher, Mon. Not. R. Astron. Soc. **326**, 1027 (2001).
22. N. Foster-Schreiber, R. Geuzel, D. Lutz, and A. Sternberg, Astrophys. J. (in press); astro-ph/0309357.
23. A. V. Tutukov, Pis'ma Astron. Zh. **9**, 160 (1983) [Sov. Astron. Lett. **9**, 86 (1983)].
24. Z. Haiman, A. A. Thoal, and A. Loeb, Astrophys. J. **464**, 523 (1996).
25. V. Bromm, P. S. Coppi, and R. Larson, Astrophys. J. **564**, 23 (2002).
26. A. Maeder, Astron. Astrophys. **264**, 105 (1992).
27. A. Tutukov and A. Fedorova, Astron. Zh. **78**, 1008 (2001) [Astron. Rep. **45**, 882 (2001)].
28. T. Di Matteo, R. Croft, V. Springel, and L. Hernquist, astro-ph/0309533.
29. D. Gudehus, Astrophys. J. **382**, 1 (1991).
30. L. Koopmans and T. Tren, Astrophys. J. **568**, 5 (2002).
31. D. S. Wiebe, A. V. Tutukov, and B. M. Shustov, Astron. Zh. **78**, 977 (2001) [Astron. Rep. **45**, 854 (2001)].

Translated by L. Yungel'son

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 374–379. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 411–417.
Original Russian Text Copyright © 2004 by Megn.

Spectral Characteristics of Components of a Model Radio-Brightness Distribution for the Radio Galaxy 3C 234

A. V. Megn

Institute of Radio Astronomy, National Academy of Sciences of Ukraine, Krasnoznamennaya ul. 4, Kharkov, 61002 Ukraine

Received June 25, 2003; in final form, November 10, 2003

Abstract—Physical processes that can, under cosmic conditions, give rise to emission whose spectrum peaks at some frequency are discussed in the context of the spectrum of the central extended component of a model brightness distribution for the radio galaxy 3C 234. This component is not detected at decameter wavelengths, probably due to the absorption of the radiation in the plasma in the source itself.
© 2004 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

The angular structure of the radio galaxy 3C 234 has been studied in detail at centimeter and decimeter wavelengths, at frequencies ranging from 408 to 15 375 MHz [1–10]. In [11], based on observations using the URAN-1 and URAN-2 radio interferometers [12], we constructed model brightness distributions for the source at 25 and 20 MHz and also derived model brightness distributions of this galaxy at a number of other frequencies from 1400 to 8400 MHz, corresponding to observations at these frequencies made on interferometers with the same baseline orientation and angular resolution as the URAN-1 and URAN-2 interferometers ($D/\lambda =$ 2820 and 3525 for URAN-1 and $D/\lambda = 10\,153$ and 12 692 for URAN-2, respectively). The high-frequency models for 3C 234 turned out to be virtually identical, consisting of five components: two compact components (southwest and northeast), with sizes smaller than $1''$ separated by a maximum projected distance of $\sim 110''$ in position angle 66°, and three extended components between them, with sizes from $4'' \times 6''$ to $16'' \times 32''$. The maximum emission among these three components was provided by the third, largest component, which was located at the center of the radio galaxy, in the vicinity of its nucleus.

However, the model brightness distribution for 3C 234 at decameter wavelengths had only four components; emission was observed only from the two compact and two outer extended components. At the same time, the radio emission of the central extended component was not detected at low frequencies.

Our aim here, which is a continuation of the study [11], is to investigate possible physical processes that could result in this behavior.

2. SPECTRAL CHARACTERISTICS OF THE COMPONENTS OF THE MODEL RADIO-BRIGHTNESS DISTRIBUTION FOR THE RADIO GALAXY 3C 234

Based on [1–11], we determined the spectra for the compact components at frequencies from 20 to 15 375 MHz and for the outer extended components from 20 to 8400 MHz. The spectrum $S(\nu)$ of the central extended component (S is the spectral flux of the emission, ν is the frequency) can be determined observationally only from 1400 to 8400 MHz.

Figure 1 shows the spectra of the entire radio galaxy $S_0(\nu)$ (curve *1*); of the combined emission of all components except for the central one, $S_1(\nu) + S_2(\nu) + S_4(\nu) + S_5(\nu)$ (curve *2*); and of the emission of the central extended (third) component $S_3(\nu)$ (curve *3*). Various symbols are used to plot the data for the corresponding frequencies. In frequency intervals where measurements are available, the spectra are drawn as solid curves, while supposed, but not measured, parts of the spectra are shown by dashed curves.

Figure 1 shows that, at frequencies $\nu > 1400$ MHz, all the above spectra are power-law functions $S \sim \nu^{-\alpha}$ (linear in $\log S - \log \nu$ coordinates) with spectral indices $\alpha \approx 1$. The spectrum of the total radio emission of the galaxy $S_0(\nu)$ is also a power law of this type when $\nu \geq 750$ MHz; at lower frequencies (40 MHz $\leq \nu \leq 750$ MHz), the spectrum is likewise power-law, but with the spectral index $\alpha \approx 0.86$. At still lower frequencies, the $S_0(\nu)$ dependence is no longer a power law and we observe a monotonic decrease in the growth of the spectral flux with decreasing frequency down to the lowest frequency at which observations have been carried out, $\nu = 14.6$ MHz. The total spectral fluxes of all four of the model

1063-7729/04/4805-0374$26.00 © 2004 MAIK "Nauka/Interperiodica"

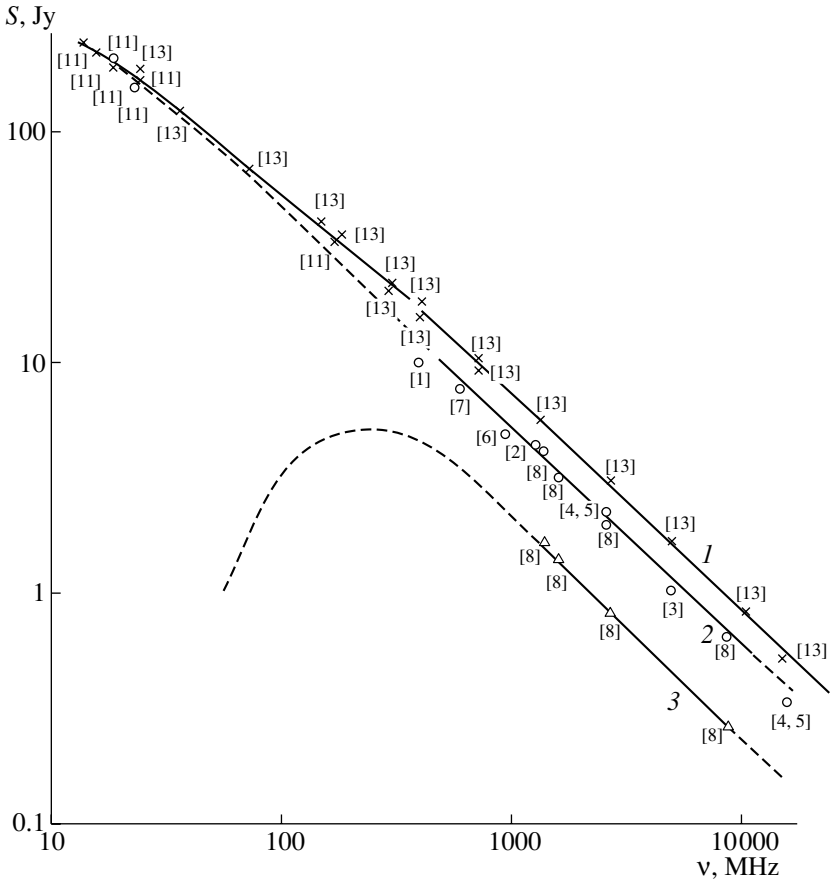

Fig. 1. Spectra of the radio galaxy 3C 234 and components of the model radio-brightness distribution: (*1*) spectrum of the entire galaxy $S_0(\nu)$ (crosses), (*2*) spectrum of the combined emission of all components except for the central one $S_1(\nu) + S_2(\nu) + S_4(\nu) + S_5(\nu)$ (circles), (*3*) spectrum of the third (central extended) component $S_3(\nu)$ (triangles). Labels are references to data from the literature.

components except for the central one are derived from observations from 15 375 to 408 MHz and from 25 to 20 MHz. In Fig. 1 (curve *2*), this spectrum is plotted at frequencies below 408 MHz based on the assumption that it remains a power law down to $\nu \approx 60$ MHz, with the same spectral index as at higher frequencies, and, at lower frequencies, is consistent with the spectral fluxes of these components measured at 25 and 20 MHz.

The spectrum of the third (central extended) component $S_3(\nu)$ outside 1400–8400 MHz has not been determined observationally and must be approximated indirectly as the difference between $S_0(\nu)$ and the total emission of the remaining model components:

$$S_3(\nu) = S_0(\nu) - [S_1(\nu) + S_2(\nu) + S_4(\nu) + S_5(\nu)]. \quad (1)$$

We suppose that the turnover of the total spectrum of the radio galaxy $S_0(\nu)$ near $\nu \approx 750$ MHz is associated with the properties of the spectrum of the third model component, $S_3(\nu)$, which is not detected at decameter wavelengths.

The mean dependence $S_3(\nu)$ for $\nu < 1400$ MHz found in this way is shown in Fig. 1 by the dashed curve. We can see in Fig. 1 that the maximum of the spectrum, $S_{3\max} = 5.25$ Jy, is observed at $\nu = 251$ MHz. However, note that, given the accuracies of the various spectral fluxes in (1), this method for determining $S_3(\nu)$ is subject to large errors, especially at $\nu < 251$ MHz, where they increase rapidly with decreasing frequency.[1]

The vertical bars on curve *1* in Fig. 2 show the intervals within which S_3 can be contained at the various frequencies. Owing to the relatively large uncertainties in the $S_3(\nu)$ estimates derived from (1), these data should be considered preliminary and in need of refinement via direct mapping of 3C 234 at

[1] When deriving $S_3(\nu)$ from (1), the difference between the estimates obtained using the two methods did not exceed 6% at frequencies from 1400 to 8400 MHz, where the spectral densities were also measured directly.

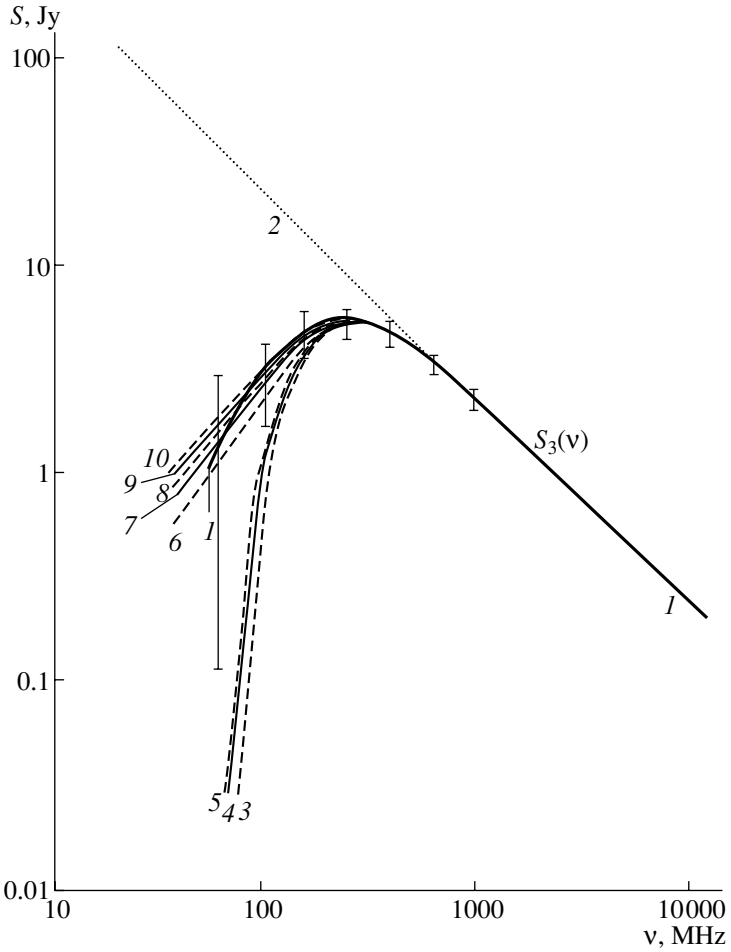

Fig. 2. Spectra of the third (central, extended) model component $S_3(\nu)$ for various absorption mechanisms that can act under cosmic conditions: (*1*) spectrum obtained from the data of [11]; (*2*) spectrum formed taking into account self-absorption; (*3*)–(*5*) spectra formed taking into account absorption of the radiation in plasma with temperatures $T = 10$, 85 000, and 1 000 000 K in the path along which the radio signal propagates; (*6*)–(*10*) spectrum formed taking into account absorption of the radiation in plasma with temperatures $T = 10$, 32, 100, 1000, and 10 000 K, respectively, within the third (extended) component.

meter wavelengths, at least at short meter wavelengths corresponding to 300 to 100 MHz.

However, since the presence of an extremum in $S_3(\nu)$ at $\nu = 251$ MHz and the decrease in the spectral flux at lower frequencies is beyond doubt, it is of interest to identify physical processes acting under cosmic conditions that could result in this type of spectrum.

Among various physical processes that give rise to spectra with an extremum at some frequency, three are the most widespread: reabsorption (synchrotron self-absorption), absorption of radio emission in clouds of cosmic plasma in the path along which the radio signal propagates, and absorption of radiation in an ionized medium within the source itself [14].

Let us briefly consider the possible effects of these physical mechanisms on the shape of the spectrum of the central extended model component (galaxy 3C 234).

3. SELF-ABSORPTION

The spectrum of a self-absorbed radio source can be calculated near the peak ($\nu = \nu_m$) and at lower frequencies based on measurements at high frequencies ($\nu \gg \nu_m$). In this case, according to [15, 16],

$$\nu_m = \left[\frac{S_1 \nu_1^\alpha H_\perp^{1/2} (1+z)^{1/2}}{\tau_0 \Delta\theta^2}\right]^{2/2\alpha+5}, \qquad (2)$$

$$S_m = S_1 \left(\frac{\nu_1}{\nu_m}\right)^\alpha \frac{1 - e^{-\tau_0}}{\tau_0},$$

where ν_m is the peak frequency in MHz, S_m is the maximum spectral flux in Janskys, S_1 is the spectral flux in Janskys at frequency $\nu_1 \gg \nu_m$ in MHz, $H_\perp$ is the magnetic-field component perpendicular to the direction of motion of the relativistic electrons in μOe, z is the redshift, $\Delta\theta$ is the angular size of the radio

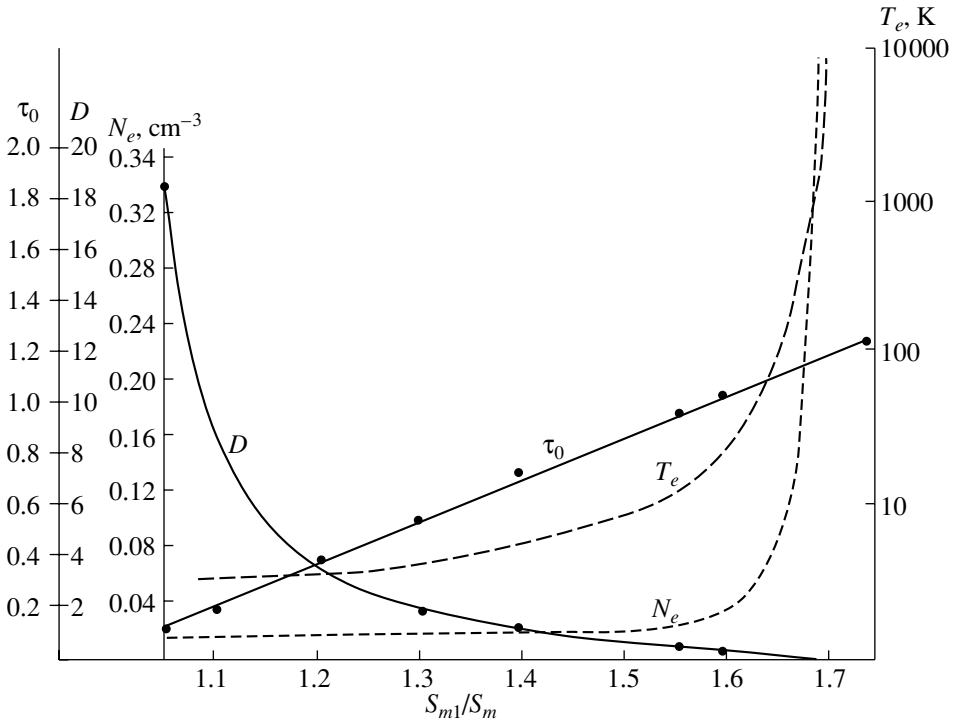

Fig. 3. Values of τ_0, D, N_e, and T_e as functions of S_{m1}/S_m taking into account absorption of the radiation in plasma of the third model component in 3C 234 ($\alpha = 1$, $\nu_m = 251$ MHz).

source (component) in arcseconds, and τ_0 is the optical depth at frequency ν_m [17]:

$$e^{\tau_0} = 1 + \frac{2\alpha + 5}{5}\tau_0. \qquad (3)$$

Note that, in the power-law section of the spectrum ($\nu \gg \nu_m$), the spectral flux S_1 can be measured at any frequency ν_1, since the value of $S_1\nu_1^\alpha$ in the above expressions is invariant and does not depend on the frequency of the measurement. In the case considered, the spectrum at any frequency is [15, 16]

$$S(\nu) = S_m \left(\frac{\nu}{\nu_m}\right)^{5/2} \frac{1 - e^{-\tau_0\left(\frac{\nu}{\nu_m}\right)^{-\alpha-5/2}}}{1 - e^{-\tau_0}}. \qquad (4)$$

Formulas (2)–(4) can be used to find ν_m and S_m in frequency intervals where observations have not been carried out and also to derive $S(\nu)$ at low frequencies $\nu < \nu_m$ for the case of a self-absorbed synchrotron source, based on measurements of its angular size and spectrum at high frequencies.

For the central extended model component in 3C 234, calculations with the above relationships and with $\alpha \approx 1$, $\nu_1 = 1600$ MHz, $S_1 = 1.41$ Jy, $\Delta\theta \approx 22''$ [11], $z = 0.185$, and $H_\perp \approx 240\,\mu$Oe [18][2] yield $\tau_0 = 0.64$, $\nu_m = 3.95$ MHz, and $S_m = 422$ Jy. The spectrum $S_{3p}(\nu)$ calculated with (4) for this case is shown in Fig. 2 by the dotted curve *2*.

Figures 1 and 2 show that both this calculated dependence and the values of ν_m and S_m are completely inconsistent with the above dependences $S_0(\nu)$ and $S_3(\nu)$, even allowing for the errors. This leads us to conclude that the action of self-absorption cannot explain the spectrum $S_3(\nu)$ found with (1).

4. ABSORPTION IN THE LINE-OF-SIGHT COSMIC PLASMA

If radiation is absorbed by clouds of plasma in the path along which it propagates, then the spectrum $S(\nu)$ at any frequency is determined as [15, 16]

$$\ln\frac{S(\nu)}{S_m} = \alpha \ln\frac{\nu_m}{\nu} + \frac{\alpha}{2+D} \times \left[1 - \left(\frac{\nu_m}{\nu}\right)^2 \left(1 + D\ln\frac{\nu_m}{\nu}\right)\right], \qquad (5)$$

where the parameter D and the optical depth τ_0 at frequency ν_m are

$$D = \frac{\alpha}{\tau_0} - 2, \quad \tau_0 = \ln\frac{S_{m1}}{S_m}. \qquad (6)$$

Here, as before, S_m is the maximum spectral flux at frequency ν_m and S_{m1} is the spectral flux at frequency

[2] Note that $H_\perp$ does not strongly affect the values of ν_m and S_m: for example, according to (2), a change in $H_\perp$ from 100 to 500 μOe produces a change in ν_m from 3.5 to 4.4 MHz and a change in S_m from 480 to 380 Jy.

ν_m, which is linearly (on a logarithmic scale) extrapolated from high frequencies.

In this case, according to [15, 16, 19, 20], we can essentially use the values of D, ν_m (in Hz), and α to find the electron temperature T_e (in Kelvin) and the emission measure $N_e^2 L$ of the absorbing plasma, where N_e is the electron density in the plasma (in cm^{-3}) and L is the size of the absorbing region (in cm):

$$17.7 + 1.5 \ln T_e = \ln \nu_m + \frac{1}{D}, \qquad (7)$$

$$\frac{9.8 \times 10^{-3} N_e^2 L}{T_e^{3/2}} = \tau_0 \nu_m^2 D.$$

Calculations using (6) and (7) with $\alpha = 1$, $S_m = 5.25$ Jy, $S_{m1} = 8.52$ Jy, and $\nu_m = 251 \times 10^6$ Hz yield $\tau_0 = 0.484$, $D = 0.065$, $T_e \approx 85 \times 10^3$ K, and $N_e^2 L \approx 1.6 \times 10^6$ cm^{-6} pc. If the region of absorbing plasma between the observer and the central extended model component is no larger than 60 kpc, this corresponds to $N_e \geq 5$ cm^{-3}.

The derived electron temperature and emission measure do not agree with measurements of the high-temperature plasma in HII regions obtained using other methods, which indicate that, depending on the type of object considered, T_e varies from 10^3 to 10^4 K; $N_e^2 L$, from 10^2 to 10^4 cm^{-6} pc; and N_e, from 1 to 10 cm^{-3} [21, 22]. It is even more important that the dependence $S_3(\nu)$ does not agree with the calculated spectrum $S_{3p}(\nu)$ obtained using (5) and the above values of α, S_m, ν_m, and D (curve *4* in Fig. 2). In addition to this dependence for $T_e \approx 85 \times 10^3$ K, Fig. 2 shows the results of similar calculations with the same values of ν_m and S_m for $T_e = 10$ K (curve *3*) and $T_e = 10^6$ K (curve *5*). It follows from Fig. 2 that varying the electron temperature over a wide range leaves the appearance of the calculated spectra for this absorption mechanism virtually unaffected. In the case of curves *3*, *4*, and *5*, the spectral flux decreases with decreasing frequency, $\nu < \nu_m$, much faster (approximately as ν^5–ν^7) than is observed for dependence *1* ($S_3(\nu) \sim \nu^{1.4}$–$\nu^{1.8}$). Thus, absorption in regions of cosmic plasma along the propagation path, likewise, probably cannot explain the spectrum $S_3(\nu)$ we have derived.

5. ABSORPTION IN PLASMA WITHIN THE COSMIC RADIO SOURCE

In the case of absorption of the radiation in the ionized medium of the radio source, the spectrum becomes [15, 16]

$$S(\nu) = S_m \left(\frac{\nu}{\nu_m}\right)^{2-\alpha} \times \frac{1}{1 + D \ln \dfrac{\nu_m}{\nu}} \frac{1 - e^{-\tau_0 \left(\frac{\nu_m}{\nu}\right)^2 \left(1 + D \ln \frac{\nu_m}{\nu}\right)}}{1 - e^{-\tau_0}}, \qquad (8)$$

where the optical depth τ_0 at frequency ν_m depends on the ratio S_{m1}/S_m as

$$\frac{\tau_0}{1 - e^{-\tau_0}} = \frac{S_{m1}}{S_m} \qquad (9)$$

and the parameter D is

$$D = \frac{\alpha (1 - e^{-\tau_0})}{1 - e^{-\tau_0} - \tau_0 e^{-\tau_0}} - 2. \qquad (10)$$

In this case, we can also find the electron temperature T_e and emission measure $N_e^2 L$ of the source plasma using (7). Calculations using (7), (9), and (10) with the same values of α, S_m, S_{m1}, and ν_m as before yield $\tau_0 = 1.06$, $D = 0.28$, $T_e \approx 32$ K, $N_e^2 L \approx 112$ cm^{-6} pc, and $N_e \approx 0.043$ cm^{-3} if the size of the central extended model component in 3C 234 is $L \approx 60$ kpc [11].

The resulting parameters for the low-temperature plasma could correspond to regions of ionized carbon CII [23].

Figure 2 shows the spectrum calculated from (8) with the resulting data (curve *7*). The same figure presents calculated values of $S_p(\nu)$ for $T_e = 10$, 100, 1000, and 10 000 K (curves *6*, *8*, *9*, and *10*, respectively). It follows from Fig. 2 that, from 251 to 60 MHz, all the dependences except for curve *6* are fairly consistent with $S_3(\nu)$ (curve *1*); although the curves diverge at lower frequencies, this could be due to the very poor accuracy in $S_3(\nu)$ for spectral fluxes less than 1 Jy.

Note that, although the values of T_e and N_e obtained in the last case are consistent with the parameters of cool cosmic plasma, we cannot judge based on these values whether this absorption mechanism is acting, since, in contrast to the spectrum $S_{3p}(\nu)$, which does not depend appreciably on the plasma parameters, these parameters themselves have large errors (especially at small values of D). Accurate estimation of these parameters requires measurement of ν_m, S_m, S_{m1}, and α with good accuracy.

By means of illustration, Figure 3 shows τ_0, D, N_e, and T_e as functions of S_{m1}/S_m, calculated with (9), (10), and (7) with $\alpha = 1$ and $\nu_m = 251$ MHz, taking into account absorption in plasma in the radio source. We can see that, when $S_{m1}/S_m < 1.4$ (for $D > 1$), N_e and T_e grow relatively slowly with increasing S_{m1}/S_m, so that they can be determined fairly accurately. When S_{m1}/S_m is large (D is small), N_e and T_e grow very rapidly with increasing S_{m1}/S_m, and it becomes difficult to find N_e and T_e to an acceptable accuracy. For example, in our case ($S_{m1}/S_m = 1.62$),

if we determine this ratio to within ± 0.01, T_e and N_e are within the limits $27 \leq T_e \leq 46$ K and $0.03 \leq N_e \leq 0.05$ cm^{-3}; i.e., when the error in S_{m1}/S_m is $\pm 0.6\%$, the errors in the estimated values of T_e and N_e are ± 26 and $\pm 23\%$, respectively. Since, in practice, the accuracy of the measured ratio S_{m1}/S_m is usually much poorer, the determination of N_e and T_e in this way for $S_{m1}/S_m > 1.5$–1.6 is subject to very large errors.

6. CONCLUSION

Thus, the spectrum $S_3(\nu)$ of the central extended component in the model radio-brightness distribution for the galaxy 3C 234, which has a peak at 251 MHz, can be explained by the absorption of radiation in a low-temperature plasma with a low electron density within the source itself. It seems natural that synchrotron self-absorption and absorption in regions of cosmic plasma along the propagation path of the radio signal cannot explain the spectrum $S_3(\nu)$, since, as a rule, self-absorption is important only in very compact radio sources and absorption in clouds of cosmic plasma along the line of sight is likewise not probable for an extended object.

To conclude, we note that, given the approximate character of the spectrum $S_3(\nu)$ derived here, it is desirable to determine it more accurately via direct radio-interferometric mapping of 3C 234 at meter wavelengths.

ACKNOWLEDGMENTS

The author is grateful to S.L. Rashkovskiĭ, N.K. Sharykin, and V.A. Shepelev for their help.

REFERENCES

1. P. N. Wilkinson, Mon. Not. R. Astron. Soc. **160**, 305 (1972).
2. J. A. Hogbom and I. Carlsson, Astron. Astrophys. **34** (3), 341 (1974).
3. F. M. Riley and G. G. Pooley, Mem. R. Astron. Soc. **80**, 105 (1975).
4. S. F. Burch, Mon. Not. R. Astron. Soc. **186**, 293 (1979).
5. S. F. Burch, Mon. Not. R. Astron. Soc. **186**, 519 (1979).
6. N. H. Bedford, A. J. Kerr, S. N. Mathur, *et al.*, Mon. Not. R. Astron. Soc. **195**, 245 (1981).
7. R. G. Strom and R. G. Conway, Astron. Astrophys., Suppl. Ser. **61**, 547 (1985).
8. P. Alexander, Mon. Not. R. Astron. Soc. **225**, 27 (1987).
9. M. J. Hardcastle, P. Alexander, G. G. Pooley, *et al.*, Mon. Not. R. Astron. Soc. **288**, 859 (1997).
10. M. J. Hardcastle, P. Alexander, G. G. Pooley, *et al.*, Mon. Not. R. Astron. Soc. **296**, 445 (1998).
11. A. V. Megn, S. Ya. Braude, S. L. Rashkovskiĭ, *et al.*, Astron. Zh. **80**, 1128 (2003) [Astron. Rep. **47**, 1038 (2003)].
12. A. V. Megn, S. Ya. Braude, S. L. Rashkovskiĭ, *et al.*, Radiofiz. Radioastron. **2** (4), 385 (1997).
13. H. Kuhr, U. Nauber, I. I. K. Pauliny-Toth, and A. Witzel, Preprint (Max-Planck-Institut fur Radioastronomie, Bonn, 1981).
14. S. Ya. Braude, I. N. Zhuk, A. V. Megn, *et al.*, Ukr. Fiz. Zh. **14** (11), 1761 (1969).
15. S. Ya. Braude and A. V. Megn, Kinemat. Fiz. Neb. Tel **13** (1), 74 (1997).
16. A. V. Megn and S. Ya. Braude, Kinemat. Fiz. Neb. Tel **13** (1), 82 (1997).
17. S. Ya. Braude, Astron. Zh. **42** (6), 1150 (1965).
18. J. W. Dreher, Astron. J. **86** (6), 833 (1981).
19. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasma* (Fizmatgiz, Moscow, 1960) [in Russian].
20. V. L. Ginzburg and S. I. Syrovatskiĭ, *The Origin of Cosmic Rays* (Izd. Akad. Nauk SSSR, Moscow, 1963) [in Russian].
21. S. A. Kaplan and S. B. Pikel'ner, *Physics of the Interstellar Medium* (Nauka, Moscow, 1979) [in Russian].
22. E. A. Abramenkov and V. V. Krymkin, Astron. Zh. **59** (2), 263 (1982) [Sov. Astron. **26**, 160 (1982)]; Astron. Zh. **60** (1), 53 (1983) [Sov. Astron. **27**, 32 (1983)]; Astron. Zh. **69** (3), 489 (1992) [Sov. Astron. **36**, 246 (1992)]; Astron. Zh. **69** (4), 731 (1992) [Sov. Astron. **36**, 374 (1992)].
23. A. A. Konovalenko, Pis'ma Astron. Zh. **10** (11), 846 (1984).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 380–392. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 418–430.
Original Russian Text Copyright © 2004 by Pashchenko, Rudnitskiĭ.

Observations of Late-Type Variable Stars in the Water Vapor Radio Line. The Long-Period Variable R Cassiopeia

M. I. Pashchenko and G. M. Rudnitskiĭ

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

Received May 20, 2003

Abstract—Observations of circumstellar maser emission from the long-period variable R Cas in the 1.35-cm water-vapor line are reported. The observations were carried out on the 22-m radio telescope of the Pushchino Radio Astronomy Observatory in 1980–2003 (JD = 2444409–2452724). Over the 23 years of observations, strong flares in the H_2O line profile were recorded in 1982 (with a peak flux density up to 400 Jy) and 1986–1989 (up to 750 Jy). Subsequently, from 1990 to March 2003, the H_2O line flux was usually below the detection threshold of the radio telescope (<5–10 Jy). Episodic small increases of the emission with peak flux densities of 20–60 Jy were observed. The variations of the H_2O line flux F are correlated with variations in the visual brightness of the star. The phase delay $\Delta\varphi$ of the F variations relative to the optical light curve of R Cas ranged from $0.2–0.3P$ during the observations ($P = 430.46^{\rm d}$ is the star's period). A model for the variability of the H_2O maser in R Cas is discussed. If the variations are due to periodic impacts by shock waves driven by the stellar pulsations, the time for the shock to travel from the photosphere to the inner boundary of the H_2O-masing shell may reach $2–4P$. The flares could be due to transient episodes of enhanced mass loss by the star or to the propagation of an exceptionally strong shock from the stellar surface. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

More than 500 giant and supergiant late-type stars are currently known to be sources of maser emission in the $6_{16}–5_{23}$ rotational line of the H_2O molecule ($\lambda = 1.35$ cm) [1]. All these stars are Mira Ceti or semiregular SRa, SRb long-period variables. The H_2O radio emission exhibits strong variations. The flux density in the 1.35 cm line sometimes varies by several orders of magnitude. From the point of view of long-term monitoring, it is interesting to compare the variations in the H_2O line and in the optical and infrared. The correlation of the variability in the H_2O line and in the infrared, as well as the accompanying changes in the structure of the H_2O line profile, characterize processes in the inner layers of the circumstellar envelopes of late-type giants and supergiants. The H_2O observations provide a unique tool for studying the region of acceleration of the material lost by the star [2]: the radius of the H_2O maser zone approximately coincides with the radius of dust condensation R_c; the main acceleration of gas in the circumstellar envelopes of late-type stars begins just near R_c.

This paper continues a series of publications [3–10] on the variability of H_2O masers associated with long-period variable stars. We present the results of observations of the Mira Ceti variable R Cas (BD+50°4202, HD 224490, IRC+50484, IRAS 23558+5106), carried out by us from 1980 to 2003. Some preliminary results of these observations were published earlier in [3, 4].

In the Fourth Edition of the General Catalog of Variable Stars [11] (GCVS), R Cas is classified as a Mira Ceti long-period variable; the maximum amplitude of its visual brightness variations is $4.7–13.5^m$, and the spectral class of the star varies in the range M6e–M10e. The GCVS gives the following light elements for R Cas:

$$\text{Max} = \text{JD}\,2444463 + 430.46^{\rm d}E. \qquad (1)$$

The light curve of R Cas over many years since 1901 was studied in [12]. The mean curve is asymmetric, with the brightness rise being faster than its decline; the degree of asymmetry is $(M - m)/P = 0.4$. We used the long-term visual observations of R Cas obtained by the French Association of Variable Star Observers (AFOEV), which we acquired via the Internet.[1] The star's mean period has not changed during the time it has been observed. This also follows from a comparison of the epochs of visual maxima calculated with the elements (1) and the observed maxima (see below, Figs. 2–4).

The $JHKLM$ infrared light curves of R Cas were analyzed in [13]; during the time covered by the observations of [13], the most probable period of the

[1] ftp://cdsarc.u-strasbg.fr/pub/afoev/cas/r.

1063-7729/04/4805-0380$26.00 © 2004 MAIK "Nauka/Interperiodica"

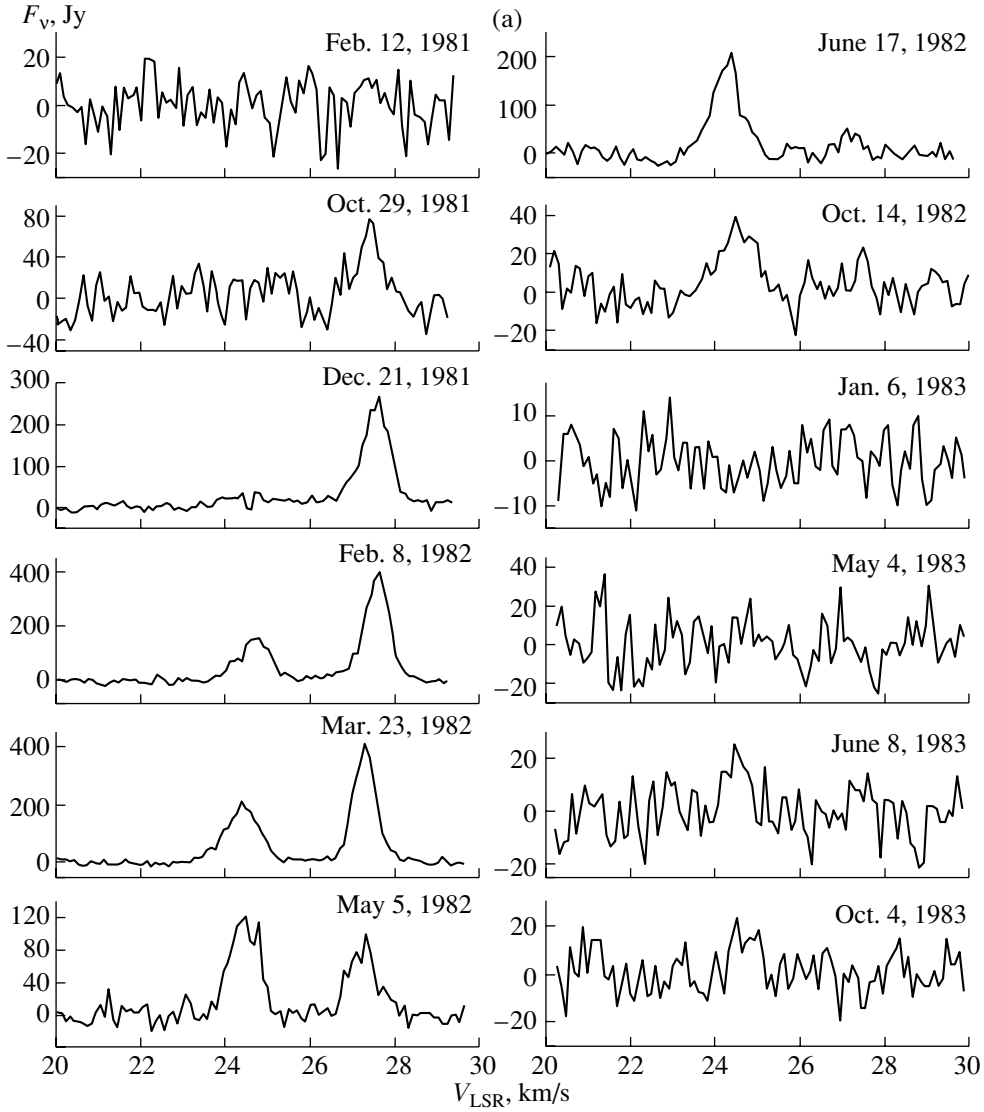

Fig. 1. Profiles of the H_2O maser line of R Cas. The horizontal and vertical axes plot the radial velocity with respect to the LSR V_{LSR} and the flux density F_ν.

infrared variability was $P \approx 429.6^d$, similar to the period in the visual. The maximum of the infrared brightness takes place $\sim 0.2P$ after the visual maximum. Like other variable red giants, R Cas is losing mass. Weigelt and Yudin [14] used the data of [13] to construct a model for the dust envelope of R Cas; they estimate that the mass-loss rate of the star is $\dot{M} \sim 6 \times 10^{-7} M_\odot$/year. Various estimates yield distances to the star ranging from $d = 107$ pc [15] (based on the HIPPARCOS data) to $d = 270$ pc [16].

Molecular radio emission is generated in the circumstellar gas–dust envelope formed during the mass-loss process. R Cas is a known source of OH [17], H_2O [18, 19], and SiO [20] maser emission. Thermal radio emission in lines of SiO, HCN [15], and CO [21] has also been observed. The radial velocity of the star relative to the Local Standard of Rest (LSR) derived from observations of the thermal molecular lines is $V_* = +26$ km/s [15].

2. OBSERVATIONS

Our long-term monitoring of R Cas in the H_2O line was performed on the 22-m radio telescope of the Pushchino Radio Astronomy Observatory (Astro Space Center, Lebedev Physical Institute, Russian Academy of Sciences) near Moscow. The receiver included a 22-GHz maser amplifier cooled with liquid helium (the system noise temperature was $T_N = 200$–300 K). In the 1993–2003 observations, we used a cooled FET 22-GHz amplifier ($T_N = 80$–120 K). The spectral backend was a 128-channel

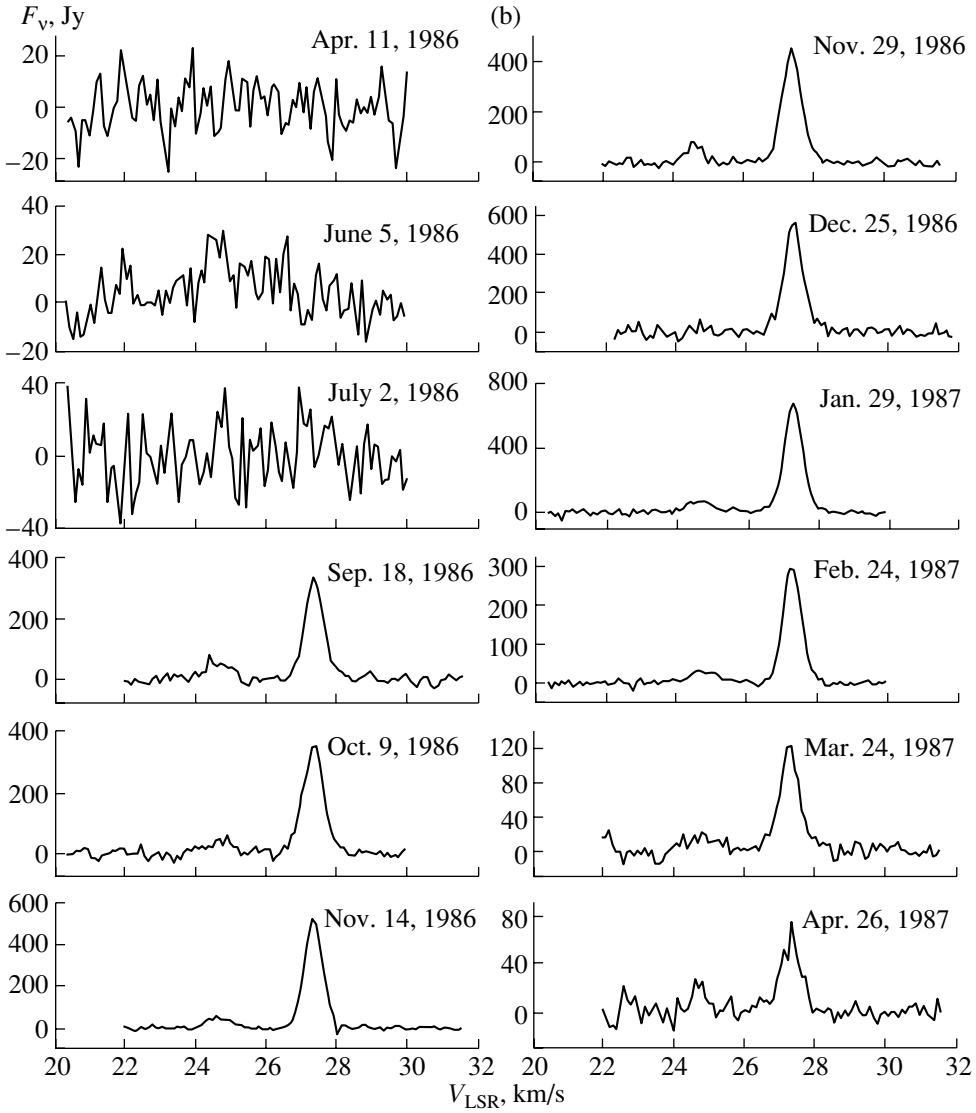

Fig. 1. (Contd.)

filter-bank spectrometer constructed at the Sternberg Astronomical Institute; its frequency resolution was 7.5 kHz (0.101 km/s in radial velocity at the frequency of the 22-GHz H_2O line). The observational technique and instrumentation are described in detail in [22]. In observations of variability of the H_2O line, special attention must be paid to correcting for absorption in the Earth's atmosphere; this problem is discussed in [7].

The results of the observations are presented in Figs. 1a–1f as spectral profiles of the H_2O line. The vertical axis plots flux density F_ν (in Jy), and the horizontal axis, the radial velocity relative to the LSR (V_{LSR}, km/s).

Figures 2–4 show the integrated flux of the H_2O maser emission F_{int} as a function of time for various radial-velocity intervals in the H_2O line profile. For comparison, the visual light curve (AFOEV data) and variability curves for the SiO maser radio emission [23, 24] are also given.

The two radial-velocity intervals in the H_2O line profile where the maser emission is concentrated are 23–25 and 25–29 km/s. Figures 2 and 3 present the integrated fluxes for these intervals. In addition, emission appeared at velocities of 29–31 km/s in 1988–1989; a plot for this range is given in Fig. 3. In 1993–1998, weak H_2O emission was only occasionally recorded at 25–29 km/s; Fig. 4 shows the data for this range. After 1998, H_2O emission was never detected in our observations. The upper limits on the flux density were of the order of 5–10 Jy.

The most conspicuous features of the plots in Figs. 2 and 3 are the strong flares of the R Cas maser emission: at the end of 1981 and beginning of

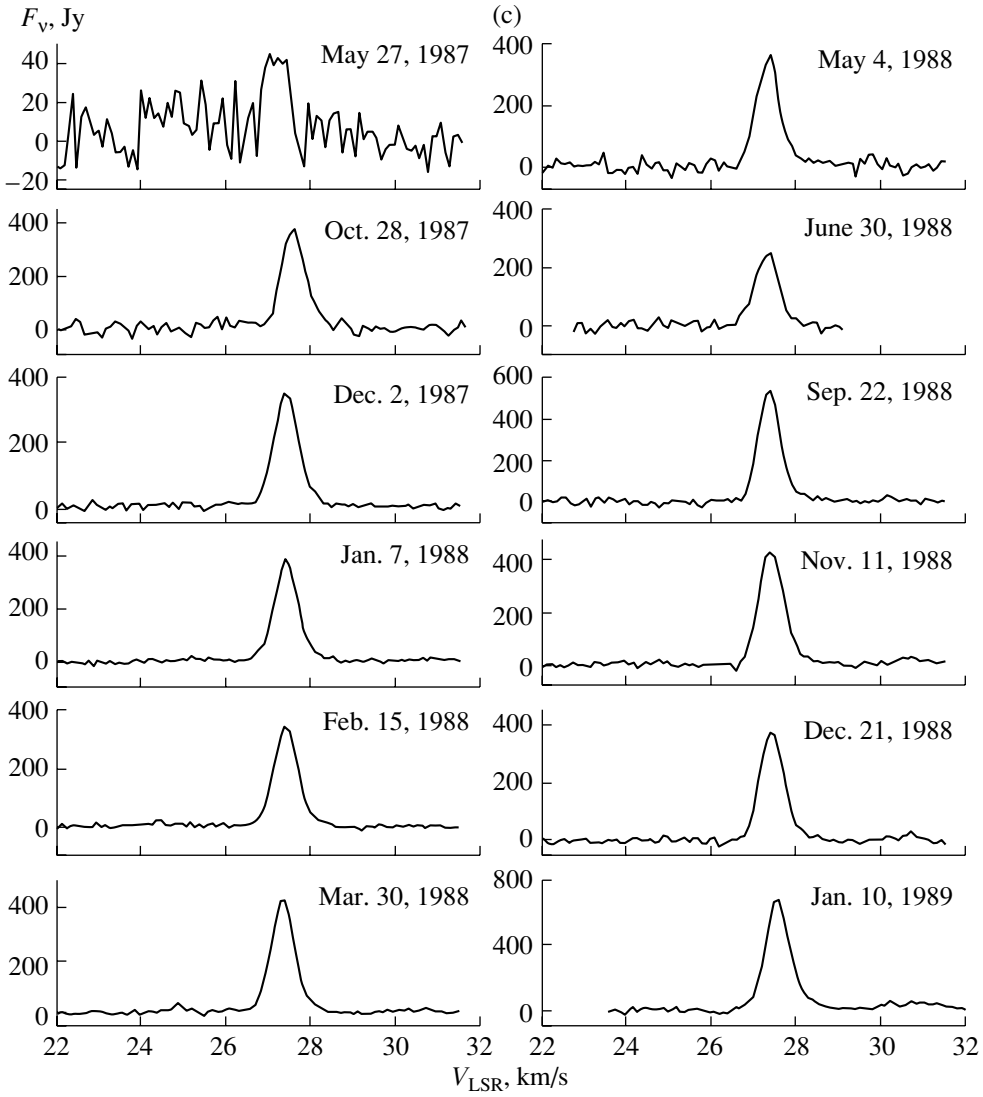

Fig. 1. (Contd.)

1982, when the flux density of the feature at $V_{LSR} = 27.6$ km/s reached 400 Jy (see line profiles in Fig. 1a), and from the end of 1986 to the beginning of 1989, when three successive flares with peak flux densities of up to 750 Jy took place at $V_{LSR} = 25$–29 km/s (Fig. 1d). The flare at the beginning of 1987 was also accompanied by a flux rise at 23–25 km/s, and the flare at the beginning of 1989, by a flux rise at 29–31 km/s. However, the data for the latter V_{LSR} interval are incomplete. Not all profiles in Fig. 1 cover this radial-velocity interval; therefore, we cannot trace the full evolution of the H_2O spectrum in this V_{LSR} range.

In time intervals when we could record the H_2O emission of R Cas, the fluxes in the various velocity ranges were correlated with the visual variations. The H_2O flux variations were delayed relative to the optical variation by $\Delta\varphi \sim (0.2$–$0.3)P$. The same effect is present in the variability of the SiO maser emission in Figs. 2 and 3. Brand *et al.* [25] also found a correlation between the H_2O maser flux of R Cas and its visual light curve and a delay of the H_2O maxima relative to the visual maxima. The observations of Brand *et al.* [25] cover the interval from September 1987 to July 1999 and partly overlap with our data; they also noted flares of the R Cas H_2O maser emission at the beginning of 1988 and 1989.

Similar correlations and phase delays $\Delta\varphi$ of the H_2O maser variations relative to the visual light curve have been observed for other stellar H_2O masers (see [5–9] and references therein). The nature of these correlations is discussed in the next section.

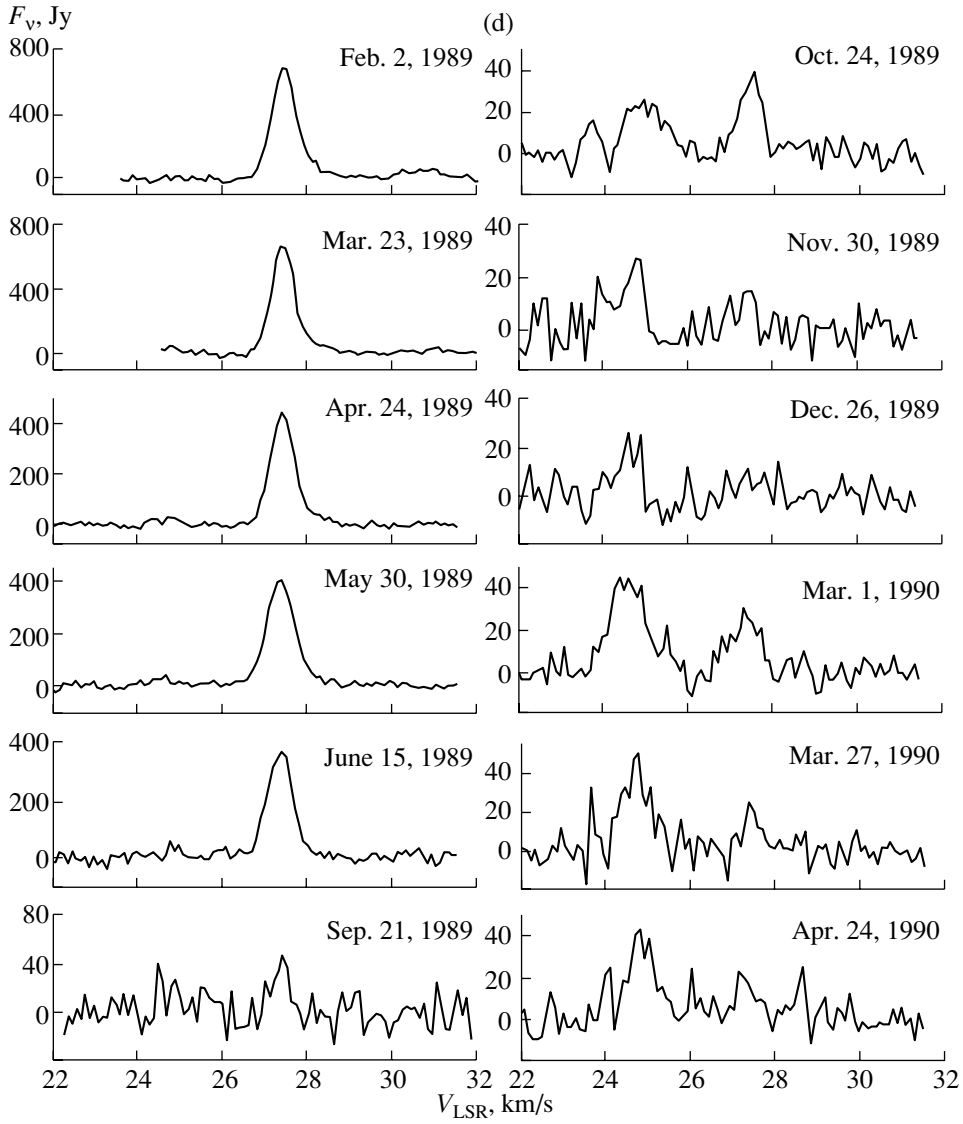

Fig. 1. (Contd.)

3. DISCUSSION

Monitoring of the H_2O maser emission of R Cas over several cycles of the optical variability was carried out in [23, 25]. Some individual H_2O line observations of R Cas have also been published. The table lists published observations of R Cas (the most detailed observations are [25], which are discussed above). The columns give (1) the dates of the observations, the (2) peak flux density, (3) radial velocity, (4) full width at half maximum (FWHM), and (5) integrated flux $F_{\rm int}$ of the features, and (6) references. As a rule, the $F_{\rm int}$ values were taken from tables in the cited papers or were estimated from the $F_{\rm peak}$ and FWHM of published line profiles. We assumed the lines had Gaussian profiles, so that $F_{\rm int} = 0.079 F_{\rm peak} {\rm FWHM}(\pi/4\ln 2)^{1/2}$ (0.079 is the factor to convert from Jy km/s to 10^{-20} W/m^2). Estimated parameters are marked with $\sim$. If two or three spectral features were observed in the profile, each is given in a separate row. When the literature data coincide in time with our observations, the agreement of the integrated fluxes is good.

The observed correlation between the visual brightness of the star and the intensity of the H_2O maser radiation is explained in the literature by two possible mechanisms: (1) periodic heating of circumstellar dust particles (and, as a consequence, of the ambient gas) by the variable infrared radiation of the star [26] and (2) the effect of shock waves driven by the stellar pulsations on the maser-generation region [2].

The former mechanism results in a small and nearly constant phase delay $\Delta\varphi$ of the $F_{\rm int}(H_2O)$ variations relative to the visual light curve (the same

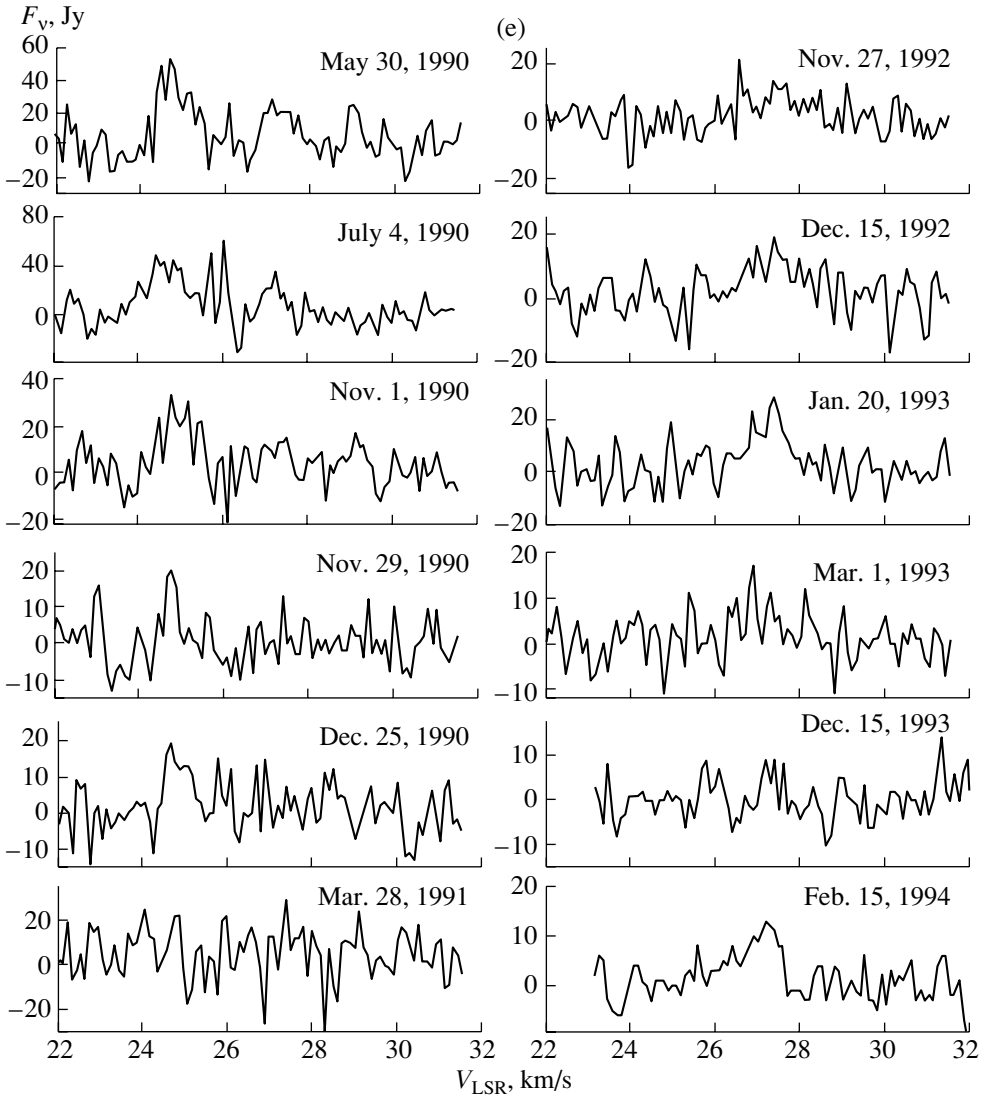

Fig. 1. (Contd.)

$\Delta\varphi$ as for the infrared curve, as is noted in the Introduction [13]). A delay is indeed observed for the H_2O masers of Mira variables [3–5] and, according to our observations and the data of [25], of R Cas. However, a more detailed analysis of the model of [26] demonstrates that it encounters difficulties because the secondary infrared radiation from dust heated by the stellar continuum disrupts the population inversion of the H_2O molecules, decreasing the maser intensity.

We consider a model in which the H_2O masers are excited by shock waves arising from pulsations of the star [2] to be more realistic. According to [2], the H_2O masers are located in the circumstellar envelope, at a distance of $r \sim 10^{14}$ cm from the center of the star, in a so-called quasi-stationary layer (QSL). The existence of QSLs has been deduced for many Miras, based on the observation of molecular absorption lines that have nearly zero radial velocity relative to the star (hence the name) [38] and also from infrared interferometry [39]. A QSL is a "temporary pause" for the material leaving the star. The material flows into the interstellar medium from the outer boundary of the QSL. At the same time, in the absence of new mass ejections, material can fall back onto the star from the inner boundary of the QSL. A QSL can disappear due to the loss of material, sometimes for several variability cycles [38]. The physical conditions in the QSL ($T_{kin} \sim 800$ K, $n \sim 10^9$ cm^{-3}) are favorable for pumping of the H_2O masers. The main excitation sources of the masers are shock waves from the star, which cross the QSL at a velocity V_{sh} that is no higher than ~10 km/s. Faster shocks would dissociate molecules and, at $V_{sh} > 15$–20 km/s, ionize the gas, making the existence of masers impossible.

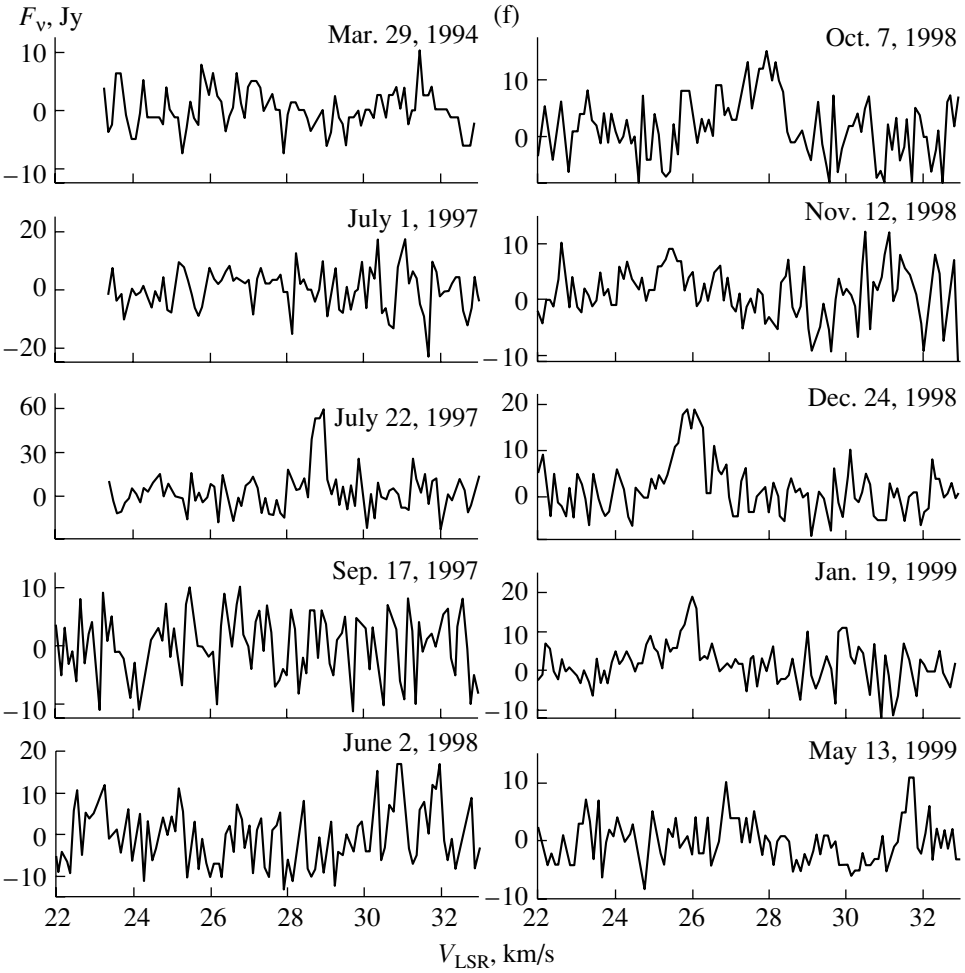

Fig. 1. (Contd.)

The propagation of a shock from the photosphere to the region of the H_2O masers is accompanied by a number of observational manifestations [8]. Near the photosphere, the shock radiates optical emission lines, including Balmer hydrogen lines. With growing distance from the photosphere, the shock weakens, and the ionization of gas ceases when $V_{sh} < 15$ km/s. The shock then affects the inner layers of the circumstellar envelope, where the SiO masers are located. The SiO masers arise from transitions between the rotational levels of excited vibrational states ($v = 1, 2, 3, 4$). Therefore, pumping of the SiO masers requires higher densities and temperatures than pumping of the H_2O masers. The most suitable conditions for this may be just behind the shock fronts. In SiO observations of R Cas [40, 41], the time when a shock arrives at the SiO maser region can be traced directly. The propagation of the shock is accompanied by broadening of the SiO line profile and rotation of the plane of linear polarization of the emission.

Interferometric observations of H_2O and SiO circumstellar masers (for R Cas obtained in, e.g., [33, 42, 43]) confirm that the SiO masers are located closer to the stellar photosphere than the H_2O masers. A comparison of the H_2O and SiO variability can be used to trace the shock's propagation in various layers of the circumstellar envelope. There is a flare in the 1.35-cm line when the shock reaches the H_2O maser region. We have traced this sequence of events in the Mira variable R Leo: after an Hα flare in the optical spectrum of R Leo in May 1996, there was an increase in the SiO and then in the H_2O maser emission [8]. Thus, the true time delay relative to the visual maximum could be $\Delta t = (\Delta\varphi + n)P$; i.e., it can be increased by an integer number of variation periods of the star. In this case, Δt is the travel time of the shock (related to the visual maximum) from the photosphere to the H_2O maser layer in the circumstellar envelope. We drew the same conclusion earlier based on data on the H_2O maser variability of the Mira RS Vir [10]. A shock with a velocity of $\sim$10 km/s can propagate to fairly large distances

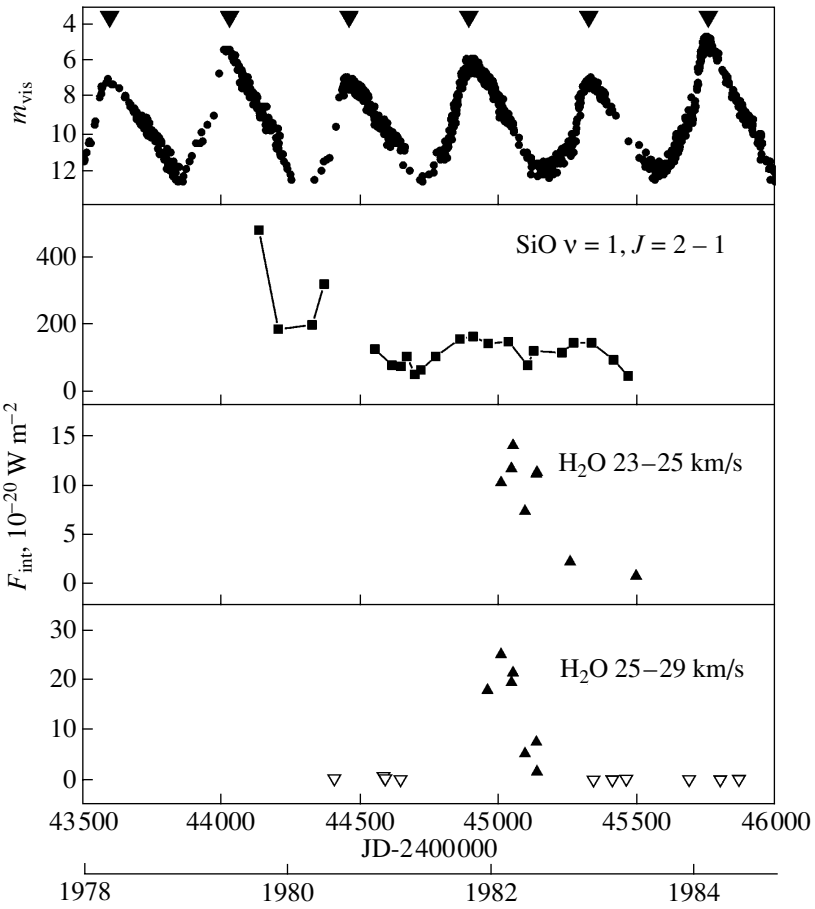

Fig. 2. Variability of the emission of R Cas in 1980–1984. Shown from top to bottom are the visual light curve [AFOEV data; the downward pointing filled triangles mark epochs of visual maxima calculated with the GCVS elements (1)]; the integrated flux in the SiO $\nu = 1$, $J = 2$–1 line [23]; and the integrated fluxes in two segments of the profile of the H_2O 6_{16}–5_{23} line (our data). The downward pointing open triangles show upper limits.

from the stellar surface, because its radiative losses are low. The main energy losses of the shock are due to the radiation of heated post-shock dust and the dissociation of molecular hydrogen (if it is present in the circumstellar medium in appreciable quantities) [44].

Compared to the optical light curve, the temporal behavior of the H_2O emission of R Cas (Figs. 2–4) indicates that the increase of $F_{int}(H_2O)$ could be associated not with the latest preceding visual maximum but instead with one of the earlier maxima. The strong flares of 1986–1989 were preceded by a bright visual maximum ($V < 5^m$) at JD $\sim$ 2445740. The maximum two periods before the flare of 1981–1982 (JD $\sim$ 2444000, $V \sim 5.5^m$) was also brighter than average. The episodes of enhanced emission in 1992–1994 and 1997–1998 also corresponded to bright visual maxima occurring $(2–3)P$ earlier. Thus, the number of variability periods n corresponding to the delay of the shock's effect on the H_2O maser region (at $R \sim (1.5 - 5) \times 10^{14}$ cm in the case of R Cas) can reach 2–4 (or 2.5–4.7 years); this matches the time for the shock's propagation from the photosphere to the QSL.

The structure of the H_2O line profile of R Cas has remained stable since the 1970s. The V_{LSR} intervals where the main H_2O emission is concentrated are located near 24.8, 27.6, and 29.5 km/s. The emission at these velocities is not always present. In particular, emission at 24.8 km/s was observed in 1982 (Fig. 1a) and then weaker emission at this velocity from 1986 to 1990. The emission at 27.6 and 29.5 km/s was strongest at the beginning of 1989. The flare of the 27.6-km/s feature at the beginning of 1989 was not accompanied by an appreciable increase of the emission at 24.8 km/s, and, likewise, the emission at 27.6 km/s remained weak as the 24.8-km/s feature grew in 1990. These peculiarities can be explained by the consecutive impact of the shock on different regions of maser emission in the circumstellar envelope (see below).

Literature data for the H_2O maser emission of R Cas

Date	$F_{\rm peak}$, Jy	$V_{\rm LSR}$, km/s	FWHM, km/s	Estimated $F_{\rm int}$, 10^{-20} W/m^2	Reference
Sep. 9, 1971	56	25.9	1.0	1.0	[18]
Sep. 16–20, 1976	<6				[27]
Nov. 28, 1976	3.8	~25	~1	0.72	[28]
Dec. 6, 1980	7.4	24.6	2.0	2.7	[23]
Dec. 6, 1980	6.4	27.4	2.2		[23]
Apr. 2, 1981	8.6	24.7	2.5	2.3	[23]
Nov. 28, 1981	14	24.7	1.9	12	[23]
Nov. 28, 1981	98	27.5	1.1		[23]
Nov. 28, 1981	5.8	29.3	3.4		[23]
June 4, 1982	180	24.7	0.9	18	[23]
June 4, 1982	67	27.5	0.9		[23]
Aug. 21, 1982	3.5	28.0	2.1		[23]
Dec. 2, 1982	4.3	24.7	1.4	0.24	[23]
June 17, 1983	22	24.6	0.9	3.4	[23]
June 17, 1983	6.4	29.6	1.2		[23]
Oct. 9, 1983	9.6	24.8	2.2	1.1	[23]
Oct. 5–9, 1983	~9	~25	~2	1.4	[29]
Mar. 22–25, 1986	<7				[30]
May 20–30, 1986	<15				[31]
June 2–6, 1986	<15				[31]
Mar. 12–24, 1986	<15				[31]
Sep. 10, 1987	66.2	27.3	2.6	4.0	[1]
Feb. 11, 1990	23	24.6	~1	2.9	[32]
Feb. 11, 1990	20	27.3	~1		[32]
June 2–3, 1990	40	24.1	~1	5.9	[33]
June 2–3, 1990	35	27.6	~ 1		[33]
Apr. 14, 1993	3.7	24.1–28.0*		0.32	[34]
Apr. 26–May 11, 1993	5.5	26.1	4.0**	1.7	[35]
Apr. 27, 1994	~6	~ 26	~2	0.9	[36]
Mar. 26, 1995				1.38	[37]
1987–1999	see text				[25]

* Emission at the ~1 Jy level in this $V_{\rm LSR}$ interval.

** The observations were carried out with a resolution of 0.7 km/s; probably, blended features.

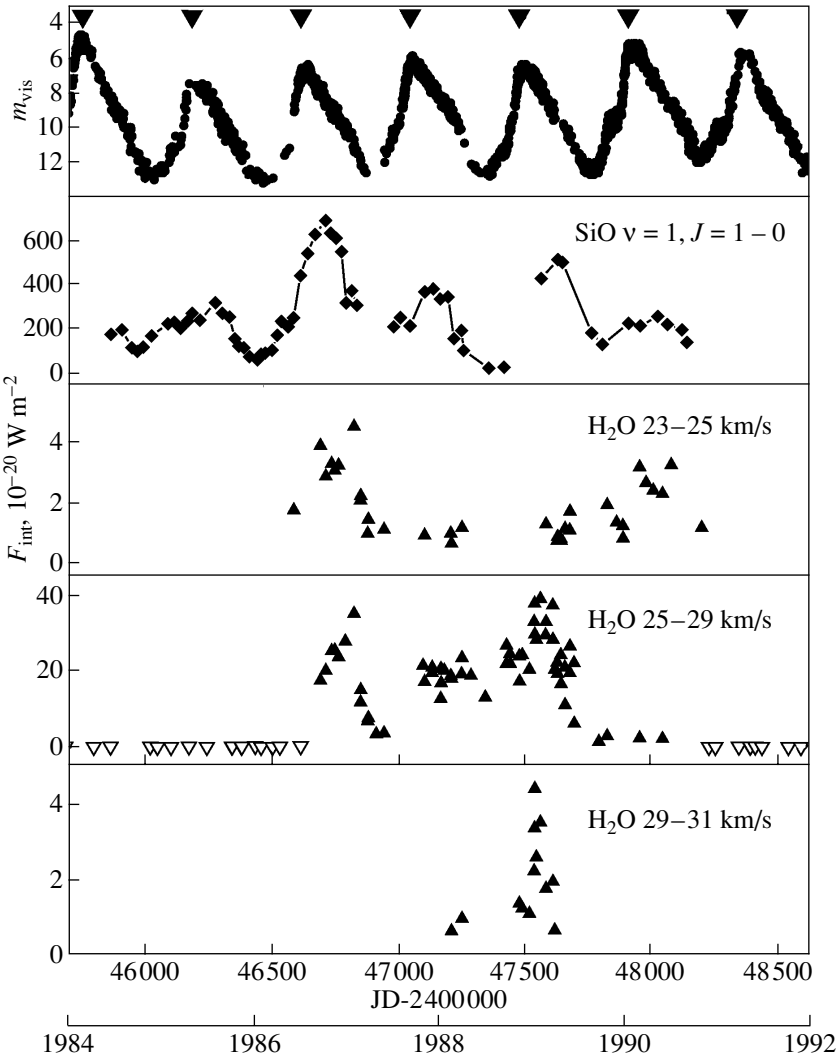

Fig. 3. Variability of the emission of R Cas in 1984–1991. Shown from top to bottom are the visual light curve (AFOEV data); the integrated flux in the SiO $\nu = 1$, $J = 1$–0 line [24]; and the integrated fluxes in three segments of the profile of the H_2O 6_{16}–5_{23} line (our data).

The emission at 27.6 km/s is of special interest. According to the interferometric observations of [36], emission at $V_{LSR} = 27.6$ km/s was observed toward the optical position of the star; i.e., the emission at this velocity was projected against the stellar disk. This is not a coincidence. In [45], we proposed a model in which the H_2O maser emission observed toward the stellar disk results from amplification of the stellar radio continuum by masing H_2O molecules. Note that the velocity 27.6 km/s is slightly more positive than the stellar velocity, $V_* = 26$ km/s. Consequently, we see material falling onto the star from the QSL when we observe the H_2O line toward the stellar disk; the 27.6-km/s feature is the image of the stellar "radio photosphere" [46], amplified by the H_2O molecules. The emission at more negative (24.8 km/s) and positive (29.5 km/s) velocities arises in the material flowing from the QSL toward the observer and toward the star, respectively.

In this connection, we note that variations of the 27.6-km/s feature precede those of the 24.8-km/s feature. This is especially visible in the line profiles (Figs. 1a, 1d). A natural explanation is that, as it propagates from the star, the shock first crosses material falling back toward the star (with more positive radial velocities) and then layers moving away from the star toward the observer (with more negative radial velocities). The variations of the 29.5-km/s feature (Fig. 3) also slightly precede those of the 27.6-km/s flux. It is possible that this is a masing region that is closer to the photosphere, which falls onto the star with a still greater velocity, so that the shock

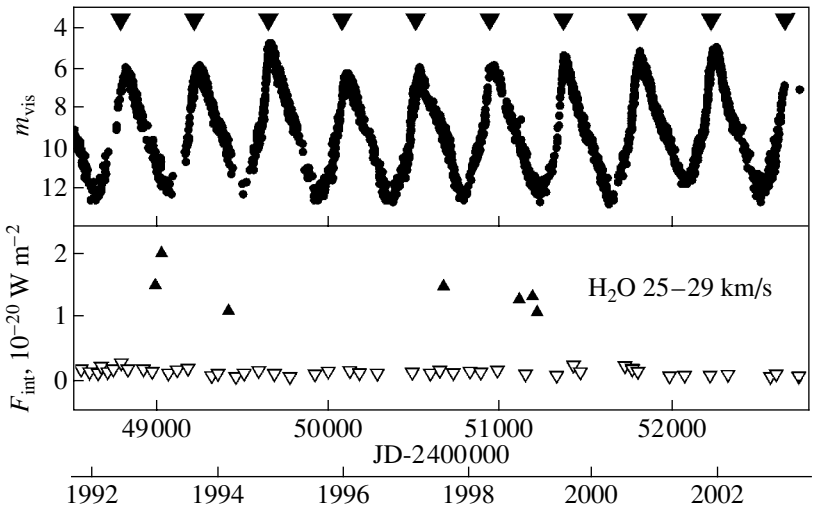

Fig. 4. Variability of the emission of R Cas in 1992–2003. Shown from top to bottom are the visual light curve (AFOEV data) and the integrated flux of the H_2O 6_{16}–5_{23} line in the range $V_{LSR} = 25$–29 km/s (our data).

reaches it earlier than the 27.6-km/s region. The variability of the radial velocity of the H_2O emission peak in R Cas was also explained by the propagation of a shock in a medium with a velocity gradient by Pashchenko [4].

The R Cas H_2O maser is a so-called "transient" maser [8]. The H_2O emission of such masers experiences strong flares from time to time, then decreases, and can fall below the detection threshold of our monitoring. Our sample includes the "transient" maser sources U Ori [9], R Leo [8], W Hya [7], R LMi, U Aur, and R Cas. Another group of stars has fairly stable masers. During our monitoring, their H_2O flux density has varied less and has never fallen below the detection threshold. Examples of such objects are RR Aql [5], RS Vir [10], R Aql, S CrB, X Hya, and others. Both groups of maser sources have a common property: in both "stable" masers and "transients" (at epochs when they are detected), the variations of the H_2O line "brightness" follow the optical variations with some phase lag.

The model of [2] explains the transience of the masers as an effect of the episodic appearance and subsequent disappearance of a QSL around the star. The QSL can arise from a one-time ejection of a large amount of material from the star, e.g., by a shock that is much stronger than average. Such shocks can arise once per several periods of the star during especially "high" brightness maxima. There is another possibility: as was shown by Wood [47], shocks propagating in the envelope of a Mira variable can overtake one another and merge. This also creates an enhanced shock and results in increased mass loss. In stable masers, either the QSL is always present or the mass-loss flux is sufficiently large to support maser emission without substantial interruptions.

4. CONCLUSIONS

We have obtained profiles of the 1.35-cm H_2O maser emission of the Mira variable R Cas for 1981–2003 and traced the correlation of the H_2O flux variations for both the line as a whole and for individual spectral components. Strong maser flares were recorded at the end of 1981 and the beginning of 1982 and from the end of 1986 to the beginning of 1989. The maser emission subsequently decreased: only isolated episodes of weak emission were observed, and the H_2O emission of R Cas was usually below the detection threshold for our observations (<5–10 Jy). Data from the literature (see the table) show that a similar interval when the R Cas H_2O maser was weak occurred in the 1970s. We classify the R Cas H_2O maser source as a "transient" maser, for which H_2O emission appears in isolated flares of up to several hundreds of Jy, followed by a gradual decrease in the maser emission lasting from several months to two to three years, when the emission can fall to very low levels, not exceeding a few Jy.

In time intervals when the H_2O emission of the R Cas maser emission was detected, the H_2O maser variations were correlated with the visual variations of the star, with a phase delay $\Delta\varphi$ of the $F_{int}(H_2O)$ relative to the optical variations of $\sim$0.2–0.3 of the variability period ($P = 430^d$). We favor a model in which the H_2O maser emission is excited by shock waves driven by the stellar pulsations [2]. In this model, the masing H_2O molecules are concentrated in a quasi-stationary layer (QSL) that represents a local density enhancement in the circumstellar envelope. The

layer can appear during episodes of intensified mass loss from the star, then disappear for some time. The decrease of the maser emission could be due to the temporary absence of a QSL. If the variations of the R Cas H_2O maser are due to the periodic effect of pulsation-driven shocks, the travel time for a shock from the photosphere to the QSL can reach $(2-4)P$. The model also explains the observed variations of the structure of the H_2O line profile as an effect of the consecutive impact of a shock first on material falling from the QSL back to the star (the Doppler features at radial velocities $V_{\rm LSR} = 27.6-29.5$ km/s $> V_*$), then on material flowing from the QSL toward the observer (the feature at 24.5 km/s).

ACKNOWLEDGMENTS

The authors are grateful to E.E. Lekht, V.A. Samodurov, and A.M. Tolmachev for help with the observations. The RT-22 radio telescope is supported by the Ministry of Industry and Science of the Russian Federation (registration number 01-10). This research made use of the visual observations of the Association Française des Observateurs d'Étoiles Variables (AFOEV) from the SIMBAD database of the Centre des Données Astronomiques de Strasbourg (France).

REFERENCES

1. G. Comoretto, F. Palagi, R. Cesaroni, *et al.*, Astron. Astrophys., Suppl. Ser. **84**, 179 (1990).
2. G. M. Rudnitskiĭ and A. A. Chuprikov, Astron. Zh. **67**, 293 (1990) [Sov. Astron. **34**, 147 (1990)].
3. I. I. Berulis, E. E. Lekht, M. I. Pashchenko, and G. M. Rudnitskiĭ, Astron. Zh. **60**, 310 (1983) [Sov. Astron. **27**, 179 (1983)].
4. M. I. Pashchenko, Astron. Tsirk., No. 1543, 9 (1990).
5. I. I. Berulis, E. E. Lekht, V. A. Munitsyn, and G. M. Rudnitskiĭ, Astron. Zh. **75**, 394 (1998) [Astron. Rep. **42**, 346 (1998)].
6. M. I. Pashchenko and G. M. Rudnitskiĭ, Astron. Zh. **76**, 363 (1999) [Astron. Rep. **43**, 311 (1999)].
7. G. M. Rudnitskiĭ, E. E. Lekht, and I. I. Berulis, Pis'ma Astron. Zh. **25**, 467 (1999) [Astron. Lett. **25**, 398 (1999)].
8. V. F. Esipov, M. I. Pashchenko, G. M. Rudnitskiĭ, and S. V. Fomin, Pis'ma Astron. Zh. **25**, 775 (1999) [Astron. Lett. **25**, 672 (1999)].
9. G. M. Rudnitskij, E. E. Lekht, J. E. Mendoza-Torres, *et al.*, Astron. Astrophys., Suppl. Ser. **146**, 385 (2000).
10. E. E. Lekht, J. E. Mendoza-Torres, G. M. Rudnitskij, and A. M. Tolmachev, Astron. Astrophys. **376**, 928 (2001).
11. P. N. Kholopov, N. N. Samus', V. P. Goranskiĭ, *et al.*, *General Catalog of Variable Stars* (Nauka, Moscow, 1985), Vol. 1 [in Russian].
12. J. E. Isles, D. R. B. Saw, and J. Brit, Astron. Assoc. **99**, 121 (1989).
13. A. E. Nadzhip, A. M. Tatarnikov, V. I. Shenavrin, *et al.*, Pis'ma Astron. Zh. **27**, 376 (2001).
14. G. Weigelt and B. F. Yudin, Astron. Zh. **78**, 594 (2001) [Astron. Rep. **45**, 510 (2001)].
15. J. H. Bieging, S. Shaked, and P. D. Gensheimer, Astrophys. J. **543**, 897 (2000).
16. V. Bujarrabal, J. Gómez-González, and P. Planesas, Astron. Astrophys. **219**, 256 (1989).
17. Nguyen-Quang Rieu, R. Fillit, and M. Gheudin, Astron. Astrophys. **14**, 154 (1971).
18. B. E. Turner and R. H. Rubin, Astrophys. J. **170**, L113 (1971).
19. W. J. Wilson, P. R. Schwartz, G. Neugebauer, *et al.*, Astrophys. J. **177**, 523 (1972).
20. N. Kaifu, D. Buhl, and L. E. Snyder, Astrophys. J. **195**, 359 (1975).
21. C. Loup, T. Forveille, A. Omont, and J. F. Paul, Astron. Astrophys., Suppl. Ser. **99**, 291 (1993).
22. E. E. Lekht, J. E. Mendoza-Torres, and R. L. Sorochenko, Astrophys. J. **443**, 222 (1995).
23. L.-Å. Nyman and H. Olofsson, Astron. Astrophys. **158**, 67 (1986).
24. J. Alcolea, J. R. Pardo, V. Bujarrabal, *et al.*, Astron. Astrophys., Suppl. Ser. **139**, 461 (1999).
25. J. Brand, L. Baldacci, and D. Engels, *IAU Symp. No. 206: Cosmic Masers: From Protostars to Black Holes,* Ed. by V. Migenes and M. Reid (ASP, San Francisco, 2002), p. 310.
26. A. M. Gómez Balboa and J. R. D. Lépine, Astron. Astrophys. **159**, 166 (1986).
27. J. H. Spencer, K. J. Johnston, J. M. Moran, *et al.*, Astrophys. J. **230**, 449 (1979).
28. F. M. Olnon, A. Winnberg, H. E. Matthews, and G. V. Schultz, Astron. Astrophys., Suppl. Ser. **42**, 119 (1980).
29. A. P. Lane, K. J. Johnston, P. F. Bowers, *et al.*, Astrophys. J. **323**, 756 (1987).
30. R. Barvainis and S. Deguchi, Astron. J. **97**, 1089 (1989).
31. I. V. Gosachinskiĭ, R. A. Kandalyan, F. S. Nazaretyan, *et al.*, Astrofizika **32**, 365 (1990).
32. K. M. Menten and G. J. Melnick, Astrophys. J. **377**, 647 (1991).
33. F. Colomer, M. J. Reid, K. M. Menten, and V. Bujarrabal, Astron. Astrophys. **355**, 979 (2000).
34. J. A. Yates, R. J. Cohen, and R. E. Hills, Mon. Not. R. Astron. Soc. **273**, 529 (1995).
35. H. Takaba, N. Ukita, T. Miyaji, and M. Miyoshi, Publ. Astron. Soc. Jpn. **46**, 629 (1994).
36. E. González-Alfonso, J. Cernicharo, J. Alcolea, and M. A. Orlandi, Astron. Astrophys. **334**, 1016 (1998).
37. M. Szymczak and A. M. Le Squeren, Mon. Not. R. Astron. Soc. **304**, 415 (1999).
38. K. H. Hinkle, W. W. G. Scharlach, and D. N. B. Hall, Astrophys. J., Suppl. Ser. **56**, 1 (1984).
39. B. Menneson, G. Perrin, G. Chagnon, *et al.*, Astrophys. J. **579**, 446 (2002).
40. F. O. Clark, T. H. Troland, F. J. Lovas, and P. R. Schwartz, Astrophys. J. **244**, L99 (1981).

41. F. O. Clark, T. H. Troland, and D. R. Johnson, Astrophys. J. **261**, 569 (1982).
42. R. B. Phillips, G. R. Sivakoff, C. J. Lonsdale, and S. S. Doeleman, Astron. J. **122**, 2679 (2001).
43. R. B. Phillips, A. H. Straughn, S. S. Doeleman, and C. J. Lonsdale, Astrophys. J. **588**, L105 (2003).
44. G. M. Rudnitskij, Astrophys. Space Sci. **251**, 259 (1997).
45. G. M. Rudnitskij, *From Miras to Planetary Nebulae: Which Path for Stellar Evolution?*, Ed. by M. O. Mennessier and A. Omont (Frontières, Gif sur Yvette, 1990), p. 268.
46. M. J. Reid and K. M. Menten, Astrophys. J. **476**, 327 (1997).
47. P. R. Wood, Astrophys. J. **227**, 220 (1979).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 393–398. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 431–437.
Original Russian Text Copyright © 2004 by Ismailov.

Periodic Variability of the T Tauri Star DI Cephei

N. Z. Ismailov

Shemakha Astrophysical Observatory, National Academy of Sciences of Azerbaijan, Shemakha, 373243 Azerbaijan

Received September 10, 2003; in final form, November 10, 2003

Abstract—Original spectroscopic and photometric observations are used together with data from the literature to compile a master light curve for DI Cep and to analyze individual data series for periodicity. The spectroscopic data reveal the period $P = 9.24^{\rm d}$. Data from different authors show the nine-day quasi-periodicity in the brightness variations fairly reliably, although the period cannot be derived from the entire master set of photometric data. The period $P = 18.28^{\rm d}$ is also derived from the V brightnesses. We find variations of the zero epoch of the nine-day period, possibly due to the changing location of an accretion-disk hot spot. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Despite the numerous observations that have been accumulated for individual T Tauri stars, the variability mechanisms for this class of star remain incompletely understood. The proposed origins of the variability are different for classical T Tauri stars (CTTS) and weak-emission T Tauri stars (WTTS) [1]. It is thought that, whereas classical T Tauri variability is usually due to changing parameters of the accretion zone, the brightness and spectral variations of weak-emission T Tauri stars are due to a spotted photosphere and rotational modulation. Only long series of observations can enable the identification of general trends that can improve our understanding of the nature of the variability of young stars. In this paper, we investigate long-term variations of the spectrum and brightness of the T Tauri star DI Cep using many years of photoelectric and spectroscopic observations.

DI Cep is a classical T Tauri star with $W(\mathrm{H}\alpha) > 100$ Å [2]. Joy [3] was the first to report the detection of emission lines of CaII, FeII, HeI, and HI and classified the spectrum as dK3e. The observation of strong CaII H and K lines is also noted in [4]. In other studies, the spectral type of DI Cep has been estimated as G5 [5] and G8V [5, 6]. Ismailov [7] observed the spectral type to vary in the range from F4 to K5, with the most probable spectral type being G5–G(7.5 ± 1.5)V.

Appreciable variations of the intensities of emission lines of DI Cep over 10–15 min were detected by Bastian and Mundt [8]. Grinin *et al.* [2] and Krasnobabtsev [6] also reported variations of the line spectrum. It was demonstrated in [6, 9] that the emission-line intensities varied with a cycle of ~16 days, interpreted as being due to the presence of active regions on the stellar surface.

Based on an analysis of the brightness of DI Cep on 240 plates, Kholopov [10] showed that the star's brightness experienced slow variations with an amplitude of 1^m per year, which were superimposed by more rapid fluctuations by 0.5^m within a day. Using observations in a photometric system free from the influence of emission lines, Petrov [11] recorded rapid brightness variations with amplitudes up to 0.8^m in both the ultraviolet and red. Flare brightness variability of DI Cep was revealed in the UBV observations of Keleman [12], who observed large-amplitude brightness variations on JD 2 445 588, by 3^m in U, 0.8^m in B, and 0.4^m in V. The long series of observations of Kardopolov and Filip'ev [13] showed brightness variations with a V-band amplitude of 0.4^m within one season.

The spectroscopic observations [8, 9, 14] revealed rapid variations of the spectrum on time scales from several minutes to an hour. Ismailov [15] also detected flare variability of the star's spectrum and brightness with a duration of about nine days and demonstrated that the brightness increase could be due to the formation of a hot spot with a temperature of about 8000 K occupying about 2% of the total surface area of the star. A similar conclusion was later reached in [16]. According to Gahm and Petrov [17], the star's variability results from variable circumstellar extinction and emission activity. A magnetospheric disk accretion model for DI Cep was suggested in [18]. Studies of the ultraviolet and visible brightness in [19] showed variations with a probable 11-day period.

1063-7729/04/4805-0393$26.00 © 2004 MAIK "Nauka/Interperiodica"

Results of Fourier analysis of the spectroscopic and photometric observations for DI Cep

No.	Observation series	Most probable frequencies, day^{-1}	Periods, days
		Spectroscopic data	
1	$W_\lambda(H\alpha)$ from [2]	0.0725, 0.1085	13.79, 9.24
2	$W_\lambda(H\beta)$ from [2]	0.1081, 0.073	9.24, 13.6
3	$W_\lambda(H\beta)$ from [9]	0.1096	9.12
4	$W_\lambda(H\delta)$ from [9]	0.1081	9.24
		V-band photometry	
1	JD 2 443 686–2 443 762	0.056, 0.112, 0.072	17.69, 8.91, 13.89
2	JD 2 444 023–2 444 168	0.050, 0.095	20.00
3	JD 2 444 393–2 444 530	0.028	35.71
4	JD 2 444 822–2 444 942	0.051, 0.013	19.16, 7.5
5	JD 2 446 593–2 446 663	0.097, 0.063	10.26, 15.8
6	JD 2 449 934–2 450 045	0.053, 0.09	18.28, 11.1

2. OBSERVATIONAL DATA AND RESULTS

We used the published spectroscopic observations of Grinin *et al.* [2] (the same data were also published by Krasnobabtsev [6]) and Ismailov [9, 14, 15] for our new analysis. The spectrograms had approximately the same mean dispersions (93–100 Å per mm) and were obtained at the Crimean Astrophysical Observatory and the Shemakha Astrophysical Observatory. The observations of [2] were acquired in 1975 (during a single season), while those in [9, 14, 15] were acquired in 1975–1987.

We searched for periods in the variations of the star's brightness and spectrum using the method of Scargle [20], which was subsequently improved in [21]. The technique is described in detail by Antokhin *et al.* [22]. For the sake of uniformity, we analyzed the observing data from each season separately. We searched for periods in the frequency range 0–1 day^{-1}. The presence of cyclic variations in the spectrum of DI Cep on time scales of about 16–18 days was already suspected in [6, 9]; variability with an 11-day period was suspected based on the data of [16, 23]. Accordingly, we used a power spectrum with a frequency resolution of 0.0005 d^{-1} to first search for significant peaks in the range 0.025–1 d^{-1}. The power spectra computed from the spectroscopic data [2, 9] revealed significant peaks exceeding the 3σ level, primarily at frequencies of 0.1–0.2 d^{-1}.

The results of our frequency analysis of the equivalent widths (W_λ) of the Hα and Hβ emission lines are presented in the first part of the table, which contains three columns containing the data-series number, the most significant frequencies in the studied range, and the corresponding periods. Our analysis shows that several series demonstrate the presence of a significant frequency corresponding to the period $P = 9.24 \pm 0.07^{\rm d}$. Our detection of this period from observations obtained within a single year make this detection very reliable.

As an example, Fig. 1 shows part of the power spectrum and the window derived using the Hα equivalent widths from [2]. We obtained a similar plot using the Hβ W_λ values for both the series [2] and our own data.

Figure 2 displays relations between the equivalent widths of individual emission lines and the phase of the derived nine-day period for data obtained by different authors. We can see that both series of spectroscopic observations display a periodic variation with phase. The zero epoch for all the plots is JD 2442616.472. The amplitude of the equivalent-width variations with phase exceeds 100%. We can see from Fig. 2b that the scatter of the data points for our data is relatively high, possibly due to the fact that our data cover several observing seasons. In addition, Fig. 2 shows a phase shift of our data relative to those of [2].

We used the data of the catalog of Herbst *et al.* [1] to analyze the brightness variations for DI Cep. We used up to 450 measurements from this catalog (the data of [13] and the unpublished data of V.S. Shevchenko), including 58 of our own original measurements, partially published in [15, 19], and 97 UBV measurements from [2]. The V brightnesses from [2] acquired during a single night were averaged.

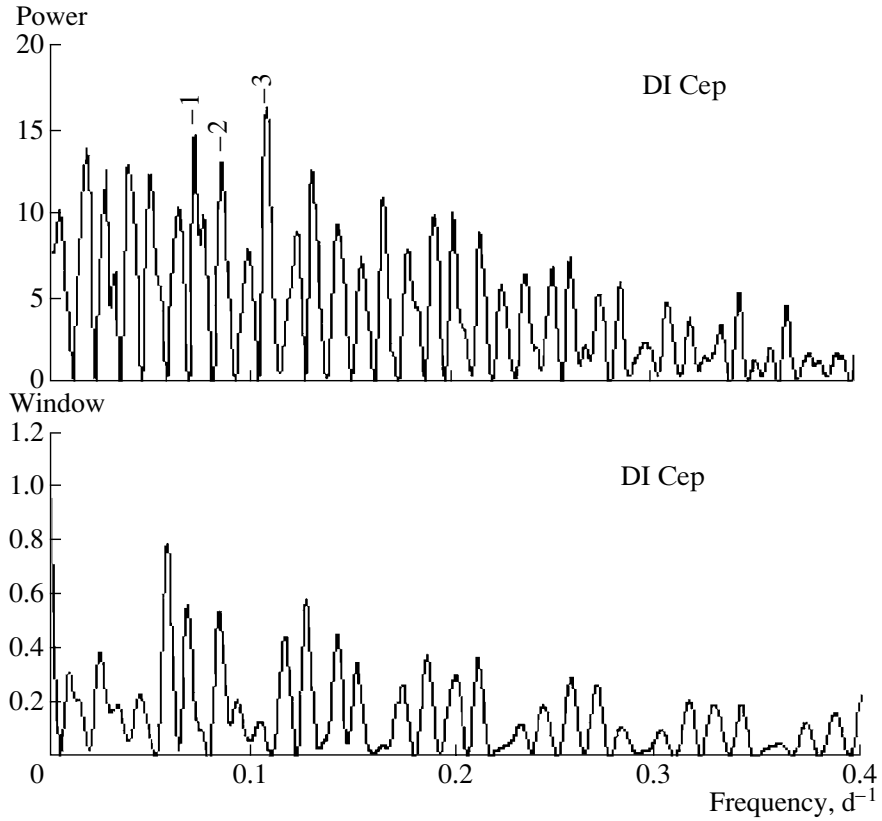

Fig. 1. Sample power spectrum and its spectral window for the data on $W_\lambda(H\alpha)$ [2]. The numbers indicate the peaks corresponding to the periods (1) 13.79^d, (2) 11.56^d, and (3) 9.24^d.

Figure 3 presents the resulting master V-band light curve. In addition to rapid variations within a single season, gradual slow variability of the mean brightness from season to season is observed. The twenty years of observations displayed in Fig. 3 demonstrate the following maximum ranges for the brightness and mean-color variations: $V \sim 11.14^m - 11.71^m$, $\Delta V = 0.57^m$, $\Delta(B-V) = 0.37^m$, $\Delta(U-B) = 0.79^m$, $\Delta(V-R) = 0.25^m$.

The mean light curve (Fig. 3) suggests the presence of a slow variation in the V brightness with a time scale $\geq$5 years. Unfortunately, the available data are not sufficient to determine the character of this slow component of the star's brightness variations.

The power spectrum computed using all the available V-band photometric data does not reveal any significant period in the frequency range under study. This is presumably due to the fact that the master data set is comprised of data obtained by different authors in different seasons. The periodicity could be biased by the presence of phase differences between data sets, as well as systematic and random errors. This may be why our period search in the frequency range 0.01–1 d^{-1} using all the data after subtracting the long-term trend did not reveal any persistent peaks. To ensure better uniformity of the data, we subdivided the data into six smaller sets containing data acquired within a single year and analyzed each of these separately.

Our analysis shows that the power spectra of the individual sets of V brightnesses often contain the frequency 0.053 ± 0.003 d^{-1}, corresponding to the period $18.28^d \pm 1.75^d$ (see the table). The power spectrum of the third set contains an extremely strong peak at a frequency of 0.028 d^{-1}, corresponding to the period 35.71^d, or twice the 18.28-day period. Some of the photometric data of [2] acquired during a single season yield a peak near a period of about nine days. The frequency corresponding to the nine-day period was also reliably observed in our photometric data [15, 19]. In addition, the data sets compiled using the data of [1] also exhibit modes corresponding to the nine-day period.

As an example, Fig. 4 presents the power spectrum and spectral window for the sixth data set for the frequency range 0.1 d^{-1}. The arrow indicates the peak corresponding to $P = 18.28^d$. As is noted above, the complete series of V values does not yield a significant periodicity at any period. Individual observations acquired within a single year indicate a

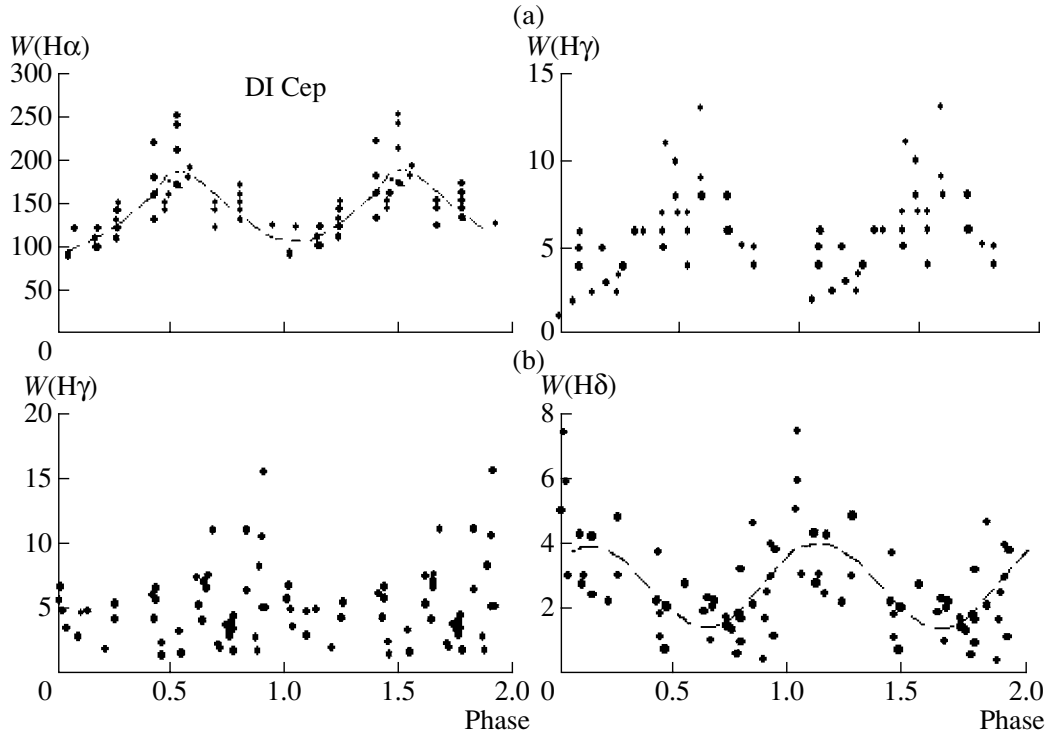

Fig. 2. Relation between the emission-line equivalent widths and the phase of the nine-day period according to data from (a) [2] and (b) [9, 14, 15]. The dashed curves correspond to fourth-order polynomials.

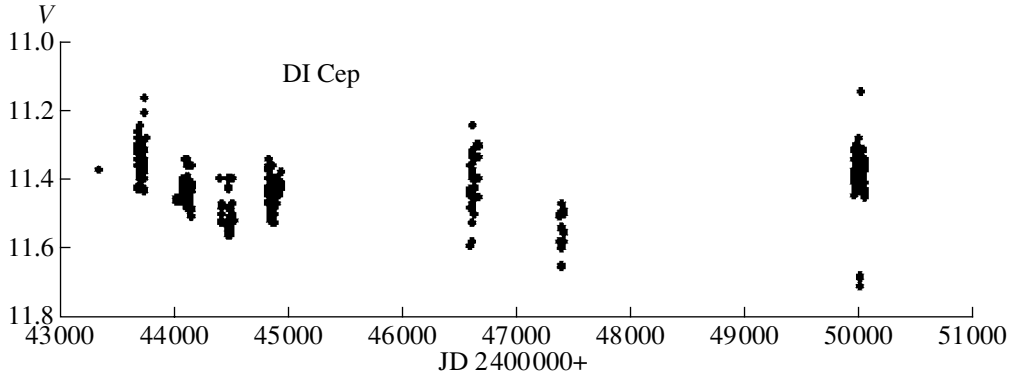

Fig. 3. Master V-band light curve of DI Cep based on data from [1, 2, 15, 19].

clearer periodicity in the star's brightness, with a period of 18 days and a V-band amplitude up to 0.04^m. Figure 5 presents the V-band light curve folded (a) with the 18-day period, based on the sixth set of [1], and (b, c) with the nine-day period, based on the author's data and the data of [2]. The data of [2] for a single night were averaged. The time intervals selected from the data sets in each case are indicated in the figure as Julian Date ranges. We can see that the amplitude for individual observations for the nine-day period reaches $0.2^m \pm 0.01^m$, whereas the amplitude for the 18-day period $0.04^m \pm 0.01^m$.

3. DISCUSSION AND CONCLUSIONS

According to the effective-temperature scale of Cohen and Kuhi [24] and assuming a spectral type G(7–8)V, we obtain for DI Cep $T_{\rm eff} = 5400$ K. The normal color index for a spectral type of G8V is $(B-V)_0 = 0.75^m$, whereas the mean value from the observations is $\langle B-V\rangle = 0.88^m$. Accordingly, we find $E(B-V) = 0.13^m \pm 0.18^m$ and can take into account the interstellar extinction, $A_v = 0.52^m$. This yields a distance modulus for DI Cep of 6^m–7^m, implying a distance to the star of 150–200 pc, in agreement with [2]. Taking into account the bolo-

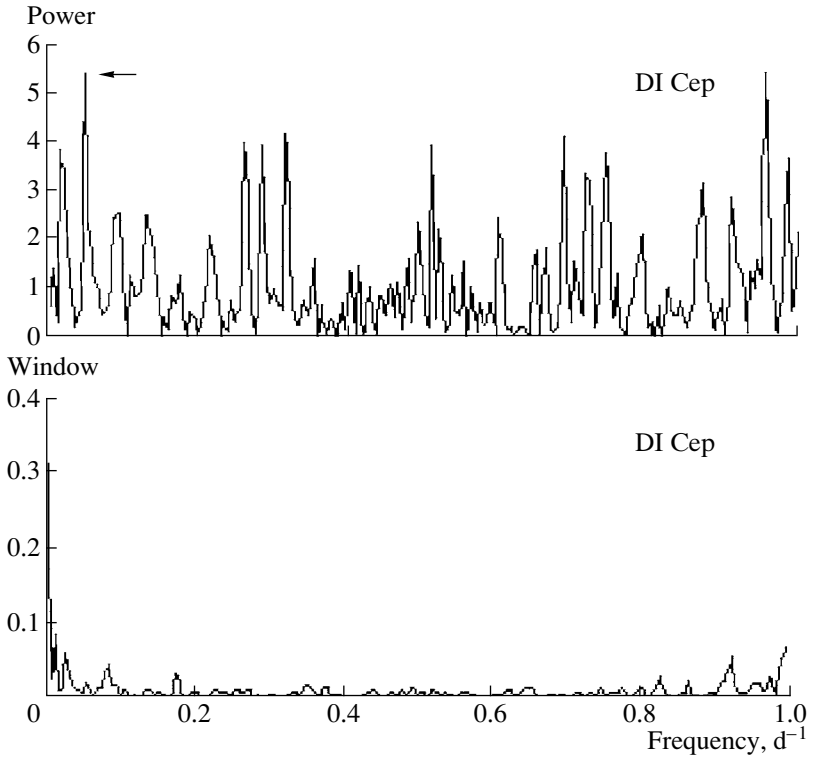

Fig. 4. Sample power spectrum and its spectral window for the sixth series of V brightnesses. The arrow marks the peak corresponding to the 18.28-day period.

metric correction, the star's absolute luminosity is $L = 5.1L_\odot$ and its age is 3×10^6 years. For a distance of 200 pc, we obtain the absolute luminosity $4.0^m \pm 0.3^m$, and, taking into account the bolometric correction, the star's radius is $R_* = 1.8 \pm 0.2R_\odot$.

The rotation rate of DI Cep was earlier estimated to be $v \sin i = 28$ km/s [18] and 23 ± 4 km/s [19]. If the derived period is the period of the star's axial rotation, we can estimate the inclination of the rotational axis to the line of sight: $i = 27° \pm 2.5°$.

The analysis of the brightness variations of weak-emission T Tauri stars [25] demonstrated that the zero epoch of this cyclic process is usually fairly stable. However, Herbst *et al.* [1] noted phase instabilities for the periodic brightness component of classical T Tauri stars. If the nine-day period in DI Cep is due to modulation by the star's axial rotation, phase variations of the nine-day period are quite plausible. Long continuous observations of the star are needed before we can reach firm conclusions about changes in the position of a hot spot. If there is indeed a phase shift in the rotational variations, this will cast doubt on the assumption made in some studies that the accretion occurs along certain specific directions, controlled by the magnetic lines of force (for example, cf. [18, 24]), since it then becomes necessary to suppose that the position of the magnetic lines of force of classical T Tauri stars change with time, for reasons that are unclear.

The variability of the color indices of DI Cep can easily be explained by the presence of a hot spot [15, 16]. Our observed nine-day period for variations of the emission spectrum and brightness and the absence of a correlation between the absorption and emission spectrum [9] suggest that the hot spot may have formed in the regions where the hydrogen emission spectrum is excited, i.e., in a circumstellar envelope or stellar disk.

It follows from our analysis that the periodograms derived from the equivalent widths of the Hα and Hβ emission lines reveal the periodic variations much more reliably than the periodograms based on the V-band brightnesses. The hydrogen emission lines are apparently more sensitive indicators of the variations than the broad-band V magnitude. We demonstrated the existence of a minimum emission-line state in DI Cep in [9]. With the hydrogen and CaII H and K emission lines becoming stronger, this "quiescent" state may become an "active" state, with the spectrum of DI Cep in quiescence not differing appreciably from those of typical weak-emission T Tauri stars. The variations of the emission spectrum may be a direct consequence of the formation of a hot spot on the stellar surface, so that we observe rotational modulation. In our opinion, we are dealing here with

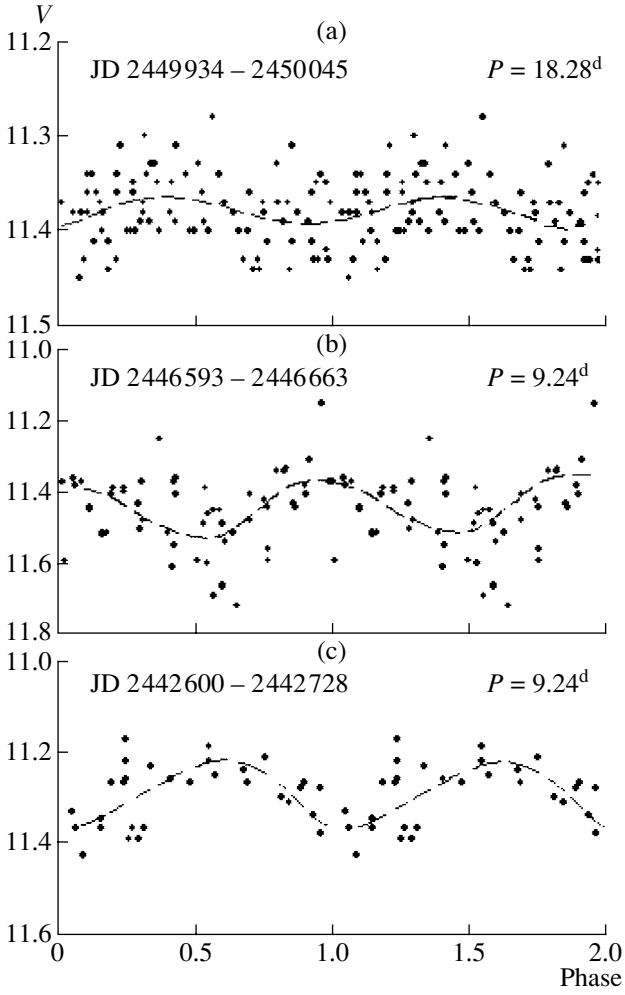

Fig. 5. Phased V-band light curves. (a) Data from the sixth data series (see the table) versus the phase of the 18-day period [1]. (b) The author's data versus the phase of the nine-day period. (c) Data from [2] versus the phase of the nine-day period. The Julian Dates of the beginning and end of the observing series are indicated. The dashed curves are fourth-order polynomials.

a fundamental difference between the circumstellar structures of weak-emission and classical T Tauri stars.

Thus, our study enables us to draw the following conclusions.

(1) Using our own data and data from the literature, we have reliably detected periodic variations of the spectrum and brightness of DI Cep for the first time. Both the spectroscopic and photometric data indicate a period of $P = 9.24^{\mathrm{d}}$. The V brightness shows this periodicity only for individual observing seasons. The equivalent widths of the hydrogen and CaII H and K emission lines in the spectrum show corresponding variations when folded with the phases of the derived period.

(2) The data of different authors show a shift in the zero epoch of the periodic spectrum and brightness variations, possibly due to instabilities of the location of accretion from the circumstellar disk. The inclination of the rotational axis to the line of sight is $27^\circ \pm 2.5^\circ$.

ACKNOWLEDGMENTS

The author is grateful to I. Antokhin for kindly providing the frequency analysis software.

REFERENCES

1. W. Herbst, D. K. Herbst, E. J. Grossman, and D. Weinstein, Astron. J. **108**,1906 (1994).
2. V. P. Grinin, Yu. S. Efimov, V. I. Krasmobabtsev, *et al.*, Perem. Zvezdy **21**, 247 (1980).
3. A. H. Joy, Astrophys. J. **110**, 424 (1949).
4. É. S. Brodskaya, Izv. Krym. Astrofiz. Obs. **7**, 133 (1951).
5. G. H. Herbig, Astrophys. J. **214**, 747 (1977).
6. V. I. Krasnobabtsev, Izv. Krym. Astrofiz. Obs. **65**, 100 (1982).
7. N. Z. Ismailov, Odessa Astron. Publs. **14**, 188 (2001).
8. U. Bastian and R. Mundt, Astron. Astrophys. **78**, 181 (1979).
9. N. Z. Ismailov, Perem. Zvezdy **22**, 489 (1987).
10. P. N. Kholopov, Perem. Zvezdy **9**, 157 (1953).
11. P. P. Petrov, Izv. Krym. Astrofiz. Obs. **44**, 25 (1972).
12. J. Keleman, Inf. Bull. Var. Stars, No. 2744, 1 (1985).
13. V. I. Kardopolov and G. F. Filip'ev, Perem. Zvezdy **22**, 103 (1985).
14. N. Z. Ismailov, Perem. Zvezdy **22**, 500 (1988).
15. N. Z. Ismailov, Perem. Zvezdy **22**, 892 (1988).
16. M. Fernandez and C. Eiroa, Astron. Astrophys. **310**, 143 (1996).
17. G. P. Gahm and P. P. Petrov, *Activity in Red-Dwarf Stars* (Reidel, 1983), p. 497.
18. F. V. Hessman and E. W. Guenther, Astron. Astrophys. **321**, 497 (1997).
19. N. Z. Ismailov, Inf. Bull. Var. Stars, No. 4470, 1 (1997).
20. J. D. Scargle, Astrophys. J. **263**, 835 (1982).
21. J. H. Horne and S. L. Balinas, Astrophys. J. **302**, 757 (1986).
22. I. Antokhin, J.-F. Bertrand, R. Lamontagne, and A. F. J. Moffat, Astron. J. **109**, 817 (1995).
23. A. I. Gomez de Castro and M. Fernandez, Mon. Not. R. Astron. Soc. **283**, 55 (1996).
24. M. Cohen and L. V. Kuhi, Astrophys. J., Suppl. Ser. **41**, 743 (1979).
25. C. M. Johns-Krull and A. P. Hatzes, Astrophys. J. **487**, 896 (1997).

Translated by N. Samus'

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 399–406. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 438–446.
Original Russian Text Copyright © 2004 by Johansson, Letokhov.

The Possibility of Resonance-Enhanced Two-Photon Ionization of Ne and Ar Atoms in Astrophysical Plasmas

S. Johansson[1] and V. S. Letokhov[2]

[1]*Lund Observatory, Lund, Sweden*

[2]*Institute of Spectroscopy, Troitsk, Russia*

Received September 10, 2003; in final form, November 10, 2003

Abstract—We consider possible schemes for the resonance-enhanced two-photon ionization (RETPI) of Ne and Ar atoms under the action of bichromatic radiation of intense resonance lines of HI, HeI, and HeII in a radiation-rich astrophysical plasma. The ionization rate is comparable to or exceeds the recombination rate in rarified astrophysical plasma, which leads to the accumulation of singly ionized ions with a subsequent transition to the higher ionization state via RETPI. We consider the RETPI reaction chains NeI $\rightarrow \ldots \rightarrow$ NeV and ArI $\rightarrow \ldots \rightarrow$ ArVI. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In some cases, low-density astrophysical plasma ($N_H < 10^{10}$ cm^{-3}) can have a very high radiative-energy density, comparable to the energy density of the free electrons. One example is gas condensations (blobs) near a hot star. The best-known case is the Weigelt blobs near the most massive and brightest star of our Galaxy, η Car [1]. Such blobs can be distinguished from the central star in observations with the Hubble Space Telescope. Such observations show that they are located only a few hundred stellar radii from the central star [1, 2] and that their size and hydrogen density imply a high optical depth to the Lyman continuum [1, 3], and quite possibly to the ionization continuum of helium as well. The photoionization–recombination cycle leads to the formation of intense HI, HeI, and HeII resonance lines with effective spectral temperatures of the order of 10 000 K, close to the photospheric temperature of the star, and with spectral widths of hundreds of cm^{-1} [4]. This unique situation differs radically from the conditions in a typical planetary nebula located orders of magnitude further from the central star, when the main role in exciting and ionizing atoms is played by the electron energy, while the photon density is neglible [5, 6]. Such gaseous blobs located nearby a hot star can be considered "radiation rich."

Another example of a radiation-rich region is the environment of an active galactic nucleus (AGN), although the origin of the high-intensity radiation in these objects has a different nature [6, 7]. In the broad-line regions of AGNs, the ionization parameter (the ratio of the density of the ionizing photons to the density of free electrons) can reach 10^{-2}–10^{-3} [7]. High-intensity lines of H, He, and HeII arise during the photoionizational reprocessing of the bright, short-wavelength radiation of the AGN into recombination lines of these species.

In such radiation-rich regions, photoionization processes can play a more dominant role than electron-collisional ionization (see, for example, [8]). Moreover, the high intensity of the VUV lines of hydrogen and helium can stimulate incoherent nonlinear photonic processes that do not require that the radiation be coherent. One such process is resonance-enhanced two-photon ionization (RETPI), which is well known in nonlinear optics [9, 10]. The possibiity of RETPI in an astrophysical plasma (blobs) near a hot star was considered in [11], including the possibility of a series of RETPI processes $X \rightarrow X^+ \rightarrow X^{2+}$ etc. under the action of VUV lines of HI, HeI, and HeII [12]. More detailed calculations for atomic C, N, and O are presented by Johansson and Letokhov [13], who considered a scheme for successive RETPI of C, N, and O atoms right to CV, NV, and OV under the action of the radiation of one or two spectral lines from a set of HI, HeI, and HeII lines from HeII Ly ε (234 Å) to HI Ly α (1215 Å). In the current paper, we consider an analogous scheme for the successive RETPI of Ne and Ar atoms up to NeV and ArVI, also under the action of the radiation of one or two intense lines of HI, HeI, and HeII. We first supplement the formulas presented in [13] with simple expressions that are useful for estimating the probability of RETPI.

2. PROBABILITY OF RETPI OF AN ATOM SUBJECT TO BICHROMATIC RADIATION

Let us consider an atomic particle X (a neutral atom or ion in ionization state N) in a field of bichro-

1063-7729/04/4805-0399$26.00 © 2004 MAIK "Nauka/Interperiodica"

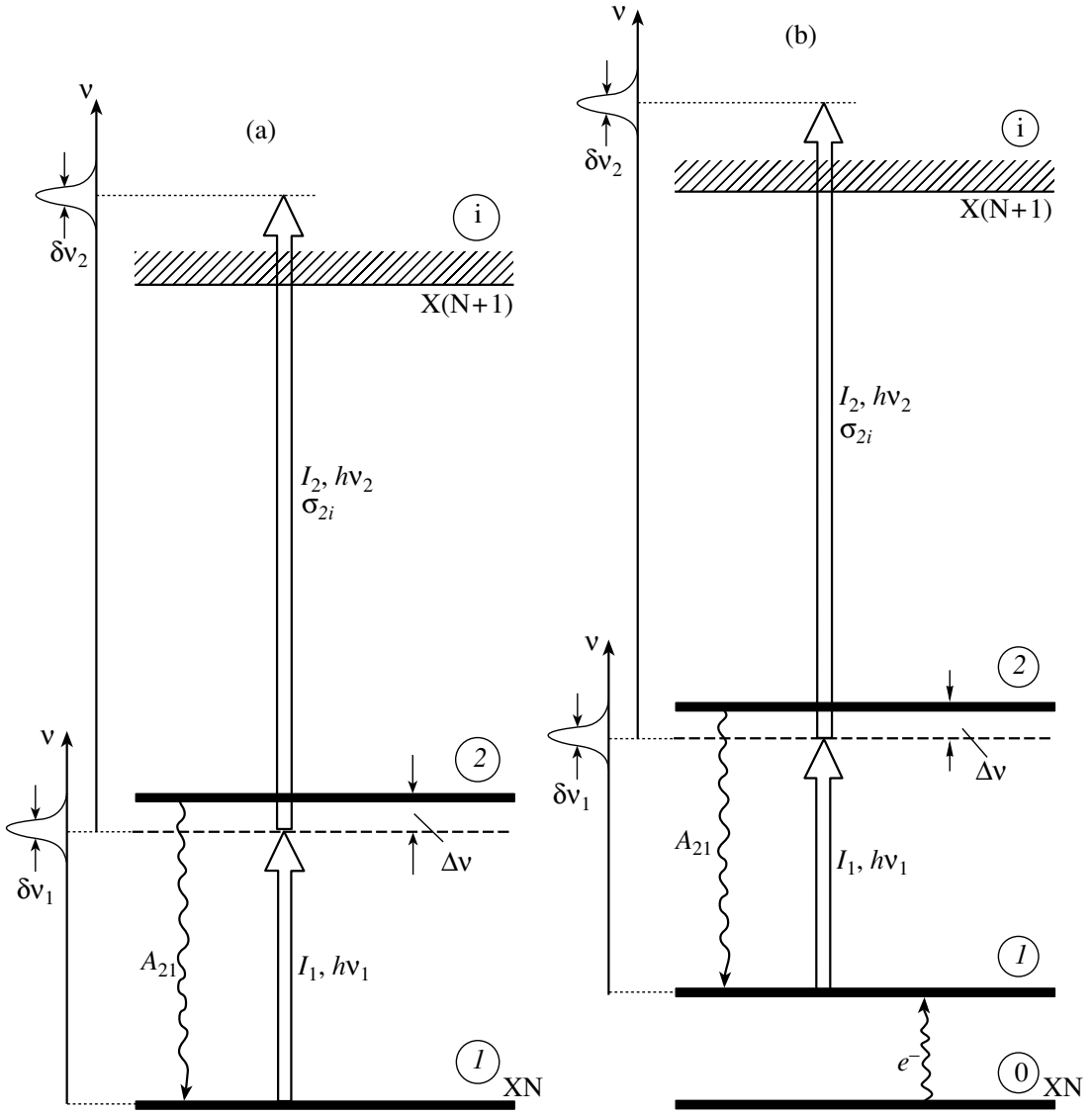

Fig. 1. General schemes for resonance-enhanced two-photon ionization (RETPI) of atomic particles XN under the action of the radiation of two intense spectral lines with widths $\delta\nu_1$ and $\delta\nu_2$ and a frequency offset $\Delta\nu$ relative to the exact intermediate resonance for a permitted spectral transition 1–2. (a) Purely photonic RETPI from the ground state. (b) RETPI from a low-lying state excited by collisions with electrons.

matic radiation with frequencies ν_1 and ν_2 (Fig. 1a), intensities I_1 and I_2 (in photon/cm^2 s), and spectral widths $\delta\nu_1$ and $\delta\nu_2$. The energy of a photon of either frequency, $h\nu_1$ or $h\nu_2$, is not sufficient to ionize the atom, but the energy $h\nu_1 + h\nu_2$ exceeds the ionization threshold of the atom. Therefore, two-photon ionization of the atom is possible in principle. We do not consider here the rare case when ν_1 coincides with the frequency for the transition ν_{12}, when two-step resonance ionization of the atom is possible [9]. We are interested in the more realistic situation when the frequency offset is $\Delta\nu = |\nu_1 - \nu_2| \gg \delta\nu_1$. In this case, RETPI is possible and has a fairly high probability due to the presence of an intermediate resonance [9] without the real excitation of level 2 (Fig. 1a). The probability of this process (in s^{-1}) is determined by the expression [13]

$$W_{1i}^{(2)} = \frac{\lambda_{21}^2}{32\pi^3}\frac{g_2}{g_1}\frac{A_{21}}{(\Delta\nu)^2}\sigma_{2i}I_1I_2, \qquad (1)$$

where A_{21} and λ_{21} are the Einstein coefficient and wavelength for the transition 1–2, g_1 and g_2 are the degeneracies of levels 1 and 2, σ_{2i} is the cross section for photoionization from the excited state at frequency ν_2, and the offset is expressed in Hz. To get some idea of the magnitude of $W_{1i}^{(2)}$, we will consider a simple numerical example. Let the transition 1–2 correspond to Ly α HI ($\lambda_{21} = 1215$ Å), with $A_{21} \simeq$

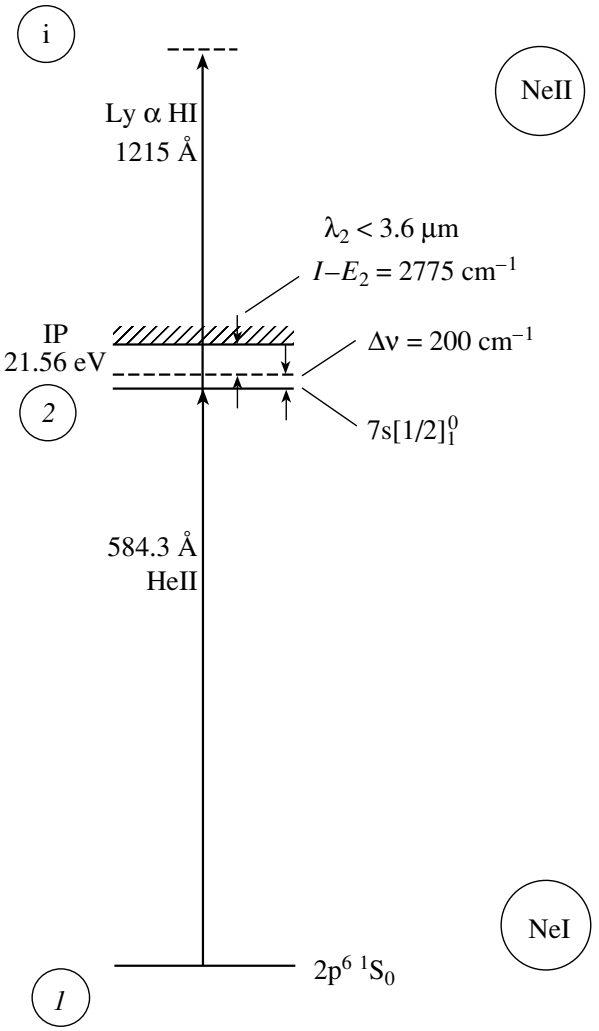

Fig. 2. RETPI scheme for NeI → NeII.

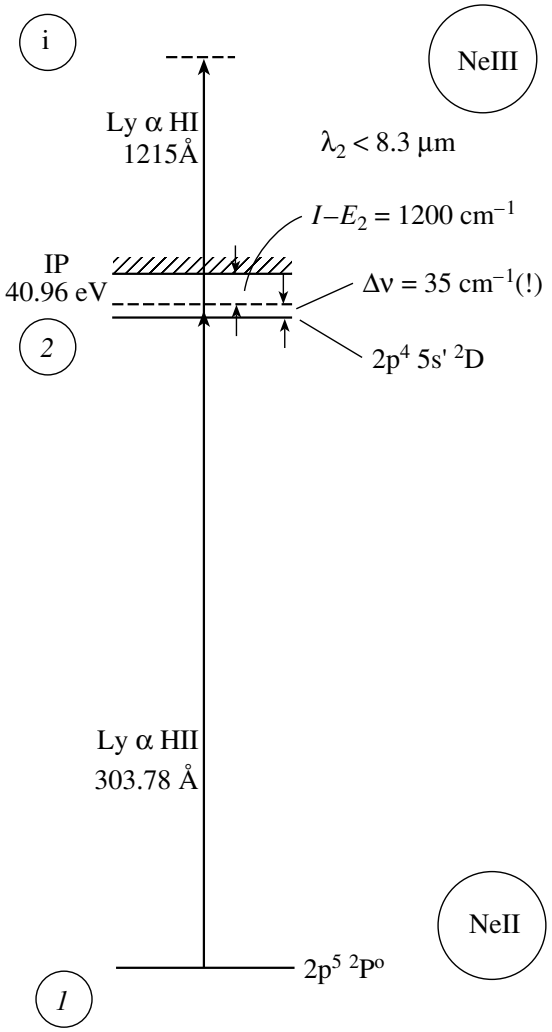

Fig. 3. RETPI scheme for NeII → NeIII.

10^9 s^{-1} and the cross section for photoionization of the excited state $\sigma_{2i} \simeq 10^{-17}$ cm^2. Let the spectral lines have the same wavelength λ_{21}, spectral widths $\delta\nu_1 = \delta\nu_2 \simeq 300$ cm^{-1} (10^{13} Hz), and effective spectral temperatures $T_{eff} \simeq 15\,000$ K. The intensities $I_{1,2}$ (photons/cm^2 s) of the spectral lines are then determined by the expression

$$I_{1,2} = \frac{8\pi}{\lambda^2} \frac{\delta\nu_{1,2}}{\left[\exp\left(\dfrac{h\nu_{1,2}}{kT_{eff}}\right) - 1\right]} \qquad (2)$$

and are equal to $I_1 = I_2 = 6.5 \times 10^{20}$ photons/cm^2 s. If the frequency offset for λ_1 relative to a resonance transition for an atom (ion) subject to RETPI is $\Delta\nu \simeq 1000$ cm^{-1} (3×10^{13} Hz), $W_{1i}^{(2)} \simeq 6 \times 10^{-7}$ s^{-1}. This is quite substantial for a rarified astrophysical plasma, for which the electron-excitation and recombination rates can be comparable to or even lower than this value. It stands to reason that we must bear in mind the quadratic dependence of $W_{1i}^{(2)}$ on the frequency offset from the intermediate resonance $\Delta\nu$ and the intensity of the radiation. Expression (1) for the probability of RETPI of an atomic particle per unit time can be written in the form

$$W_{1i}^{(2)} = \beta^{(2)} I_1 I_2, \qquad (3)$$

where the coefficient $\beta^{(2)}$ (in cm^4 s) is

$$\beta^{(2)} \simeq \frac{\lambda_{12}^2}{32\pi^3} \frac{g_2}{g_1} \frac{A_{21}}{(\Delta\nu)^2} \sigma_{2i}. \qquad (4)$$

Expressions (1) and (2) for $W_{1i}^{(2)}$ are presented for the case when the line width $\delta\nu_1$ at frequency ν_1 satisfies the condition $\delta\nu_1 \ll \Delta\nu$ and the photoionization cross section σ_{2i} varies only slightly within the width of the line at ν_2, as is shown in Fig. 1a.

The cross sections for two-photon absorption (in cm^2) at the frequencies ν_1 and ν_2 are given by

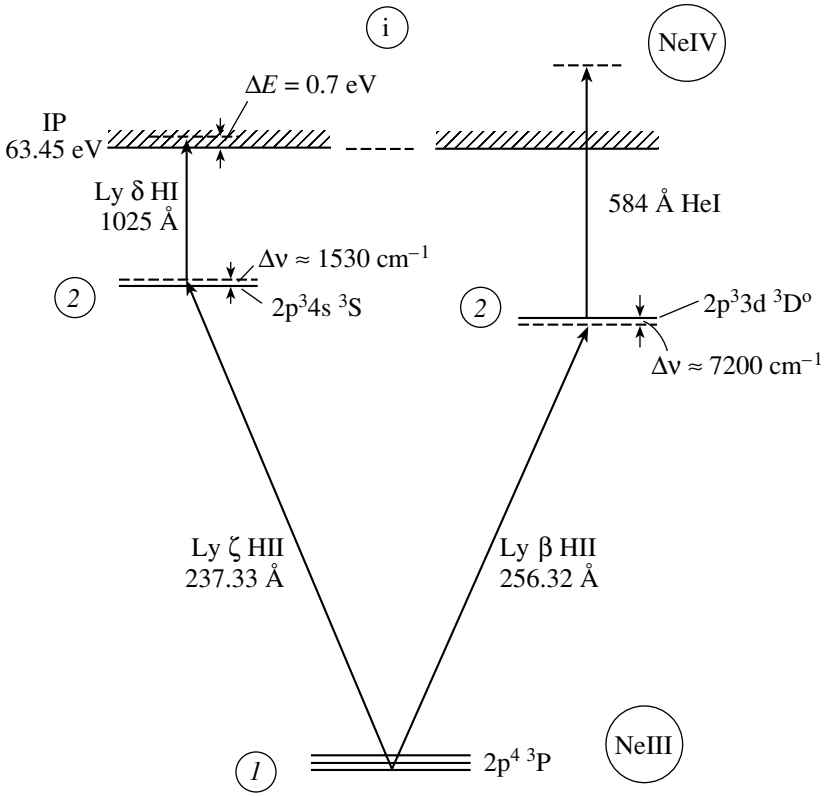

Fig. 4. RETPI scheme for NeIII → NeIV.

the expressions

$$\sigma_1^{(2)} = \beta^{(2)} I_2 \tag{5a}$$

and

$$\sigma_2^{(2)} = \beta^{(2)} I_1. \tag{5b}$$

Accordingly, the coefficients for two-photon absorption per unit length are given by the expressions

$$\chi_1^{(2)} = \sigma_1^{(2)} N_0 = \beta^{(2)} I_2 N_0 \tag{6a}$$

and

$$\chi_2^{(2)} = \sigma_2^{(2)} N_0 = \beta^{(2)} I_1 N_0, \tag{6b}$$

where N_0 is the density of two-photon-absorbing particles (in cm^{-3}). The numbers of absorbed photons n_1 and n_2 per unit time per unit volume at frequencies ν_1 and ν_2 are the same and are given by the expression

$$n_1 = n_2 = \chi_1^{(2)} I_1 = \chi_2^{(2)} I_2 = \beta^{(2)} I_1 I_2 N_0 = W_{1i}^{(2)} N_0. \tag{7}$$

For example, if the probability $W_{1i}^{(2)} \simeq 10^{-5}$–$10^{-7}\ \mathrm{s}^{-1}$ and the density of two-photon-absorbing particles is $N_0 \simeq 10^3\ \mathrm{cm}^{-3}$, the coefficients for the two-photon absorption are much smaller than $10^{-20}\ \mathrm{cm}^{-1}$.

3. SCHEMES FOR SUCCESSIVE RETPI FOR Ne AND Ar

Analysis of the permitted photon transitions for neutral Ne and Ar atoms and their ions indicates the possibility of RETPI under the action of appropriate lines or pairs of lines of HI, HeI, and HeII in transitions to the ground state in the EUV; these lines usually have large optical depths, intensities, and line widths. Using the Grotrian diagrams for neon and argon [14] and original data for NeII [15], NeIII [16, 17], NeIV [18], ArI and ArII [19], ArIII [20], and ArIV [21], we selected among the large number of quasi-resonance coincidences those having the closest resonances with some permitted photon transition for Ne or Ar atoms or ions, as was done in [13].

Figures 2–5 present the RETPI schemes for Ne ions up to NeV selected in this way. There is no suitable resonance with EUV spectral lines of HeII for NeV. Thus, NeV lies at the end of the successive-RETPI chain: NeI → NeII → NeIII → NeIV → NeV. In all cases, the offset from the intermediate resonance does not exceed 1500 cm^{-1}. In the RETPI reactions NeII → NeIII and NeIV → NeV, the offset $\Delta\nu$ is very small, and we expect a shift in the ionization equilibrium toward NeIII and NeV relative to the densities of NeII and NeIV. In addition, virtually excited NeI and NeII have very low photoionization

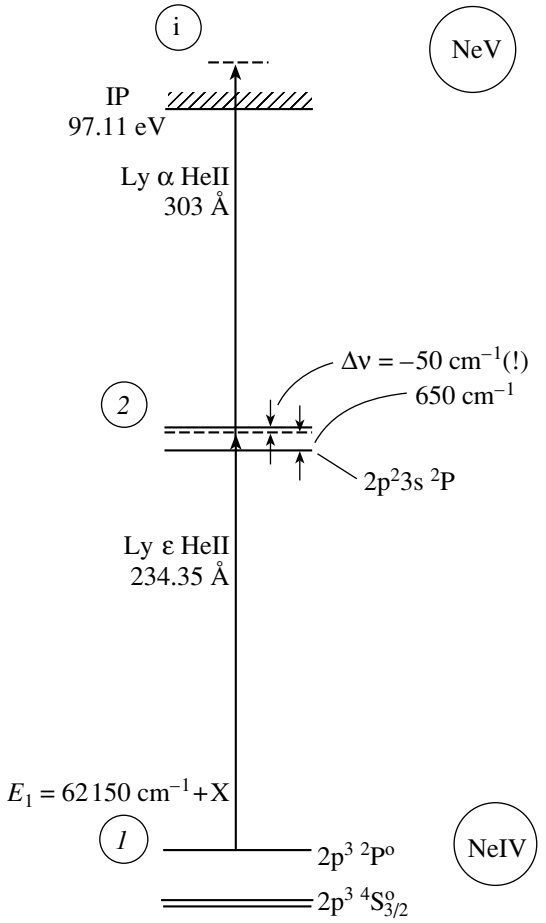

Fig. 5. RETPI scheme for NeIV → NeV.

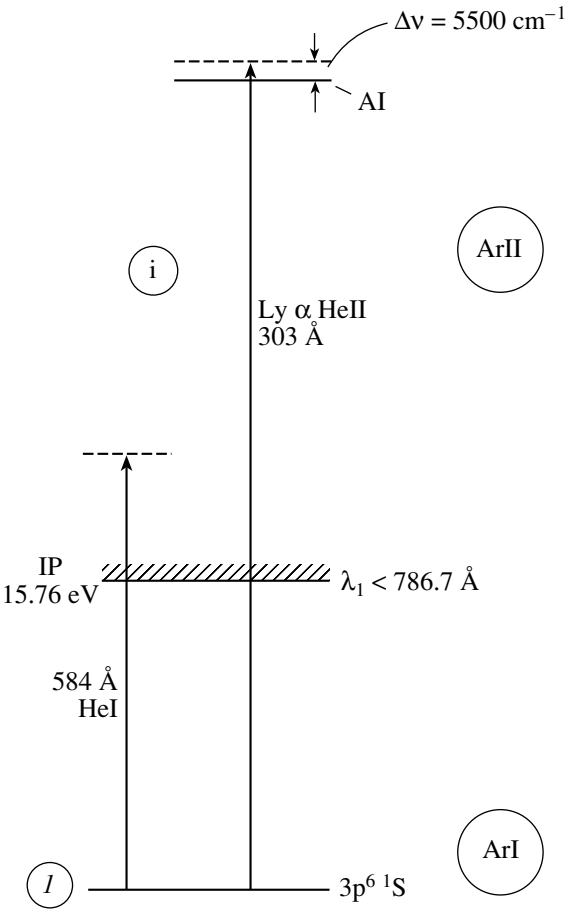

Fig. 6. RETPI scheme for ArI → ArII (AI is an autoionization state).

energies, and nearly any radiation can participate in the second RETPI stage: both radiation from within the plasma and weaker radiation, from the central star, for example. In the case of NeIV, RETPI is possible only from a low-lying excited level (see the scheme in Fig. 1b) whose position is not known precisely. In this case, the participation of electron collisions is required to excite this low-lying level, since recombinational population of the level is not possible in the absence of NeIV. This should limit the rate of formation of NeV in a plasma with a low electron density; on the other hand, the formation of NeV via RETPI requires intense optically thick Ly α and Ly ε lines of HeII, which can be radiated in a plasma with a high density of HeII and, consequently, of photoelectrons.

There are more possibilities for Ar compared to Ne. First, the photoionization of ArI can be realized via the absorption of a single EUV photon radiated by the central star, which gives rise to the photoconversion of HeI/HeII, as well as the Ly α 303 Å recombination line of HeII (Fig. 6). However, there are quite a few possibilities for RETPI of ArII, shown in Fig. 7, both with the participation of high lines of the Ly HI series (Fig. 7a) with a comparatively small offset $\Delta\nu = 180$ cm^{-1} and with the participation of the 515.6 Å, 522.2 Å, and 584.3 Å EUV lines of HeI (Figs. 7b–7d). Especially efficient is the scheme with the participation of the intense 584.3 Å line, for which the offset is very small ($\Delta\nu \simeq 40$ cm^{-1}, which can be comparable to the Doppler width of the HeI line). ArIV can be formed in two ways (Fig. 8). First, the intense 303 Å HeII line can photoionize ArIII with one photon whose energy slightly exceeds the ionization threshold. Second, RETPI can occur from a low-lying $3p^4\ ^1D$ excited state that is populated by electron collisions (the case of Fig. 1b). This is also applicable to RETPI of ArIV from low-lying excited states with small offsets $\Delta\nu$ (Fig. 9). Finally, there is a very favorable situation for the RETPI reaction ArV→ArVI (Fig. 10), which has a very small offset ($\Delta\nu = 50$ cm^{-1}). This offset is comparable to the Doppler width of the 522 Å HeI line, which is in-

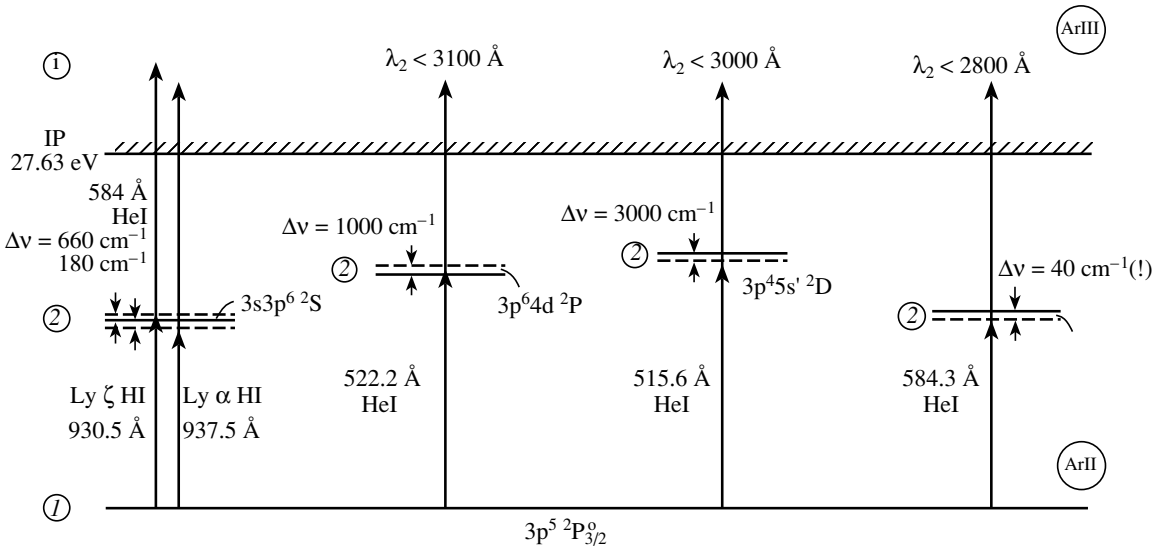

Fig. 7. Possible RETPI schemes for ArII → ArIII.

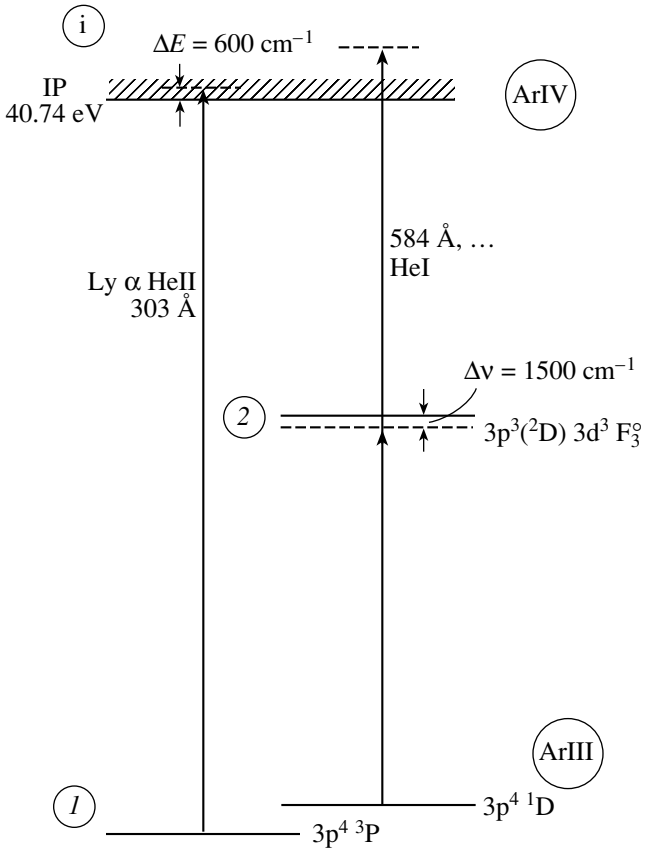

Fig. 8. RETPI process for ArIII → ArIV initiated from a low-lying excited state.

creased by Doppler diffusion of the radiation during resonance scattering [6]. Thus, in the case of Ar, the chain of successive RETPI reactions can continue right up to ArVI.

All these schemes are summarized in the table, which presents the chains of successive RETPI reactions for the atoms and ions considered under the action of radiation in one or two spectral lines with

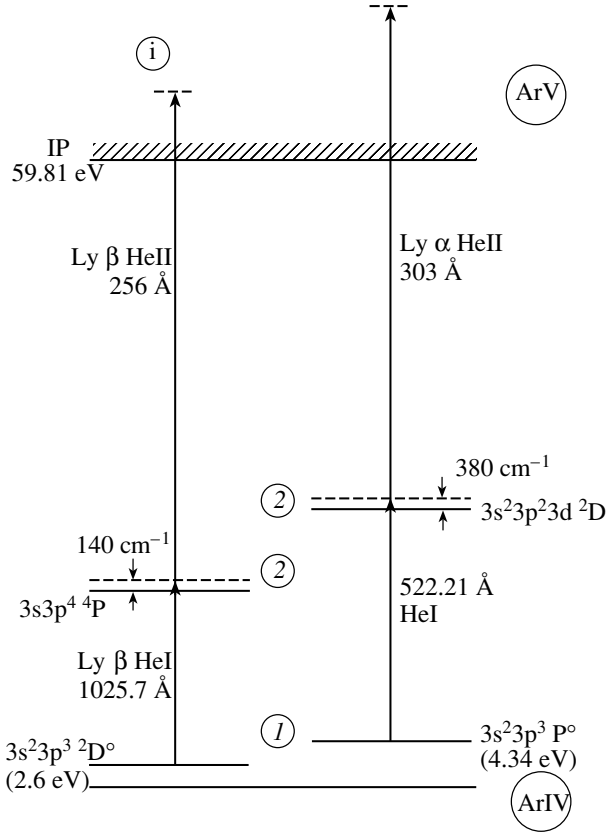

Fig. 9. RETPI process for ArIV → ArV initiated from a low-lying excited state.

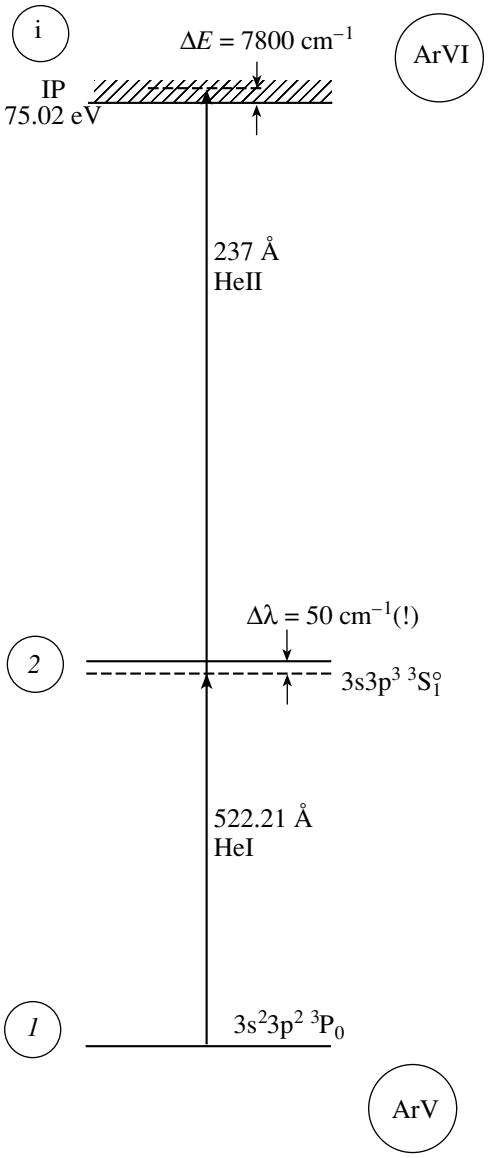

Fig. 10. RETPI scheme for ArV → ArVI with a very small frequency offset $\Delta\nu$.

progressively increasing photon energies: HI + HeI, HI + HeII, HeI + HeII, and HeII + HeII. The resulting set of successive-RETPI schemes for Ar strongly resembles that for Ne, although it includes other combinations of intense EUV spectral lines.

4. CONCLUSIONS

First and foremost, we conclude that the elementary RETPI processes we have considered should be taken into account in ionization-balance calculations for specific spectral lines, especially for radiation-rich regions. We expect some nonmonotonicity in the appearance of ions of certain multiplicities due to the existence of favorable near coincidences between the frequencies of intense exciting lines and of absorption lines of the atoms and ions. The resulting nonmonotonicity in the appearance of successive ions can lead to deviations from standard models if electron temperatures are estimated without including the effects of RETPI. The same is true in connection with estimates of the densities of elements based on emission lines that arise via the recombination of ions whose formation rate is enhanced due to RETPI of ions of some multiplicity. In contrast to electron-collisional ionization, the probability of this process displays a nonmonotonic dependence even on the effective temperature of the radiation, since $W_{1i}^{(2)}$ is sensitive to the offset $\Delta\nu$, which has various values for various ions.

Finally, we note that, strictly speaking, RETPI is a purely collisionless photon process only when the RETPI process is initiated from the ground state (Fig. 1a). However, in some of the schemes considered, the RETPI is initiated from a low-lying excited level with an energy less than a few eV (Fig. 1b). In this case, some links in the RETPI chain rely on collisions. However, it is unlikely that these links will become a bottleneck for the chain, since the cross sections for the electronic excitation of the low-lying levels are much greater than the cross sections for the electronic ionization of the corresponding ions in the

Chains of successive-RETPI schemes for Ne and Ar based on HI, HeI, and HeII lines in the far VUV

Element	Pairs of photoionizing HI, HeI, and HeII lines			
	HI + HeI	HI + HeII	HeI + HeII	HeII + HeII
Ne	NeI → NeII (21.56 eV) 584 Å + 1215 Å	NeII → NeIII (40.96 eV) 303 Å + 1215 Å NeIII → NeIV (63.45 eV) 237 Å + 1025 Å	NeIII → NeIV (63.45 eV) 256 Å + 584 Å	NeIV→NeV (97.11 eV) 234 Å + 303 Å
Ar	ArI → ArII (15.76 eV) 584 Å ArII → ArIII (27.63 eV) 930 Å + 584 Å 584 Å + 1215 Å 522 Å + 1215 Å 515 Å + 1215 Å	ArI → ArII (15.76 eV) 303 Å ArIV → ArV (59.81 eV) 1025 Å + 256 Å	ArIII → ArIV (40.74 eV) 584 Å + 584 Å 303 Å ArIV → ArV (59.81 eV) 522 Å + 303 Å ArV → ArVI (75.02 eV) 522 Å + 237 Å	

collision-dependent chain. Nonetheless, this remains a subject for a specialized analysis of specific models for ionization balance in radiation-rich astrophysical plasmas.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project code 03-02-16377). V.S.L. thanks the Lund Observatory (Sweden) for hospitality.

REFERENCES

1. K. Davidson and R. M. Humphreys, Ann. Rev. Astron. Astrophys. **35**, 1 (1997).
2. T. Gull, private communication.
3. S. Johansson and V. S. Letokhov, Astron. Astrophys. **378**, 266 (2001).
4. V. V. Klimov, S. Johansson, and V. S. Letokhov, Astron. Astrophys. **385**, 313 (2002).
5. S. R. Pottasch, *Planetary Nebulae* (Reidel, Dordrecht, 1984).
6. D. E. Osterbrock, *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (Univ. Sci. Books, Sausalito, Ca., 1989).
7. I. Robson, *Active Galactic Nuclei* (Wiley, Chichester, 1986).
8. A. V. Filippenko, Astrophys. J. **289**, 475 (1985).
9. V. S. Letokhov, *Laser Photoionization Spectroscopy* (Nauka, Moscow, 1987; Academic, Orlando, 1987).
10. N. B. Delone, *Interaction of Laser Radiation with Matter* (Nauka, Moscow, 1989) [in Russian].
11. S. Johansson and V. S. Letokhov, Science **291**, 625 (2001).
12. S. Johansson and V. S. Letokhov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 135 (2001) [JETP Lett. **73**, 118 (2001)].
13. S. Johansson and V. S. Letokhov, Astron. Astrophys. **375**, 319 (2001); Astron. Astrophys. **395**, 345 (2002).
14. S. Bashkin and J. O. Stoner, Jr., *Atomic Energy Levels and Grotrrian Diagrams* (North-Holland, Amsterdam, 1975), Vol. 1.
15. W. Persson, Phys. Scripta **3**, 133 (1971).
16. W. Persson, C.-G. Wahlström, L. Jönsson, and H. O. DiRocco, Phys. Rev. A **43**, 4791 (1991).
17. A. E. Livingston, R. Buttner, A. S. Zacarias, *et al.*, J. Opt. Soc. Am. B **14**, 522 (1997).
18. A. E. Kramida, T. Bastin, E. Bimont, *et al.*, Eur. Phys. J. D **7**, 525 (1999).
19. G. Norlin, Phys. Scripta **8**, 249 (1973).
20. J. E. Hansen and W. Persson, J. Phys. B **20**, 693 (1987).
21. F. Bredice, M. Gallardo, and J. G. Reyna Almandos, Phys. Scripta **51**, 446 (1995).

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 407–417. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 447–458.
Original Russian Text Copyright © 2004 by Chertok, Slemzin, Kuzin, Grechnev, Bugaenko, Zhitnik, Ignat'ev, Pertsov.

Analysis of a Solar Eruptive Event on November 4, 2001, Using CORONAS-F/SPIRIT Data

I. M. Chertok[1], V. A. Slemzin[2], S. V. Kuzin[2], V. V. Grechnev[3], O. I. Bugaenko[4], I. A. Zhitnik[2], A. P. Ignat'ev[2], and A. A. Pertsov[2]

[1] *Institute of Terrestrial Magnetism, Ionosphere, and Radio Wave Propagation, Troitsk, 142190 Russia*

[2] *Lebedev Institute of Physics, Leninskiĭ pr. 53, Moscow, 117924 Russia*

[3] *Institute of Solar-Terrestrial Physics, Irkutsk, Russia*

[4] *Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119899 Russia*

Received October 13, 2003; in final form, November 10, 2003

Abstract—We analyze large-scale solar activity following the eruption of a very powerful, geoeffective coronal mass ejection in the 23rd solar cycle, observed at 175, 284, and 304 Å on November 4, 2001, using data from the CORONAS-F/SPIRIT telescope. In particular, we have shown that the restructuring of the magnetic field above the eruption center was accompanied by the formation of a multicomponent post-eruptive arcade, which was observed in all three bands over many hours and had an extent of the order of 0.5 $R_\odot$. Two kinds of dimmings were observed, i.e., compact dimmings on either side of this arcade and channeled dimmings along some extended features beyond the active region. The intensity in the dimmings decreased by several tens of percent. The enhanced emission observed at the top of the post-eruptive arcade can be due to energy release in the course of magnetic reconnection high in the corona at the relaxation stage of the perturbed magnetic field to a new equilibrium state with a closed configuration. It can also be due to an enhanced emission measure because of the oblique direction of the line of sight crossing both loop tops and footpoint regions. The spatial coincidence of the main dimmings in lines corresponding to different temperatures indicates that a plasma outflow from the transition region and coronal structures with opened field lines are responsible for these dimmings. Variations in the plasma temperature associated with coronal mass ejections probably play an important role for some dimmings, which appear different in different lines.
© *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known (see the reviews [1–5] and references therein) that a coronal mass ejection (CME) and the associated restructuring of the magnetic field in an extensive region of the solar atmosphere are accompanied by a number of large-scale phenomena. In particular, so-called dimmings and post-eruptive arcades with lifetimes of several hours to a day are observed on the solar disk in the soft X-ray and extreme ultraviolet (EUV) after large halo-type CMEs.

Dimmings, or transient coronal holes, are regions of decreased intensity, which are formed after a CME near the eruption center and occupy a significant part of the solar disk [1, 4, 6–9]. As a rule, the locations and structures of dimmings adjacent to the eruption center coincide in lines corresponding to different temperatures. This suggests that dimmings result from the total or partial opening of coronal magnetic fields, which leads to an evacuation of material and a corresponding decrease in the intensity. Direct evidence for the outflow of material from dimmings located near the eruption center was revealed in [10] based on the observed Doppler shifts of some of the EUV lines. However, the narrow and extended channeled dimmings that are observed under the conditions of the complex global solar magnetosphere [8] can appear different in different lines, which suggests that an important role is played by plasma-temperature variations in the formation of some dimming structures [9].

A bright post-eruptive arcade frequently appears at the site of pre-existed sigmoidal (twisted) sheared structures. The arcade is formed due to the relaxation of the large-scale coronal magnetic field strongly perturbed by a CME to an equilibrium state through magnetic reconnection. The reconnection occurs in extended current sheets high in the corona. It is accompanied by prolonged energy release and the formation of a relatively simple loop system that increase in size and occasionally display features of cusplike structures [2–7, 11–15].

In recent years, dimmings and post-eruptive arcades associated with CMEs have been studied

1063-7729/04/4805-0407$26.00 © 2004 MAIK "Nauka/Interperiodica"

mainly using soft X-ray heliograms obtained with the Yohkoh/SXT telescope [16], SOHO/EIT EUV data [17], and, to a lesser extent, TRACE data [18]. We analyze here the big eruptive event of November 4, 2001 (after 16^h UT) using data recorded in three EUV bands with the Ritchey–Chrétien (284 Å) and Herschel (175 and 304 Å) telescopes of the SPIRIT complex [19] onboard the CORONAS-F spacecraft [20]. Our choice of this event and the considerable interest in it are due to two additional circumstances. First, there were no SOHO/EIT observations at that time due to maintenance tests, while TRACE observed activity in a restricted region far from the eruption center. Second, this event is among the biggest and most geoeffective events in the 23rd solar cycle. Based on a number of its features and parameters, this event is comparable to the well known "Bastille Day" event of July 14, 2000 [21]. In addition to substantial solar phenomena (see below), this event was accompanied by an appreciable increase in the flux of high-energy particles detected by ground-based neutron monitors (GLE), a flux of protons with energies $E > 10$ MeV reaching 2×10^3 cm^{-2} s^{-1} sr^{-1} for several hours after the event and exceeding 10^4 cm^{-2} s^{-1} sr^{-1} when the interplanetary disturbance arrived at the Earth, an extreme geomagnetic storm with $D_{st} \approx -300$ nT, and a Forbush decrease in the Galactic cosmic rays detected at high-latitude stations exceeding 10% (see http://sgd.ngdc.noaa.gov/sgd/jsp/solarindex.jsp).

Section 2 presents general information about the SPIRIT telescopes and describes the observational data and reduction techniques used. Section 3 describes the situation preceding the eruptive event and some features of the event itself. Sections 4 and 5 consider the main analyzed objects—the post-eruptive arcade and dimmings. In Section 6, we discuss the results obtained and present concluding remarks.

2. OBSERVATIONAL DATA AND TECHNIQUES OF ANALYSIS

The SPIRIT complex [20] (see also the Web site http://www.xras.lebedev.ru) includes two EUV telescopes: the Herschel two-channel telescope–coronagraph, which observes at 175 and 304 Å, and the Ritchey–Chrétien four-channel telescope, which observes at 171, 195, 284 and 304 Å (similar to the SOHO EIT telescope [17] in its spectral characteristics). In addition, SPIRIT includes spectroheliographs operating at 177–207 Å and 285–335 Å and a crystal spectroheliograph operating at 8.42 Å (see the table).

The FeIX–XI coronal lines emitted by plasma with a temperature of $T_e \approx 1.5$ MK dominate at 175 ± 5 Å. The 284 ± 8 Å band contains a high-temperature ($T_e \approx 2.0$ MK) FeXV line. The 304 ± 12 Å band contains both the transition-region HeII ($T_e \approx 0.05$ MK) line and the lower-intensity coronal SiXI ($T_e \approx 1.6$ MK) line.

SPIRIT has been routinely observing the Sun starting from August 15, 2001. The observations include regular synoptic sessions with mean durations of 10–48 min performed in all spectral bands one to four times per day, as well as sessions scheduled for special programs: studies of the dynamics of active regions with a time resolution up to 10 s, observations of the solar corona to distances reaching $3R_\odot$, long-term (up to 20 days) observations in nonshadowed orbits with a resolution of 50–100 s, flare observations with a resolution of 7 s, spectroscopic observations, etc.

On November 4, 2001, SPIRIT was performing synoptic observations (12:19–12:20 UT) and carrying out observations every 5 min in several spectral channels of the 284 Å (18:38–19:04 UT), 175 Å (17:03–17:29 UT), and 304 Å (20:13–20:39 UT) bands. Note that the Herschel telescope obtained the 175 and 304 Å images synchronously withing 10 ms and an exposure time of 2 s. In addition, several images were obtained at these same wavelengths in other orbits. In particular, heliograms corresponding to 11:11 UT were obtained at both 175 and 304 Å.

According to the GOES data, an X-ray flare of magnitude X1.0 occurred from 16:03 to 16:57 UT but was not observed by SPIRIT because the satellite was in shadow. Images at 284, 175, and 304 Å were obtained four to five hours before the event, as well as four hours after the flare maximum. Despite the data, we are able to analyze long-term CME features such as post-eruptive arcade and dimmings.

For several hours after the flare, the arcade is brighter than the unperturbed portions of the disk by several orders of magnitude. To demonstrate the structure of the bright arcade and low-contrast features simultaneously in the heliograms, the corresponding images are shown on a nonlinear scale using a power function.

The dimmings are clearly visible in the difference images obtained by subtracting a heliogram preceding the event from those obtained during or after the event. The dimmings are visible as dark objects with decreased intensity. However, a simple subtraction of images, especially images separated by many hours, inevitably yields some bright and dark artifacts (false dimmings), because the structures observed on the disk are displaced due to solar rotation. To avoid this,

Spectral ranges observed by SPIRIT and excitation temperatures of solar-plasma ions emitting in these ranges [19, 20]

Channel	Spectral range, Å	Field of view, arcmin	Pixel size, arcsec	Ions	T, MK
Herschel telescope–coronagraph	175 ± 5	45×50	2.6	FeIX–FeXI	1–1.6
	304 ± 15	Corona at $(2–5)R_\odot$		HeII, SiXI	0.05, 1.6
Ritchey–Chrétien 4-channel telescope	171 ± 3	42×48	2.5	FeIX–FeX	1–1.3
	195 ± 6			FeXII	1.6
	284 ± 8			FeXV	2
	304 ± 8			HeII, SiXI	0.05, 1.6
MgXII spectroheliograph	8.418–8.423	> 60	4.2	MgXII	10
EUV spectroheliograph	177–207	> 60	5.2×93	OIV, FeIX–FeXXIV, CaXIV–CaXVII	0.3–16
	285–335		7.5×140	HeII, SiXI, FeXV–FeXVI, MgVIII, NiXVII, CaXVII	0.05–5

we use a procedure to correct for the solar rotation before the subtraction (similar to the technique of [8, 9]); namely, we rerotate the images at each wavelength to a single time prior to the event. We apply this rerotation to the entire visible hemisphere by rotating a spherical surface located at a heliocentric distance of $r \approx 1R_\odot$. This correction for the solar rotation is not ideal, since the three-dimensional structures observed in the EUV occupy some range of altitudes at larger heliocentric distances. However, our estimates show that the errors introduced by rotating the sphere with radius $r \approx 1R_\odot$ are negligible and cannot appreciably change the observed dimmings, at least for events in the central sector of the solar disk.

In addition to the SPIRIT data, we also used Hα heliograms, SOHO/MDI magnetograms [22], and Yohkoh/SXT soft X-ray heliograms. These heliograms are also used to analyze the post-eruptive arcade. These and other additional data related to the event under study, including information about the CME obtained in white light by the SOHO/LASCO coronagraph [23] and the CORONAS-F/SPIRIT heliograms can be found at the web site http://helios.izmiran.troitsk.ru/lars/Chertok/ in the form of images and movies.

3. SITUATION PRECEDING THE EVENT

The eruption occurred in the large active region AR 9684 (coordinates N06 W18) situated to the northwest of the disk center (Fig. 1a). A flocculus observed in Hα was intersected by an extended filament, FF. The SOHO/MDI magnetogram (Fig. 1b) shows a line separating the polarities of the photospheric magnetic field within the active region that corresponds to the filament. A comparison with the image obtained by the same instrument shows that the main sunspots had northern polarity and were located to the west of the filament. According to more detailed data, leading and trailing sunspots situated a small distance from each other were surrounded by a common penumbra, forming a typical preflare δ configuration. The magnetic field lines associated with the compact sunspot group probably closed on the scattered magnetic fields to the east of the filament.

The Yohkoh/SXT soft X-ray image (Fig. 1c) shows a large, high-temperature coronal arcade with a twisted (sigmoid) structure (see http://solar.physics.montana.edu/nuggets/2001/011109/011109.html). As was mentioned in the Introduction (see also [24]), such a structure is suggestive of a sheared magnetic configuration, providing evidence for stored excess energy and a high probability of eruption. A similar sigmoid structure was observed by SPIRIT in the hot Mg XII line before the event.

Figures 1d–1f present the CORONAS-F/SPIRIT heliograms obtained at 175 and 304 Å at 11:11 UT and at 284 Å at 12:19 UT, approximately four to five hours before the event. They show two brightened regions in the 175 and 284 Å coronal and 304 Å transition-region lines, which correspond to chromospheric flocculi on either side of the optical filament (line separating the field polarities). The preeruptive X-ray arcade is almost invisible in all three EUV bands, but there are compact bright knots located

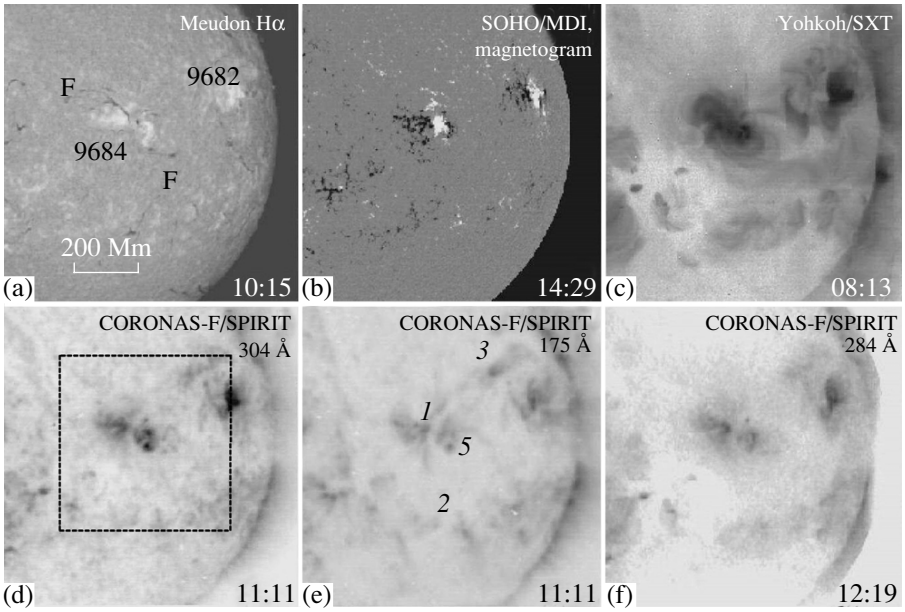

Fig. 1. Heliograms for the western disk sector illustrating the situation preceding the eruptive event of November 4, 2001. (a) Meudon Hα heliogram and (b) SOHO/MDI longitudinal magnetogram (light and dark colors correspond to northern and southern polarities, respectively). (c) Negative Yohkoh/SXT soft X-ray and CORONAS-F/SPIRIT 304, 175, and 284 Å EUV images (in order of increasing temperature).

at the ends of the central part of the arcade, which probably mark fragments of the most intense loops. In addition, there are the narrow bright (especially at 175 Å) structures 1-2 and 1-3. The first bright knot corresponds to the western edge of the bright loops bounding the X-ray arcade from the south, while the second represents a magnetic connection between the region AR 9684 and the northern neighborhood of the region AR 9682 to the west.

The event itself (Fig. 2) started with the disappearance of the southern portion of the Hα filament noted above, and included a prolonged flare of magnitude 3B/X1.0 with its maximum at 16:20 UT. The halo-type CME was observed starting from 16:35 UT with the LASCO/C2 coronagraph. The CME displayed a bright, large-scale loop structure ascending above the western limb with a high velocity in the plane of the sky ~1800 km/s. By 16:40 UT, a region of lower intensity but rapidly growing CME brightness enveloped the eastern limb sector as well, and the CME became visible around the entire occulting disk. After 17:00 UT, the CME front was visible for all positional angles of the LASCO/C3 coronagraph against an intense background of numerous protons arriving at the Earth with energies of tens of MeV (see http://cdaw.gsfc.nasa.gov/CME_list/).

4. POST-ERUPTIVE ARCADE

Figure 3 shows the post-eruptive arcade that was formed after the flare maximum observed in soft X-rays at 16:20 UT, which was detectable over many hours. The earliest CORONAS-F/SPIRIT images obtained at 304 and 175 Å at 17:03 UT (Figs. 3a and 3d) show that the most intense (i.e., dark in the negative heliograms in Fig. 3) central part of the arcade imaged in the transition region and low-temperature coronal lines consisted of a system of loops extending north–south. The brightest loops and their elements were observed at the ends of this structure (1-4 and 2-5) and also at its eastern (1-2) and western (4-5) sides. These features of the arcade, especially the predominance of the western edge, became increasingly distinct during the subsequent 26 min of observations at 304 and 175 Å (Figs. 3b and 3e). The heliograms also show the presence of some additional diffuse loops with much lower intensity to the northeast, east, and southwest of the central part of the arcade.

The heliogram obtained at 18:38 UT in the high-temperature 284 Å band (Fig. 3g) explains the multi-component structure and subsequent evolution of the arcade. We can clearly see that the brightest structure (1-4–5-2) bounds the northern and southern ends, and also the western edge of the central part of the arcade. To the south, this component of the arcade becomes a less intense but broader and more distinct system of loops (5-6–7-2). In addition, the bright loops 8-9 and surrounding halo were observed at the northern end of the arcade. The arcade component situated to the west of the bright edge 4-5 was less

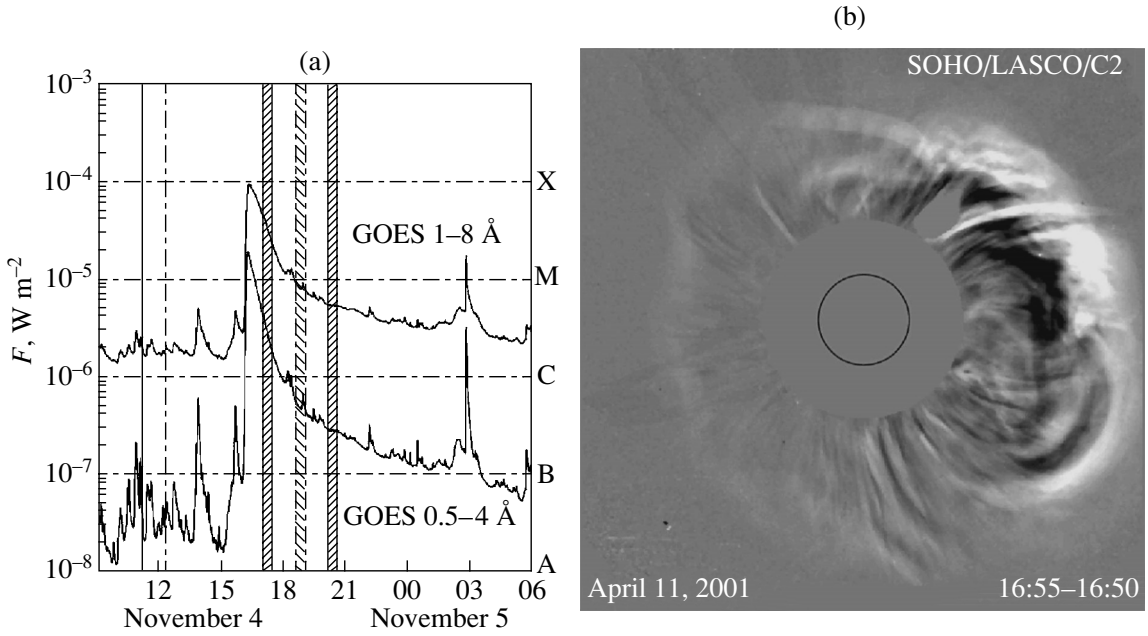

Fig. 2. (a) Evolution of soft X-ray emission detected by the GOES satellites and (b) difference CME image obtained with the SOHO/LASCO/C2 coronagraph. The vertical lines in (a) mark the times and intervals of the SPIRIT observations discussed in the text; the solid line and heavy shading mark the 175 and 304 Å bands, while the dashed line and light shading mark the 284 Å band. The central circle in (b) corresponds to the optical solar limb.

distinct. Whereas the identification of the eastern edge 1-2 with the bases of loops of the central part of the arcade seems quite obvious, there are two possible interpretations of the bright western edge 4-5: it could represent either the footpoint regions or the bright tops of loops (the so-called "spine" of the post-eruptive arcade [25, 26]).

To find an answer, we consider the Yohkoh/SXT soft X-ray image displaying high-temperature plasma with $T_e > 2.5$ MK at 18:45 UT (Fig. 3j). Along with the eastern branch (1-4) of the looplike structure, it also partially shows its western branch (4-9) at that time. We also see that its center (i.e., top 4 of this looplike structure) is the origin of the bright edge 4-5. Moreover, this edge shows a distinct southeastern extension in the form of the bright ray 5-10 passing through the central zone of the southern component of the arcade. Consequently, this suggests that both the bright edge 4-5 (observed at 304, 175, 284 Å, and in soft X-rays) and its extension 5-10 represent the bright tops of loops forming the post-eruptive arcade. Due to the observed orientation of the arcade, the "spine" in the central part of the arcade projected onto the plane of the sky may coincide with a ribbon of the western footpoints of the loops.

The subsequent arcade evolution observed in the low-temperature 304 and 175 Å bands (Figs. 3c, 3f) after 3 hours displayed a considerable decrease in the brightness of the tops of the loops of the central component of the arcade, and the extended bright structures 1-2-7 and 9-5-6 corresponding to the eastern and western footpoints of the post-eruptive loops became the most visible. Our analysis shows that the shapes and locations of these structures coincide with those for the diverging postflare Hα ribbons typical of eruptive events. The enhanced brightness at the tops of the loops in the central and southern components of the arcade was maintained over at least tens of minutes in the high-temperature 284 Å band (Figs. 3h, 3i) and over four hours in soft X rays (Figs. 3k, 3l).

Note also that the post-eruptive arcade evolved above the Hα filament and the magnetic separation line at the location of the twisted preeruptive structure (Fig. 1). The bright central component of the post-eruptive arcade was observed within the active region, while the more diffuse southern component of the arcade was observed above the zone of partial disappearance of the optical filament. The extent of the bright central part of the arcade was about 110 000–130 000 km and reached 330 000 km when the comparatively weak southern and northern components observed at 284 Å were included (Fig. 3g–3i). Such arcade lengths are typical for the largest-scale eruptive events occurring in active regions. For example, the EUV post-eruptive arcade for the Bastille Day

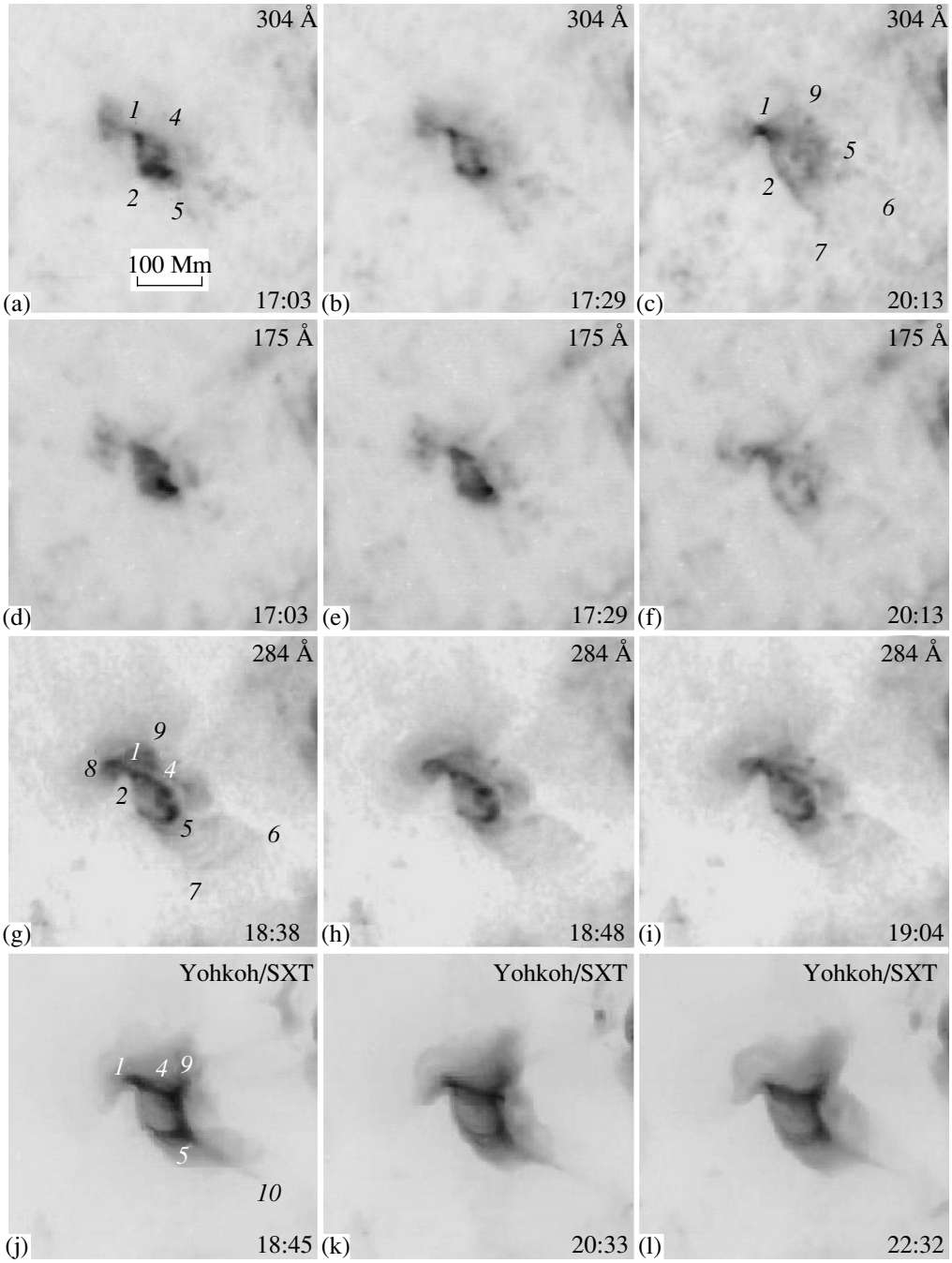

Fig. 3. Post-eruptive arcade in the negative CORONAS-F/SPIRIT heliograms obtained in the (a–c) 304 Å, (d–f) 175 Å, and (g–i) 284 Å EUV bands using the Yohkoh/SXT and (j–l) in soft X rays. All images show the area marked by the dashed frame in Fig. 1d.

event of July 14, 2000, had a total length of about 200 000 km [21].

5. DIMMINGS

The rerotated difference heliograms presented in Fig. 4 for the same western sector of the disk as in Fig. 1 illustrate the considerable size and multicomponent structure of the post-eruptive arcade. Here, the images of the central and northern components of the arcade are saturated, since the intensity thresholds were chosen for optimal representation of the dimmings accompanying the event. The images at 11:11 UT for 175 and 304 Å and at 12:19 UT for 284 Å were used as the reference preeruptive images to be compared with other heliograms obtained by the SPIRIT telescope during the event. Despite the long

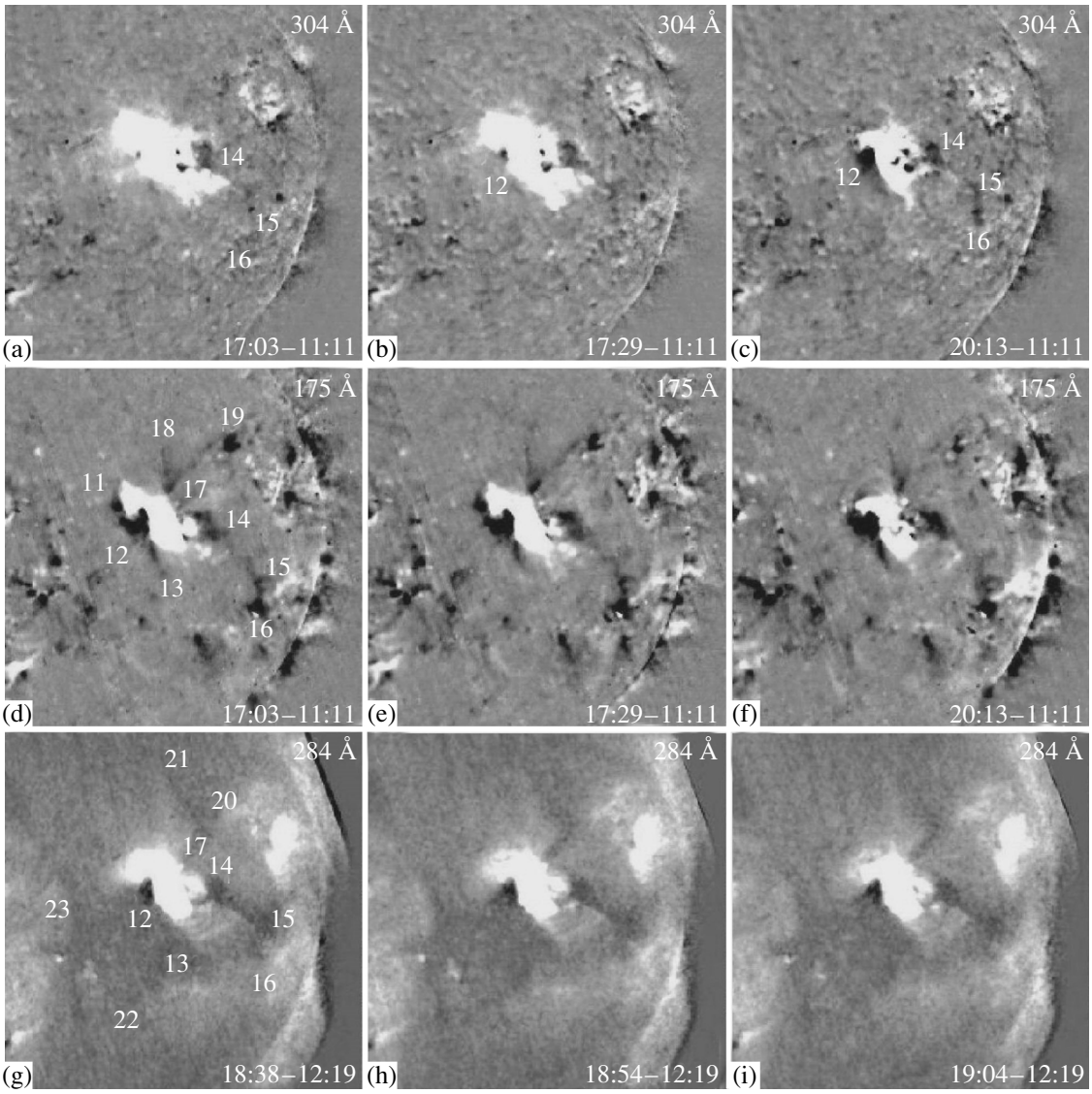

Fig. 4. Rerotated difference CORONAS-F/SPIRIT images demonstrating the bright post-eruptive arcade and dark dimmings in the (a–c) 304 Å, (d–f) 175 Å, and (g–i) 284 Å EUV bands for the same western sector of the disk as in Fig. 1.

time interval between these images and the onset of the event (four to five hours), they were chosen as the base images because there were no significant flares or eruptions on the Sun, in particular, in the western half of the disk, during this interval. Therefore, the main changes detected in the difference images of the active region AR 9684 and surrounding area can be interpreted as phenomena associated with the eruptive event analyzed. We emphasize again that, as in [8, 9], obtaining correct difference images for such a long time interval requires a preliminary rerotation of the heliograms to the same single time. In our case, the correction for the solar rotation was carried out by rerotating the heliograms to the times of the first images obtained by the SPIRIT telescope after the event maximum, namely, 17:03 UT for the images at 175 and 304 Å and 18:38 UT for those at 284 Å.

Let us start our analysis of the dimmings with the 175 Å data, since, according to [9], the most distinct dimmings are manifest in coronal lines with moderate excitation temperatures. Figure 4d, which corresponds to 17:03 UT, shows, along with the bright arcade, some persistent dark structures with intensities that are significantly lower than in the reference image. These structures are dimmings. The most prominent are the comparatively compact dimmings (11-12-13 and 14) immediately adjacent to the post-eruptive arcade at its eastern and western sides. These are accompanied by several narrow, extended channeled dimmings. One of these

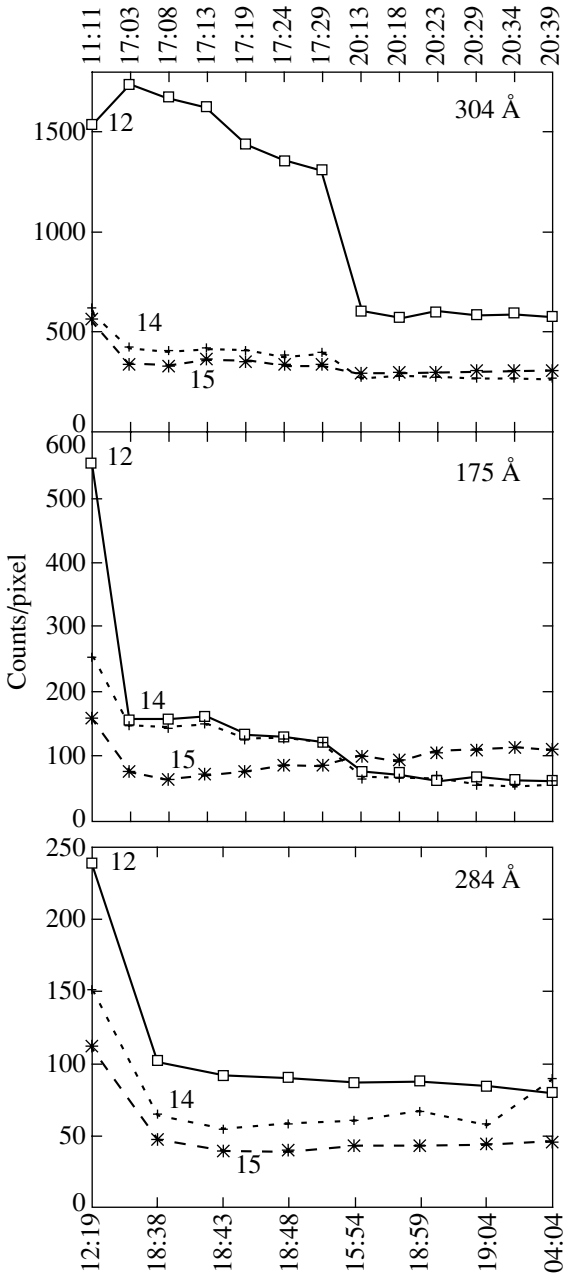

Fig. 5. Intensity variations within 22″ × 22″ areas at the centers of the main dimmings labeled in Fig. 4 by the corresponding numbers, depending on the observation times of the images.

(14-15-16) extends from the western compact dimming to the southwestern limb. Two other channeled dimmings (17-18 and 17-19) extend to the north and northwest from the dimming node at the northern end of the arcade. The length of these channeled dimmings is comparable to or greater than the extent of the post-eruptive arcade. There are other dark structures in this difference image, in particular, located some distance to the east and south of the arcade, and also near the western limb. However, they are not directly connected to the arcade, and their origin and possible relation to the event under study remain open to question. The difference heliograms for 17:29 and 20:13 UT (Figs. 4e, 4f) show that the main dimmings observed at 175 Å persisted with almost no variability over the three subsequent hours. This is true, in particular, for the compact dimmings 12-13 and 14, as well as the southwestern channeled dimming 14-15-16. The northern and northwestern dimming channels 17-18 and 17-19 became somewhat weaker, although they maintained their form and location.

With some small differences, the main dimming structures are also present in the analogous difference images for the highest-temperature band, 284 Å (Figs. 4g–4i). Here, we see the central (12) and, partially, the southern (12-13) components of the eastern dimming, while there is some brightening in place of the northern fragment of this dimming at 175 Å. At 284 Å, we can also see the western compact dimming 14 and southwestern dimming channel 14-15-16, which originates from it and passes around the southern component of the post-eruptive arcade. In place of the northwestern channeled dimmings, there is the relatively weak bent dimming structure 17-20-21. The heliograms for 284 Å also show that the transient coronal hole 12-13-22-23, which is relatively weak but extended in area, apparently adjoins the eastern compact dimming 12.

The difference heliograms for the low-temperature 304 Å band (Figs. 4a, 4b) also show the compact western dimming 14 adjacent to the arcade and, partially, the southwestern dimming channel 14-15-16. However, there is almost no sign of the northwestern channeled dimmings. The most important difference from the dimming features observed at 175 and 284 Å is that, at 304 Å, the central portion of the eastern compact dimming 12 initially displayed only a small dark element (Fig. 4a). The area of this dimming gradually increased during 17:03–17:29 UT (Fig. 4b). The dimming takes approximately the same shape and area as those observed in the earliest images at 175 and 284 Å (Figs. 4d and 4g) only until 20:13 UT (Fig. 4c). At 304 Å, the southwestern channeled dimming 14-15-16 is also the most distinct in the heliogram for 20:13 UT. Note that the dark elements visible at 304 Å within the central part of the post-eruptive arcade resulted from the fading or shifting of bright features observed in the reference heliogram for 11:11 UT (Fig. 1d).

When analyzing the heliograms for 304 Å, we should keep in mind that, in principle, the SiXI coronal line can contribute to the observed intensity variations along with the HeII transition-region line, al-

though it is obvious that the contribution of the silicon line will not dominate. The distinctions between the dimmings described above observed at 304 Å, on the one hand, and 175 and 284 Å, on the other hand, support of this statement. In addition, comparison with the preeruptive heliograms (Fig. 1) shows that it is primarily structures localized either at the very edge of or beyond the active region that were subject to dimmings as a result of the eruption. In particular, the compact dimmings 12-13 and 14, which coincided in all three bands (i.e., in all three corresponding temperatures), at least at the end of the event, were localized along the eastern edge 1-2 and southwestern end 5 of the preeruptive arcade, while the southwestern dimming channel 14-15-16 and the channeled dimming 17-19, observed mainly at 175 Å, were identified with bright structures extending from AR 9684 to the northwestern and southeastern limb sectors of the disk.

The extent of the intensity decrease in the dimmings compared to the preeruptive level, that is, the dimming depth, is presented in Fig. 5. The variations in the radiation intensity integrated over a $22'' \times 22''$ areas located in the central portion of the main dimmings labeled by the corresponding numbers in Fig. 4 are shown here for the three EUV bands. Since the time intervals between neighboring images differ widely (see Section 2 and Fig. 2), the time when the sequenced images were obtained is indicated along the horizontal axis. The curves in Fig. 5 show that the dimming depth reaches tens of percent in all three EUV bands. The channeled dimmings were somewhat shallower than the dimmings at the eruption center. The eastern (12) and western (14) compact dimmings (Figs. 5b, 5c) in the 175 and 248 Å coronal lines were the deepest (40–70%). The depth of these dimmings in the 304 Å transition-region line (Fig. 5a) is also significant (30–60%). In addition, we can see that, at 304 Å, the eastern dimming 12 appreciably lags behind the other dimmings in both area (see above) and depth. While the main decrease in the emission of the dimmings in the coronal lines and the western dimming (14) in the transition-region line was detected in the first images following the onset of the event (17:03 UT for 175 and 304 Å and 18:38 UT for 284 Å), the emission of the eastern dimming 12 at 304 Å significantly decreased between 17:29 and 20:13 UT. Figure 5 also indicates that the decreased intensities in both the coronal and transition-region dimmings were maintained over several hours.

6. DISCUSSION AND CONCLUSIONS

Our analysis of CORONAS-F/SPIRIT data at 175, 284, and 304 Å, which correspond to different temperatures, for a large-scale, geoeffective CME displays the long-term and large-scale phenomena typical of the post-eruptive phase of such events, namely, a post-eruptive arcade and dimmings.

The structure of the post-eruptive arcade provides evidence for a transition from the complex, sheared magnetic structure observed in soft X rays before the event to a topologically simpler system of hot loops with a different orientation, transverse to the extended filament passing through the active region [3]. Thus, the magnetic configuration after the eruption differs significantly from the preeruptive configuration. This represents a convincing argument in favor of magnetic reconnection. In our case, the total extent of the arcade was almost 0.5 $R_\odot$. The brightest central and northern parts were observed within the active region (i.e., in the region of strong magnetic fields), while the broad but much weaker southern component formed above the region where the filament disappeared beyond the active region. As a whole, the post-eruptive arcade was probably generated in the region of structures involved in the CME and subjected to the largest disturbances.

The loop tops rather than the footpoint regions in the post-eruptive arcade displayed the maximum brightness over several hours, especially in the high-temperature 284 Å band and in soft X rays. Similar bright arcade spines have been detected in numerous other eruptive events [12, 26]. In particular, a spine was observed between two bright ribbons corresponding to the footpoints of numerous loops constituting the post-eruptive arcade of the Bastille Day event of July 14, 2000, which occurred in the central sector of the disk [14, 15, 27]. This phenomenon is explained by heating of the plasma due to the prolonged post-eruptive energy release in current sheets high in the corona [15, 28, 29], when magnetic field lines opened or stretched in the course of CME eruption relax to a new closed equilibrium state via magnetic reconnection. The hot loops are first formed if energy release occurs in this manner. For this reason, some elements of the arcade observed in our event, i.e., its southern component in our case, are most clearly visible in the high-temperature bands. Another possible reason for the enhanced emission in the upper portion of the arcade is an increase in the longitudinal emission measure when the line of sight passes through the loop at its top. Numerous observations of limb events with distinct brightening at loop tops (see, for example, [30]) support this possibility. For the location and orientation of our arcade, the lines of sight passing through the upper portions of loops can also intersect the western footpoints of the loops, where the plasma density is increased due to the evaporation of chromospheric material.

Our analysis of dimmings displaying long-term and deep post-eruptive darkening was possible due to the technique of compensating for the solar rotation by forming rerotated difference images (see [8, 9]) using heliograms separated by time intervals of many hours. In this case, similar to other eruptive events occurring in the complex global magnetosphere typical of the solar maxima, there were two types of dimmings: (i) compact, clearly distinguishable dimmings adjacent to the eruption center and (ii) narrow, extended channeled dimmings passing from the eruption center to distant regions of the disk. The compact dimmings adjacent to the eruption center and some channeled dimmings observed in lines with different excitation temperatures in the 175, 284, and 304 Å bands mainly coincide in their structure and location. This suggests that these dimmings were formed from long-lived structures located on either side of the post-eruptive arcade, with its magnetic-field lines becoming either entirely or partially open during the CME. The plasma outflow from these structures along open field lines decreases the plasma density, resulting in a decrease in the intensity of the EUV emission [1, 4, 6–10]. An additional argument in favor of this model for the formation of the dimmings is that the lifetimes of dimmings are many hours and the recovery to a closed configuration is a very slow process (see, for example, [31, 32]). The behavior of individual portions of compact and channeled dimmings that appeared different in different coronal and transition-region lines may be associated with variations in the plasma temperature during the CME.

The presence of dimmings in both the coronal and transition-region lines indicates that their formation due to either the opening of magnetic field lines or plasma-temperature variations affected not only the corona but also the comparatively cool transition-region plasma. Analyzing the eastern and western compact dimmings 12 and 14, whose behavior at 304 Å was different, we can see that the involvement of the transition region in the CME process and the formation of dimmings there can occur either synchronously with associated coronal dimmings or with a considerable time delay of the order of tens of minutes.

Overall, the presence of a large-scale post-eruptive arcade and extended channeled dimmings provides evidence that the CME affected not only various magnetospheric layers above the eruption center but also more global structures, on scales significantly exceeding the size of the active region.

ACKNOWLEDGMENTS

The authors are grateful to the scientific leaders of the CORONAS-F project, especially V.N. Oraevskiĭ and I.I. Sobel'man, who supervised the project, and the team members of the DLR Remote Sensing Data Center (Neustrelitz, Germany) and IZMIRAN Cosmic Information Technology Center (Troitsk, Russia) for the observational data used in our study. The authors also thank the staffs of the Yohkoh/SXT telescope, SOHO/MDI magnetograph, SOHO/LASCO coronagraph, GOES project, and Paris-Meudon Observatory for additional data used. This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-17272, 03-02-16049, and 03-02-16591) and the Federal Ministry of Industry and Science (project nos. NSh 477.2003.2 and 1445.2003.2).

REFERENCES

1. H. S. Hudson and D. F. Webb, in *Coronal Mass Ejections*, Ed. by N. Crooker, J. Joselyn, and J. Feynman, AGU Geophys. Monograph Ser. **99**, 27 (1997).
2. D. F. Webb, J. Atmos. Sol.-Terr. Phys. **62**, 1415 (2000).
3. A. C. Sterling, J. Atmos. Sol.-Terr. Phys. **62**, 1427 (2000).
4. N. Gopalswamy and B. J. Thompson, J. Atmos. Sol.-Terr. Phys. **62**, 1427 (2000).
5. H. S. Hudson and E. W. Cliver, J. Geophys. Res. **106**, 25199 (2001).
6. B. J. Thompson, S. P. Plunkett, J. B. Gurman, *et al.*, Geophys. Res. Lett. **25**, 2465 (1998).
7. D. M. Zarro, A. C. Sterling, B. J. Thompson, *et al.*, Astrophys. J. **520**, L139 (1999).
8. I. M. Chertok and V. V. Grechnev, Astron. Zh. **80**, 162 (2003) [Astron. Rep. **47**, 139 (2003)].
9. I. M. Chertok and V. V. Grechnev, Astron. Zh. **80**, 1013 (2003) [Astron. Rep. **47**, 934 (2003)].
10. L. K. Harra and A. C. Sterling, Astrophys. J. **561**, L215 (2001).
11. S. Kahler, Astrophys. J. **214**, 891 (1977).
12. J. A. Klimchuk, L. W. Acton, K. L. Harvey, *et al.*, in *X-ray Solar Physics from Yohkoh*, Ed. by Y. Uchida, T. Watanabe, K. Shibata, and H. S. Hudson (Universal Acad. Press, Tokyo, 1994), p. 181.
13. A. C. Sterling and H. S. Hudson, Astrophys. J. **491**, L55 (1997).
14. L. Fletcher and H. Hudson, Sol. Phys. **204**, 71 (2001).
15. M. J. Aschwanden and D. Alexander, Sol. Phys. **204**, 93 (2001).
16. S. Tsuneta, L. Acton, M. Bruner, *et al.*, Sol. Phys. **136**, 37 (1991).
17. J.-P. Delaboudinière, G. E. Artzner, J. Brunaud, *et al.*, Sol. Phys. **162**, 291 (1995).
18. B. N. Handy, L. W. Acton, C. C. Kankelborg, *et al.*, Sol. Phys. **187**, 229 (1999).
19. V. N. Oraevskiĭ and I. I. Sobel'man, Pis'ma Astron. Zh. **28**, 457 (2002) [Astron. Lett. **28**, 401 (2002)].
20. I. A. Zhitnik, O. I. Bougaenko, J.-P. Delaboudinière, *et al.*, in *Proceedings of the 10th European Solar Physics Meeting*, ESA SP **506**, 915 (2002).

21. *Topical Issue on the 2000 Bastille Day Flare Event*, Ed. by O. Engvold and Z. Švestka, Sol. Phys. **204** (2001).
22. P. H. Scherrer, R. S. Bogart, R. L. Bush, *et al.*, Sol. Phys. **162**, 129 (1995).
23. G. E. Brueckner, R. A. Howard, M. J. Koomen, *et al.*, Sol. Phys. **162**, 357 (1995).
24. R. C. Canfield, H. S. Hudson, and D. E. McKenzie, Geophys. Res. Lett. **26**, 627 (1999).
25. Z. Švestka and E. Cliver, in *Eruptive Solar Flares*, Ed. by Z. Švestka *et al.* (Springer, New York, 1992), p. 1.
26. A. H. McAllister, Y. Uchida, J. I. Khan, and K. Shibata, in *X-ray Solar Physics from Yohkoh*, Ed. by Y. Uchida, T. Watanabe, K. Shibata, and H. S. Hudson (Universal Acad. Press, Tokyo, 1994), p. 189.
27. I. M. Chertok, V. V. Fomichev, A. A. Gnezdilov, *et al.*, Sol. Phys. **204**, 141 (2001).
28. J. M. McTiernan, S. R. Kane, J. V. Loran, *et al.*, Astrophys. J. **416**, L91 (1993).
29. M. A. Livshits, O. G. Badalyan, and A. V. Belov, Astron. Zh. **79**, 659 (2002) [Astron. Rep. **46**, 597 (2002)].
30. P. T. Gallagher, B. R. Dennis, S. Krucker, *et al.*, Sol. Phys. **210**, 341 (2002).
31. V. G. Fainshtein, G. V. Rudenko, and V. V. Grechnev, Sol. Phys. **181**, 133 (1998).
32. S. W. Kahler and H. S. Hudson, J. Geophys. Res. **106**, 29239 (2001).

Translated by V. Badin

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 418–432. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 459–474.
Original Russian Text Copyright © 2004 by Pipin.

Variations in the Solar Luminosity, Radius, and Quadrupole Moment as Effects of a Large-Scale Dynamo in the Solar Convection Zone

V. V. Pipin

Institute of Solar–Terrestrial Physics, Siberian Division, Russian Academy of Sciences, P.O. Box 4026, Irkutsk, 664033 Russia

Received October 9, 2003; in final form, November 10, 2003

Abstract—The effect of large-scale magnetic fields generated by the solar dynamo on the irradiance of the Sun and stratification of the solar convection zone is studied using a numerical model of a spherical axisymmetric dynamo. This model provides a joint description of the generation of large-scale magnetic fields, differential rotation, and convective heat transfer taking into account energy transformations associated with the large-scale magnetic fields, as well as the stratification of the convection zone. The model further develops a previously suggested self-consistent approach to analyzing solar luminosity variations, based on the conservation of the energy of the large-scale magnetic fields and turbulent flows. The results indicate that the increase in the solar luminosity near the maximum of the cycle is mainly due to the dissipation of the energy of magnetic fields escaping to above the photosphere, with the partial conversion of this energy into radiation. In addition, near-photospheric magnetic fields strongly affect the latitudinal nonuniformity of the cyclic variations in the radiative flux. The large-scale magnetic field also influences the hydrostatic equilibrium of the convection zone and gives rise to 11-year variations in the sound speed with a relative amplitude of 10^{-3}. The cyclic magnetic activity generates oscillations in the quadrupole moment with an amplitude of $4.5 \times 10^{-9}(GM_\odot/R_\odot)$. According to our estimates, the variations in the solar radius are very small, $\sim 10^{-6}\ R_\odot$. Our numerical model is used to estimate the variations in the orbital periods of close binaries whose primaries have the same spectral class as the Sun. Modulation of the centrifugal force by torsional oscillations can provide a plausible explanation for variations in the orbital periods of the companion stars in these systems. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A number of factors could be responsible for variations in the luminosity and radius of the Sun. Quasi-regular oscillations of the luminosity and, possibly, radius—both 11-year cyclic and secular—are of considerable interest. The luminosity increases by about 0.1% from the minimum to the maximum of the magnetic cycle (see, e.g., [1, 2]). Comparisons of data for the magnetic activity of the Sun with data for its luminosity and radius suggest that the 11-year and secular variations in the irradiance are most likely due to modulation of the heat flux by magnetic fields generated in the solar convection zone.

Extensive information on possible physical mechanisms for the magnetic modulation of the solar luminosity and radius has been accumulated in the literature. We will briefly note the most promising mechanisms, directing the reader to appropriate reviews and papers for details [1, 3–6]. According to the analyses presented in [3] and [5–7], the most plausible mechanisms for the modulation of the solar irradiance and radius are (1) variations in the strength of heat sources and sinks with the activity cycle due to the dissipation and generation of global magnetic fields and flows in the solar convection zone; (2) the magnetic-shadow effect (the magnetic field reduces the effective thermal conductivity, suppressing heat flow to the solar surface); and (3) the contribution of magnetic fields localized above the photosphere. Analytical estimates and detailed numerical calculations of these effects [5, 6] indicate that the third mechanism is most significant. Thermal variations produced in the convection zone by the first and second mechanisms are efficiently smoothed (smeared) over the entire zone due to the enormous thermal diffusivity. In other words, the solar interior possesses a high thermal inertia, which, however, decreases toward the surface. Therefore, the closer the heat source or sink to the surface, the larger its contribution to luminosity variations. Sources located above the surface in a transparent region will be most efficient. It is for this reason that the main contribution is made by magnetic fields that have escaped into the region above the photosphere.

The history of measurements of the solar radius is more than three centuries long. Most researchers who have studied this problem are convinced that the solar radius is not constant in time. We can only currently

1063-7729/04/4805-0418$26.00 © 2004 MAIK "Nauka/Interperiodica"

state with certainty that variations in the radius during the solar cycle do not exceed 0.005% [8]. Many researchers, such as Basu [9], Rozelot [10], and Neto *et al.* [11], have suggested that variations in the solar radius and the sunspot cycle are interrelated. In addition, there is no universal opinion as to whether the radius variations are in phase or antiphase with the solar cycle. For example, Gilligand [12] and Delache *et al.* [13] (see also [8]) report an anticorrelation between variations in the radius and magnetic activity. At the same time, measurements of the apparent solar diameter based on images in a neutral-iron line [14] suggest the opposite. The results concerning secular variations in the solar radius are more consistent [9, 10]. At secular activity minima, the solar radius is likely to be smaller than normal. Note that increased nonuniformity of rotation and slower overall rotation of the surface layers of the Sun are also characteristic of secular minima [9, 10, 15]. This may indicate that a significant role in the mechanism giving rise to solar-radius variations is played by the rate of rotation of the surface layers—for example, via variations in the centrifugal potential of the Sun. We will consider the role of this mechanism in the modulation of the solar radius and quadrupole moment in more detail below.

Here, we quantitatively study effects related to the modulation of the solar luminosity and radius by the large-scale magnetic field generated by the solar dynamo. As is shown in [5, 6], this problem can be self-consistently formulated in the framework of mean-field magnetohydrodynamics. The first calculations implementing this type of scheme [16] proved this to be a promising approach. The current study generalizes the previously obtained results. We take into account departures from hydrostatic equilibrium in the convection zone induced by the large-scale magnetic field. This enables us to analyze modulations of the radius and quadrupole moment with the solar activity cycle.

The outline of the article is as follows. The governing equations of the model will be considered in Section 2, the numerical results will be analyzed in Section 3, and the article will conclude with a brief discussion of the results.

2. GOVERNING EQUATIONS

The governing equations describing the evolution of large-scale fields in a turbulent medium can be obtained using the techniques of mean-field magnetohydrodynamics. The general assumptions and derivation of these equations are given in [5, 16]. We present below a brief exposition of the basic equations and boundary conditions adopted in our numerical calculations. We consider the following main physical variables: the induction vector of the axisymmetric large-scale magnetic field

$$\mathbf{B} = \mathbf{e}_\phi B(r,\theta) + \mathrm{curl}\left(\frac{\mathbf{e}_\phi A(r,\theta)}{r\sin\theta}\right), \qquad (1)$$

the large-scale velocity field

$$\mathbf{V} = \mathbf{e}_\phi r\sin\theta\Omega(r,\theta) \qquad (2)$$

(we neglect meridional circulation), the specific entropy $S(r,\theta)$, and the pressure $P(r,\theta)$. We use a mixing-length approximation to describe convective heat transfer and the basic turbulence parameters, such as the rms velocity and the correlation time for convective flows.

2.1. Dynamo Equations

The generation of the large-scale magnetic field is described by the equation

$$\frac{\partial \mathbf{B}}{\partial t} = \mathrm{curl}\left\{\mathcal{E} + \mathbf{V}\times\mathbf{B}\right\}, \qquad (3)$$

where $\mathcal{E}$ is the turbulent electromotive force (emf), which represents the averaged effect of correlations between the fluctuations in the large-scale velocity field $\mathbf{u}'$ and the magnetic field $\mathbf{B}'$, and $\mathcal{E} = \langle \mathbf{u}'\times\mathbf{B}'\rangle = \mathcal{E}_{gen} + \mathcal{E}_{pump} + \mathcal{E}_{diff}$, where $\mathcal{E}_{gen}$, $\mathcal{E}_{pump}$, and $\mathcal{E}_{diff}$ are the components of the turbulent emf responsible for the generation (α effect), transfer, and dissipation of the large-scale magnetic field, respectively. It is assumed for the Sun that nonuniformity of the density and the rotation-induced anisotropy of the turbulence make the most important contributions to the turbulent emf; the turbulent emf for such conditions was calculated in [17, 18]. Summary expressions for $\mathcal{E}_{gen}$, $\mathcal{E}_{pump}$, and $\mathcal{E}_{diff}$ can be found in [19]. We present below the final form of the dynamo equations in a spherical coordinate system, together with some commentary.

Thus, after substituting the known expressions for the components of the turbulent emf taking into account (1) and (2), we obtain

$$\frac{\partial B}{\partial t} = \frac{1}{r}\left\{\frac{\partial(\Omega, A)}{\partial(r,\theta)} + \frac{\partial r\mathcal{E}_\theta}{\partial r} - \frac{\partial \mathcal{E}_r}{\partial\theta}\right\}, \qquad (4)$$

$$\mathcal{E}_r = \eta_T\psi_\eta\left(B\left(-\frac{\alpha_M}{3\ell_c\gamma}\phi_1 + \frac{\phi_{||}}{r}\right)\sin\theta\cos\theta\right.$$
$$\left. - \frac{(\phi + \phi_{||}\sin^2\theta)}{r\sin\theta}\frac{\partial \sin\theta B}{\partial\theta} + \phi_{||}\sin\theta\cos\theta\frac{\partial B}{\partial r}\right)$$
$$+ C'_\alpha\frac{\alpha_M\phi_\alpha}{3\ell_c\gamma}\left((\mathcal{A}_1 + \mathcal{A}_2 + \mathcal{A}_4\cos^2\theta)\frac{\cos\theta}{r^2\sin\theta}\frac{\partial A}{\partial\theta}\right.$$
$$\left. + (\mathcal{A}_2 + \mathcal{A}_4\cos^2\theta)\frac{1}{r}\frac{\partial A}{\partial r}\right),$$

$$\mathcal{E}_\theta = \eta_T\psi_\eta\Bigg(B\Bigg(r(\phi_2 - 2\phi_1\sin^2\theta)\frac{\alpha_M}{3\ell_c\gamma} - \phi_{||}\cos^2\theta\Bigg) + (\phi + \phi_{||}\cos^2\theta)\frac{\partial rB}{\partial r} - \phi_{||}\sin\theta\cos\theta\frac{\partial B}{\partial\theta}\Bigg) + C'_\alpha\frac{\alpha_M\phi_\alpha}{3\ell_c\gamma} \times \Bigg((\mathcal{A}_2 + \mathcal{A}_4\cos^2\theta)\frac{1}{r^2}\frac{\partial A}{\partial\theta} + (\mathcal{A}_1 + \mathcal{A}_4\sin^2\ \theta)\frac{\cos\theta}{r\sin\theta}\frac{\partial A}{\partial r}\Bigg),$$

$$\frac{\partial A}{\partial t} = \eta_T\psi_\eta\Bigg((\phi + \phi_{||}\cos^2\theta)\frac{\partial^2 A}{\partial r^2} + \frac{\phi + \phi_{||}\sin^2\theta}{r^2}\frac{\partial}{\partial\theta}\frac{1}{\sin\theta}\frac{\partial A}{\partial\theta} + \phi_{||} \times \Bigg(\frac{\sin^2\theta}{r}\frac{\partial A}{\partial r} + \frac{3\sin\theta\cos\theta}{r^2}\frac{\partial A}{\partial\theta} - \frac{\sin 2\theta}{r}\frac{\partial^2 A}{\partial r\partial\theta}\Bigg) - \frac{\alpha_M}{3\ell_c\gamma}\Bigg((\cos^2\theta(\phi_1 - \phi_3) + \phi_1 + \phi_2)\frac{\partial A}{\partial r} - (\phi_1 - \phi_3)\frac{\sin 2\theta}{2r}\frac{\partial A}{\partial\theta}\Bigg)\Bigg) + C'_\alpha\eta_T\frac{\alpha_M}{3\ell_c\gamma}\psi_\alpha\mathcal{A}_1 r\sin\theta\cos\theta B, \quad (5)$$

where $C'_\alpha = C_\alpha\dfrac{\eta_0}{R_\odot\alpha_0}$ and C_α is a dimensionless quantity characterizing the power of the generation sources due to the α effect; η_0 and α_0 are the maxima of the coefficients of magnetic diffusivity and the α effect; $\alpha_M = H_p/\ell_c$ is a parameter that appears in the mixing-length theory and relates the pressure scale height H_p to the correlation length of turbulent convective flows ℓ_c; and $\phi_{n,||}$ and $\mathcal{A}_n$ are functions of the Coriolis number defined in [17, 18] (see also [19]), which describe the influence of rotation on the transfer, dissipation, and generation of the magnetic field.

The Coriolis number is defined as $\Omega^* = 2\tau_c\Omega_0$, where τ_c is the correlation time for the turbulence and Ω_0 is the basic rotation rate of the Sun. The functions ψ_η and ψ_α obtained in [17, 18] depend on $\beta = B/\left(\sqrt{4\pi\rho}u_c\right)$ and take into account the weakening of the dissipation and the generation of the field due to the nonlinear action of the magnetic fields on the turbulence. We determine the eddy diffusivity η_T using the following relationships from mixing-length theory:

$$u_c = \frac{\ell_c}{2}\sqrt{-\frac{g}{c_p}\frac{\partial S}{\partial r}}, \quad \eta_T = P_m\nu_T, \quad \nu_T = \frac{u_c\ell_c}{3}, \quad (6)$$

where u_c is the rms velocity of the convective flows and S is the entropy. We have also introduced an additional parameter—the magnetic Prandtl number $P_m = \nu_T/\eta_T$, where ν_T is the eddy viscosity. The generation of the toroidal magnetic field (4) is directly related to the differential-rotation law, which can be found from the angular-momentum balance equation for the solar convective envelope written in the next subsection.

2.2. Differential Rotation

Detailed descriptions of modern approaches to studying the differential rotation of the Sun are given in [20, 21]. We use a simple approximation that disregards meridional circulation. Since differential rotation plays a secondary role in the problem at hand, we will parameterize the sources of differential rotation (the so-called Λ effect). The main reason for parameterizing the Λ effect is to obtain a rotation law that, by and large, is consistent with helioseismological data. We already employed such an approach in [22, 23]. Thus, the angular-momentum balance in the solar convection zone can be written as

$$\rho r\sin\theta\frac{\partial\Omega}{\partial t} = \frac{1}{r^3}\frac{\partial}{\partial r}r^3 \times \left(T_{r\phi} + \frac{B}{4\pi r^2\sin\theta}\frac{\partial A}{\partial\theta}\right) + \frac{1}{r\sin^2\theta}\frac{\partial}{\partial\theta} \times \sin^2\theta\left(T_{\theta\phi} - \frac{B}{4\pi r\sin\theta}\frac{\partial A}{\partial r}\right), \quad (7)$$

where $T_{r\phi}$ and $T_{\theta\phi}$ are the turbulent angular-momentum fluxes, which can be expressed in the form given in [23].

2.3. Heat Transfer and Stratification

It follows from the discussion above that all coefficients describing the effects of small-scale turbulence and parameters taking into account the influence of the mean fields on turbulence, such as Ω^* and β, are ultimately determined by the parameters of the "background" turbulence, such as the rms convection velocity u_c. The velocity of convective flows depends on the distribution of the gradient of the specific entropy $S = c_v\ln(P/\rho)$.

An entropy equation taking into account energy sources and sinks due to large-scale fields was obtained by Pipin and Kitchatinov [5]:

$$\rho T\frac{\partial S}{\partial t} = \nabla_{\mathbf{i}}\rho T\chi_{ij}\nabla_{\mathbf{j}}S - T_{ij}\nabla_{\mathbf{i}}V_j - \frac{1}{4\pi}\left(\mathcal{E}\cdot\mathbf{curlB}\right), \tag{8}$$

where χ_{ij} is the thermal-diffusivity tensor for the turbulent convective flows. Using the expression for χ_{ij} given in [17], Eq. (8) can be rewritten as

$$\rho T\frac{\partial S}{\partial t} = -\frac{1}{r^2}\frac{\partial}{\partial r}r^2F_r - \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin^2\theta F_\theta + \left(T_{\theta\phi}\frac{\partial\Omega}{\partial\theta} + rT_{r\phi}\frac{\partial\Omega}{\partial r}\right)\sin\theta - \frac{1}{4\pi}\left(\mathcal{E}\cdot\mathbf{curlB}\right), \tag{9}$$

where the heat flux $\mathbf{F}$ is the sum of the convective and radiative fluxes: $\mathbf{F} = \mathbf{F}_c + \mathbf{F}_{rad}$. According to [20], the components of $\mathbf{F}_c$ and $\mathbf{F}_{rad}$ can be written in a spherical coordinate system as

$$F_{cr} = -\rho T\frac{\chi_T}{2}\left((\phi_1 + \phi_{||}\cos^2\theta)\frac{\partial S}{\partial r} - \phi_{||}\frac{\sin 2\theta}{2r}\frac{\partial S}{\partial\theta}\right), \tag{10}$$

$$F_{c\theta} = -\rho T\frac{\chi_T}{2}\left(\frac{(\phi_1 + \phi_{||}\sin^2\theta)}{r}\frac{\partial S}{\partial\theta} - \phi_{||}\frac{\sin 2\theta}{2}\frac{\partial S}{\partial r}\right), \tag{11}$$

$$F_{rad\,r} = \frac{16\sigma T^3}{3\kappa\rho^2 c_p}\left(\frac{\partial S}{\partial r} + \frac{\gamma-1}{\gamma}\frac{\partial P}{\partial r}\right),$$

$$F_{rad\,\theta} = -\frac{16\sigma T^3}{3\kappa\rho^2 c_p}\frac{1}{r}\frac{\partial S}{\partial\theta}.$$

The functions of the Coriolis number ϕ_1 and $\phi_{||}$ reflect the anisotropy of turbulent heat transfer in the rotating medium. Formulas for these functions can be found in [17]. The eddy thermal diffusivity $\chi_T = \nu_T$ can be determined from (6).

The temperature and density in (9) are determined by the equation of state for a perfect gas and by the stratification of the medium. These parameters can be found from the solution to the equations of hydrostatic equilibrium linearized around the basis state. Since the stratification of the medium deep in a stellar convection zone differs little from an adiabatic medium, we will use a spherically symmetric adiabatic distribution for all the thermodynamic parameters (except the entropy) in the solar convection zone, unperturbed by magnetic fields and rotation, as the reference stratification. We will also assume that most solar material is localized below the convection zone.

Thus, the reference stratification is determined by the equations

$$\frac{\partial T_{(0)}}{\partial r} = -\frac{\rho_{(0)}}{C_p}\frac{\partial}{\partial r}\Psi_{(0)}, \tag{12}$$

$$\rho_{(0)} = \rho_e\left(\frac{T_{(0)}}{T_e}\right)^{\frac{\gamma}{\gamma-1}},$$

$$\Psi_{(0)} = -\frac{GM_\odot}{r},$$

where ρ_e and T_e are specified at the outer boundary and taken from a model for the internal structure of the Sun. The influence of rotation and magnetic fields results in small deviations of the thermodynamic parameters from the reference stratification.

We represent the thermodynamic parameters as the sums of the reference distributions and perturbations, that is,

$$P = P_{(0)} + P_{(1)},\quad \rho = \rho_{(0)} + \rho_{(1)},\quad \Psi = \Psi_{(0)} + \Psi_{(1)}, \tag{13}$$

etc. These decompositions must be substituted into the general equation for the conservation of angular momentum in a rotating spherical convective envelope with a large-scale magnetic field:

$$-\nabla_{(p)}P - \rho\nabla_{(p)}\left(\Psi - \frac{1}{2}|\Omega\times\mathbf{r}|^2\right) + \frac{1}{4\pi}(\mathbf{curlB}\times\mathbf{B})_{(p)} + \mathrm{div}\mathbf{T}_{(p)} = 0, \tag{14}$$

$$\nabla^2\Psi = 4\pi\rho G,$$

where the subscript (p) denotes the poloidal components of the differentiation operator and vectors and $\mathbf{T}$ is the turbulent-stress tensor.

The full expression for $\mathbf{T}$ is fairly complex (see, e.g., [21]). Most important in our problem are the diagonal terms of $\mathbf{T}$, which have the form of a "turbulent" pressure; the nondiagonal components of the stress tensor are important when considering angular-momentum transfer and meridional circulation. The latter is not included in the problem at hand. Obviously, the azimuthal components of $\mathbf{T}$ do not contribute to (14) because of the axial symmetry that has been imposed. We will proceed in our calculations with the simplest representation of the diagonal components of $\mathbf{T}$ in the form $T_{ij} = \delta_{ij}\frac{\rho}{3}\left(u_c^2 + \frac{h^{(0)2}}{4\pi\rho}\right)f_\beta$, where $h^{(0)}$ is the background field of magnetic fluctuations excited by the small-scale dynamo and f_β allows for the effect of the large-scale magnetic fields on turbulence. It is reasonable to assume equipartition of the energy of the convective flows and small-scale magnetic fields $\left(u_c^2 \approx h^{(0)2}/4\pi\rho\right)$. The function

f_β was previously calculated in [24]. Further, the reference stratification (12) does not include the effect of the turbulent pressure; therefore, we normalize the known expression for f_β to zero, introducing the new function $f'_\beta = f_\beta - 1$, so that $f'_\beta = 0$ in the absence of magnetic fields. We thus obtain $T_{ij} = \frac{2\rho}{3} u_c^2 f'_\beta \delta_{ij}$, where

$$f'_\beta = \frac{3}{8\beta^2}\left(\frac{\beta^2-1}{\beta^2+1} + \frac{\beta^2+1}{\beta}\arctan\beta\right) - 1.$$

Substituting (13) into (14), applying (12), and neglecting self-gravitation effects yields the following equations for the perturbations of the pressure and potential:

$$\nabla^2\left(P_{(1)} + \frac{2\rho_{(0)}}{3}u_c^2 f'_\beta + \frac{B^2}{8\pi}\right) + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\rho_{(1)}\right) = \frac{1}{r^2}\frac{\partial}{\partial r}r^2 f_r + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta r f_\theta, \quad (15)$$

$$\nabla^2\Psi_{(1)} = 4\pi G\rho_{(1)},$$

where f_r, f_θ are the components of the external force, that is, the sum of the large-scale Lorentz force and the centrifugal force,

$$f_r = \Omega^2\rho_{(0)} r\sin^2\theta - \frac{B^2}{4\pi r},$$

$$r f_\theta = \Omega^2\rho_{(0)} r^2\sin\theta\cos\theta - \frac{B^2\cos\theta}{4\pi\sin\theta}.$$

Our formulation disregards the nondiagonal components of the Lorentz force, since we assume that the large-scale poloidal field is much smaller than the azimuthal field. The variations of the density and pressure are related as follows:

$$\frac{\rho_{(1)}}{\rho_{(0)}} = \frac{P_{(1)}}{\gamma P_{(0)}} - \frac{\delta S}{c_p}. \quad (16)$$

The Lagragian variations of the radius, $\xi_{(r)}$, can be estimated from the expression for the radial pulsations once the variations of the potential, $\Psi_{(1)}$, are known:

$$\xi_{(r)} = -\frac{1}{4\pi G\rho_{(0)}}\frac{\partial}{\partial r}\Psi^{(0)}_{(1)}, \quad (17)$$

where $\Psi^{(0)}_{(1)}$ is the zeroth mode in the multipole expansion of the perturbations of the gravitational potential.

3. ESTIMATE OF POSSIBLE VARIATIONS IN THE RADIUS

As Spruit notes [3], there are two types of variations in radius related to magnetic activity. The first type is perturbations due to the direct effect of magnetic fields and flows on the distributions of the plasma density and pressure in the region occupied by the field. This type of perturbation can be considered "mechanical." According to [3], the effect of a mechanical perturbation on the modulation of the radius can be directly estimated from the size of the region of perturbed density. For example,

$$\frac{\delta R}{R} \approx -\frac{d_{CZ}}{R}\frac{\rho_{(1)}}{\rho_{(0)}} = -0.3\frac{\rho_{(1)}}{\rho_{(0)}}, \quad (18)$$

where d_{CZ} is the depth of the solar convection zone.

Assume now that a 1-kG field is present in the convection zone of the Sun. The density is lower in the region occupied by the field, so that large-scale magnetic fields tend to increase the solar radius, and the relative amplitude of this action is $\frac{\delta R}{R} \approx \frac{d_{CZ}}{R}\frac{B^2}{8\pi P_{(0)}} = 2\times10^{-7}$; we set $P_{(0)} = P_e$, where P_e is the pressure at the outer boundary (see below). The modulation of the centrifugal force by torsional oscillations with a relative amplitude of $\delta\Omega/\Omega \sim 10^{-3}$, which corresponds to variations of the linear velocity of about 10 m/s, results in a radius modulation of about $\frac{\delta R}{R} \approx -\frac{d_{CZ}}{R}\frac{\rho\Omega^2R^2}{P_{(0)}}\frac{\delta\Omega}{\Omega} = \mp5\times10^{-7}$. This contribution depends on the sign of the velocity variations. A region that rotates more rapidly has a lower density relative to the ambient matter, which leads to a relative increase in the diameter. In addition, since the centrifugal force is maximum near the equator, this type of perturbation plays a special role in modulating the oblateness and quadrupole moment of the Sun.

Another mechanical-type contribution to perturbations of the radius results from the action of the magnetic fields on the turbulent stresses in the medium, in particular, on the turbulent pressure $\frac{2\rho u_c^2}{3}f'(\beta)$. The turbulent pressure is reduced in a magnetic field due to inhibition of the turbulence, which leads to a negative effect of this factor on modulations of the radius: $\frac{\delta R}{R} \approx -\frac{d_{CZ}}{R}\frac{2\rho u_c^2}{3P_{(0)}}f'_\beta = -2\times10^{-7}$.

Thermal effects of the magnetic fields and flows on variations of the radius are associated with the expenditure of energy to generate fields and flows,

which cools the solar plasma. Thus, thermal perturbations of the radius decrease the size of the Sun in the growth phase of the large-scale field. Thermal perturbations efficiently level off due to the enormous thermal inertia of the convection zone. Therefore, this mechanism seems to be efficient only in surface layers of the convection zone. According to our estimates, the entropy variations near the surface are $\delta S/c_p \sim 10^{-6}$ and the corresponding variations in the radius are $\dfrac{\delta R}{R} \approx \dfrac{d_{CZ}}{R}\dfrac{\delta S}{c_p} = 3 \times 10^{-7}$.

Order-of-magnitude estimates show that the modulation of the centrifugal force by torsional oscillations together with the direct influence of the large-scale field on the hydrostatic balance should play the determining role. This mechanism controls the quasi-quadrupole variations in the structure of the solar convection zone, since the centrifugal force obviously depends on the latitude. This conclusion is confirmed by the calculations presented below.

4. BOUNDARY CONDITIONS

We assume the following boundary conditions for the above equations: a superconductivity condition for the magnetic field at the base of the convection zone, i.e.,

$$\mathcal{E}_\theta = A = 0;$$

vanishing radial stresses for the angular-momentum balance (7), i.e.,

$$T_{r\phi} + \frac{B}{4\pi r^2 \sin\theta}\frac{\partial A}{\partial \theta} = 0; \qquad (19)$$

and vanishing perturbations of the potential, $\Psi_{(1)} = 0$, and vanishing convective flux, $F_c = 0$.

For the magnetic field at the surface, we assume the "partial reflection" of magnetic energy suggested previously by Kitchatinov *et al.* [19]; i.e., we introduce vacuum boundary conditions for the vector potential A and determine the azimuthal component of the large-scale field B from the relation

$$\mathcal{E}_\theta = -\frac{\eta_T}{R_\odot}\frac{B^2}{B_{(0)}},$$

where $B_{(0)}$ is the magnetic field for which the rate of escape of magnetic energy reaches $\eta_T/R_\odot$. If $B_{(0)} \to 0$, these boundary conditions become vacuum boundary conditions. For $B \ll B_{(0)}$, we obtain closed boundary conditions. The parameter $B_{(0)}$ is chosen in accordance with [19]: $B_{(0)} = 150$ G. The outer boundary condition for the equation describing the differential rotation is the same as the condition at the base of the convection zone and implies vanishing stresses (19). We smoothly match the perturbation of the potential at the surface with the external gravitational potential of the star in the form of a multipole expansion:

$$\Psi_{out} = -\frac{GM_\odot}{R_\odot} \times \left(\frac{R_\odot}{r} - \sum_{n=1}^{\infty} J_{2n}\left(\frac{R_\odot}{r}\right)^{2n+1} P_{2n}(\theta)\right).$$

As the heat-transfer equation (9) is solved, the thermal flux from the surface is approximated by the flux from a blackbody:

$$F_r = \frac{L_\odot}{4\pi r^2}\left(1 + 4\frac{T_e}{T_{eff}}\frac{\delta S}{c_p}\right),$$

where T_{eff} is the effective temperature of the photosphere and T_e is the temperature at the outer boundary of the integration domain.

5. BASIS CHARACTERISTICS AND PARAMETERS OF THE MODEL

Some basis parameters of our model are listed below. The following parameters are specified *a priori*. The value $\alpha_M = 1.7$ of the parameter of mixing-length theory corresponds to the standard value [25]. The base of the convection zone in our model, which is determined from the condition $L_\odot/4\pi r_b^2 = F_{rad\,r}$, is situated at $r_b = 0.71\ R_\odot$. We adopt $r_i = 0.715\ R_\odot$ for the inner boundary of the integration domain to avoid large gradients of the turbulence parameters near the base of the convection zone. The outer boundary of the integration domain is at $r_e = 0.96\ R_\odot$, which nearly corresponds to the base of the supergranulation layer on the Sun. In the region from r_i to r_e, the distribution of thermodynamic parameters is close to adiabatic. The reference stratification of the thermodynamic parameters can be determined if the temperature T_e and density ρ_e are specified at the outer boundary. These values can be taken from a model for the internal structure of the Sun. We use the parameters for the model of Stix [25]: $T_e = 1.9308 \times 10^5$ K and $\rho_e = 3.6 \times 10^{-3}$ g cm^{-3}.

Another model parameter is the basis sidereal rate of the solar rotation, $\Omega_0 = 2.86 \times 10^{-6}$ s^{-1}. We used the standard values for other parameters, such as $R_\odot$, $L_\odot$, and $M_\odot$.

Two control parameters appear in the dynamo equations—the magnetic Prandtl number $P_m = \nu_T/\eta_T$ and the dimensionless parameter C_α characterizing the power of the generation sources due to the α effect. The kinematic viscosity can be computed using the heat-transfer equation and the relationships of mixing-length theory described above. The calculations show (see below) that the maximum

kinematic eddy viscosity in the convection zone is $\nu_T \sim 2 \times 10^{13}$ cm^2 s^{-1}. In our computations, we used $P_m = 20$; this means that the maximum magnetic eddy diffusivity corresponds to $\eta_T \sim 10^{12}$ cm^2 s^{-1}. This η_T leads to a period of the magnetic cycle of $\sim$8–9 yr. The parameter C_α was chosen to be slightly above the minimum threshold value specifying the existence of solutions that do not decay with time: $C_\alpha \approx P_m^{-1}/4$.

Let us say a few words concerning the influence of the chosen model parameters on the results. The most important parameter is the magnetic Prandtl number. Variations in P_m affect the cycle period most strongly. As P_m is reduced, the period decreases proportionally. We note that there is virtually no qualitative difference between the cases $P_m = 20$ and $P_m = 10$ in terms of the behavior of the large-scale fields, e.g., in the direction of the drift of these fields with the cycle. No reasonably reliable procedure for the self-consistent determination of P_m is currently known. Simulations of the decay of sunspots suggest that the magnetic eddy diffusivity should be $\sim 10^{12}$ cm^2 s^{-1} [26]. In this case, with a viscosity of $\nu_T \sim 2 \times 10^{13}$ cm^2 s^{-1}, the choice of $P_m = 10$–20 is justified.

The system made up of (4), (5), (7), (9), and (15) was integrated numerically using a finite-difference method. We employed a conservative numerical scheme with 40 grid points in the radial direction and 80 in the latitudinal direction. In addition, a Legendre polynomial decomposition was used to solve (15). The evolution of the system was followed from the initial state until a certain well-established regime of magnetic activity. The initial state had an entropy, angular-velocity distribution, and stratification that were unperturbed by the magnetic field; we obtained these by solving the corresponding equations neglecting the effects of magnetic fields. Some of the results will be discussed briefly below.

Figure 1 presents the basis characteristics of our model (the solution unperturbed by magnetic fields): the distributions of the angular velocity, viscosity, rms convective velocity, and Coriolis number. The basis angular-velocity distribution in the convection zone agrees both qualitatively and quantitatively with the results of the helioseismological inversion reported by Schou *et al.* [27]. The characteristics of the turbulence shown in the center and the entropy distribution shown on the right are consistent with earlier results obtained by Kitchatinov *et al.* [19].

Figure 2 demonstrates the main properties of the solution to the dynamo equations for the distributions of angular velocity and entropy in Fig. 1. The effects of magnetic fields on the transfer of heat and angular momentum are neglected. The generation reaches saturation only because of the magnetic suppression of the α effect. The figure is plotted in latitude–time coordinates and represents the Maunder diagrams for the evolution of the large-scale toroidal magnetic field near the outer boundary of the convection zone (top). The evolution of the radial magnetic field is shown at the bottom. It is clear from the figure that the model is in good qualitative and quantitative agreement with the observations. Deficiencies of the model include the fact that the maxima of the toroidal field are located at a latitude of $\sim 40^\circ$ rather than $\sim 20^\circ$, as is observed. Moreover, the reversal of the radial field takes place near the equator shortly before the activity maximum, while it is observed, as a rule, after the maximum.

Depending on the degree of self-consistency and the physical processes involved, our model can be subdivided into partial subproblems, some of which have been considered earlier. For example, the interaction between the differential rotation and magnetic field can be described in terms of the so-called $\alpha\Lambda$ dynamo [23] (see also [22]). The dynamo is named differently from the standard $\alpha\Omega$ dynamo, because the differential rotation (Ω effect) is not specified *a priori* and is determined from the angular-momentum balance equations, in which the Λ effect—the source of generation of large-scale nonuniformity of the angular velocity—plays a determining role. The influence of the generation of large-scale fields on heat transfer in the convection zone and the radiative flux at its outer boundary has also been partially analyzed in [5–7, 28]. These results will be generalized below, taking into account variations in the convection-zone stratification.

Figure 3 shows the computed Maunder diagrams in time–latitude coordinates for variations of the toroidal magnetic field (near the surface), the departure of the surface angular velocity from the mean, the convective radiation flux $\dfrac{L_\odot}{\pi r_b^2}\dfrac{T_e}{T_{eff}}\dfrac{\delta S}{c_p}$, the total energy flux, and the gas pressure at the outer boundary. The evolution of the magnetic field does not display substantial differences from the simplest case presented in Fig. 2, since the dynamo being considered is subcritical. The second diagram from the top in Fig. 3 represents the evolution of the torsional oscillations in time–latitude coordinates. The maximum amplitude of the variations is $\sim$10 m/s.

The variations of the rotational velocity have two drift branches—polar and equatorial. These are due to the two basic mechanisms through which the magnetic fields influence the angular-momentum balance in the convection zone. The first is the Lorentz force produced by the large-scale magnetic field, and the second is the modification of the sources of differential rotation (Λ effect) in the magnetic field. These mechanisms were considered separately by Kueker

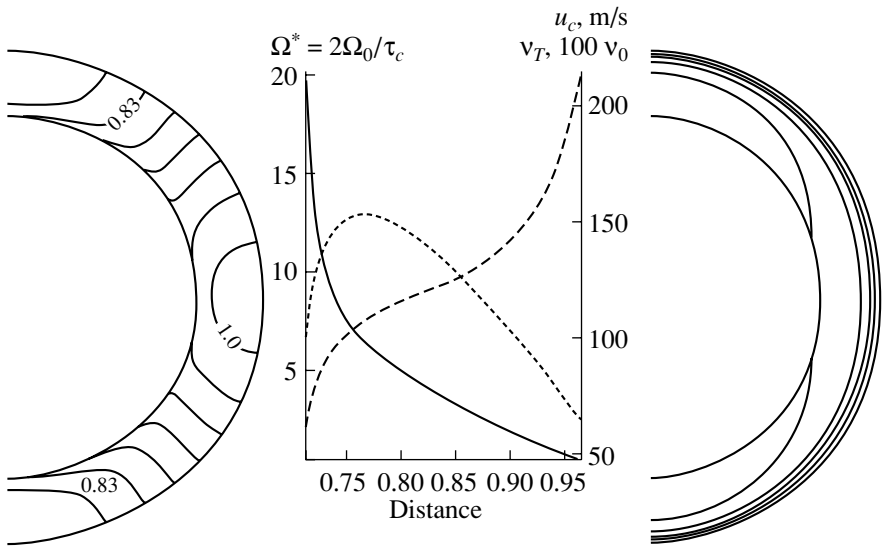

Fig. 1. Left: contours of angular velocity in relative units; the contour increment is $\sim 0.03\Omega_0$. Center: radial dependences of the turbulent characteristics of the convection zone; plotted are the convective velocity u_c (dashed curve, right scale), the kinematic viscosity ν_T (dotted curve, right scale, in units of $100\nu_0$, where $\nu_0 = 2 \times 10^{13}$ cm^2 s^{-1}), and the Coriolis number $\Omega^* = 2\Omega_0\tau_c$ (solid curve, left scale). Right: the reference distribution of the entropy.

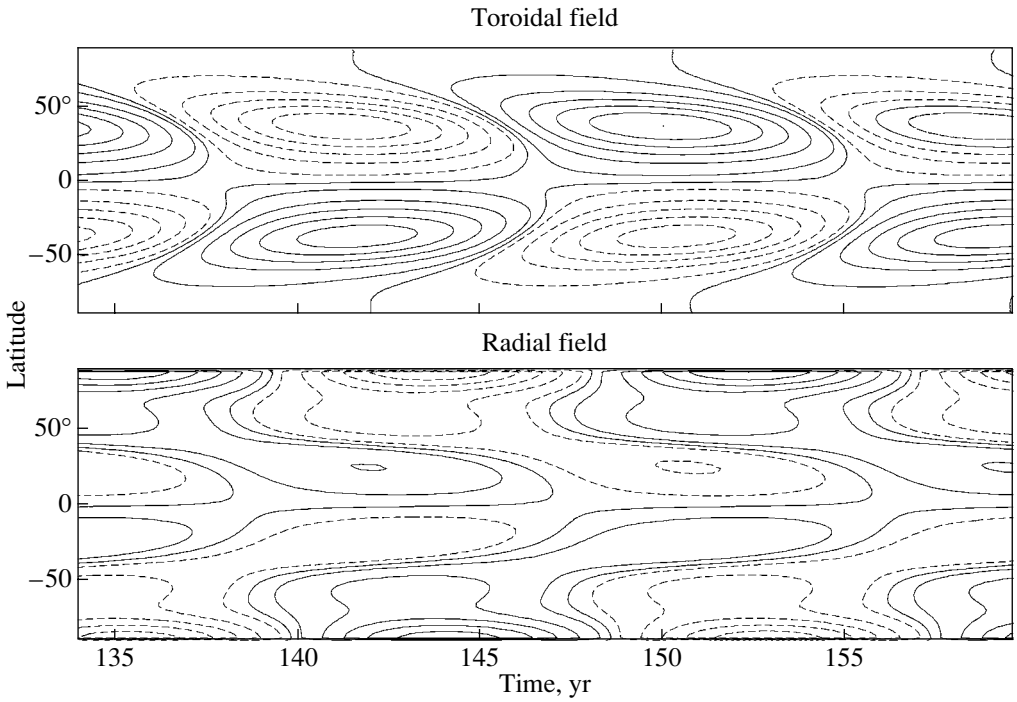

Fig. 2. Maunder diagrams representing the evolution of the large-scale toroidal magnetic field near the outer boundary of the convection zone (top) and the radial magnetic field (bottom).

et al. [29] and in [22, 23]. The polar branch of the drift of the angular-velocity variations is mainly due to large-scale Lorentz forces, since the radial field is maximum in the polar region and drifts poleward at high latitudes. The equatorial branch of the drift is supported by the reduction and amplification of the Λ effect by the large-scale toroidal magnetic field. A comparison of the evolution of the magnetic field and angular velocity shows that the maximum magnetic activity is localized in the region of maximum equatorial acceleration. This has been noted repeatedly based on analyses of Doppler measurements of

variations in the rotational velocity, e.g., by LaBonte and Howard [30]. Note also that the pattern of the torsional oscillations is in good qualitative and quantitative agreement with the results of a helioseismological inversion of a long time series of observations of solar acoustic oscillations [31].

The middle diagram in Fig. 3 shows the variations in the heat flux at the outer boundary, $\frac{L_\odot}{\pi r_b^2}\frac{T_e}{T_{eff}}\frac{\delta S}{c_p}$. The heat flux is in antiphase with the magnetic field. This is primarily due to the magnetic–shadow effect—the reduction of the convective flux by the magnetic field, which is pronounced in sunspots. The relative amplitude of the heat-flux variations is $\sim 10^{-4}$. Other mechanisms responsible for variations in the amount of heat in the convection zone and related to transformations of the energy of large-scale magnetic fields and flows weakly affect the flux of thermal energy at the outer boundary, since they are efficiently shielded from an outer observer due to the enormous inertia of the solar interior. As the surface is approached, the thermal inertia of the solar gas decreases; therefore, the effect of surface magnetic fields dominates over the phenomena that occur deep in the solar convection zone. According to tentative estimates by Spruit [3] and numerical calculations [5, 6], magnetic fields escaping the convection zone should make the largest contribution to the 11-yr luminosity variations.

Our model includes a description of physical processes *inside* the convection zone, and computing the contribution to the solar luminosity of magnetic fields above the photosphere goes beyond the scope of our analysis. The *maximum* amplitude of variations in the solar luminosity can be estimated from the computations if we assume that the energy of escaping magnetic fields is completely converted into radiation. In this case, the amplitude of variations in the solar radiation flux is determined by oscillations of the magnetic-energy flux from the convection zone.

The evolution of the total energy flux (magnetic + thermal) at the outer boundary is illustrated by the second to last diagram in Fig. 3. The energy flux varies in phase with the activity of the large-scale azimuthal magnetic field with a relative amplitude of $\sim 8 \times 10^{-4}$. The variations in the energy flux depend strongly on latitude. The maximum positive flux variation drifts toward the equator, in accordance with the drift of the magnetic-field activity. Thus, our model predicts 11-yr variations in the radiation flux from the convection zone that are uniform in latitude.

We emphasize that the question of the intrinsic origin of the increase in the solar luminosity at the magnetic-activity maximum cannot be resolved using our formulation of the problem. We can only obtain an energy estimate for possible luminosity variations by assuming that the escaping magnetic energy is completely dissipated into heat. Our formulation does not take into account other possible mechanisms by which the magnetic fields could influence the outgoing energy from the convection zone. It is known, for example, that photospheric magnetic fields can form channels for the enhanced removal of energy from the photosphere via excess emission in thin flux tubes of the magnetic network or MHD oscillations propagating from the photosphere into the corona.

The large-scale magnetic field generated in the convection zone gives rise to variations in the hydrostatic equilibrium, which can be characterized by the variations in the gas pressure and the density of material in the convection zone.

The last diagram in Fig. 3 represents the gas-pressure variations at the outer boundary. The maximum relative amplitude of the deviations of the pressure from the mean is 5×10^{-4}. Note that the cyclic pressure variations are almost quadrupolar. By and large, the computations show that the departures of the pressure from the mean are negative in the growth phase of the magnetic activity and positive in the decline phase. It also follows from the above estimates that these pressure variations result primarily from the modulation of the centrifugal force due to the acceleration of the equatorial region at the activity maximum, as well as from the direct contribution of the large-scale magnetic fields to the pressure perturbations. Note that both sources act in the same way, producing negative departures of the pressure from the mean in the growth phase of the magnetic field. The pressure variations are closest to being quadrupolar in the decline phase of the magnetic activity.

Figure 4 shows contours of physical quantities in a meridional section of the convection zone for five times. The initial time is chosen arbitrarily and nearly corresponds to the growth phase of the magnetic activity. A period from 0 to 8 yr is covered, with the time step between plots being 2 yr. Time runs along each column in Fig. 4 from top to bottom. At the activity minimum (fourth row of plots), the magnetic configuration is nearly dipolar and the differential rotation at the surface is minimum. At the same time, the total thermal energy of the gas is reduced in deep layers of the convection zone due to energy that is spent on the generation of the magnetic fields of the new cycle. The variations in the sound speed are shown in the last column in Fig. 4; these were found from the

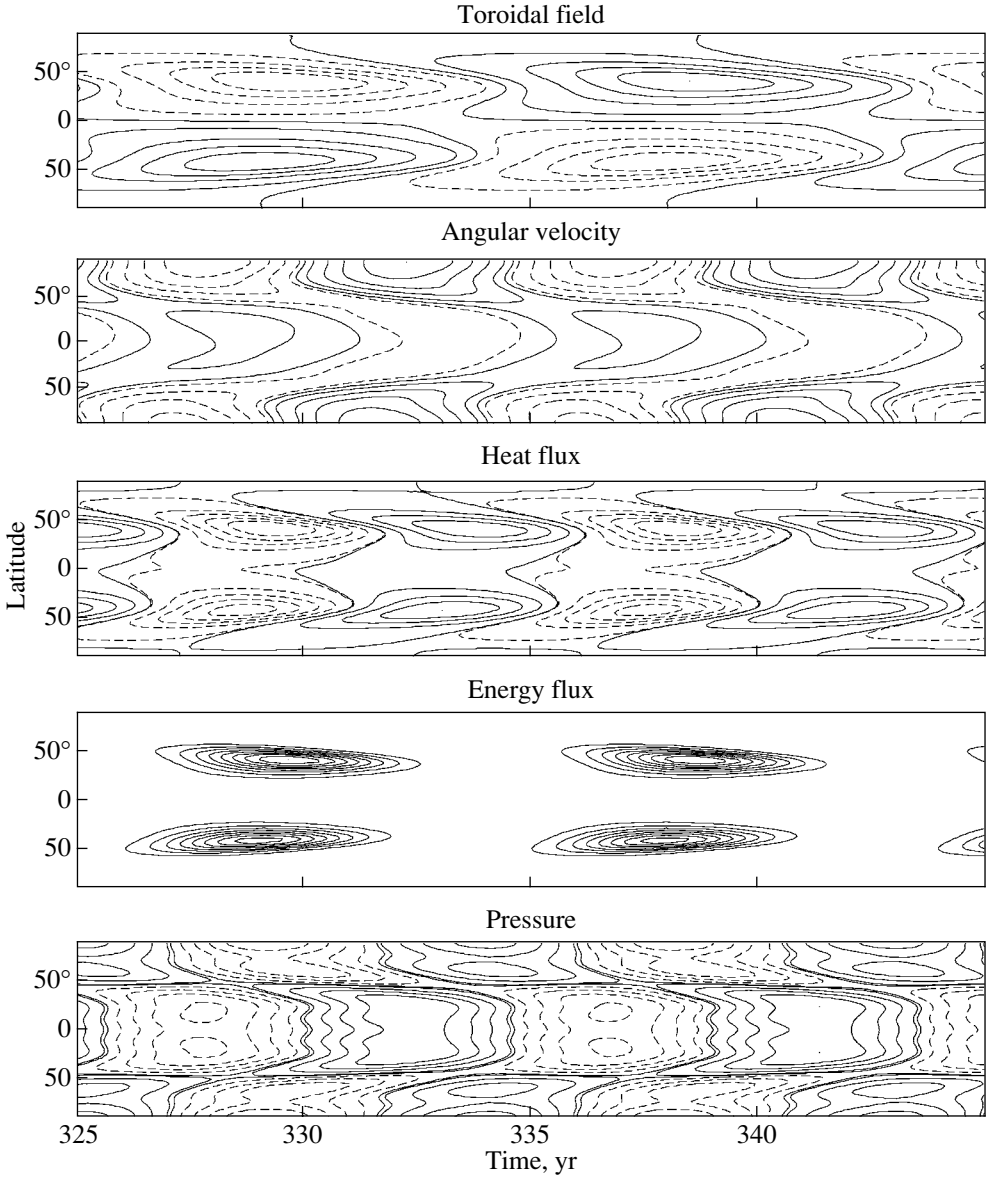

Fig. 3. Maunder evolutionary diagrams for the toroidal magnetic field (top), departure of the angular rotational velocity from the mean (second from top), heat flux from the convection zone (central), total energy flux (second from bottom), and pressure (bottom). All quantities are taken at the outer boundary of the convection zone. The amplitudes of the oscillations of the field and rotation rate are 400 G and 10 m/s. The relative amplitudes of the oscillations are 10^{-4} for the heat flux, 10^{-3} for the energy flux, and 5×10^{-4} for the pressure.

relationship

$$\delta C^2 = \frac{P_{(1)}}{\rho_{(0)}} - \frac{P_{(0)}}{\rho_{(0)}^2}\rho_{(1)}.$$

We can see that the computed departures of the sound speed from the mean near the surface are positive at the cycle minimum. The magnetic-activity maximum is somewhere between the first and second rows of plots in Fig. 4. The latitudinal nonuniformity of the differential rotation is maximum during this period. In addition, the heat-flux variations are negative in surface layers of the convection zone, mainly due to the thermal-shadow effect. The computed sound speed also displays maximum negative departures from the mean during the period of maximum magnetic activity. The amplitude of the variations is $|\delta C| \sim 10^{-3}$ and is maximum in surface layers of the convection zone. During periods of maxima, the reduction of the sound speed can give rise to a "redshift" in the eigenfrequencies of acoustic modes of oscillations with an amplitude of $\sim 10^{-6}$. This effect had been noted earlier (see, e.g., [32]). Note also that

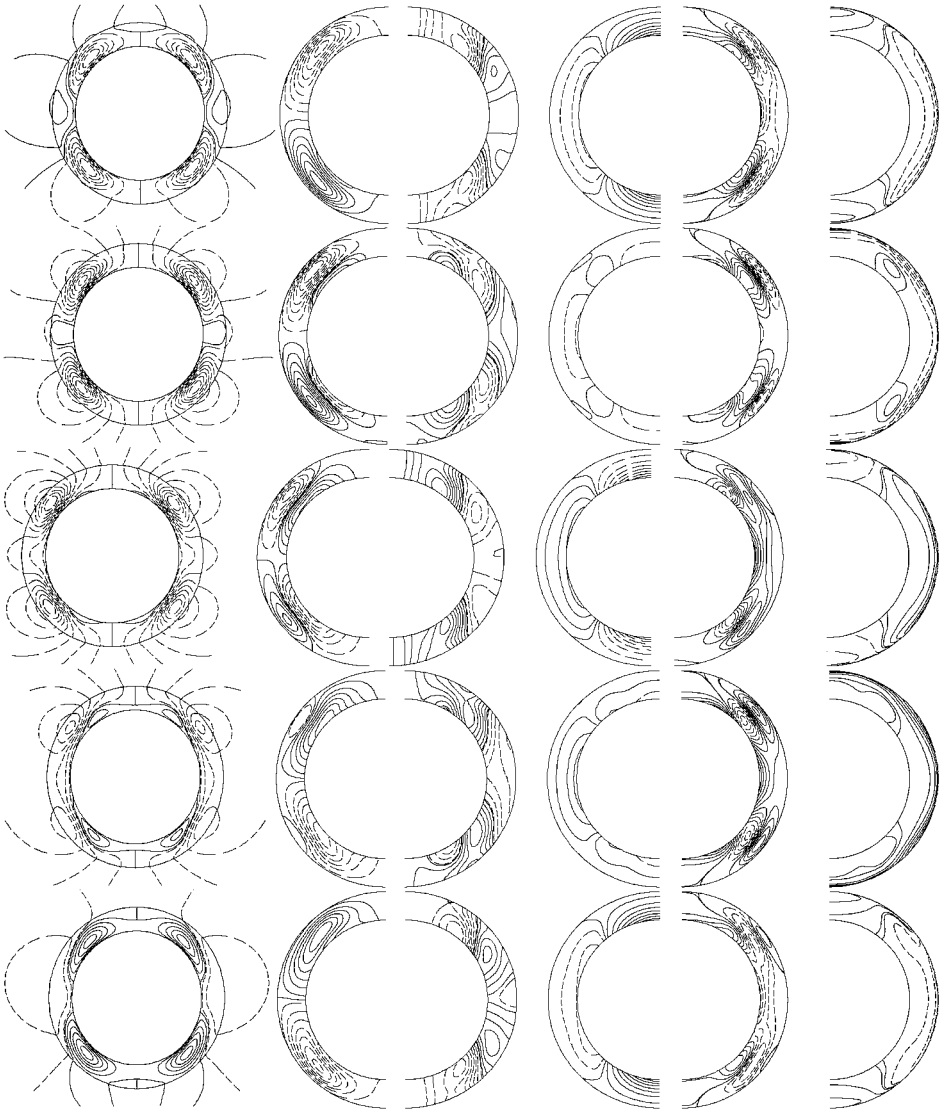

Fig. 4. Computed variations of various physical quantities over a half cycle. From left to right, the columns show (1) the configuration of the poloidal magnetic field, (2) contours of the toroidal magnetic field (left) and variations in the angular velocity (right), (3) contours of the variation in pressure (left) and thermal energy (right), and (4) contours of the relative variations in the sound speed. Time runs from top to bottom in steps of 2 yr. The maximum of the cycle is between the first and second row.

the modulation of the pressure inside the convection zone basically follows the variations in the angular velocity.

To analyze the results, we will need some integrated characteristics of the model. We characterize the overall magnetic activity by the integral of the modulus of the magnetic flux in the convection zone, i.e.,

$$F_M = \int_0^{\pi}\int_{r_i}^{r_e} |B(r,\theta)| \sin\theta r \mathrm{d}\theta \mathrm{d}r,$$

and characterize the type of magnetic-field symmetry with respect to the equator via the parity parameter [19]

$$P = \frac{E^{(S)} - E^{(A)}}{E^{(S)} + E^{(A)}},$$

where $E^{(S)}$ and $E^{(A)}$ are the integrated magnetic energies of the quadrupolar and the dipolar components of the large-scale magnetic field, respectively. The luminosity variations $\delta L_\odot$ are made up of variations of the integrated heat flux and the integrated energy flux of the magnetic fields going outward from the

convection zone to the surface, S_M,

$$\frac{\delta L_\odot}{L_\odot} = 4\int\limits_0^\pi \left(\frac{T_e}{T_{eff}}\frac{\delta S}{c_p} + \pi r_b^2 \frac{S_M}{L_\odot}\right)\sin\theta \mathrm{d}\theta, \quad (20)$$

$$S_M = -\frac{1}{4\pi}\mathcal{E}_\theta B\Big|_{r_b},$$

provided that all the escaping energy is converted into radiation. Formula (20) does not include any description of possible seasonal variations due to the inclination of the solar rotational axis to the plane of the Earth's orbit. This contribution is not significant in our model, since (i) the inclination angle of the axis is small and (ii) the largest contribution from this effect should be made by high-latitude regions, where variations in the radiative flux are small (Fig. 3).

Variations in the rotational velocity represent one of the main mechanisms for modulating the stratification of the convection zone. For our analysis, it is convenient to represent the angular velocity at the surface as a Legendre polynomial expansion:

$$\Omega(r_e,\theta) = \Omega_0\left(\omega_0 + \omega_2 P_2(\theta) + \omega_4 P_4(\theta)\right). \quad (21)$$

Figure 5 shows the evolution of the integrated characteristics of the model. The computed luminosity variations almost coincide in phase with the variations of the magnetic energy and flux in the convective envelope. The small lag is due to the fact that we are considering the modulus of the magnetic flux *integrated* over a cross section of the convection zone, while the luminosity variations are primarily due to the surface magnetic fields. The flux of the magnetic-field energy is directly related to the power of processes generating the large-scale magnetic fields. This power can be estimated from the half-cycle amplitude of the variations in the magnetic energy: $\sim\frac{1.1\times10^{39}}{4}\mathrm{erg/yr} = 8.72\times 10^{30}\mathrm{erg/s} \sim 2.25\times10^{-3}L_\odot$. Thus, according to our model, about half of the generated magnetic energy leaves the convection zone. At the same time, the amplitude of the relative variations in the power spent on generating the magnetic fields coincides with the magnetic-energy flux from the convection zone. The amplitude of variations in the integrated magnetic flux is 1.5×10^{24} Mx. The calculation of the magnetic-flux budget done by Schrijver and Harvey [33] for photospheric magnetic fields yields similar flux variations, $\sim10^{24}$ Mx. The computed luminosity variations are 1.2×10^{-3}, and the heat-flux variations constitute only 10% of this figure. The luminosity modulation agrees with the values indicated by observations of 11-year variations of the solar irradiance [2, 34]. The estimated luminosity variations during the cycle are based on the assumption that the energy of the escaping magnetic fields is completely converted into emission. The model does not take into account other mechanisms for the enhancement of solar emission in the presence of magnetic fields. Some such mechanisms were previously considered in [3, 7].

The variations in the radius computed from the gravitational-potential variations (17) comprise $\sim1.5\times10^{-6}\,R_\odot$. The maximum solar radius nearly corresponds to periods of activity growth and activity maximum of the surface magnetic fields. No reasonably trustworthy data on variations of the solar radius with the 11-yr cycle are available. Our computations support the view of Basu [9] and Rozelot [10]. We conclude that oscillations of the solar radius with the activity cycle are primarily due to the disturbance of the stratification in the outer layers of the convection zone produced by modulations of the centrifugal force and the direct influence of magnetic fields on the density distribution in the convection zone. This follows from a comparison of the results presented in Figs. 3 and 5. The maximum radius corresponds to periods of the largest differential rotation. There is also observational evidence for a similar interrelation between oscillations of the radius and the differential rotation of the Sun [10, 15].

The modulation of the quadrupole moment due to magnetic activity lags behind the variations of the radius in phase. The maximum quadrupole moment corresponds to the phase of magnetic-activity decline. It also correlates with the modulation of the zeroth mode in the Legendre polynomial expansion of the surface rotational velocity (21). Note that the maxima of the variations in ω_0 correspond to periods when regions of faster rotation are present at both the pole and the equator, i.e., to periods of the maximum velocity of the overall rotation of the surface, and, therefore, of the maximum contribution of the centrifugal force to the quadrupole potential.

The relative variation in the quadrupole potential during the activity cycle is $\Delta J_2 = 4.5\times10^{-9}\frac{GM_\odot}{R_\odot}$. In comparison, the tidal potential of Jupiter at the solar surface [35] is equal to $\sim\frac{M_\Psi}{M_\odot}\left(\frac{R_\odot}{a}\right)^3\frac{GM_\odot}{R_\odot} = 6.92\times10^{-13}\frac{GM_\odot}{R_\odot}$, where a is the semimajor axis of Jupiter's orbit. Thus, the contribution of the magnetic activity to the modulation of the quadrupole moment is comparable to the tidal potential of a solar-type star moving in Jupiter's orbit. According to [36], the quadrupole moment of the inertia tensor varies with the amplitude $\Delta Q = -\frac{2}{3}\Delta J_2 M_\odot R_\odot^2 \approx 2.9\times10^{46}\,\mathrm{g\,cm^2}$.

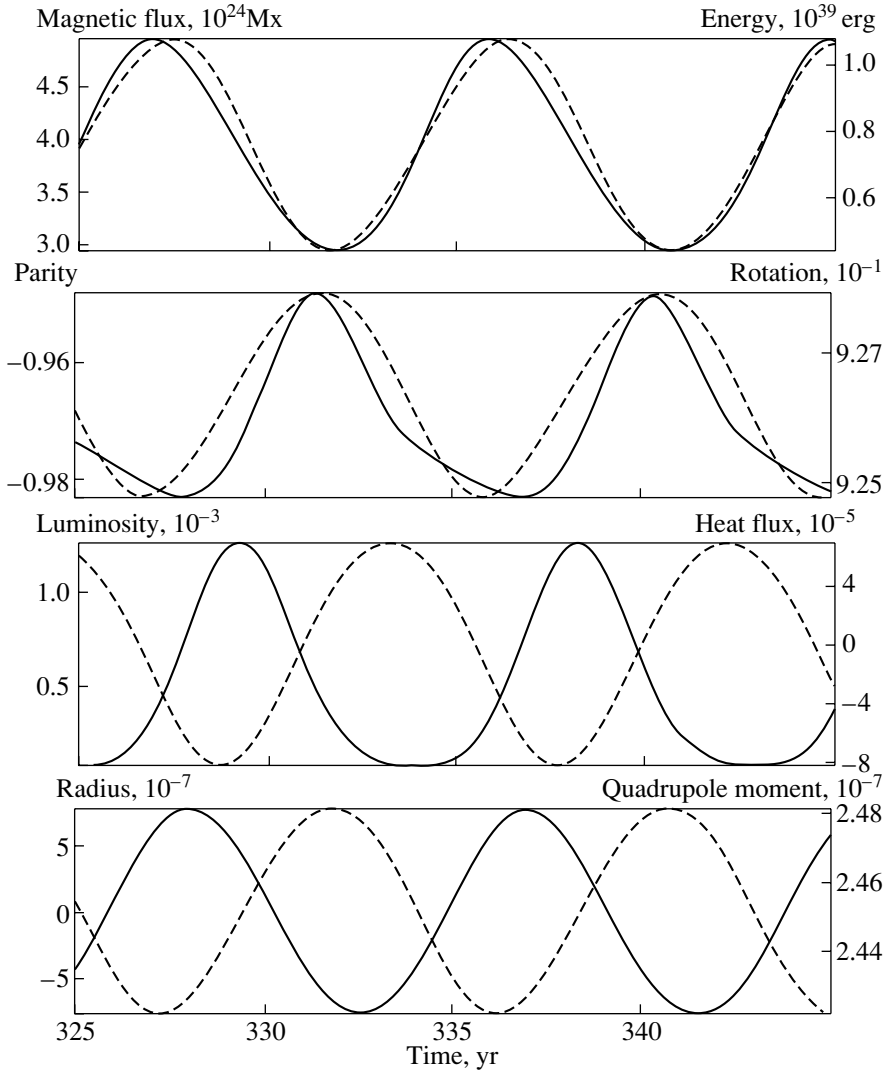

Fig. 5. Computed variations in the integrated characteristics of the model with the activity cycle. The left scale corresponds to the solid curves, and the right scale, to the dashed curves. The dashed curve in the second graph from the top indicates the variations of the zeroth coefficient ω_0 in the decomposition of the angular velocity (21).

The modulation of the quadrupole moments of magnetoactive stars on the lower part of the main sequence has interesting implications, which appear to be most pronounced in close binaries. These are variations in the orbital period of the companion in a binary system in which the primary is magnetoactive and has a late spectral class. The mechanism for the variations in the orbital period was previously considered in [30, 36–38]. According to [36], the relative variation in the orbital period due to variations in the quadrupole moment of the primary is

$$\frac{\Delta\mathcal{P}}{\mathcal{P}} = -9\left(\frac{R}{a}\right)^2 \frac{\Delta Q}{MR^2}. \qquad (22)$$

Here, $\mathcal{P}$ is the orbital period of the companion, R and M are the radius and mass of the primary, and a is the semimajor axis of the companion's orbit. The variation in the quadrupole moment due to modulations of the differential rotation, $\Delta\Omega/\Omega$, is

$$\Delta Q = \frac{4}{9}k_2\frac{R^5\Omega^2}{G}\left(\frac{\Delta\Omega}{\Omega}\right), \qquad (23)$$

where k_2 is the apsidal constant [35]. We see that the amplitude of the small oscillations of the quadrupole moment is a function of the fifth power of the radius and the second power of the angular velocity; at the same time, it depends linearly on the relative variations in the differential rotation. Studies of the modulation of the orbital period due to magnetic activity in close binaries requires a special analysis, and care must be taken when extrapolating the results obtained to the case at hand.

Parameters of stars taken from [38]

Object	$\frac{R}{a}$	$\frac{R}{R_\odot}$	$\frac{M_\odot}{M}$	$\frac{\Omega}{\Omega_\odot}$	$\frac{\Delta\mathcal{P}}{\mathcal{P}}$	$\left(\frac{\Delta\mathcal{P}}{\mathcal{P}}\right)_{obs}$
WW Dra (G2IV)	0.24	3.89	0.74	5.6	2.3×10^{-6}	6×10^{-6}
AR Lac (G2IV)	0.34	3.1	0.77	12.9	1.2×10^{-5}	2.4×10^{-5}
SV Cam (G3V)	0.26	1.1	1.07	41.9	4.6×10^{-6}	7.2×10^{-6}
U Peg (G2V)	0.34	1.4	0.55	67.8	2.1×10^{-5}	6.6×10^{-6}

Let us consider a binary system in which the primary has a nearly solar spectral class, such as the WW Dra system—a cataclysmic RS CVn binary. The rotational period of the primary, which is a G2 subgiant, is synchronized with the companion's orbit and comprises $\mathcal{P}_* = 4.5$ days. As is estimated in [20, 39], the latitudinal nonuniformity of the rotation between a pole and the equator is $\delta\Omega_* = \Delta\Omega/\Omega \approx 0.03$, where $\Delta\Omega$ is the difference in the rotational velocities at the pole and equator. According to [19], the dimensionless parameter $\mathcal{D}$ characterizing the strength of the dynamo mechanism in this star is estimated to be

$$\mathcal{D}_* = \frac{\Delta\Omega\alpha_{\phi\phi}R^3}{\eta_T^2} = \mathcal{D}_\odot\frac{\delta\Omega_*}{\delta\Omega_\odot}\left(\frac{\mathcal{P}_\odot}{\mathcal{P}_*}\right)^{1.5} \approx 1.36\mathcal{D}_\odot.$$

This means that the amplitude of torsional oscillations during the magnetic cycle expected for this star does not differ appreciably from our result for the solar dynamo. Therefore, we conclude that the amplitude of variations in the quadrupole moment (23) is mainly determined by the radius and rotational velocity of the star.

We now rewrite (23) in terms of solar quantities:

$$\Delta Q = \left(\frac{R}{R_\odot}\right)^5\left(\frac{\Omega}{\Omega_\odot}\right)^2\Delta Q_\odot, \qquad (24)$$

$$\frac{\Delta\mathcal{P}}{\mathcal{P}} = 6\left(\frac{R}{a}\right)^2\left(\frac{M_\odot}{M}\right)\left(\frac{R}{R_\odot}\right)^3\left(\frac{\Omega}{\Omega_\odot}\right)^2\Delta J_{2\odot}. \qquad (25)$$

Substituting the parameters of WW Dra into (24) and (25) yields for the modulation of the orbital period $\Delta\mathcal{P}/\mathcal{P} = 2.25 \times 10^{-6}$, while the observed value is $(\Delta\mathcal{P}/\mathcal{P})_{obs} = 6 \times 10^{-6}$. These figures agree in order of magnitude. The data given in the table and taken from Lanza and Rodono [38] demonstrate that this is not a simple coincidence.

The computed and observed variations in the orbital period are given in columns 6 and 7 of the table, respectively. Binaries with nearly solar-type primaries were selected. Note that the type of dynamo could differ substantially from the solar dynamo in any of these stars. The concentration of magnetic fields in active longitudes is much more pronounced in solar-type binaries of this type than on the Sun. The main goal of the above estimates was to estimate the applicability of variations in the angular-momentum distribution within the magnetoactive companion to explain the modulation of the orbital period. The oscillation amplitude of the period is most strongly affected by the rotational velocity of the star, which is a much more important factor than the amplitude of torsional oscillations. This primarily stems from the assumption of the smallness of these oscillations introduced to derive (23).

6. CONCLUSIONS

Let us summarize the basic results of our study. We have considered a numerical model for the solar dynamo that jointly describes the generation of magnetic fields, angular-momentum balance, heat transfer, and hydrostatic equilibrium of the solar convective envelope in the presence of a large-scale magnetic field. The results of our calculations demonstrate that the turbulent dynamo of the large-scale magnetic field that operates deep in the convection zone modulates various physical quantities during the course of the activity cycle. Among the effects considered, the cyclic modulation of the differential rotation with an amplitude of $\sim$10 m/s is noteworthy. A similar result was previously obtained in [23, 29] based on a simpler model (without heat transfer and stratification). The calculations reveal an increase in the solar luminosity at the maximum of the cycle with an amplitude of $\sim 1.2 \times 10^{-3} L_\odot$. This modulation is mainly due to the escape of magnetic energy from the dynamo region. Since this is the maximum possible estimate for the power of the outgoing magnetic fields converted into radiation, this result confirms the results of [3, 4, 7] that the increase in luminosity at the maximum of the cycle could also be due to the indirect influence on the emergent radiation of magnetic fields localized above the photosphere. In this case, the surface magnetic fields can be regarded as additional channels for the outward transfer of energy from the solar interior. The excess radiation observed in elements of the magnetic

network and the dissipation of MHD waves excited by photospheric hydrodynamic flows and propagating into the corona are relevant examples.

According to our calculations, the cyclically varying centrifugal force modulated by torsional oscillations seems to be the main source of the 11-yr variations of the sound speed in the convection zone. Departures of the sound speed from the mean are maximum near the outer boundary of the dynamo zone. The model demonstrates small variations in the solar radius with the activity cycle with an amplitude of $1.5 \times 10^{-6}\ R_\odot$ and variations of the quadrupole moment with an amplitude of $4.5 \times 10^{-9} \dfrac{GM_\odot}{R_\odot}$. The radius is maximum in the growth phase and during the period of maximum differential rotation of the Sun. The quadrupole potential and the oblateness of the Sun are maximum during the period of magnetic-activity decline, which is related to the growth of the overall rotational velocity of the surface layers of the Sun in this period. The estimated variations in the orbital periods of close binaries whose primaries have the same spectral class as the Sun are promising for applying the torsional-oscillation modulations of the centrifugal force to explain the variations of the orbital period of the companions in such systems.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 02-02-16199, 02-02-39027, 02-02-16044, 03-02-16384), INTAS (grant 2001-0550), and the Ministry of Industry, Science, and Technology of the Russian Federation (grant of the President of the Russian Federation for the support of leading scientific schools NSh-733.2003.2).

REFERENCES

1. R. C. Wilson, Space Sci. Rev. **38**, 203 (1984).
2. R. C. Wilson and A. V. Mordvinov, Geophys. Res. Lett. **26**, 3613 (1999).
3. H. Spruit, Space Sci. Rev. **94**, 113 (2000).
4. M. Stix, Astron. Astrophys. **93**, 339 (1981).
5. V. V. Pipin and L. L. Kichatinov, Astron. Zh. **77**, 872 (2000) [Astron. Rep. **44**, 771 (2000)].
6. L. L. Kitchatinov, A. V. Mordvinov, and V. V. Pipin, Soln.-Zemnaya Fiz. **2**, 3 (2002).
7. N. J. Balmforth, D. O. Gough, and W. J. Merryfield, Mon. Not. R. Astron. Soc. **278**, 437 (1996).
8. J. Pap, J. P. Rozelot, S. Godier, and F. Varadi, Astron. Astrophys. **372**, 1005 (2001).
9. D. Basu, Sol. Phys. **183**, 291 (1998).
10. J. P. Rozelot, J. Atmos. Sol.-Terr. Phys. **63**, 375 (2001).
11. E. R. Neto, A. H. Andrei, J. L. Penna, *et al.*, Sol. Phys. **212**, 7 (2003).
12. R. L. Gilligand, Astrophys. J. **248**, 1144 (1981).
13. P. Delache, F. Laclare, and H. Sadsaoud, Nature **317**, 416 (1985).
14. R. K. Ulrich and L. Bertello, Nature **377**, 214 (1995).
15. E. Ribes, J. C. Ribes, and R. Barthalot, Nature **326**, 52 (1987).
16. F. Krause and K.-H. Rädler, *Mean Field Magnetohydrodynamics and Dynamo Theory* (Pergamon Press, Oxford, 1980).
17. L. L. Kitchatinov, V. V. Pipin, and G. Rüdiger, Astron. Nachr. **315** (2), 157 (1994).
18. G. Rüdiger and L. L. Kitchatinov, Astron. Astrophys. **269**, 581 (1993).
19. L. L. Kitchatinov, M. V. Mazur, and M. Jardine, Astron. Astrophys. **359**, 531 (2000).
20. L. L. Kitchatinov and G. Rüdiger, Astron. Astrophys. **344**, 911 (1999).
21. G. Rüdiger, *Differential Rotation and Stellar Convection* (Akademie-Verlag, Berlin, 1989).
22. L. L. Kitchatinov, V. V. Pipin, V. I. Makarov, and A. G. Tlatov, Sol. Phys. **189**, 227 (1999).
23. V. Pipin, Astron. Astrophys. **346**, 295 (1999).
24. L. L. Kitchatinov and V. V. Pipin, Astron. Astrophys. **274**, 647 (1993).
25. M. Stix, *The Sun. An Introduction*, 2nd ed. (Springer, Berlin, 2002).
26. G. Rüdiger and L. L. Kitchatinov, Astron. Nachr. **321**, 75 (2000).
27. J. Schou, S. Antia, R. S. Basu, *et al.*, Astrophys. J. **505**, 390 (1998).
28. A. Brandenburg, D. Moss, and I. Tuominen, Astron. Astrophys. **265**, 328 (1992).
29. M. Kueker, G. Ruediger, and V. V. Pipin, Astron. Astrophys. **312**, 615 (1996).
30. B. J. LaBonte and R. Howard, Sol. Phys. **75**, 161 (1982).
31. S. V. Vorontsov, J. Christensen-Dalsgaard, J. Shou, *et al.*, Science **296**, 101 (2002).
32. Y. Elsworth, R. Howe, G. R. Isaak, *et al.*, Nature **345**, 322 (1990).
33. C. J. Schrijver and K. L. Harvey, Sol. Phys. **150**, 1 (1984).
34. A. V. Mordvinov and R. C. Willson, Sol. Phys. **215**, 5 (2003).
35. J.-P. Zahn, Astron. Astrophys. **57**, 383 (1977).
36. J. H. Applegate, Astrophys. J. **385**, 621 (1992).
37. A. F. Lanza, M. Rodono, and R. Rosner, Mon. Not. R. Astron. Soc. **296**, 893 (1998).
38. A. F. Lanza and M. Rodono, Astron. Astrophys. **349**, 887 (1999).
39. R. A. Donahue, S. H. Saar, and S. L. Baliunas, Astrophys. J. **466**, 384 (1996).

Translated by A. Getling

Astronomy Reports, Vol. 48, No. 5, 2004, pp. 433–438. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 5, 2004, pp. 475–480.
Original Russian Text Copyright © 2004 by I. Podgorny, A. Podgorny, Minami, Morimoto.

An MHD Model for a Heliospheric Current Sheet

I. M. Podgorny[1], A. I. Podgorny[2], S. Minami[3], and M. Morimoto[3]

[1] *Institute of Astronomy, Russian Academy of Sciences, ul. Pyatnitskaya 48, Moscow, 109017 Russia*

[2] *Lebedev Institute of Physics, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia*

[3] *Osaka University, Osaka, Japan*

Received May 8, 2003; in final form, November 10, 2003

Abstract—A numerical solution of the full set of MHD equations shows the generation of a heliospheric current sheet during the thermal expansion of the corona. Calculations were performed for a compressible plasma taking into account dissipative terms and anisotropy of the thermal conductivity of the magnetized plasma. It is shown that the current sheet is not magnetically neutral. The sheet contains a normal component of the magnetic field, which plays a fundamental role during the formation of the sheet and in the stationary state. The sheet is stable against MHD perturbations, which are apparently carried away by the plasma flow. For the numerical scheme chosen, the minimum sheet thickness is determined by the length of the spatial integration step. The PERESVET code was used for the calculations.
© *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The generation of a heliospheric current sheet is explained by the stretching of magnetic field lines during the thermal expansion of the coronal plasma. Descriptions of current sheets [1, 2] usually refer to Pneuman and Kopp's [3] calculations as evidence that heliospheric current sheets do not contain a normal component of the magnetic field; i.e., they are neutral. In fact, the absence of a normal field component was a result of the methods used for these calculations, which were performed for a half-plane, assuming a symmetric magnetic field [3]. The plasma was assumed to be incompressible and ideally conducting, and the equations neglected dissipation and the term responsible for the gravitation. All magnetic field lines were divided into two classes: closed and open. The plasma within the region occupied by closed magnetic lines was assumed to be in hydrostatic equilibrium, while a free expansion of the corona was assumed for the open lines. The solution also assumed a force balance that made any transverse plasma motions impossible; that is, any stationary magneto-hydrodynamical (MHD) plasma transport of the magnetic field was eliminated. A point above which the normal field was precisely equal to zero was assumed to lie on the axis of the current sheet. This condition was used as the boundary condition for calculating the field from the equation $\mathrm{curl}\mathbf{B} = 4\pi\mathbf{j}/c$. Thus, the absence of a normal component of the magnetic field in the current sheet in [3] is not a result of the calculations but was instead *a priori* introduced as a boundary condition. The incorrectness of these calculations was previously discussed in [4].

A stationary flow of magnetized coronal plasma was studied in [5], taking into account the solar rotation but assuming an isothermal corona. However, the normal component of the magnetic field in the current sheet was assumed to be zero, as in [3].

A number of theoretical works appearing after 1963 [6] indicated that neutral current sheets are rery unstable, so that a neutral sheet could be created (if at all) only over a time shorter than the time for the growth of the instability. This argues that no neutral current sheet can exist in a stationary or quasi-stationary state. In fact, every current sheet observed experimentally or simulated numerically contains a normal component of the magnetic field.

A more realistic approach to MHD models for heliospheric current sheets was used in [7, 8], which modeled the stationary MHD configuration of the magnetic field and the solar wind. The total region for the solution was divided into two subregions: Region I near the Sun, where the flow velocity approaches zero, and Region II, a distant region where the velocity exceeds both the Alfven and magneto-acoustic speeds at each point of the flow streaming away from the Sun. In Region I, a stationary solution was found via the relaxation of a solution for the non-stationary problem. The solution for Region II was obtained by integrating along a radial variable having the same meaning as the time variable in the non-stationary equation. A potential field constructed in accordance with the observed photospheric field was

1063-7729/04/4805-0433$26.00 © 2004 MAIK "Nauka/Interperiodica"

used for the initial magnetic field. The initial density and velocity in Region I were specified in accordance with a Parker flow, which determines to a considerable degree the form of the solution obtained. The MHD equations used did not contain dissipative terms, and the magnetic viscosity of the numerical scheme was substantial for the current sheet modeled. The solution obtained in [7] gave no particular evidence in support of neutrality of the current sheet, but the results of [8] clearly indicate the presence of a normal component of the magnetic field, while such a component is absent from the results of [3]. However, certain frequently used simplifications were applied in [7, 8]. In particular, Joule heating and anisotropy of the thermal conductivity were neglected. The contradiction between the results of [3] and [8] in connection with such an important question as the structure of a heliospheric current sheet leads to the need for numerical experiments to study the generation of a heliospheric current sheet during the expansion of the coronal plasma at an epoch when the solar magnetic field is dipolar, i.e., the configuration near solar minimum.

The main difference between the calculations presented here and the studies [7, 8] is that we impose no *a priori* assumptions about the initial flow. We not only study the steady-state regime but also try to model the expansion of the hot corona into a region with magnetic field but almost no material and to obtain a stationary solution based on this expansion. A brief summary of the preliminary calculations was presented in [9].

2. DESCRIPTION OF THE NUMERICAL EXPERIMENT

In our modeling, we used the PERESVET code to solve the full set of MHD equations including all dissipative terms for a compressible plasma. The energy equation takes into account the anisotropy of the thermal conductivity in magnetic fields [9]. Calculations both including and neglecting gravitation were carried out. The following quantities were used as dimensionless parameters: the density $\rho_0 = 10^4 m_i = 1.67 \times 10^{-20}$ g cm^{-3}, the temperature $T_0 =$ 20 eV, the magnetic field $B_0 = 1$ G, the velocity $V_0 = B_0/\sqrt{4\pi\rho_0} \approx 2 \times 10^9$ cm/s, the length $L_0 = 8R_\odot = 6 \times 10^{11}$ cm, and the time $t_0 = L_0/V_0 \sim 300$ s. Here, $m_i = 1.67 \times 10^{-24}$ g is the mass of an ion in the corona and $R_\odot = 6.9 \times 10^{10}$ cm is the solar radius. The set of dimensionless, three-dimensional MHD equations takes the form

$$\frac{\partial \mathbf{B}}{\partial t} = \mathrm{curl}(\mathbf{V} \times \mathbf{B}) - \frac{1}{\mathrm{Re}_m}\mathrm{curl}\left(\frac{\sigma_0}{\sigma}\mathrm{curl}\mathbf{B}\right), \quad (1)$$

$$\frac{\partial \rho}{\partial t} = -\mathrm{div}(\mathbf{V}\rho), \quad (2)$$

$$\frac{\partial \mathbf{V}}{\partial t} = -(\mathbf{V}, \nabla)\mathbf{V} - \frac{\beta_0}{2\rho}\nabla(\rho T) \quad (3)$$

$$- \frac{1}{\rho}(\mathbf{B} \times \mathrm{curl}\mathbf{B}) + \frac{1}{\mathrm{Re}\rho}\Delta\mathbf{V} + G_g\mathbf{G},$$

$$\frac{\partial T}{\partial t} = -(\mathbf{V}, \nabla)T - (\gamma - 1)T\mathrm{div}\mathbf{V} \quad (4)$$

$$+ (\gamma - 1)\frac{2\sigma_0}{\mathrm{Re}_m\sigma\beta_0\rho}(\mathrm{curl}\mathbf{B})^2 - (\gamma - 1)G_q L'(T)\rho$$

$$+ \frac{\gamma - 1}{\rho}\mathrm{div}\big[\mathbf{e}_{||}\kappa_{dl}(\mathbf{e}_{||}, \nabla T) + \mathbf{e}_{\perp 1}\kappa_{\perp dl}(\mathbf{e}_{\perp 1}, \nabla T)$$

$$+ \mathbf{e}_{\perp 2}\kappa_{\perp dl}(\mathbf{e}_{\perp 2}, \nabla T)\big].$$

The computational domain is a cube with a side of $8R_\odot$, or $0 \le x \le 1$, $0 \le y \le 1$, $0 \le z \le 1$ in dimensionless units. The locations of the Sun, corona, and computational domain in the $y = 0.5$ plane are shown in Fig. 1. The current density and dipole moment are expressed in units of $j_0 = cB_0/4\pi L_0$ and $M_0 = B_0 L_0^3$, respectively. The energy density is expressed in units of $B_0^2/4\pi$. The dimensionless expressions for the densities of the kinetic and magnetic energies and for the plasma pressure take the forms $\rho V^2/2$, $B^2/2$, and $\beta_0\rho T/2$, respectively. All the computational results are presented in these dimensionless units. The principles used to select the dimensionless parameters are described in [10, 11].

In (1)–(4), $\mathrm{Re}_m = L_0V_0/\nu_{m0}$ is the magnetic Reynolds number, $\nu_{m0} = c^2/4\pi\sigma_0$ is the magnetic viscosity for the conductivity σ_0 at the temperature T_0, and σ is the conductivity, with $\sigma/\sigma_0 = T^{3/2}$. However, the quantity influencing the solution in practice is the numerical magnetic viscosity, which is determined by the length of the spatial step and whose Reynolds number is ~50, rather than the high Re_m chosen. Although $\beta_0 = 8\pi n_0 kT_0/B_0^2$ (where $n_0 = \rho_0/m_i$, m_i is the ion mass) is a well-known dimensionless parameter, in this form, it is not the ratio of the plasma pressure to the magnetic pressure for a given spatial point, but is simply a dimensionless parameter expressed in terms of the dimensionless units selected for the calculation. The viscosity term does not significantly effect the results but is used to provide higher stability of the finite difference scheme employed. Here, $\mathrm{Re} = \rho_0 L_0 V_0/\eta$ is the Reynolds number, η is the viscosity, $G_q = L(T_0)\rho_0 t_0/T_0$, $L(T_{\mathrm{dimensional}})$ is the radiation function for ionization equilibrium in the corona, $T_{\mathrm{dimensional}} = T_0 T$, and $L'(T) = L(T_{\mathrm{dimensional}})/L(T_0)$ is the dimensionless radiation function. For the problem at hand, the effect of radiation was not significant. Here, $\mathbf{e}_{||}$, $\mathbf{e}_{\perp 1}$, and $\mathbf{e}_{\perp 2}$ are orthogonal unit

vectors parallel and perpendicular to the magnetic field; $\kappa_{dl} = \kappa/(\Pi\kappa_0)$ is the dimensionless coefficient of thermal conductivity along the magnetic field; $\Pi = \rho_0 L_0 V_0/\kappa_0$ is the Peclet number; κ_0 is the thermal conductivity for the temperature T_0 and κ is the thermal conductivity, with $\kappa/\kappa_0 = T^{5/2}$; $\kappa_{\perp dl} = [(\kappa\kappa_0^{-1}\Pi^{-1})(\kappa_B\kappa_{0B}^{-1}\Pi_B^{-1})]/[(\kappa\kappa_0^{-1}\Pi^{-1})+(\kappa_B\kappa_{0B}^{-1}\Pi_B^{-1})]$ is the dimensionless coefficient of thermal conductivity perpendicular to the magnetic field; and $\Pi_B = \rho_0 L_0 V_0/\kappa_{0B}$ is the Peclet number for thermal conductivity across a strong magnetic field (when the cyclotron radius is much smaller than the mean free path). The thermal conductivity across the strong magnetic field is denoted as κ_B, while κ_{0B} is the conductivity for the temperature T_0; ρ_0 and B_0 are the plasma density and magnetic field; $\kappa_B/\kappa_{0B} = \rho^2 B^{-2} T^{-1/2}$; $G_g\mathbf{G}$ is a dimensionless gravitational acceleration, $G_g = t_0^2/L_0$, which is insignificant for our calculations; and γ is the ratio of specific heats.

We chose the following parameters for our computations: $\gamma = 5/3$, $\mathrm{Re}_m = 8\times10^4$, $\mathrm{Re} = 10^4$, $\beta_0 = 8\times10^{-6}$, $\Pi = 2$, $\Pi_B = 2\times10^6$, $G_q = 0$, and $G_g = 0$. The principles used to select the dimensionless parameters are presented in [10, 11]. We also performed computations for $G_q = 10^{-4}$ and $G_g = 1.5\times10^{-8}$, but the solutions did not differ significantly due to the weak effects of radiation and gravitation.

We solved the MHD equations (1)–(4) in the three-dimensional computational domain using the PERESVET code. We employed a fully implicit finite-difference scheme that conserves the magnetic flux, which was solved iteratively. The numerical technique also includes an automatic multilevel decrease of the time step for regions with high gradients. These capabilities of the scheme make it possible to stabilize most numerical instabilities. The computations were performed for a $41\times41\times41$ grid, so that the numerical magnetic Reynolds number was $\sim$50.

The locations of the Sun, corona, and computational domain are shown in Fig. 1, which presents the intersection of the domain by the $y = 0.5$ plane. The magnetic dipole providing a field of $B_0 \approx 0.8$ G at the solar surface is parallel to the Z axis and has a magnetic moment in dimensionless units of $M_1 = \{M_{1x} = 0,\ M_{1y} = 0,\ M_{1z} = 9.6\times10^{-2}\}$. The dipole center is at the point $R_1 = \{x_1 = -0.217, y_1 = 0.5, z_1 = 0.5\}$. The magnetic field in the $x = 0$ plane at the point $y = 0.5$, $z = 0.5$ is 0.15. The lines of the initial (dipole) magnetic field

$$\mathbf{B} = \frac{3(\mathbf{R}-\mathbf{R}_1)(\mathbf{M}_1,(\mathbf{R}-\mathbf{R}_1)) - \mathbf{M}_1(\mathbf{R}-\mathbf{R}_1)^2}{(\mathbf{R}-\mathbf{R}_1)^5}$$

are shown in Fig. 2a. The vacuum medium cannot be described using MHD approximations, and we

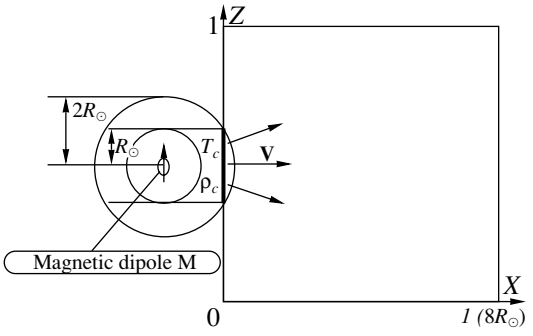

Fig. 1. Sketch of the computational domain. The thick curve at the boundary $X = 0$ shows the circle on which the coronal density ρ_c, temperature T_c, and inflow velocity are specified.

assumed an extremely low density of 10^{-1} cm^{-3} for the computational domain at $t = 0$, whose effect on the expanding coronal plasma was insignificant. A low temperature of 20 eV was assumed for the computational domain at the initial time.

The thermal expansion of the corona starts at $t = 0$. The coronal parameters ($\rho_c/m_i = 2\times10^7$ cm^{-3} and $T_c = 200$ eV) are specified in the $x = 0$ plane in the circle corresponding to the intersection of this plane with a spherical surface (Fig. 1) of radius $2R_\odot$ centered at the point (−0.217, 0.5, 0.5). The center of this circle is located at the point ($y = 0.5$, $z = 0.5$), and its radius is 0.125. The velocity of the plasma outflow from the corona is determined by the continuity equation, ensuring that the mass flux corresponds to the mass carried away by the solar wind ($\sim10^{-14}\,M_\odot$/year). This condition specifies the inflow velocity at the boundary $x = 0$ inside the circle of radius 0.125 centered at ($y = 0.5$, $z = 0.5$), $V_x = 2.5\times10^{-4}$. The self-consistent ρ, T, and $\mathbf{V}$ for the thermal expansion of the coronal plasma are automatically established in the numerical solution of the MHD equations. The current in the current sheet can freely flow in and out through the $y = 0$ and $y = 1$ planes.

Recall that our parameter β_0 did not correspond to the ratio of the pressures at a given point; in the flow of the expanding solar wind, $\beta = 8$ at $x = 0$, $y = 0.5$, $z = 0.5$. At the initial time, this parameter was small at this point in the computational domain (for approximately vacuum conditions): $\beta = 4\times10^{-9}$.

The most difficult aspect of formulating the boundary conditions is specifying the magnetic field at the right-hand boundary. During the simulation and the stretching of the field lines along the X axis, all the B_x, B_y, and B_z field components vary strongly, and the inclination of the field to the $X = 1$ plane changes as well, so that we cannot set the magnetic field at the

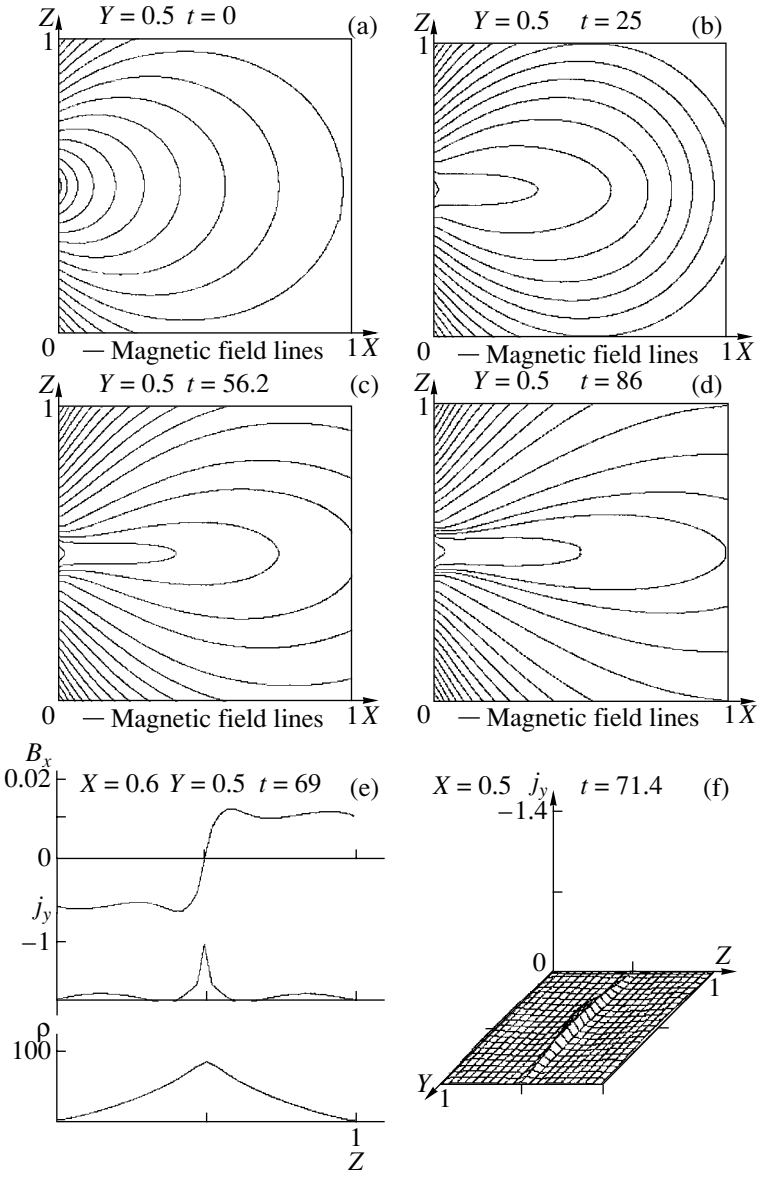

Fig. 2. (a) Magnetic dipole field lines; (b–d) magnetic lines stretched by the plasma flow at various times; (e) vertical profiles for the magnetic field, current, and plasma densities in the vicinity of the current sheet; (f) current densities j_y.

boundary $X = 1$. We likewise cannot assume $j = 0$, because the current sheet stretching the magnetic lines reaches the right-hand boundary. The best approximation is to use the condition $\partial j/\partial x = 0$. To convince ourselves that the solution obtained corresponds to the conditions in the solar wind, we can eliminate a layer with a thickness of two or three current-sheet thicknesses from consideration.

3. COMPUTATION RESULTS

Our numerical experiment assumes that the magnetic dipole field deforms during the thermal expansion of the coronal plasma. The purpose of the simulations was to clarify whether a current sheet can be generated. The following factors can facilitate the formation of the current sheet.

(1) The plasma can freely expand along the magnetic lines of the dipole field.

(2) The minimum magnetic field for a given line is located at the equator. During an increase in the dynamical pressure, $nkT + mnV^2$, the stretching of the magnetic-field lines must first start at the equator, with the resulting magnetic field acquiring opposite directions above and below the equatorial plane. The fields directed toward and away from the Sun should be separated by a thin current sheet with an azimuthal current.

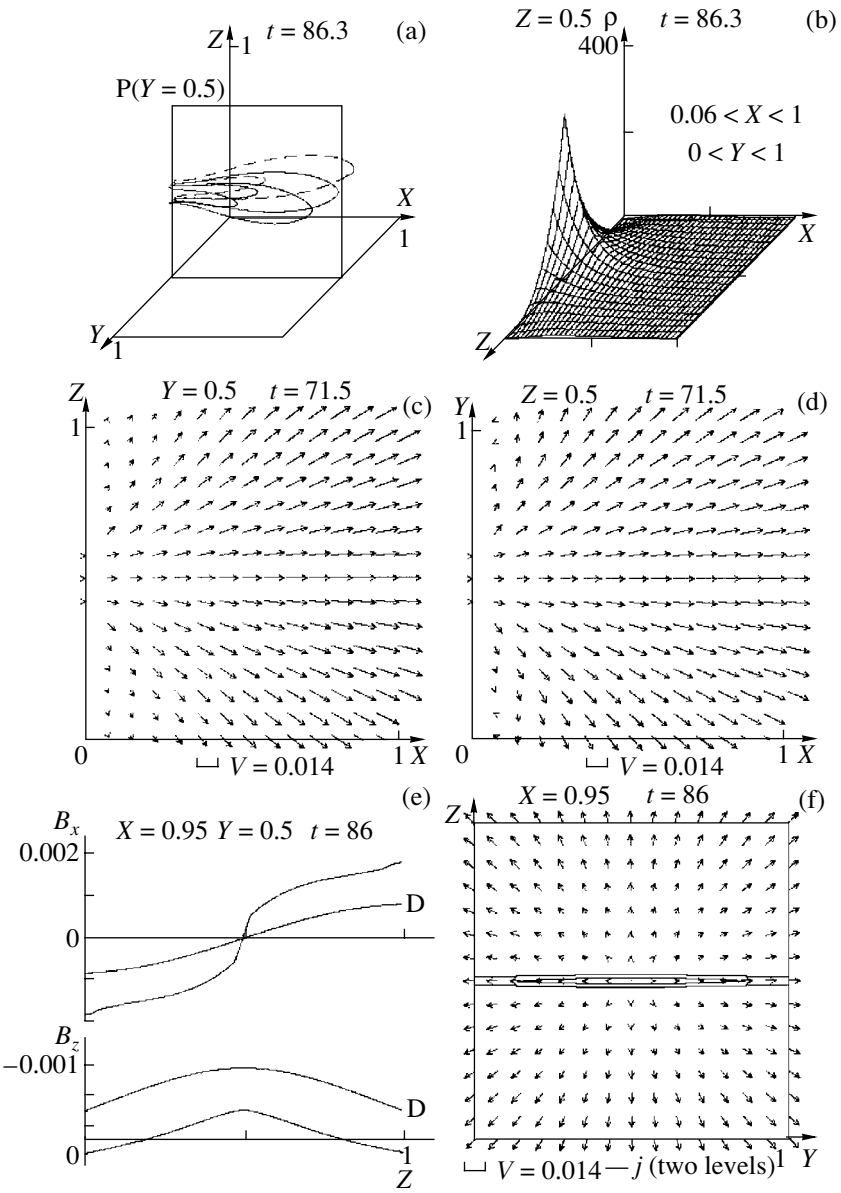

Fig. 3. (a) Stretched magnetic field lines in the three-dimensional space; (b) plasma densities in the plane perpendicular to the current sheet; (c, d) plasma velocity vectors; (e) longitudinal and transverse components of the magnetic dipole (D) field and of the field in the presence of the current sheet; (f) $j_y = \text{const}$ lines for the current sheet and velocity vectors in the YZ plane.

Figure 2a shows the magnetic lines of the solar dipole field in the plane containing the dipole axis. This field was used as the initial conditions. The density ρ and temperature T of the coronal plasma and the initial velocity calculated from the continuity equation, $4\pi(2\,R_\odot)^2\rho V = dM_\odot/dt$, were specified in the $X = 0$ plane for the time $t = 0$. Here, $dM_\odot/dt \sim 2 \times 10^6$ g/s and $M_\odot$ is the mass of the Sun. The temporal evolution of the magnetic field is presented in Figs. 2a–2d. The field changes only slightly at $t > 100$; that is, Fig. 2d shows the pattern for the steady-state field. Comparing Figs. 2d and 1a, we can see that the magnetic-field lines become stretched along the X axis due to the generation of a thin current sheet in the equatorial plane, with the current directed perpendicular to the plane of the figure. Although the magnetic-field lines in the vicinity of the current sheet are strongly stretched along the axis, the sheet is not neutral: it contains a normal field component. Moreover, this normal component of the magnetic field B_z and the V_x velocity component are responsible for maintaining the Lorentz electric field $\mathbf{V} \times \mathbf{B}/c$ that drives the azimuthal current in the heliospheric current sheet. The distributions of the current density, magnetic field B_x, and plasma density are presented in Fig. 2e, while the pattern of the current sheet, with

the current flowing in and out through the boundaries $Y = 0$ and $Y = 1$, respectively, is presented in Fig. 2f. No signs of any instability of the current sheet were demonstrated in the computations.

Figure 3a shows the three-dimensional pattern of the magnetic field containing the current sheet. For ease of visualization, we have introduced a semi-transparent plane P ($Y = 0.5$). The magnetic-field lines in front of and in the plane are shown by solid curves, while the curves corresponding to lines behind the plane are dashed. Compared to the dipole field, all the field lines are stretched along the X axis. In a steady flow, the plasma density decreases with distance from the Sun, with a "hump" of increased density formed around the current sheet (Fig. 3b). The flow in which an equilibrium density distribution has been established demonstrates a fairly high degree of spherical symmetry (Figs. 3c, 3d, 3f), and the expanding plasma velocity increases with distance from the Sun (Figs. 3c, 3d). Figure 3e presents the behavior of the steady longitudinal and transverse magnetic-field components in the vicinity of the current sheet, and Fig. 3f shows the thickness of the sheet, where two levels of $j = \text{const}$ separated by $\Delta j = j_{\max}/4$ are shown.

4. CONCLUSIONS

Comparing the calculations that included and neglected gravitation, we see that the gravitational force does not exert a determining influence on the magnetic field in the equatorial plane. The behavior of the current-sheet formation is similar in both cases. The high thermal velocity of hydrogen atoms at distances of $\sim 2R_\odot$ from the Sun—$\sim 3 \times 10^7$ cm/s, approaching the escape velocity—is responsible for this similarity. There is an appreciable difference in the plasma densities only above the poles, where the radial force $\mathbf{j} \times \mathbf{B}/c$ is zero. The current sheet is not neutral and contains a normal component of the magnetic field and plasma flowing away from the Sun. Both these factors play a decisive role in the formation of a stable current sheet. The electric field VB_z/c that results in the current is generated by the plasma flow along the sheet, while the force $j_y B_z/c$ retards the plasma motion, transferring momentum from the flow to the magnetic field and stretching the field lines. The plasma flow carries away any occasional MHD perturbations, preventing the development of such perturbations and providing a high degree of stability of the sheet. The thickness of the sheet is comparable to the spatial integration step; i.e., it is limited from below by the numerical viscosity. Additional studies are required to clarify the physical mechanisms limiting the thickness of the heliospheric sheet, in particular, calculations on a smaller grid in a two-fluid approximation.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-16186) and the Federal Science and Technology Program "Astronomy."

REFERENCES

1. E. R. Priest, *Solar Magnetohydrodynamics* (Reidel, Dordrecht, 1982; Mir, Moscow, 1985).
2. C. T. Russell, in *Space Weather. Geophysical Monograph* (AGU, 2001), Vol. 125, p. 73.
3. G. W. Pneuman and R. A. Kopp, Sol. Phys. **18**, 258 (1971).
4. S. Koutchmy and M. Livshits, Space Sci. Rev. **61**, 393 (1992).
5. Yu. V. Pisanko, Geomagn. Aeron. **25**, 17 (1985).
6. H. P. Furth, J. Killeen, and M. N. Rosenbluth, Phys. Fluids **6**, 459 (1963).
7. A. V. Usmanov, Sol. Phys. **143**, 345 (1993).
8. A. V. Usmanov, Doctoral Dissertation (St. Petersburg, 1999).
9. S. Minami, M. Morimoto, A. I. Podgorny, and I. M. Podgorny, in *Proceedings of the IAU. 8th Asian-Pacific Regional Meeting* (Japan, 2002), Vol. 2, p. 455.
10. A. I. Podgorny and I. M. Podgorny, *Solar Drivers of Interplanetary and Terrestrial Disturbances,* Ed. by K. S. Balasubramaniam, S. L. Keil, and R. N. Smartt, ASP Conf. Ser. **95**, 66 (1996).
11. A. I. Podgorny and I. M. Podgorny, Astron. Zh. **75**, 132 (1998) [Astron. Rep. **42**, 116 (1998)].

Translated by V. Badin