Astronomy Reports, Vol. 48, No. 1, 2004, p. 1. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, p. 3.

Astronomicheskiĭ Zhurnal is 80 Years Old

The history of astronomical research in Russia began in 1701, when Yakov Bryus, a brother-in-arms of Peter I, conducted observations of heavenly bodies from the Sukharev Tower in Moscow. By 1819, an Astronomy Department was established at the Main Pedagogical Institute in Saint Petersburg, and Vasiliĭ Yakovlevich Struve founded the oldest of the still existing astronomical observatories in Pulkovo in 1839.

Over the past three hundred years, Russian astronomers have made a whole series of discoveries that have substantially enriched the world of science. Unfortunately, before 1924, Russian astronomers were able to report their results only in limited-circulation publications of astronomical societies or in foreign scientific journals. Acknowledging the unprecedented growth in astronomical research in the USSR, V.G. Fesenkov founded *Astronomicheskiĭ Zhurnal* (published in English as *Astronomy Reports*), in 1924, with the support of S.A. Blazhko and A.A. Mikhaĭlov.

And now, *Astronomichskiĭ Zhurnal* has already been serving the needs of Russian science for 80 years. All the most eminent astronomical discoveries made in Russia over that time have been published in *Astronomicheskiĭ Zhurnal*. Over the entire history of the journal, the editors have done everything possible to enhance the quality of its articles and, as a consequence, all astronomical science in the country. Due to the joint efforts of the authors, reviewers, and members of the editorial board, *Astronomichskiĭ Zhurnal* is well respected in the astronomical world. This is testified to by the consistently large number of subscribers (both in Russia and abroad) and the high "impact index" of the journal.

In the name of all Russian astronomers, I wish *Astronomichskiĭ Zhurnal* a glorious anniversary year, the editors and authors of the journal further creative successes and accomplishments, and all its readers many new and interesting articles.

Chairman of the Scientific Council
on Astronomy, Academician N.S. Kardashev

1063-7729/04/4801-0001$26.00

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 2–8. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 4–10.

Observational Constraints on Λ Inflation

E. V. Mikheeva and V. N. Lukash

Astro Space Center, Lebedev Physical Institute, Moscow, Russia

Received May 8, 2003; in final form, August 8, 2003

Abstract—Observational data on the large-scale structure of the Universe and the anisotropy of the cosmic background radiation are used to derive constraints on the parameters of an inflationary model for the early Universe (Λ inflation). © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The appreciable increase in the accuracy of cosmological observations in recent years has not only led to an increase in the number of measured cosmological parameters, but has even demanded an increase in the precision of theoretical predictions. The most important achievements in observational cosmology are associated with the detection of anisotropy in the cosmic background radiation (CMB) on scales from several arcminutes to tens of degrees (WMAP [1], ARHEOPS [2], VSA [3], CBI [4], DASI [5], BOOMERanG [6], MAXIMA [7]) and studies of the spatial distribution of galaxies and galaxy clusters using the most complete and deep catalogs of these objects [8–10].

In modern analyses of observational data, models of the Universe can include as many as ten parameters. The set of measured parameters that are considered is determined by the character of the observational data. For example, data on the anisotropy of the CMB can be used to determine the amplitude and slope of the density-perturbation spectrum, the amplitude of primordial gravitational waves, the cosmological densities of all components of matter, the value of the Hubble constant, and the redshift and optical depth of reionization. So-called a priori hypotheses are widely used to enhance the accuracy with which parameters are determined. One example is the assumption that the spectrum of density perturbations or cosmological gravitational waves can be approximated by a power law. The results of other studies are often included in the form of a priori information about the expected values and possible ranges for various cosmological parameters. While the anisotropy of the CMB serves as a primary test for theories, other supplementary (or complementary) sources of information include studies of large-scale structure, data on the density of baryons in the Universe derived from the theory of primary nucleosynthesis, measurements of the Hubble constant derived from the nearest type Ia supernovae, and statistical data on more distant supernovae of this type, which make it possible to decrease the errors in estimates of the cosmological density of matter, and accordingly in the value of the cosmological constant.

Another important aid for tests of cosmological models would be the expansion of the range of measurements of the density-perturbation power spectrum to provide overlap between the measurements made by various experiments. In studies of the anisotropy of the CMB, this requires a transition to ever smaller angular scales, and in studies of the spatial distribution of galaxies, a transition to samples and catalogs containing larger linear sizes. Overlap of sections of the spectra reconstructed from various observational data was achieved in 2002, making it possible to determine cosmological parameters of the Universe at $z \sim 0$ independent of assumptions about the primordial (post-inflationary) density-perturbation spectrum [10]. These results provide convincing evidence that our understanding of the origin and evolution of large-scale structure in the Universe is correct, and that the computed values of cosmological parameters are close to their real values.

However, while it is possible to investigate the quantitative characteristics of matter in the Universe even with the current accuracy of observational data, the determination of the primordial spectra of density perturbations and cosmological gravitational waves is a much more complex problem. At first glance, this task is made simpler by the fact that the two spectra are related to each other by the parameters of the inflaton potential. In fact, for many models this is only seemingly true. As an example, let us compare the predictions of two models for chaotic inflation, one based on the theory of a massive scalar field (the potential has the form $V(\varphi) = m^2\varphi^2/2$), and the other based on the theory of a self-interacting scalar field ($V(\varphi) = \lambda\varphi^4/4$). In the former model, the slope of the density-perturbation spectrum (n_S) on scales

1063-7729/04/4801-0002$26.00

of ~ 1000 Mpc is 0.98 and the relative amplitude of primordial gravitational waves (T/S) is 0.1, while, in the latter model, these quantities are equal to 0.96 and 0.2, respectively. However, the available accuracy of the observations and methods for analysis of the observational are insufficient to distinguish spectra with such similar slopes, as well as to detect cosmological gravitational waves with such small amplitudes. Thus, these inflationary models still remain indistinguishable.

The difficulties standing in the way of direct testing of existing models of inflation have prompted many researchers to use a phenomenological approach to constructing models for the early Universe, in which the amplitudes and slopes of the spectra for various types of perturbations are initially treated as independent parameters, after which the observational constraints are compared with the theoretical predictions of a broad group of models. The main difficulty of this approach is that it is necessary during the analysis of the observational data to specify at least the functional dependence of the perturbation spectra on the wave number, which appreciably narrows the generality of the conclusions. It is obvious that, if this functional dependence is chosen incorrectly, even high-quality observational data will be inconsistent with the theoretical predictions, and if data with substantial statistical errors are used to test theories, the inflationary model that was actually realized in the early stages of the evolution of the Universe could be rejected.

Here, we consider the results of observational tests of one inflationary model—Λ inflation [12, 13]. This family of models is characterized by non-power-law spectra for cosmological perturbations: the density-perturbation spectrum has a global minimum whose position relative to the current horizon (k_{cr}) is a free parameter of the model, while the slope of the primordial gravitational-wave spectrum is negative on large scales ($k < k_{cr}$) and zero on small scales ($k > k_{cr}$). Observational constraints on the Λ-inflation parameters have been considered earlier only in models of the Universe with zero cosmological constant [14].

Although the consideration of non-power-law cosmological spectra is not fundamentally new [15–17], tests for such spectra using observational data have been limited, since reasonably trustworthy data on the qualitative and quantitative content of dark matter and energy have become available only recently. Attempts to detect deviations of the density-perturbation spectrum from a power law in phenomenonological analyses have, thus far, not yielded convincing results [1, 11].

2. Λ INFLATION

Λ inflation is a model for the early Universe based on a scalar field theory with a potential of the form [12–14]

$$V(\varphi) = V_0 + \frac{1}{\kappa}\lambda_\kappa \varphi^\kappa, \tag{1}$$

where V_0, λ_κ, and κ are constants. We will concentrate here on models with $\kappa = 4$.

In the literature, this model is often called "hybrid" inflation, but we reserve this name for the special case of Λ inflation with a certain specified method for the decay of the effective Λ term [18].

It is convenient to introduce two new parameters in order to simplify the description of the perturbation spectra. The first, c, is a combination of already determined constants of the inflaton potential:

$$c \equiv \frac{1}{2}\sqrt{\frac{V_0}{\lambda}} \tag{2}$$

(here and below, $\lambda \equiv \lambda_4$). The parameter c is related to the critical value of the inflaton for which the two terms of the potential become equal ($\varphi_{cr} \equiv (4V_0/\lambda)^{1/4} = 2\sqrt{c}$). The second parameter is the critical scale mentioned in the Introduction, k_{cr}; the inflaton φ takes on its critical value φ_{cr} when $k = k_{cr}$. When $c \gg 1$, the slow-rolling approximation is applicable for any values of the field φ, and the spectrum of the scalar perturbations (the S-mode of the metric perturbation) takes the form [19, 13] (see also the Appendix)

$$q_k = \frac{H}{2\pi\sqrt{2\gamma}} = \frac{\sqrt{2\lambda/3}}{\pi}\left(c^2 + x^2\right)^{3/4}, \tag{3}$$

and the gravitational-wave spectrum (T modes) takes the form

$$h_k = \frac{H}{\pi\sqrt{2}} = \frac{2c\sqrt{\lambda/3}}{\pi}\left(1 + \frac{x}{\sqrt{c^2 + x^2}}\right)^{-1/2}, \tag{4}$$

where $H \equiv \dot{a}/a$ is the Hubble parameter during inflation, $\gamma \equiv -\dot{H}/H^2$, $x \simeq \ln(k/k_{cr})$, $k = aH$ is the Hubble frequency (the scale of the horizon in the inflationary stage), a is the scale factor, and a dot denotes a derivative with respect to physical time. Figure 1 illustrates the appearance of both spectra.

The ratio of the power spectra (r_k) and their slopes (n_S and n_T for the density-perturbation and gravitational-wave spectra, respectively) have the form

$$r_k \equiv \frac{h_k^2}{q_k^2} = 4\gamma = -2n_T, \tag{5}$$

$$n_S - 1 \equiv 2\frac{d\ln q_k}{d\ln k} \simeq \frac{3x}{c^2 + x^2},$$

$$n_T \equiv 2\frac{d\ln h_k}{d\ln k} \simeq -\frac{\sqrt{c^2 + x^2} - x}{c^2 + x^2}.$$

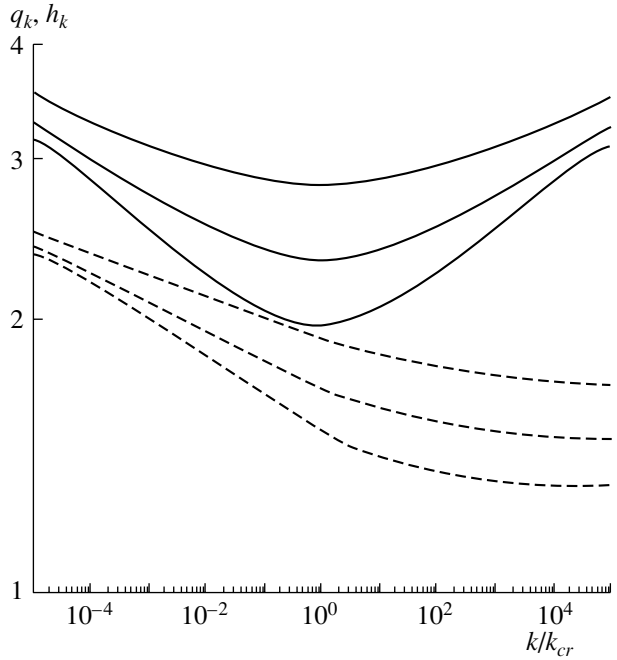


Fig. 1. Density-perturbation spectra (solid curves) and gravitational-wave spectra (dashed curves) for $c = 7$, 9, and 12 (lower, middle, and upper curves, respectively). The normalization of the spectra is arbitrary.

3. THE COSMOLOGICAL MODEL

It is obvious that a full analysis of a model for the Universe should include several observational tests that are sensitive to various cosmological parameters and operate on various scales. When used jointly, some tests can reveal the values of parameters describing the quantitative characteristics of matter (the density of "cold" dark matter Ω_c, the density of massive neutrinos Ω_ν and the number of types of massive neutrinos N_ν, and the baryonic density Ω_b, with the total density of matter $\Omega_m = \Omega_c + \Omega_\nu + \Omega_b$ and the value of the cosmological constant $\Omega_\Lambda = 1 - \Omega_m$ for a spatially flat Universe). Other types of observational tests can indicate the form of perturbation spectra (in the case of Λ inflation, these are the parameters c, k_{cr}, and λ). A third group of tests are sensitive to physical processes occurring in the cosmic plasma (for example, the reionization of hydrogen) and other parameters that are not included in the first two groups of tests, such as the Hubble constant $H_0 = 100h$ km s^{-1}Mpc^{-1} and others. There is no doubt that the task of fully testing a cosmological model is unwieldy. However, it can be simplified if we restrict our analysis to some subset of the parameters, for example, the parameters of the inflationary model.

Let us suppose that our understanding of the current Universe and the structure of matter in it correspond to reality: the spatial curvature of the Universe is zero (or indistinguishable from zero) and about one-third of the total density of matter is determined by various types of matter, while two-thirds are provided by the cosmological constant or some similar manifestation (quintessence). For specificity, we adopt the following values for the most important cosmological parameters: $\Omega_\Lambda = 0.68$ ($\Omega_m = 0.32$), $\Omega_c = 0.27$, $\Omega_b = 0.05$, $\Omega_\nu = 0$, and $h = 0.68$, which are the most probable values according to the analyses of the WMAP group [1] and are consistent with other available observational data.

Having limited the number of parameters to be studied, we can also limit the number of tests for the determination of the remaining parameters—in this case, those that describe the inflationary model. It is obvious that the necessary observational data should primarily be sensitive to the amplitude of both types of perturbation spectra on various scales, with the range of scales being as broad as possible. These properties are displayed by two fundamental tests. The first is the amplitude of the anisotropy of the CMB on angular scales $\sim 10^\circ$ (C_{10}), which is the sum of the contributions of the density and primordial gravitational-wave perturbations, and the second is the value of the parameter σ_8, which is essentially an independent normalization of the density-perturbation spectrum.

Below, we adopt without change those values of these two quantities that correspond to the so-called "basis" cosmological model—with a scale-invariant density-perturbation spectrum ($n_S = 1$) and zero amplitude for the primordial gravitational waves (T/S $= 0$).[1]

4. ANALYSIS OF THE MODEL

The contribution of primordial gravitational waves to the large-scale anisotropy of the CMB can be expressed as follows [13]:

$$\mathrm{T/S} \equiv \frac{C_{10}^T}{C_{10}^S} \simeq 3r_{k_1} \simeq -6n_T. \qquad (6)$$

(For the basis model $n_T = 0$ and T/S $= 0$.) The linear scale corresponding to the tenth harmonic of the expansion in spherical functions (see (6)) is $k_1 \simeq 1.5 \times 10^{-3} h$ Mpc^{-1}, while the scale σ_8 is accordingly $k_2 \simeq 1.5 \times 10^{-1} h$ Mpc$^{-1} \simeq 100 k_1$.

[1] Note that 0.97 is named as the most probable value of the slope of the perturbation spectrum in [1]. However, in this case, the value of σ_8 computed using the CMBFAST [21] is only 0.81 (recall that $\sigma_8 = 0.9 \pm 0.1$ is indicated [1]), while $\sigma_8 = 0.87$ in the case of the scale-invariant spectrum, which is appreciably closer to the median of the distribution of values. For this reason and to simplify the equations introduced in the following section, we have adopted $n_S = 1$ for the phenomenological basis model.

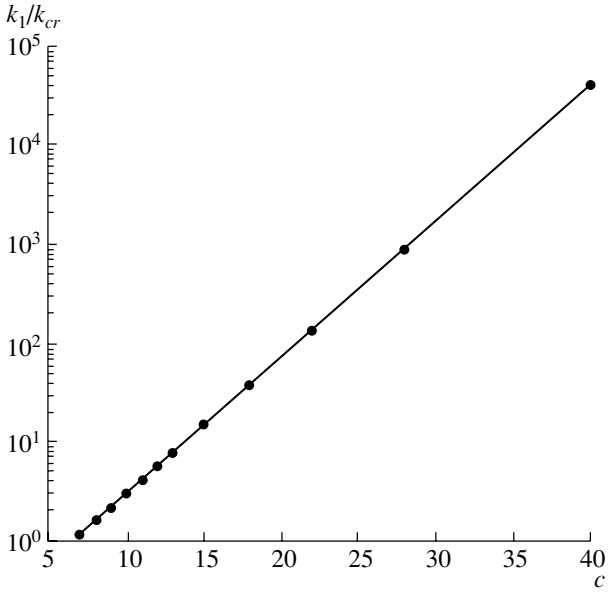


Fig. 2. Numerical solution of (10) (points) and a linear approximation to this solution.

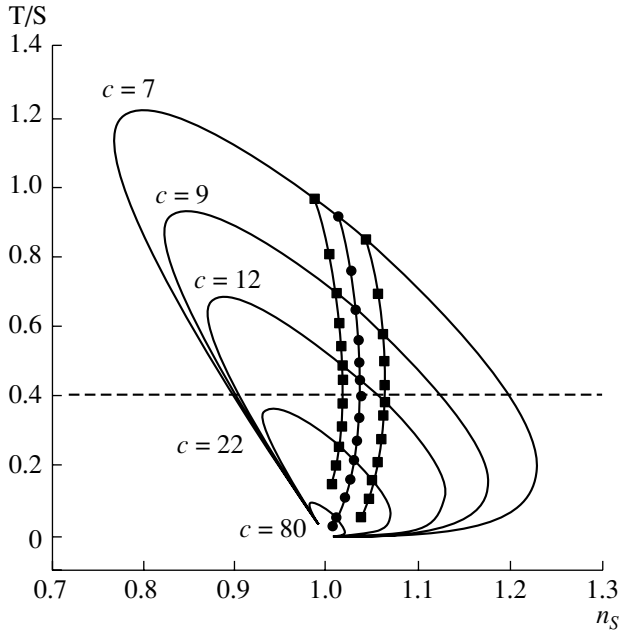


Fig. 3. The ratio T/S as a function of n_S and c (k/k_{cr} is the affine parameter along the curves). The circles (numerical calculations) denote models satisfying both of the observational tests considered (see (10)).

In the general case, the power spectrum of density-perturbations can be expressed as follows (see the Appendix):

$$P(k) = g_0^2 k q_k^2 T_k^2, \tag{7}$$

where g_0 is a constant that is independent of k and T_k is a scale-dependent transfer function taking into account the evolution of the density perturbations below the horizon. The quantities g_0 and T_k are determined by cosmological parameters describing matter in the Universe.

On the other hand, in the case of the basis model, when the density-perturbation spectrum is scale invariant ($q_k = q_0 = \text{const}$)

$$P(k)^{(0)} = k q_1^2 T_k^2, \tag{8}$$

where $q_1 = g_0 q_0^2$ is a constant.

Thus, taking into account the quadratic dependence of T/S on the perturbation spectrum and the fact that the basis model satisfies both observational tests, we can write the following system of equations:

$$\begin{cases} (1 + \text{T/S})\, P(k)|_{k=k_1} = P(k)^{(0)}|_{k=k_1} \\ P(k)|_{k=k_2} = P(k)^{(0)}|_{k=k_2}. \end{cases} \tag{9}$$

Furthermore, using the definitions of $P(k)$ and $P(k)^{(0)}$ and dividing the first equation by the second, we obtain the simple relation

$$(1 + \text{T/S}) \left(\frac{q_{k_1}}{q_{k_2}} \right)^2 = 1. \tag{10}$$

This equation contains no dependence on the structure of the dark matter and energy in the late Universe. It is obvious that, when generalizing (10) to the case when the basis model has a nonflat spectrum or there exist differences between the transfer functions of the basis and tested models, the right-hand side will become different from unity. In the case of the basis model, (10) degenerates to an identity.

The joint numerical solution to (3), (5), (6), and (10) can be approximated by the linear function (Fig. 2)

$$\log \frac{k_1}{k_{cr}} = 0.14c - 0.91, \tag{11}$$

where the statistical error of the approximation is < 0.01 for both constants.

An important consequence of this solution is that in the region in which this approximation is applicable ($c \gg 1$) the critical scale for the perturbation spectra is close to or beyond the current horizon ($k_{cr} \lesssim k_1 \sim 10^{-3} h\ \text{Mpc}^{-1}$).[2]

Figure 3 illustrates the relation between T/S, $n_S|_{k=k_1}$, and c for inflationary models satisfying both tests considered. Recall that we have estimated T/S using (6). In the region between the lines marked

[2] This solution does not include the branch (3) corresponding to chaotic inflation ($c < 1$), since we have from the beginning limited our consideration to $c \gg 1$.

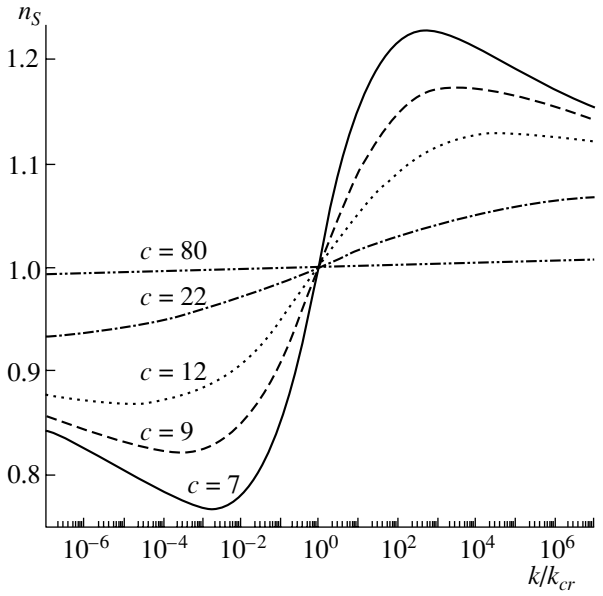


Fig. 4. n_S as a function of k/k_{cr} for various values of c.

by squares, the right-hand side of (10) varies from 0.89 to 1.11, in accordance with the relative error in measurements of σ_8 based on the WMAP data.

Thus, of all the variety of possible realizations of Λ inflation, only those models lying along the line marked by circles are permitted. Supposing that T/S is not large[3] (T/S < 0.4) we obtain a constraint on the slope of the perturbation spectrum on a scale of the order of the current horizon ($1 < n_S < 1.05$) and on the inflation parameter ($c > 12$).

Since the slope of the density-perturbation spectrum in Λ inflation depends on the scale, when $k = k_2$, it could differ appreciably from the slope of the spectrum for $k = k_1$. We can see from Fig. 4 that the slope of the spectrum $n_S(k_2)$ can vary from 1 (for $c \gg 1$) to 1.2 (for small values $c \sim 7$).

5. CONCLUSION

We have shown that, in order for Λ inflation to be compatible with the observational data, a version leading to a weakly blue density-perturbation spectrum ($n_S \gtrsim 1$) within the horizon must be realized. The extent of the necessary deviation from a flat spectrum is compatible with available constraints, but the accuracy of these constraints is currently insufficient to enable unambiguous arguments in favor of or against this model (see, for example, [24]).

[3] Note that all existing constraints on the value of T/S are based on the assumption that the density-perturbation spectrum is a power-law (see, for example, [22, 23]).

Thus, if it becomes clear as a result of future studies that the density-perturbation spectrum is blue ($n > 1$), this will become a conclusive argument in favor of Λ inflation as only this model (and models derived from it, such as hybrid inflation) can lead to such a density-perturbation spectrum.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-12274) and the Federal Scientific and Technological Program "Studies and Development in Priority Directions in Science and Technology" (40.022.1.1.1106) and the Leading Scientific School (NSh1653.2003.2).

APPENDIX

Recall the gauge-invariant determination of the perturbation spectra of the metric in the scalar (S) and tensor (T) modes [19, 13].

In the absence of a vector (V) mode, the general representation of the gravitational potentials in the linear perturbation theory has the form

$$ds^2 = (1+2D)\,dt^2 + 2C_{,\alpha}dtdx^\alpha - a^2\left[(1+2A)\,\delta_{\alpha\beta} + 2B_{,\alpha\beta} + 2G_{\alpha\beta}\right]dx^\alpha dx^\beta, \quad \text{(A.1)}$$

where any two of the four scalar potentials A, B, C, and D are arbitrary functions of all coordinates (choice of calibration), and the tensor $G_{\alpha\beta}$ ($G^\alpha_\alpha = G^\beta_{\alpha,\beta} = 0$)[4] describes cosmological gravitational waves.

For media with isotropic pressure, the full energy–momentum tensor including the zeroth- and first-order terms of the perturbation theory have the form

$$T_{ik} = (\epsilon + p)\,u_i u_k - p g_{ik}, \quad \text{(A.2)}$$

where $u_i = (1 + D; v_{,\alpha})$ is the four-velocity of matter, g_{ik} is the metric tensor (A.1), and ϵ and p are the energy density and pressure of the medium, respectively.

In the zeroth-order perturbation theory,[5]

$$\epsilon = \frac{3H^2}{8\pi G}, \quad \epsilon + p = \frac{2}{3}\gamma\epsilon. \quad \text{(A.3)}$$

In the first-order perturbation theory, the adiabatic (S) mode of the density perturbations is fully

[4] All manipulations with Greek-letter subscripts and superscripts are carried out in a Euclidean three-dimensional space using the unit tensor $\delta_{\alpha\beta}$, and a comma denotes a partial derivative with respect to the corresponding spatial coordinate.

[5] The Λ term is included in the definition of ϵ and p.

described by the four-scalar q, which is a gauge-invariant superposition of perturbations of the scale factor (A) and the total potential of the velocity of matter (v) [19]:

$$q = -A + Hv. \tag{A.4}$$

Physically, the scalar q can represent

—the curvature of space (a perturbation of the local scale factor relative to the background) in a comoving coordinate system ($v = 0$: $q = -A$),

—the dimensionless potential of the velocity of matter in a Euclidean coordinate system ($A = 0$: $q = Hv$).

In other coordinate systems (for example, synchronous systems), these interpretations remain valid on large or small scales relative to the Hubble radius, respectively.

The relation of the initial S-mode potentials with the scalar q follows from the Einstein equations [19, 14]:

$$v = \frac{q + A}{H}, \quad \delta\epsilon_c = \frac{\Delta\Phi}{4\pi G a^2}, \tag{A.5}$$

$$a^2\dot{B} + C = \frac{\Phi + A}{H},$$

$$D = \gamma q + \left(\frac{A}{H}\right)^{\cdot} = \Phi + \left(\frac{\Phi + A}{H}\right)^{\cdot} = \dot{v} - \frac{\dot{q}}{H},$$

where $\Phi = \frac{H}{a}\int a\gamma q dt$ is the Newtonian gravitational potential, and $\delta\epsilon_c = \delta\epsilon + 3Hv(\epsilon + p)$ is a density perturbation in a comoving coordinate system.

The tensor mode of the perturbations includes two polarizations of the gravitational waves:

$$G_{\alpha\beta} = h_+ e^+_{\alpha\beta} + h_\times e^\times_{\alpha\beta}, \tag{A.6}$$

where $e^{a\alpha}_{\alpha} = e^{a\beta}_{\alpha,\beta} = 0$, $e^a_{\alpha\beta} e^{a'\alpha\beta} = \delta^{aa'}$, $a = +, \times$.

The corresponding spectra q_k and h_k are determined on scales larger than the horizon in the early Universe, and do not depend on time:

$$\langle q^2\rangle = \int_0^\infty q_k^2 \frac{dk}{k}, \tag{A.7}$$

$$\langle G_{\alpha\beta}G^{\alpha\beta}\rangle\langle h_+^2\rangle + \langle h_\times^2\rangle \int_0^\infty h_k^2 \frac{dk}{k},$$

where angular brackets $\langle\ldots\rangle$ denote averaging over an ensemble of random realizations of the fields q and h. In late stages of the expansion of the Universe, when the pressure of matter can be neglected, the scalar q ceases to depend on time, and the density perturbation ($\delta \equiv \delta\epsilon_c/(\epsilon + p)$) is factorized (see (A.5)):

$$\langle\delta^2\rangle = \int_0^\infty P(k)k^2 dk = g^2 \int_0^\infty q_k^2 T_k^2 k^3 dk,$$

where $g = g(a) = \dfrac{1}{a^3\gamma H}\int \dfrac{\gamma}{H} da$, and T_k is a transfer function taking into account the evolution of perturbations in the early Universe.

REFERENCES

1. D. N. Spergel, L. Verde, H. V. Peiris, *et al.*, astro-ph/0302209.
2. A. Benoit, P. Ade, A. Amblard, *et al.*, astro-ph/0210306.
3. J. A. Rubino-Martin, R. Rebolo, P. Carreira, *et al.*, astro-ph/0205367.
4. B. S. Mason, T. J. Pearson, A. C. S. Readhead, *et al.*, astro-ph/0205384.
5. C. Pryke, N. W. Halverson, E. M. Leith, *et al.*, Astrophys. J. **568**, 46 (2002).
6. P. de Bernardis, P. A. R. Ade, J. J. Bock, *et al.*, Nature **404**, 955 (2000).
7. A. Balbi, P. Ade, J. Bock, *et al.*, Astrophys. J. **545**, L1 (2000).
8. O. Lahav, I. J. Baldry, C. M. Baugh, *et al.*, astro-ph/0205382.
9. C. J. Miller and D. J. Batuski, Astrophys. J. **551**, 635 (2001).
10. M. Tegmark and M. Zaldarriaga, Phys. Rev. D **66**, 103508 (2002).
11. S. Hannestad, S. H. Hansen, F. L. Villante, *et al.*, Astropart. Phys. **17**, 375 (2002).
12. V. N. Lukash and E. V. Mikheeva, Gravit. Cosmol. **2**, 247 (1996).
13. V. N. Lukash and E. V. Mikheeva, Int. J. Mod. Phys. A **15**, 3783 (2000).
14. V. N. Lukash, E. V. Mikheeva, V. Müller, *et al.*, Mon. Not. R. Astron. Soc. **317**, 795 (2000).
15. A. A. Starobinskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **55**, 477 (1992) [JETP Lett. **55**, 489 (1992)].
16. P. Ivanov, P. Naselsky, and I. Novikov, Phys. Rev. D **50**, 7173 (1994).
17. L. V. Semig and V. Müller, Astron. Astrophys. **308**, 697 (1996).
18. A. Linde, Phys. Rev. D **49**, 748 (1994).

19. V. N. Lukash, Zh. Éksp. Teor. Fiz. **79**, 1601 (1980) [Sov. Phys. JETP **52**, 807 (1980)].

20. S. Dodelson, L. Knox, and E. W. Kolb, Phys. Rev. Lett. **72**, 3444 (1994).

21. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).

22. A. Melchiorri and C. J. Odman, Phys. Rev. D **67**, 021501(R) (2003).

23. A. Melchiorri, P. Bode, N. A. Bahcall, *et al.*, Astrophys. J. **586**, L1 (2003).

24. B. Novosyadlyj, R. Durrer, S. Gottlöber, *et al.*, Astron. Astrophys. **356**, 418 (2000).

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 21–28. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 24–32.

Polarization in Resonance Lines: Diffuse Reflection

V. M. Loskutov

Sobolev Astronomical Institute, St. Petersburg State University, St. Petersburg, Russia

Received December 10, 2002; in final form, August 8, 2003

Abstract—The linear polarization of resonance radiation diffusely reflected from a semi-infinite atmosphere is analyzed, including a full account of the azimuthal dependence. The radiating atom is assumed to have two levels and the radiation to be fully redistributed in frequency. A new approach to the transformation of the equations of polarized radiative transport is proposed based on factorization of the full (3×3) Rayleigh phase matrix. The emergent radiation field is expressed in terms of a matrix analog for the Chandrasekhar H function. The polarization properties of radiation at the frequencies of resonance lines diffusely reflected from a scattering atmosphere are discussed. The degree of polarization of the reflected radiation can be very high. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The formation of spectral lines taking into account the polarization of radiation is a problem that has attracted attention in recent years in connection with the detection of appreciable polarization in lines observed at the edge of the solar disk [1–3]. This question is also relevant for studies of Wolf–Rayet stars, binary and symbiotic stars, and other objects (see, for example, [4, 6]).

Three problems can be modeled in the classical theory of radiative transport in scattering, semi-infinite atmospheres [7–9]: the Milne problem, radiative transport in an atmosphere with uniformly distributed sources, and radiative transport with diffuse reflection. Only the first two of these have been considered for linearly polarized radiation (see, for example, [10–18]), while the third has remained unstudied. It is this third problem that is the subject of this paper.

2. FORMULATION OF THE PROBLEM. FACTORIZATION OF THE PHASE MATRIX

Let partially polarized radiation described by the Stokes vector $\mathbf{i}_0(x_0, \mu_0, \varphi_0)$ at the dimensionless frequency x_0 relative to the center of a line be incident on a semi-infinite atmosphere consisting of two-level atoms at an angle $\arccos \mu_0$ to the outward normal at azimuth φ_0. We will assume that the radiation in the spectral line undergoes a total redistribution in frequency during the scattering. In this case, the matrix function for the redistribution, $\mathbf{R}(x, \mu, \varphi; x', \mu', \varphi')$, describing the change in the Stokes vector during the scattering (see, for example, [19]) can be approximated using the following expression derived in [11]:

$$\mathbf{R}(x, \mu, \varphi; x', \mu', \varphi') = \phi(x)\phi(x')\mathbf{P}(\mu, \varphi; \mu', \varphi'), \quad (1)$$

where primed and unprimed quantities describe the frequency and direction of propagation of the radiation before and after scattering, $\phi(x)$ is the absorption-coefficient profile, and $\mathbf{P}(\mu, \varphi; \mu', \varphi')$ is the phase matrix. An explicit expression for this function in terms of the four Stokes parameters (I, Q, U, V) is given in [10]. If we restrict our consideration to the central parts of the line, we can describe $\phi(x)$ using a Doppler profile $\phi(x) = \pi^{-1/2}e^{-x^2}$, where the frequency x is measured in Doppler widths. The matrix $\mathbf{P}(\mu, \varphi; \mu', \varphi')$ is the 4×4 Rayleigh phase matrix. It is well known [7] that, in the case of both Rayleigh and resonance scattering, the Stokes parameter V describing the elliptical polarization of the radiation can be considered independently of the other three parameters I, Q, and U describing the intensity and linear polarization. This is also the case for the transport of linearly polarized radiation in an atmosphere with a weak magnetic field (the Hanley effect) [20]. Therefore, we will further consider only the first three Stokes parameters. In this case, $\mathbf{P}$ will be a 3×3 matrix.

In contrast to the first two modeling problems referred to above, we must take into account the azimuthal dependence of the radiation field when considering the diffuse reflection of the radiation. Usually, following Chandrasekhar [7], the phase matrix $\mathbf{P}(\mu, \varphi; \mu', \varphi')$ is expanded in azimuthal Fourier harmonics whose arguments are even multiples of $\varphi - \varphi'$. We then obtain a 2×2 matrix equation for the azimuth-independent component of the radiation field, and appropriate manipulation yields two separate independent scalar equations for the azimuthal components.

1063-7729/04/4801-0021$26.00

We will use another approach. It is well known that substantial progress can often be made in both analytical and numerical problems if it is possible to carry out a rigorous or approximate separation of variables. At the base of our approach lies the following important fact, which has gone unnoticed until now. It turns out that the phase matrix $\mathbf{P}(\mu,\varphi;\mu',\varphi')$ can be written as the product of two factors, one dependent only on quantities describing the direction of the incident radiation and one dependent only on angles characterizing the direction of the scattered radiation:

$$\mathbf{P}(\mu,\varphi;\mu',\varphi') = \mathbf{A}(\mu,\varphi)\mathbf{A}^T(\mu',\varphi'), \qquad (2)$$

where "^{T}" denotes the transpose of a matrix. Here, $\mathbf{A}$ denotes the following 3×6 matrix:

$$\mathbf{A}(\mu,\varphi) = \begin{pmatrix} 1 & w_0(1-3\mu^2) & w_1\mu\sqrt{1-\mu^2}c_1 & -w_1\mu\sqrt{1-\mu^2}s_1 & w_2(1-\mu^2)c_2 & w_2(1-\mu^2)s_2 \\ 0 & w_0 3(1-\mu^2) & w_1\mu\sqrt{1-\mu^2}c_1 & -w_1\mu\sqrt{1-\mu^2}s_1 & -w_2(1+\mu^2)c_2 & -w_2(1+\mu^2)s_2 \\ 0 & 0 & w_1\sqrt{1-\mu^2}s_1 & w_1\sqrt{1-\mu^2}c_1 & -w_1\mu s_2 & w_1\mu c_2 \end{pmatrix}, \qquad (3)$$

where for the sake of brevity we have denoted

$$c_1 = \cos\varphi, \quad s_1 = \sin\varphi, \quad c_2 = \cos 2\varphi, \quad s_2 = \cos 2\varphi, \qquad (4)$$

and the constants w_i can be expressed in terms of the depolarization coefficient W:

$$w_0 = \sqrt{\frac{W}{8}}, \quad w_1 = \sqrt{\frac{3W}{2}}, \quad w_2 = \sqrt{\frac{3W}{8}}. \qquad (5)$$

The matrix $\mathbf{A}^T(\mu',\varphi')$ is 6×3.

Note that our matrix $\mathbf{A}(\mu,\varphi)$ differs from the analogous matrix in [21] (constructed using formulas from [22]) by a factor of $\sqrt{2}$ for the azimuthal components and in the sign of the elements containing sines. However, those papers were concerned with a different problem—the formation of lines in scattering atmospheres with weak magnetic fields that give rise to the Hanley effect. In that case, the matrix $\mathbf{A}(\mu,\varphi)$ played an auxiliary role, appearing only in the last stage of the analysis, in the transition from the Fourier vector derived in [23] to the physical Stokes vector. As follows from (2), our matrix $\mathbf{A}(\mu,\varphi)$ factorizing the phase matrix $\mathbf{P}(\mu,\varphi;\mu',\varphi')$ should play a key role in *all problems* connected with multiple resonance and molecular scattering.

Factorization independent of the azimuth of a component of the phase matrix was carried out long before [24]. It has been established that this factorization is not unique (see the discussion of this question in [25, 26]). The most convenient form was presented in [27], and has the form of the product of the matrix that is the upper left (2×2) block of $\mathbf{A}(\mu,\varphi)$ and the transpose matrix.

The relationship between our formalism and the traditionally used approach of Chadresekhar is of interest. We represent $\mathbf{A}(\mu,\varphi)$ in the form of the block row matrix

$$\mathbf{A} = (\mathbf{A}_0(\mu), \mathbf{A}_1(\mu,\varphi), \mathbf{A}_2(\mu,\varphi)), \qquad (6)$$

where the $\mathbf{A}_i (i=0,1,2)$ is a 3×2 matrix and is made up of $\mathbf{A}_0$ (the first two columns), $\mathbf{A}_1$ (the next two columns), and $\mathbf{A}_2$ (the last two columns of $\mathbf{A}$). The transpose of this matrix $\mathbf{A}^T(\mu,\varphi)$ will clearly be a block column matrix:

$$\mathbf{A}^T = (\mathbf{A}_0^T, \mathbf{A}_1^T, \mathbf{A}_2^T)^T. \qquad (7)$$

It is then easy to see from (2) that the phase matrix $\mathbf{P}(\mu,\varphi;\mu',\varphi')$ is presented as a sum of three terms, each of which describes a separate Fourier harmonic:

$$\mathbf{P}(\mu,\varphi;\mu',\varphi') = \mathbf{P}_0(\mu;\mu') + \mathbf{P}_1(\mu,\varphi;\mu',\varphi') + \mathbf{P}_2(\mu,\varphi;\mu',\varphi') \qquad (8)$$

and

$$\mathbf{P}_0(\mu;\mu') = \mathbf{A}_0(\mu)\mathbf{A}_0^T(\mu'), \qquad (9)$$
$$\mathbf{P}_i(\mu,\varphi;\mu',\varphi') = \mathbf{A}_i(\mu,\varphi)\mathbf{A}_i^T(\mu',\varphi') \quad (i=1,2).$$

The dimensionality of $\mathbf{A}_0$, $\mathbf{A}_0^T$, and $\mathbf{P}_0$ is 2×2.

Thus, we obtain the Chadresekhar representation of the phase matrix $\mathbf{P}(\mu,\varphi;\mu',\varphi')$.

3. TRANSPORT EQUATION AND ITS TRANSFORM

In the absence of continuum absorption, the equation of transport for the Stokes vector of the scattered radiation $\mathbf{i}(\tau,x,\mu,\varphi)$, where τ is the mean optical depth over the line profile, has the form

$$\mu\frac{\partial\mathbf{i}(\tau,x,\mu,\varphi)}{\partial\tau} = \phi(x)\mathbf{i}(\tau,x,\mu,\varphi) - \mathbf{s}(\tau,x,\mu,\varphi). \qquad (10)$$

Here, $\mathbf{s}(\tau, x, \mu, \varphi)$ is the vector source function. In the approximation of a total redistribution in frequency [formula (1)], it has the form

$$\mathbf{s}(\tau, x, \mu, \varphi) = \frac{\lambda}{4\pi}\phi(x)\int\limits_{-\infty}^{\infty}\phi(x')dx' \qquad (11)$$

$$\times \int \mathbf{P}(\mu, \varphi; \mu', \varphi')\mathbf{i}(\tau, x', \mu', \varphi')d\omega'$$

$$+ \frac{\lambda}{4\pi}\phi(x)\mathbf{P}(\mu, \varphi; -\mu_0, \varphi_0)\mathbf{i}_0(x_0, \mu_0, \varphi_0)e^{-\frac{\phi(x_0)}{\mu_0}\tau},$$

where λ is the albedo for single scattering and the inner integral is taken over all directions (4π).

When using the factorization (2), the source function $\mathbf{s}(\tau, x, \mu, \varphi)$ can be written

$$\mathbf{s}(\tau, x, \mu, \varphi) = \phi(x)\mathbf{A}(\mu, \varphi)\mathbf{s}(\tau) \qquad (12)$$

and $\mathbf{s}(\tau)$ is

$$\mathbf{s}(\tau) = \frac{\lambda}{4\pi}\int\limits_{-\infty}^{\infty}\phi(x)dx\int \mathbf{A}^T(\mu, \varphi) \qquad (13)$$

$$\times \mathbf{i}(\tau, x, \mu, \varphi)d\omega + \mathbf{s}_0(\tau),$$

while

$$\mathbf{s}_0(\tau) = \frac{\lambda}{4\pi}\phi(x_0)\mathbf{A}^T(-\mu_0, \varphi_0)\mathbf{i}_0(x_0, \mu_0, \varphi_0)e^{-\frac{\phi(x_0)}{\mu_0}\tau}. \qquad (14)$$

It follows from (10) and (12) that the vector $\mathbf{i}(\tau, x, \mu, \varphi)$ can be written in the form

$$\mathbf{i}(\tau, x, \mu, \varphi) = \mathbf{A}(\mu, \varphi)\mathbf{i}(\tau, z), \qquad (15)$$

where $z = \mu/\phi(x)$. Note that the vectors $\mathbf{i}(\tau, z)$ and $\mathbf{s}(\tau)$ have six components each. Substituting (15) into (10) and (13) yields

$$z\frac{\partial \mathbf{i}(\tau, z)}{\partial \tau} = \mathbf{i}(\tau, z) - \mathbf{s}(\tau), \qquad (16)$$

$$\mathbf{s}(\tau) = \frac{\lambda}{2}\int\limits_{-\infty}^{\infty}\phi(x)dx\int\limits_{-1}^{1}\mathbf{\Psi}(\mu)\mathbf{i}(\tau, \mu/\phi(x))d\mu + \mathbf{s}_0(\tau), \qquad (17)$$

where *the characteristic matrix* $\mathbf{\Psi}(\mu)$ can readily be expressed in terms of $\mathbf{A}(\mu, \varphi)$; namely, it is the azimuth-averaged product $\mathbf{A}^T\mathbf{A}$:

$$\mathbf{\Psi}(\mu) = \frac{1}{2\pi}\int\limits_{0}^{2\pi}\mathbf{A}^T(\mu, \varphi)\mathbf{A}(\mu, \varphi)d\varphi. \qquad (18)$$

Note that this same characteristic matrix $\mathbf{\Psi}(\mu)$ appears in connection with the transport of linearly polarized radiation in an atmosphere with a weak magnetic field (the Hanley effect) [28].

It is easy to see that the matrix $\mathbf{\Psi}(\mu)$ has the following structure. Its nonzero elements occupy the upper left 2×2 block, from which a "tail" trails to the right and downward along the main diagonal. Thus, only eight elements of $\mathbf{\Psi}(\mu)$ differ from zero:

$$\Psi_{11} = 1, \qquad (19)$$

$$\Psi_{12} = \Psi_{21} = \sqrt{\frac{W}{8}}(1 - 3\mu^2),$$

$$\Psi_{22} = \frac{W}{4}(5 - 12\mu^2 + 9\mu^4),$$

$$\Psi_{33} = \Psi_{44} = \frac{3}{4}W(1 - \mu^2)(1 + 2\mu^2),$$

$$\Psi_{55} = \Psi_{66} = \frac{3}{8}W(1 + \mu^2)^2.$$

We denote

$$\boldsymbol{\epsilon} = \mathbf{E} - \frac{\lambda}{2}\int\limits_{-1}^{1}\mathbf{\Psi}(\mu)d\mu, \qquad (20)$$

where $\mathbf{E}$ is the unit matrix. It is easy to see that

$$\boldsymbol{\epsilon} = \mathrm{diag}\{\varepsilon_I, \varepsilon_Q, \varepsilon_Q, \varepsilon_Q, \varepsilon_Q\} \qquad (21)$$

and

$$\varepsilon_I = 1 - \lambda, \quad \varepsilon_Q = 1 - \frac{7}{10}W\lambda. \qquad (22)$$

As we will see below, the matrix $\boldsymbol{\epsilon}$ plays an important role in the problem at hand. It can be viewed as a matrix generalization of the usual scalar probability for the loss of a photon during scattering $\varepsilon = 1 - \lambda$ (compare with the discussion in [29]).

If $\phi(x)$ monotonically decreases when $x \geq 0$ and $\phi(-x) = \phi(x)$, as is always the case in reality, by exchanging the order of the integration in (17) we obtain [9, §2.7]: (for more detail see [29])

$$\mathbf{s}(\tau) = \frac{\lambda}{2}\int\limits_{-\infty}^{\infty}\mathbf{G}(z)\mathbf{i}(\tau, z)dz + \mathbf{s}_0(\tau), \qquad (23)$$

where

$$\mathbf{G}(z) = 2\int\limits_{x(z)}^{\infty}\phi^2(x)\mathbf{\Psi}(z\phi(x))dx \qquad (24)$$

and the function $x(z)$ is determined by the relations

$$x(z) = 0, \quad |z| \leq 1/\phi(0), \qquad (25)$$

$$\phi(x(z)) = 1/|z|, \quad |z| > 1/\phi(0).$$

The matrix $\mathbf{G}(z)$ has the same nonzero elements as $\mathbf{\Psi}(\mu)$.

Elements I_{33} and I_{55} (Doppler profile, $\lambda = 1, W = 1$)

z	I_{33}	I_{55}	z	I_{33}	I_{55}
0	1.0000	1.0000	1.0E+01	1.6376	1.5910
1.0E-01	1.0586	1.0396	1.2E+01	1.6591	1.6162
2.0E-01	1.0978	1.0687	1.4E+01	1.6758	1.6361
3.0E-01	1.1300	1.0937	1.6E+01	1.6893	1.6523
4.0E-01	1.1577	1.1160	1.8E+01	1.7003	1.6657
5.0E-01	1.1822	1.1362	2.0E+01	1.7096	1.6770
6.0E-01	1.2042	1.1547	2.5E+01	1.7274	1.6989
7.0E-01	1.2242	1.1719	3.0E+01	1.7402	1.7148
8.0E-01	1.2424	1.1878	3.5E+01	1.7499	1.7270
9.0E-01	1.2593	1.2027	4.0E+01	1.7574	1.7366
1.0E+00	1.2749	1.2167	5.0E+01	1.7686	1.7508
1.2E+00	1.3030	1.2423	6.0E+01	1.7765	1.7609
1.4E+00	1.3277	1.2654	7.0E+01	1.7824	1.7685
1.6E+00	1.3497	1.2862	8.0E+01	1.7870	1.7744
1.8E+00	1.3694	1.3052	9.0E+01	1.7906	1.7792
2.0E+00	1.3873	1.3226	1.0E+02	1.7936	1.7831
2.5E+00	1.4254	1.3606	1.5E+02	1.8030	1.7954
3.0E+00	1.4565	1.3923	2.0E+02	1.8080	1.8020
3.5E+00	1.4824	1.4193	3.0E+02	1.8134	1.8091
4.0E+00	1.5045	1.4427	4.0E+02	1.8161	1.8128
4.5E+00	1.5236	1.4631	5.0E+02	1.8179	1.8151
5.0E+00	1.5403	1.4812	1.0E+03	1.8215	1.8200
5.5E+00	1.5550	1.4974	2.0E+03	1.8235	1.8227
6.0E+00	1.5681	1.5119	5.0E+03	1.8248	1.8244
7.0E+00	1.5905	1.5369	1.0E+04	1.8252	1.8251
8.0E+00	1.6089	1.5579	5.0E+04	1.8256	1.8256
9.0E+00	1.6244	1.5757	1.0E+05	1.8257	1.8257

Following the ideas developed in [26, 30–32], we move from the *vector* equations (16) and (17) to the *matrix* equation

$$z\frac{\partial \mathbf{I}(\tau, z, z_0)}{\partial \tau} = \mathbf{I}(\tau, z, z_0) - \mathbf{S}(\tau, z_0) \qquad (26)$$

and

$$\mathbf{S}(\tau, z_0) = \frac{\lambda}{2}\int\limits_{-\infty}^{\infty} \mathbf{G}(z)\mathbf{I}(\tau, z, z_0)dz + \mathbf{E}e^{-\tau/z_0}, \quad (27)$$

where $\mathbf{I}(\tau, z, z_0)$ is a 6×6 matrix satisfying the boundary condition

$$\mathbf{I}(0, z, z_0) = \mathbf{0}, \quad z < 0, \qquad (28)$$

while $\mathbf{S}(\tau, z_0)$ is the corresponding 6×6 source matrix and $z_0 = \mu_0/\phi(x_0)$. In (26)–(28), we have explicitly indicated the dependence of all quantities on the *parameter* z_0.

A comparison of (16, 17) and (26, 27) taking into account (14) and (15) shows that the Stokes vector $\mathbf{i}(\tau, x, \mu, \varphi)$, which also depends on the three parameters x_0, μ_0, and φ_0, can be expressed in terms of the matrix $\mathbf{I}(\tau, z, z_0)$ as follows:

$$\mathbf{i}(\tau, x, \mu, \varphi) = \frac{\lambda}{4\pi}\mathbf{A}(\mu, \varphi)\mathbf{I}(\tau, z, z_0) \qquad (29)$$
$$\times\, \mathbf{A}^T(-\mu_0, \varphi_0)\phi(x_0)\mathbf{i}_0(x_0, \mu_0, \varphi_0),$$

where

$$z = \mu/\phi(x), \qquad z_0 = \mu_0/\phi(x_0). \qquad (30)$$

Note that equations (26), (27) have *precisely* the same form as the corresponding equations for the azimuth-independent component of the radiation field from [16]. Therefore, we can use all the apparatus laid out in that work.

It is easy to see that such an approach is also applicable for other initial sources (not necessarily with an exponential dependence on τ).

4. REFLECTED RADIATION

We are interested in the polarization of radiation reflected from a semi-infinite atmosphere. The matrix $\mathbf{I}(0, z, z_0)$ for the emergent radiation is related to the source matrix $\mathbf{S}(\tau, z_0)$ in the usual way

$$\mathbf{I}(0, z, z_0) = \int\limits_0^{\infty} \mathbf{S}(\tau, z_0)e^{-\tau/z}\frac{d\tau}{z}. \qquad (31)$$

It is shown in [16] (see also [31]) for the azimuth-independent component of the radiation field that the Stokes matrix $\mathbf{I}(0, z, z_0)$ can be expressed in terms of the matrix $\mathbf{I}(z)$, which is a generalization of the Ambartsumian–Chandrasekhar function H, as follows:

$$\mathbf{I}(0, z, z_0) = \frac{\mathbf{I}(z)\mathbf{I}^T(z_0)}{z + z_0}z_0. \qquad (32)$$

This expression remains valid in our case. All our manipulations formally repeat those presented in [16]. The only difference is that the dimension of the matrices is now 6×6 rather than 2×2. The matrix $\mathbf{I}(z)$ satisfies the nonlinear integral equation

$$\mathbf{I}(z)\left(\boldsymbol{\epsilon}^{1/2} + \frac{\lambda}{2}\int\limits_0^{\infty} \mathbf{I}^T(z')\mathbf{G}(z')\frac{z'dz'}{z + z'}\right) = \mathbf{E}, \quad (33)$$

where

$$\boldsymbol{\epsilon}^{1/2} = \mathrm{diag}\left\{\varepsilon_I^{1/2}, \varepsilon_Q^{1/2}, \varepsilon_Q^{1/2}, \varepsilon_Q^{1/2}, \varepsilon_Q^{1/2}, \varepsilon_Q^{1/2}\right\} \quad (34)$$

and

$$\varepsilon_I^{1/2} = \sqrt{1-\lambda}, \quad \varepsilon_Q^{1/2} = \sqrt{1 - 7/10W\lambda}. \quad (35)$$

Adopting $\tau = 0$ in (29) and using (32), we finally obtain

$$\mathbf{i}(0, x, \mu, \varphi) = \frac{\lambda}{4\pi}\mathbf{A}(\mu, \varphi)\frac{\mathbf{I}(z)\mathbf{I}^T(z_0)}{z + z_0} \times \mathbf{A}^T(-\mu_0, \varphi_0)\mu_0\mathbf{i}_0(x_0, \mu_0, \varphi_0). \quad (36)$$

Formula (36) can be used to find the Stokes vector of radiation reflected from a semi-infinite atmosphere without computing the radiation field inside the atmosphere (in particular, without the need to first find the source function from the transport equation). According to (36), the problem reduces to finding the matrix $\mathbf{I}(z)$ from (33).

Formula (36) clearly shows that our approach has led to an enormous simplification: to obtain the vector function $\mathbf{i}(0, x, \mu, \varphi)$, which depends on the *three variables* (x, μ, φ) *and three parameters* (x_0, μ_0, φ_0), it is sufficient to find the matrix $\mathbf{I}$, which depends on only *one* variable. We will see in the next section that the structure of $\mathbf{I}(z)$ is such that it is straightforward to obtain it from (33).

The remarkable separation of variables expressed by (36) has its roots first in our use of the approximation of a total redistribution in frequency [formula (1)] and second in the possibility of factorizing the phase matrix [formula (2)].

5. CALCULATION RESULTS

5.1. The Matrix $\mathbf{I}(z)$

Thus, in order to determine the intensity and polarization of reflected radiation, we must know the matrix $\mathbf{I}(z)$. In addition to satisfying the nonlinear integral equation (33), as is shown in [16] for the azimuth-independent 2×2 analog, this matrix also satisfies the linear integral equation

$$\mathbf{T}(z)\mathbf{I}(z) = \boldsymbol{\epsilon}^{1/2} - \frac{\lambda}{2}\int_0^\infty \mathbf{G}(z')\mathbf{I}(z')\frac{z'dz'}{z - z'}, \quad (37)$$

where the dispersion matrix $\mathbf{T}(z)$ is determined by the expression

$$\mathbf{T}(z) = \mathbf{E} - \lambda z^2 \int_0^\infty \frac{\mathbf{G}(z')dz'}{z^2 - z'^2}. \quad (38)$$

As follows from this last expression, the structure of the matrix $\mathbf{T}(z)$ is similar to that of $\mathbf{G}(z)$. This

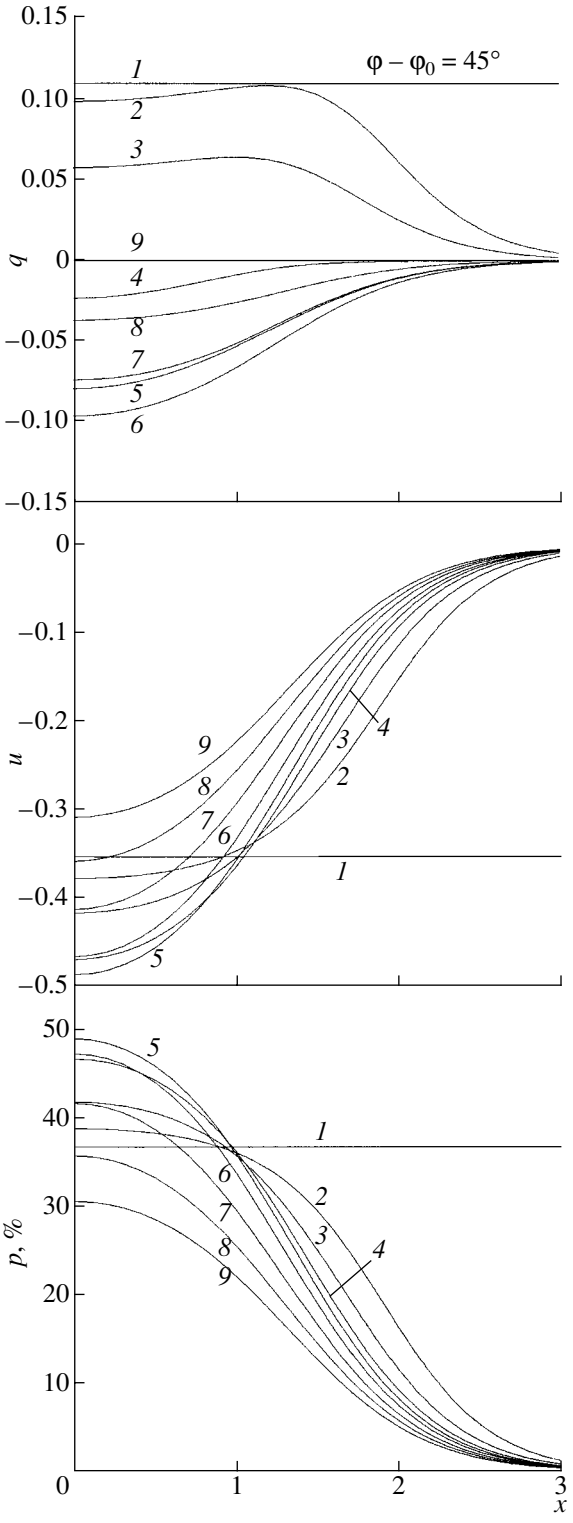


Fig. 1. Spectral profiles of the parameters q, u, and p of reflected radiation in the plane $\varphi - \varphi_0 = 45°$ for various angular distances from the zenith $\theta = \arccos\mu$. Unpolarized radiation at the central frequency is incident on the atmosphere at the angle $\arccos\mu_0 = 0.5$. The numbers next to the curves indicate the values of μ: (*1*) $\mu = 0$, (*2*) $\mu = 0.10$, (*3*) $\mu = 0.25$, (*4*) $\mu = 0.50$, (*5*) $\mu = 0.70$, (*6*) $\mu = 0.85$, (*7*) $\mu = 0.95$, (*8*) $\mu = 0.99$, (*9*) $\mu = 1.0$.

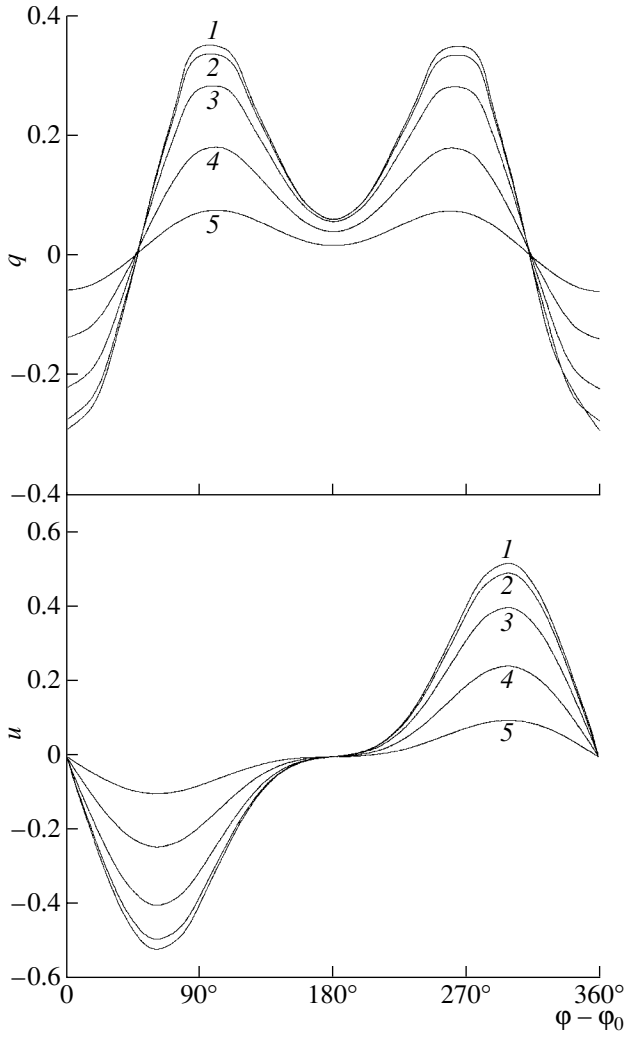


Fig. 2. Azimuth dependence of the parameters q and u for reflected radiation with $\arccos\mu = 0.5$ for several frequencies within a line. The numbers next to the curve indicate the frequencies x relative to the line center: (*1*) $x = 0$, (*2*) $x = 0.5$, (*3*) $x = 1.0$, (*4*) $x = 1.5$, (*5*) $x = 2.0$. The incident radiation is the same as in Fig. 1.

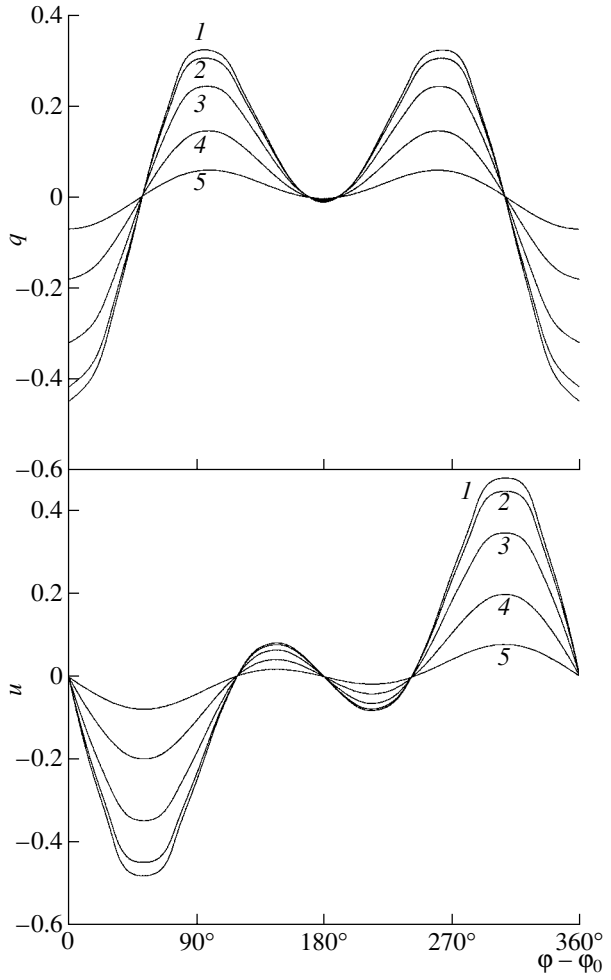


Fig. 3. Same as Fig. 2 for $\mu = 0.8$.

means that we can immediately conclude from (37) that the matrix $\mathbf{I}(z)$ has the same structure: only the eight elements indicated in (19) differ from zero, with $I_{33} = I_{44}$ and $I_{55} = I_{66}$. This structure of $\mathbf{I}$ (6×6) leads to radical simplifications. The task of finding this matrix separates into two parts. First, we must solve a 2×2 equation of the form (33) to obtain the elements $I_{ik}(z)$ with $i, k = 1, 2$. This was already done in earlier studies of the azimuth-averaged radiation fields [17]. See [17] for the values of these elements $I_{ik}(z)$ for the case of conservative resonance scattering ($\lambda = 1, W = 1$) with a Doppler profile. Second, we must solve two scalar equations of the form (33), namely,

$$I_{ii}(z)\left(\varepsilon_Q^{1/2} + \frac{\lambda}{2}\int_0^\infty I_{ii}(z')G_{ii}(z')\frac{z'dz'}{z+z'}\right) = 1, \quad i = 3, 5, \tag{39}$$

which does not present difficulties since the iteration converges rapidly even for $\lambda = 1$. The values $I_{33}(z)$ and $I_{55}(z)$ for conservative ($\lambda = 1$), dipolar ($W = 1$) scattering with a Doppler profile are presented in the Table.

In accordance with the structure of $\mathbf{I}(z)$ described above, this matrix can be represented in the block form

$$\mathbf{I}(z) = \begin{pmatrix} \mathbf{I}_0(z) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_1(z) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_2(z) \end{pmatrix}, \tag{40}$$

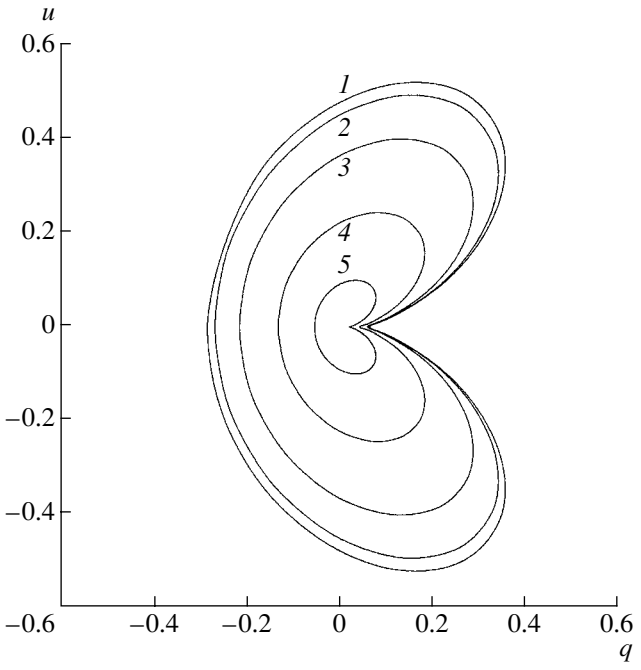


Fig. 4. Azimuthal "trajectories" in the $q-u$ plane. The parameter values are the same as in Fig. 2.

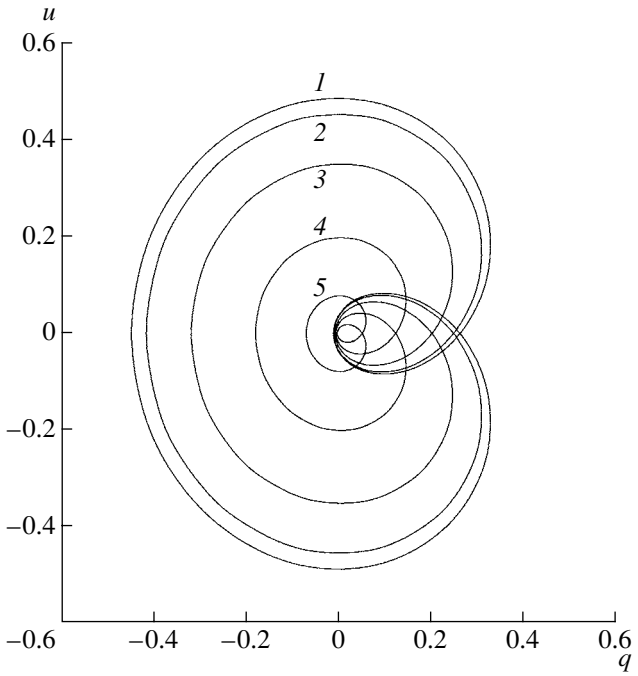


Fig. 5. Same as Fig. 4 for the parameters in Fig. 3.

where all the blocks are 2×2, $\mathbf{0}$ is the 2×2 zero matrix, and

$$\mathbf{I}_1(z) = I_{33}(z)\mathbf{E}, \quad \mathbf{I}_2(z) = I_{55}(z)\mathbf{E}. \tag{41}$$

Here, $\mathbf{E}$ is the 2×2 unit matrix.

Substituting (40) and (41) into (36) using (9) leads to a Fourier-harmonic expansion of $\mathbf{i}(0, x, \mu, \varphi)$, in accordance with the approach of Chandrasekhar:

$$\mathbf{i}(0, x, \mu, \varphi) = \frac{\lambda}{4\pi}\Bigg(\mathbf{A}_0(\mu)\frac{\mathbf{I}_0(z)\mathbf{I}_0^T(z_0)}{z+z_0}\mathbf{A}_0^T(-\mu_0) + \frac{I_{33}(z)I_{33}(z_0)}{z+z_0}\mathbf{P}_1(\mu,\varphi;-\mu_0,\varphi_0) + \frac{I_{55}(z)I_{55}(z_0)}{z+z_0}\mathbf{P}_2(\mu,\varphi;-\mu_0,\varphi_0)\Bigg)\mu_0\mathbf{i}_0(x_0,\mu_0,\varphi_0). \tag{42}$$

5.2. Polarization of Reflected Radiation

Because of the large number of parameters involved, it is not possible to present our results in detail in a form that is easy to take in. We therefore limit our presentation of results to certain characteristic examples.

Let unpolarized radiation at the central frequency of a line ($x_0 = 0$) be incident onto an atmosphere at the angle $\arccos \mu_0 = 0.5$, with pure scattering occurring in the line ($\lambda = 1$). Let us first consider the change in the relative Stokes parameters

$$q = \frac{Q}{I}, \quad u = \frac{U}{I}, \quad p = \frac{\sqrt{Q^2+U^2}}{I} \tag{43}$$

in the line frequencies for various μ at a fixed azimuth. Figure 1 gives an impression of the general character of the variations in the indicated parameters (p is expressed in percent). Note the nonmonotonic behavior of the polarization curves as μ is varied. It turns out that the parameter q can reach its maximum at some distance from the line center. Nevertheless, the polarization parameters tend to zero at sufficiently large distances from the line center (if $\mu \neq 0$). This is due to the fact that the radiation in the line wings comes from larger depths with larger values of x. At large depths, multiple scatterings isotropize the radiation field, suppressing polarization. When $\mu = 0$, the polarization is independent of frequency.

Figures 2 and 3 present the azimuth dependences of q and u for two values of μ. The symmetric behavior of q and antisymmetric behavior of u about the angle $\varphi - \varphi_0 = 180°$ is clearly visible. Note also the presence of secondary extrema in Fig. 3.

The "phase" portraits of the azimuthal behavior of the polarization in the $q-u$ plane are of some interest, and are presented in Figs. 4 and 5 in correspondence with Figs. 2 and 3. The appearance of secondary extremes leads to self-intersection of the curves in

Fig. 5. As $\varphi - \varphi_0$ grows from zero, motion along the curves begins from the extreme left point on the curves and continues in a clockwise fashion.

6. CONCLUSION

We have presented a new formalism for the analysis of the transport of polarized radiation taking into account the azimuth dependence during Rayleigh and resonance scattering. This formalism is based on factorization of the 3×3 phase matrix using a 3×6 matrix and its transpose matrix [formulas (2) and (3)]. This approach makes it possible to represent the transport equation in matrix form and fully make use in the general (azimuth dependent) case of the apparatus developed earlier for the azimuth-independent component of the radiation field. We have investigated the azimuth dependence of the polarization of reflected radiation.

ACKNOWLEDGMENTS

I thank S.I. Grachev for useful discussions and V.V. Ivanov for attention to this work and support. This work was partially financially supported by the Russian Foundation for Basic Research (project no. 00-02-16870), the program "Leading Scientific Schools of Russia" (project no. 00-15-96607), and the Federal Scientific and Technology Program "Astronomy" (theme "Studies of Energy-Generation Mechanisms in Astrophysical Objects and Behavior of Matter under Extreme Conditions").

REFERENCES

1. J. O. Stenflo, C. U. Keller, and A. Gandorfer, Astron. Astrophys. **355**, 789 (2000).
2. A. Gandorfer, *The Second Solar Spectrum* (Hochshulverlag AG ETH, Zurich, 2000).
3. V. Bommier and G. Molodij, Astron. Astrophys. **381**, 241 (2002).
4. R. Kurosawa, D. J. Hillier, and J. M. Pittard, Astron. Astrophys. **388**, 957 (2002).
5. G. W. Collins and S. R. Cranmer, Mon. Not. R. Astron. Soc. **253**, 167 (1991).
6. R. Antonucci, Ann. Rev. Astron. Astrophys. **34**, 222 (1993); astro-ph/0103048 (2001).
7. S. Chandrasekhar, *Radiative Transfer* (Oxford Univ. Press, Oxford, 1950; Inostr. Lit., Moscow, 1953).
8. V. V. Sobolev, *Radiation Transfer in Stellar and Planetary Atmospheres* (GITTL, Moscow, 1956) [in Russian].
9. V. V. Ivanov, *Radiative Transfer and the Spectra of Celestial Bodies* (Nauka, Moscow, 1969) [in Russian].
10. J. O. Stenflo, Astron. Astrophys. **46**, 61 (1976).
11. S. Dumont, A. Omont, J. C. Pecker, and D. Rees, Astron. Astrophys. **54**, 675 (1977).
12. D. E. Rees, Publ. Astron. Soc. Jpn. **30**, 455 (1978).
13. M. Faurobert, Astron. Astrophys. **194**, 268 (1988).
14. M. Faurobert-Sholl and H. Frisch, Astron. Astrophys. **219**, 338 (1989).
15. V. V. Ivanov, Astron. Zh. **67**, 1223 (1990) [Sov. Astron. **34**, 621 (1990)].
16. V. V. Ivanov, S. I. Grachev, and V. M. Loskutov, Astron. Astrophys. **318**, 315 (1997).
17. V. V. Ivanov, S. I. Grachev, and V. M. Loskutov, Astron. Astrophys. **321**, 968 (1997).
18. M. Faurobert-Sholl, H. Frisch, and K. N. Nagendra, Astron. Astrophys. **322**, 896 (1997).
19. H. Domke and I. Hubeny, Astrophys. J. **334**, 527 (1988).
20. E. Landi Degl'Innocenti, Solar Phys. **102**, 1 (1985).
21. S. I. Grachev, Astron. Zh. **78**, 1092 (2001) [Astron. Rep. **45**, 960 (2001)].
22. H. Frisch, Astron. Astrophys. **338**, 683 (1998).
23. M. Faurobert-Scholl, Astron. Astrophys. **246**, 469 (1991).
24. Z. Sekera, in *Electromagnetic Scattering*, Ed. by M. Kerker (Pergamon, Oxford, 1963), p. 547.
25. H. C. van de Hulst, *Multiple Light Scattering* (Academic, New York, 1980).
26. V. V. Ivanov, Astron. Astrophys. **303**, 609 (1995).
27. D. N. Rachkovskiĭ, Izv. Krym. Astrofiz. Obs. **67**, 78 (1983).
28. H. Frisch, in *Proc. of the Second Solar Polarization Workshop*, Ed. by K. N. Nagendra and J. O. Stenflo (Kluwer, Dordrecht, 1999), pp. 97–113.
29. M. A. Heaslet and R. F. Warming, J. Quant. Spectrosc. Radiat. Transf. **8**, 1101 (1968).
30. V. V. Ivanov, A. M. Kasaurov, V. M. Loskutov, and T. Viik, Astron. Astrophys. **303**, 621 (1995).
31. V. V. Ivanov, Astron. Astrophys. **307**, 319 (1996).
32. V. V. Ivanov, A. M. Kasaurov, and V. M. Loskutov, Astron. Astrophys. **307**, 332 (1996).

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 29–34. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 33–38.

Excitation of PtI Transitions Ending at the Levels of the Main Term

Yu. M. Smirnov

Moscow Power Engineering Institute, ul. Krasnokazarmennaya 17, Moscow, 111250 Russia

Received March 20, 2003; in final form, May 8, 2003

Abstract—We have experimentally studied the excitation of transitions of the platinum atom ending in levels of the ^3D main term. The atoms were excited by a 30 eV monoenergetic electron beam. The lines studied are located at wavelengths of 204–367 nm. The largest of the measured excitational cross sections exceed 10^{-16} cm^2. The optical-excitation functions measured at electron energies of 0–200 eV have complex structures. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The study of the spectrum of the platinum atom and its atomic constants is of great interest for both physics (first and foremost, for atomic-structure theory and atomic spectroscopy) and astrophysics. In particular, stars exhibiting anomalously high platinum abundances have recently been discovered [1]. The interpretion of the spectra of these stars requires accurate and reasonably complete data on the energy levels of atomic platinum and the radiative transitions of PtI. This applies to a considerable degree to transitions in the EUV part of the spectrum with wavelengths $\lambda < 300$ nm. Astronomical observations in this part of the spectrum became possible only fairly recently owing to the emergence of extra-atmospheric spectroscopic instruments, and have proved to be very fruitful.

Over several decades, a major source of information about the spectrum of atomic platinum has been the work of Livingood [2], who performed a detailed analysis of the UV spectrum of PtI and interpreted almost 80 energy levels of this atom. In addition, Corliss and Bozeman [3] experimentally determined the transition probabilities for 160 spectral lines of PtI. These two works were based on photographic spectra.

Studies of the spectrum and atomic constants of atomic platinum intensified in the last quarter of the twentieth century, along with the development of new spectroscopic instruments and the emergence of new theoretical methods. Ramanujam and Andersen [4] used the method of delayed coincidences to record the optical signal in a photon-counting mode in order to measure the radiative lifetimes of four energy levels of atomic platinum. Platinum atoms in the gaseous phase were obtained by dispersing the metal by bombarding it with argon ions with energies of 200–500 keV. The data obtained agree well with those computed using transition probabilities adopted from [3].

Lotrian and Guern [5] used hollow-cathode discharges to determine the branching factors of nine PtI levels belonging to the $5d^8 6s6p$ and $5d^9 6p$ configurations. They computed the transition probabilities for 15 spectral lines using values of A_{ki} determined from the radiative lifetimes of three PtI energy levels obtained earlier by Ramanujam and Andersen [4]. Somewhat later, Gough *et al.* [6] published the radiative lifetimes of 15 odd levels of PtI; the new values for the four levels analyzed earlier by Ramanujam and Andersen [4] differ from those previously computed by less than a factor of 1.7. The branching factors determined by Lotrian and Guern [5] were used to compute the transition probabilities for six spectral lines with the common upper level 44 444 cm^{-1}.

Engleman [7] refined the energies of 30 even and 52 odd levels of PtI using a hollow-cathode lamp as a source of radiation and a Fourier spectrometer. The new energy levels are given with eight decimal digits; the configurations and terms for 17 even and one odd level are given, as well as configurations alone for another 27 levels. All these parameters are listed in accordance with the compilation [8].

Sansonetti *et al.* [9] made the next important step in the study of the spectrum and energy levels of PtI. Using photographic and photoelectric methods, they recorded a normal-incidence vacuum spectrum, measured the wavelengths and intensities of 5600 lines with wavelengths of 113–433 nm, and published an atlas of the spectrum. Photographic and photoelectric recording techniques enable wavelength measurements with errors of ± 0.00020 and ± 0.001 nm, respectively, for $\lambda < 203$ nm, and ± 0.002 nm at longer wavelengths. The errors of the measured relative intensities are estimated to be 20%.

1063-7729/04/4801-0029$26.00

In their paper published simultaneously with Sansonettin *et al.* [9], Blaise *et al.* [10] performed an extensive theoretical analysis of the energy levels of atomic platinum based on spectroscopic data obtained by three groups, including the works of Engleman [7] and Sansonetti *et al.* [9]. The tables presented by Blaise *et al.* [10] contain twice as many levels as those of Engleman [7], a different configuration for one level, and additional terms for four levels. These differences may be due to the fact that the levels of PtI are characterized by strong mixing of configurations, which is difficult to adequately reflect in terms of the fairly simple symbolism of level notation. A similar situation exists for the odd levels, the list of which was expanded by nearly a factor of two compared to the results of [7] and, moreover, for all levels below 65 000 cm^{-1} designated by terms that are absent from [7]. The configuration designations were changed for three levels.

In spite of the fact that important progress has been achieved in studies of the spectrum and energy levels of PtI and that data of satisfactory accuracy (although limited quantities) have been obtained on the transition probabilities and radiative lifetimes of the platinum atom, it is striking that there is a complete lack of information about the collisional properties of atomic platinum. These properties are necessary when applying coronal models, and in general for the solution of non-LTE problems, which are being encountered more and more frequently with the development of theoretical astrophysics, especially in studies of nonstationary objects. In this paper, we analyze for the first time the excitation of atomic platinum by monoenergetic electrons with energies of 0–200 keV, which is most important for problems in the physics of low-temperature plasmas.

2. SETUP OF THE EXPERIMENT AND CONDITIONS UNDER WHICH IT WAS CARRIED OUT

We used the method of extended intersecting beams, for which there is currently no alternative in studies of the excitation of atoms with intermediate and high evaporation temperatures ($T_{ev} > 1500$ K). Detailed discussions of this method and its technical implementation can be found in [11–13].

We evaporated 99.93%-pure platinum (GOST 18389-73) from an automatic crucible oven. The metal was in the form of bars 3 mm in diameter and about 30 mm in length, and was placed onto a graphite base whose design ensured efficient thermal decoupling of the evaporated material and cooled elements of the apparatus. The platinum was heated by an electron beam. When the temperature of the molten surface was 2260 K, the density of platinum atoms in the region where the atomic and electron beams intersected reached 5.4×10^9 cm^{-3}. The metal had to be substantially overheated above the melting point, since the pressure of saturated platinum vapor at the melting point is only 0.01 Pa. The geometry of the atomic beam in the region where it intersected with the electron beam was determined by three consecutive diaphragms cooled by running water.

In most cases, the evaporation of metals is accompanied by thermal population of the lower levels. The role of this factor can be neglected only for metals with very low evaporation temperatures, or those having an isolated ground level. However, the evaporation temperature of platinum is comparatively high and, moreover, this element has several low levels that can be thermally populated with high efficiency. The excitational cross sections (both the absolute cross sections and their dependence on the electron energy) are determined to a considerable degree by the initial state from which the level considered is excited. Therefore, to enable a correct comparison of the experimental data with theoretical results (once the latter become available), experimental studies must take into account the distribution of the atoms over the initial states.

The emergence of an atom from the surface of a molten metal into vacuum (evaporation) is a difficult problem to analyze rigorously. Here, we restrict our analysis to a simple estimation of the populations of low levels of the evaporated platinum atoms assuming that they follow a Boltzmann distribution. Under the experimental conditions indicated above, our estimates yield the following populations for the total atom density in the beam (the numbers in parentheses give the energy of the levels in cm^{-1}): $5d^9 6s^3 D_3$ (0)—45.2%, $5d^9 6s^1 D_2$ (775)—19.7%, $5d^8 6s^2\ ^3F_4$ (823)—34.4%, $5d^{10}\ ^1S_0$ (6140)—0.13%, and $5d^9 6s^3 D_2$ (6567)—0.50%. The closest upper level has an energy higher than 10 000 cm^{-1}, and its population is negligibly small. It follows from these estimates that an experiment with platinum excitation actually involves three levels including the ground level, whose populations differ only slightly. Currently available experimental instruments do not enable separate estimation of the contributions of electron-collisional excitation for each of the three levels, and these contributions can be separated only via a theoretical analysis of the processes studied.

Other important experimental factors include the spectral resolution of the instrument, which is about 0.1 nm for sufficiently intense lines (taking into account the speed of the recording system). The width of the energy distribution of the electron beam is 0.9 eV for an electron energy of 100 eV and 1.0 eV for electron energies of 20 and 200 eV (for 90% of all

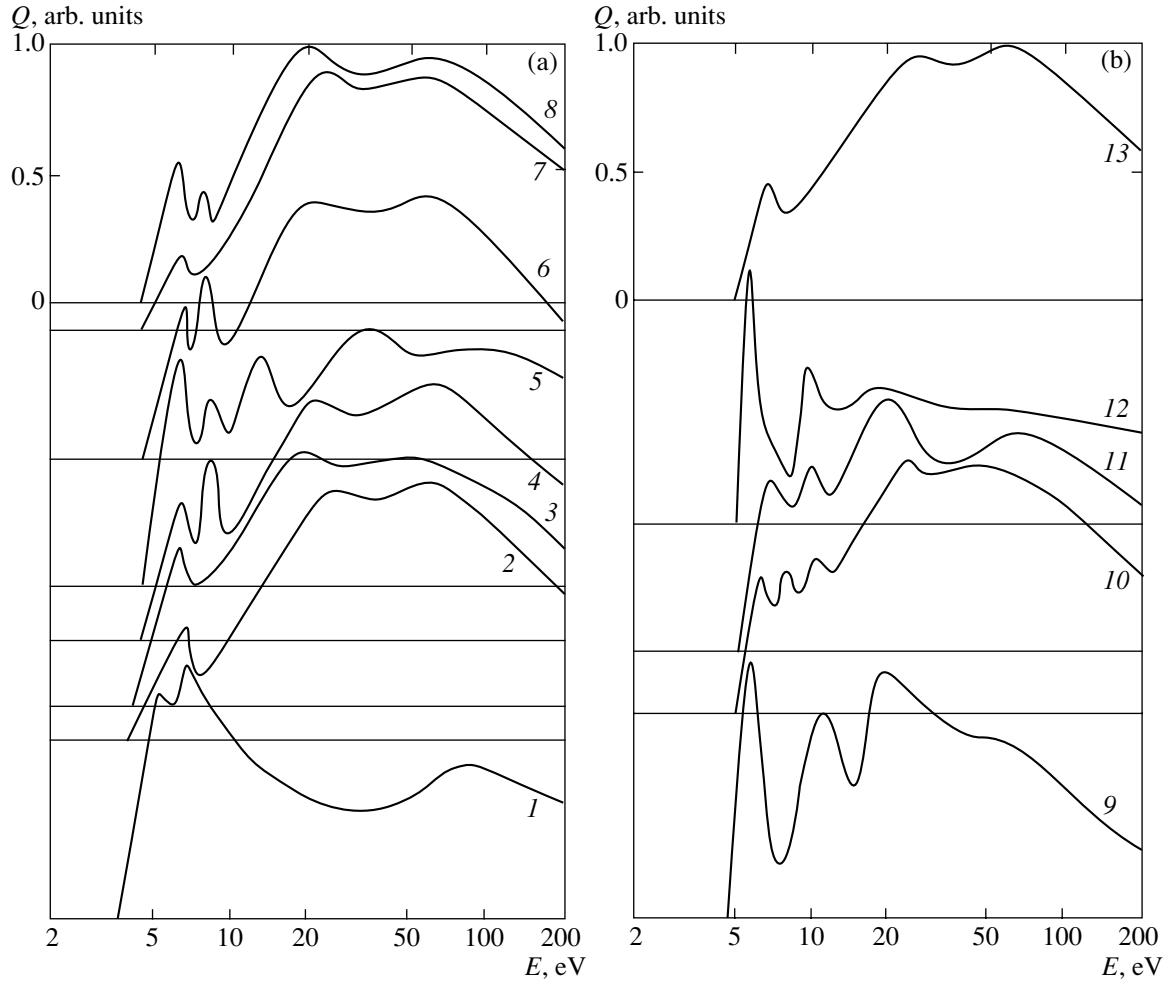


Fig. 1. Optical excitation functions of atomic platinum.

electrons). The current density of the electron beam did not exceed 1.0 mA/cm^2, and we smoothed the potential relief to drastically decrease the effect of the negative volume charge (especially for electron energies lower than 15 eV) [14].

The error of the fractional cross sections was 5–20%. The highest errors correspond to low-intensity lines and to all lines with wavelengths shorter than 230 nm, since the sensitivity of the instrument decreases substantially toward the vacuum ultraviolet. We determined the absolute cross sections with accuracies of ±18 to ±33%. One of the important factors decreasing the errors of the derived absolute excitational cross sections of platinum is the high inertness of this element and the absence of chemical processes in the evaporated metal film.

3. RESULTS AND DISCUSSION

The spectrum of platinum excited by monoenergetic 30-eV electrons was recorded from 197 to 417 nm. For about half of all the lines studied, we measured the dependence of the cross sections on the energy of the exciting electrons, i.e., their optical excitation functions (OEFs), for electron energies varying from the excitation threshold to 200 eV.

The ground state of the platinum atom—the $5d^9 6s^3 D$ triplet—exhibits considerable splitting, like the other low-lying $5d^8 6s^2\,{}^3F$ triplet state. The $5d^9 6s^3 D_3$ ground level is separated from other levels of the same term with $J = 2$, 1 by intervals of 6567 and 10 131 cm^{-1}, respectively. We analyzed the excitation of transitions ending at levels of the main term.

The table summarizes our results together with the necessary spectroscopic information. The table gives the wavelength λ, the transition, the inner quantum number J, and the energies of the lower ($E_{\rm low}$) and upper ($E_{\rm up}$) levels according to [9]. The following columns give the excitational cross sections for an electron energy of 30 eV Q_{30} and the maximum of the OEF $Q_{\max}$, the position of the maximum

Excitational cross sections of atomic platinum

λ, nm	Transition	J	$E_{\rm low}$, $\rm cm^{-1}$	$E_{\rm up}$, $\rm cm^{-1}$	Q_{30}, $10^{-18}\,\rm cm^2$	$Q_{\max}$, $10^{-18}\,\rm cm^2$	$E(Q_{\max})$, eV	OEF number
1	2	3	4	5	6	7	8	9
204.939	$5d^96s^3D-5d^86s6p^3D^\circ$	3–3	0	48779	30.1	–	–	–
207.094	$5d^96s^3D-5d^86s6p^3D^\circ$	2–3	6567	54839	11.0	–	–	–
212.859	$5d^96s^1D-5d^96p^1P^\circ$	2–1	775	47740	49.0	–	–	–
212.863	$5d^96s^3D-5d^86s6p^5D^\circ$	3–4	0	46963				
216.521	$5d^96s^3D-5d^96p^3F^\circ$	3–2	0	46170	13.8	–	–	–
224.931	$5d^96s^3D-5d^86s6p^5F^\circ$	3–2	0	44444	10.3	–	–	–
232.611	$5d^96s^3D-5d^96p^3D^\circ$	2–1	6567	49544	13.4	–	–	–
234.018	$5d^96s^3D-5d^86s6p^5P^\circ$	2–3	6567	49286	12.2	–	–	–
244.006	$5d^96s^3D-5d^86s6p^5G^\circ$	3–3	0	40970	32.0	34.1	60	13
245.097	$5d^96s^3D-5d^86s6p^5P^\circ$	3–2	0	40787	2.75	4.80	5.6	12
246.740	$5d^96s^3D-5d^86s6p^5D^\circ$	3–2	0	40516	9.07	12.1	20	11
248.717	$5d^96s^1D-5d^86s6p^5G^\circ$	2–3	775	40970	46.5	50.0	23	10
	$5d^96s^3D-5d^86s6p^5F^\circ$	3–4	0	40194				
249.581	$5d^96s^3D-5d^86s6p^3D^\circ$	2–3	6567	46622	6.67	–	–	–
250.850	$5d^96s^3D-5d^86s6p^5D^\circ$	2–2	6567	46419	4.46	–	–	–
251.503	$5d^96s^3D-5d^86s6p^5P^\circ$	1–2	10131	49880	5.60	–	–	–
252.431	$5d^96s^3D-5d^96p^3F^\circ$	2–2	6567	46170	6.35	–	–	–
253.649	$5d^96s^3D-5d^86s6p^5P^\circ$	1–1	10131	49544	6.17	–	–	–
260.314	$5d^96s^3D-5d^86s6p^5G^\circ$	1–2	10131	48535	10.6	–	–	–
261.957	$5d^96s^3D-5d^86s6p^3D^\circ$	2–3	6567	44730	2.73	–	–	–
263.935	$5d^96s^3D-5d^86s6p^5F^\circ$	2–2	6567	44444	7.83	–	–	–
264.688	$5d^96s^3D-5d^96p^3D^\circ$	3–3	0	37769	14.9	16.8	20	8
265.945	$5d^96s^3D-5d^96p^3F^\circ$	3–4	0	37590	160.	169.	23	7
267.457	$5d^96s^3D-5d^86s6p^5P^\circ$	2–3	6567	43945	5.67	–	–	–
267.715	$5d^96s^3D-5d^96p^3P^\circ$	3–2	0	37342	13.5	14.1	60	6
272.991	$5d^96s^3D-5d^86s6p^5D^\circ$	2–1	6567	43187	4.20	–	–	–
275.385	$5d^96s^3D-5d^86s6p^5D^\circ$	1–0	10131	46433	8.13	–	–	–
275.491	$5d^96s^3D-5d^86s6p^5D^\circ$	1–2	10131	46419				
277.400	$5d^96s^3D-5d^96p^3F^\circ$	1–2	10131	46170	3.51	–	–	–
283.029	$5d^96s^3D-5d^86s6p^5D^\circ$	3–3	0	35321	62.0	69.7	60	4
290.590	$5d^96s^3D-5d^86s6p^5G^\circ$	2–3	6567	40970	3.34	3.56	60	13
291.225	$5d^86s^{23}F-5d^86s6p^5F^\circ$	3–2	10116	44444	8.73	–	–	–
291.354	$5d^96s^3D-5d^86s6p^5F^\circ$	1–2	10131	44444				
292.138	$5d^96s^3D-5d^86s6p^5P^\circ$	2–2	6567	40787	2.77	4.81	5.6	12
292.979	$5d^96s^3D-5d^96p^3F^\circ$	3–3	0	34122	39.8	42.5	18	3
302.429	$5d^96s^3D-5d^86s6p^5D^\circ$	1–1	10131	43187	1.27	–	–	–
306.471	$5d^96s^3D-5d^96p^3P^\circ$	3–2	0	32620	139.	146.	60	2
310.003	$5d^96s^3D-5d^86s6p^3F^\circ$	2–2	6567	38815	5.02	6.31	5.6	9
315.656	$5d^96s^3D-5d^86s6p^3D^\circ$	1–1	10131	41802	8.11	–	–	–
320.404	$5d^96s^3D-5d^96p^3D^\circ$	2–3	6567	37769	11.1	12.5	20	8
325.198	$5d^96s^3D-5d^86s6p^3P^\circ$	1–0	10131	40873	2.95	–	–	–
329.022	$5d^96s^3D-5d^86s6p^5D^\circ$	1–2	10131	40516	3.72	4.96	20	11
330.186	$5d^96s^3D-5d^96p^3P^\circ$	2–1	6567	36844	12.3	12.9	36	5
331.504	$5d^96s^3D-5d^86s6p^5D^\circ$	3–4	0	30156	3.41	7.84	6.8	1
348.526	$5d^96s^3D-5d^86s6p^3F^\circ$	1–2	10131	38815	8.17	10.2	5.6	9
367.405	$5d^96s^3D-5d^96p^3P^\circ$	1–2	10131	37342	2.45	2.55	60	6

$E(Q_{\max})$, and the number of the OEF according to the numbering of the curves in Fig. 1. The layout of Fig. 1 is appropriate for atomic OEFs, with a logarithmic scale along the horizontal axis and a linear scale along the vertical axis. Each curve is normalized to unity at its maximum and has its own zero level on the vertical axis to avoid intersection of the different curves.

The table contains four unresolved pairs of lines. In two pairs, the separations between the lines are about 0.1 nm, but these lines have low intensities and are barely resolved by our instruments. Another pair of lines has a separation of only 0.004 nm, which makes it virtually impossible to resolve them in experiments aimed at studies of electron–atom collisions. Finally, the fourth pair of transitions merge into a single line, since their wavelengths coincide in all known digits. This line (248.717 nm) corresponds to two transitions from quintet levels. One of these transitions ends at a triplet level and the other at a singlet level.

Figure 2 shows a partial diagram of the states of atomic platinum with the studied transitions. We can see that these transitions can be divided into two equal sets: transitions occurring in the triplet system of terms and intercombinational transitions. The latter are represented exclusively by transitions from quintet levels; we recorded no transitions from a singlet level, although there is one $5d^96s^1D_2$ singlet with a population of about 20% for excitation under the conditions of our experiment. It is possible that the combination of quintet levels with levels of the triplet main term is determined by the considerable mixing of the quintet configurations. Although Blaise *et al.* [10] present only the leading components without specifying the impurities and their percentages in the composition, the more detailed version of their paper [15] gives all the necessary information. It follows from the data in [15] that the percentage comprised by the leading component exceeds 40% in only three quintet levels among those studied here; this value is lower than 30% for eight quintet levels and lower than 20% for four levels. The lower levels of the transitions considered exhibit an almost opposite pattern: the leading component comprises almost 100% of the $5d^96s^3D_{1,3}$ levels and 42.4% of the $5d^96s^3D_2$ level.

One of the transitions studied occurs as a result of excitation of the $5d^97p^3P_2^\circ$ level, whereas the remaining triplet levels belong to the $5d^96p$ and $5d^86s6p$ configurations. Spontaneous transitions from the levels of the first and second groups involve the electron transitions $6p \to 6s$ and $6p \to 5d$, respectively. Both these transitions are allowed and, moreover, the upper levels of all the studied transitions in the triplet zone are P°, D°, and F° levels, so that their combinations

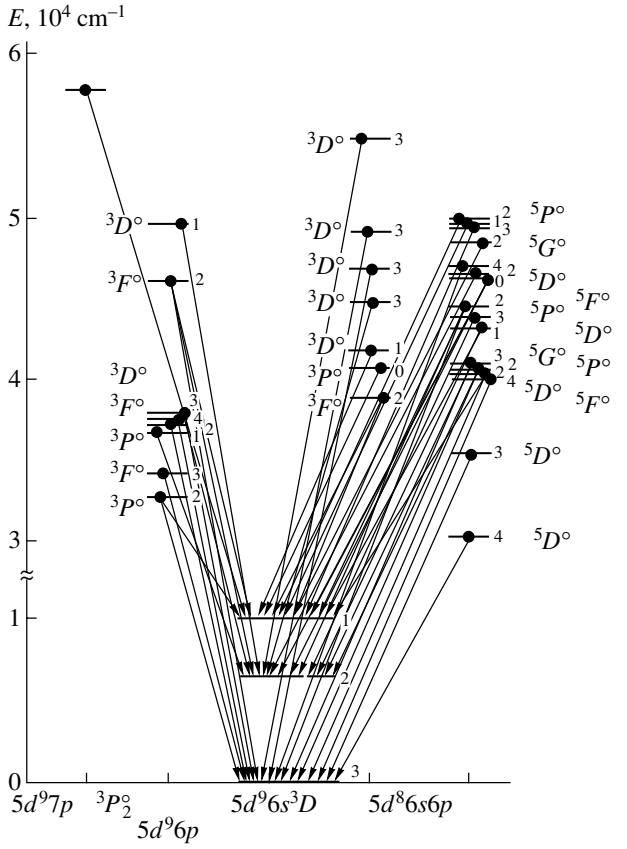


Fig. 2. Partial diagram of the states of PtI with the transitions studied. The small numbers to the right of the levels give the J values.

with the ground-term level correspond to the selection rule $\Delta L = 0, \pm 1$. These are allowed transitions and therefore have comparatively large excitational cross sections, which exceed 1×10^{-16} cm^2 for two resonance transitions.

In all the quintet levels listed in the table, the leading components belong to the $5d^86s6p$ configuration. When combined with the levels of the main term, the allowed transition $6p \to 5d$ occurs, so that the excitational cross sections of these transitions are only slightly smaller than those for transitions in the triplet zone of the terms, despite the change in the multiplet order. We found three transitions from the $^5G^\circ$ level (apart from the 248.717 nm line) for which $\Delta L = -2$, and the cross sections of two transitions exceed 1×10^{-17} cm^2.

The OEFs all have more or less well-defined complex structures. It is clear that many OEFs contain identical elements, and OEFs 2, 3, 7, and 13; 4 and 6; and 8, 9, and 12 are very similar both in terms of their general form and the ratios of the amplitudes of each local maximum. This structure of the OEFs can be interpreted only by analyzing the excitation

of PtI using methods based in the modern theory of electron–atom collisions.

4. CONCLUSIONS

We have analyzed for the first time the excitation of atomic platinum by monoenergetic electrons. The excitational cross sections of the most intense transitions exceed 1×10^{-16} cm^2, and the differences between the cross sections for allowed transitions in the triplet system of terms and for intercombinational transitions are small. All the measured OEFs have complex structures with several local maxima. The results obtained can be applied in non-LTE problems in the physics of stellar atmospheres, including studies of stars with anomalously high platinum abundances.

REFERENCES

1. D. S. Leckrone, G. M. Wahlgren, and S. G. Johansson, Astrophys. J. **377**, L37 (1991).
2. J. J. Livingood, Phys. Rev. **34**, 185 (1929).
3. C. H. Corliss and W. R. Bozeman, *Experimental Transition Probabilities for Spectral Lines of Seventy Elements* (Government Prin. Off., Washington, 1962; Mir, Moscow, 1968).
4. P. S. Ramanujam and T. Andersen, Astrophys. J. **226**, 1171 (1978).
5. J. Lotrian and Y. Guern, J. Phys. B **15**, 69 (1982).
6. D. S. Gough, P. Hannaford, and R. M. Lowe, J. Phys. B **15**, L431 (1982).
7. R. Engleman, Jr., J. Opt. Soc. Am. B **2**, 1934 (1985).
8. C. E. Moore, *Atomic Energy Levels*, NBS Circ. No. 467 (US Govt. Printing, Washington, 1958), Vol. 3.
9. J. E. Sansonetti, J. Reader, C. J. Sansonetti, and N. Acquista, J. Res. of NIST **97**, 1 (1992).
10. J. Blaise, J. Verges, J.-F. Wyart, and R. Engleman, Jr., J. Res. of NIST **97**, 213 (1992).
11. Yu. M. Smirnov, *Physics of Electron and Atomic Collisions* (Fiz. Tekh. Inst. Akad. Nauk SSSR, Leningrad, 1985) [in Russian].
12. Yu. M. Smirnov, J. Phys. (France) **4**, 23 (1994).
13. Yu. M. Smirnov, Phys. Scripta **49**, 689 (1994).
14. A. N. Kuchenev, E. A. Samsonova, and Yu. M. Smirnov, Avtometriya, No. 5, 109 (1990).
15. J. Blaise, J. Verges, J.-F. Wyart, and R. Engleman, Jr., J. Phys. (France) **2**, 947 (1992).

Translated by A. Dambis

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 35–38. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 39–42.

Identification of Gamma-ray Sources with Wolf–Rayet Stars

A. M. Gal'per and B. I. Luchkov

Moscow Physical Engineering Institute, Moscow, Russia

Received February 20, 2003; in final form, May 8, 2003

Abstract—An analysis of unidentified discrete sources of gamma-rays with energies E > 100 MeV demonstrates that the spatial characteristics of this group of gamma-ray sources coincides with those of Wolf–Rayet stars. It is concluded that Wolf–Rayet stars are potential steady sources of high-energy gamma rays with mean luminosities $L(>100\ \mathrm{MeV}) \approx 10^{35}$ erg/s. © *2004 MAIK "Nauka/Interperiodica".*

More than one hundred unidentified sources of gamma-rays with energies E > 100 MeV are currently known [1]. These are point-like gamma-ray sources at unknown distances for which the accuracy in their positions ($\approx 30'$) is not sufficient to enable firm identification with known astrophysical objects. The most complete data on unidentified gamma-ray sources have been obtained by the EGRET telescope of the orbiting COMPTON Gamma-Ray Observatory [2].

We have investigated the spatial distribution of groups of unidentified gamma-ray sources and compared these with the distributions of active stars that are possible sources of gamma-ray emission. The first results of this work were published in [3]. The current paper contains more complete and detailed results of our analysis.

It has been pointed out that unidentified gamma-ray sources do not represent a single group of objects [3]. The distributions on the celestial sphere are appreciably different for "bright" (fluxes $F > 5 \times 10^{-7}$ photons/cm^2 s) and "weak" ($F < 5 \times 10^{-7}$ photons/cm^2 s) sources. The bright gamma-ray sources are concentrated primarily toward the Galactic plane (Fig. 1), demonstrating that they are fairly distant—at distances $r > 1$ kpc; while the distribution of "weak" gamma-ray sources is essentially isotropic, suggesting that they are comparatively close objects ($r < 100$ pc). We cannot rule out the possibility that there are some extragalactic objects among the unidentified gamma-ray sources, but their distributions on the celestial sphere—with enhanced source densities toward the Galactic center and anticenter in the case of the weak sources and a concentration toward the Galactic plane in the case of the bright sources—indicate that the fraction of such objects is small. Objects with such high-energy gamma-ray radiation could be active stars on (or near) which particle acceleration occurs, accompanied by the generation of gamma-rays via electron bremsstrahlung and the decay of neutral pions formed by protons in nuclear collisions.

Our analysis below is concerned only with the bright unidentified gamma-ray sources.

As possible stellar counterparts of the unidentified gamma-ray sources, we consider Wolf–Rayet (WR) stars, which are suitable candidates in terms of their activity and spatial distribution [4]. They are also concentrated toward the Galactic plane, being distant Galactic objects. WR stars are massive stars in the final stage of their evolution before their explosion as supernovae (their mean age is 4×10^5 yrs) [5]. WR stars are born in the spiral arms of the Galaxy, where there is a relatively high concentration of gas, facilitating star formation.

We have used a qualitative test of the possible relationship between the unidentified gamma-ray sources and WR stars. These stars avoid certain intervals of Galactic longitude: $l = 140$–$220°$, 30–50°, 250–260°, and 270–280° (Fig. 2), due to the structure of the spiral arms near the Sun (Fig. 3). For example, the largest "empty" longitude interval, 140–220°, is a sector in which there is no spiral arm in the line of sight from the Earth (there is a discontinuity in the Orion spiral arm), so that no WR stars have formed in this direction. If the unidentified gamma-ray sources are associated with WR stars, they should avoid the same "forbidden" longitude intervals.

Figure 4 shows the longitude distributions of the studied objects. We can see that both the WR stars and unidentified gamma-ray sources avoid the longitude interval $l = 140$–$220°$. There are likewise no unidentified gamma-ray sources in other "forbidden" intervals. Thus, the spatial correlation of the distributions of WR stars and unidentified gamma-ray

1063-7729/04/4801-0035$26.00

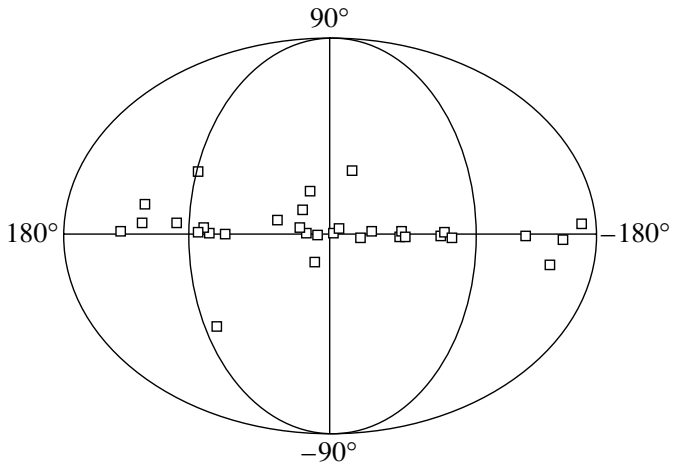


Fig. 1. Map of the distribution of bright (fluxes $> 5 \times 10^{-7}$ photons/cm^2 s) unidentified gamma-ray sources (E $>$ 100 MeV) in Galactic coordinates.

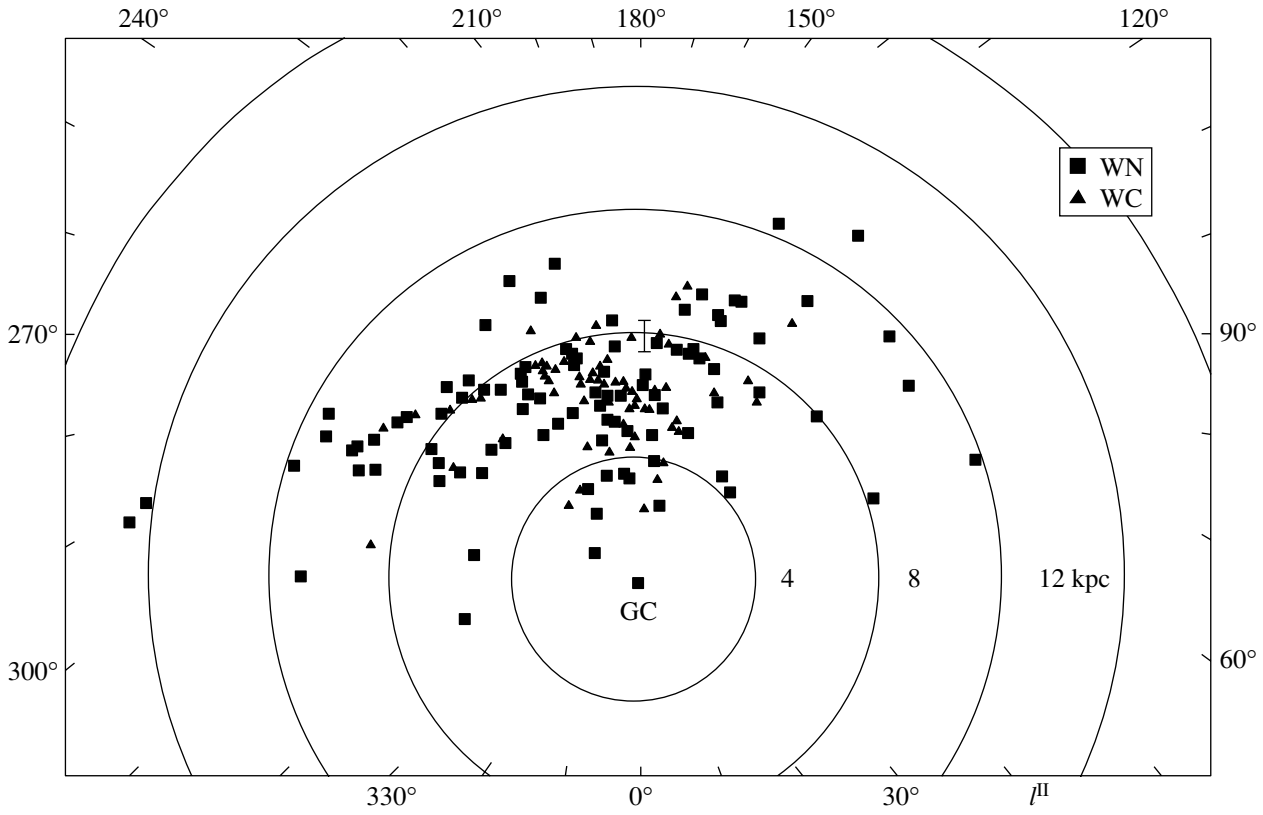


Fig. 2. Distribution of WR stars in projection onto the Galactic plane (coordinates l, b).

sources is fairly obvious, supporting the hypothesis that they are related. At the very least, we can conclude that the counterparts of the unidentified gamma-ray sources and the WR stars represent similar stellar populations that are located in the same areas of the Galaxy.

We obtained a more rigorous demonstration of the identification of unidentified gamma-ray sources with WR stars by comparing their coordinates. When performing this comparison, we must bear in mind, first, that the coordinate error boxes for the gamma-ray sources are much larger than the positional accuracy of the star positions, which is, on average, about 1° and, second, that the list of unidentified gamma-ray sources selected for the correlation analysis with the WR stars is far from complete, and is limited

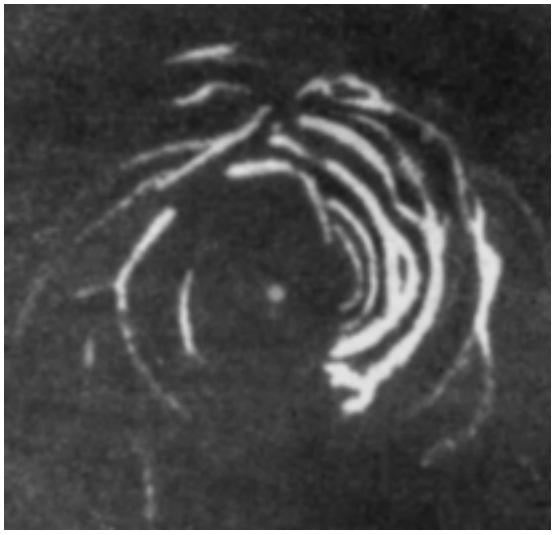


Fig. 3. Spiral structure of the Galaxy derived from 21-cm line measurements.

to sources with $F > 5 \times 10^{-7}$ photons/cm^2 s. The catalog of WR stars [4] is likewise limited to relatively close stars (within 5–10 kpc). For these reasons, we should not expect unambiguous individual identifications between the stars and gamma-ray sources. Nevertheless, we obtained a statistically significant probability for coincidences between the gamma-ray sources and WR stars. An example of such close coincidences is demonstrated in Fig. 5, which shows the distributions of unidentified gamma-ray sources and of WR stars in the strip $-2^\circ \le b \le 2^\circ$, $l = 280$–320°. We can see that four of the five unidentified gamma-ray sources in this strip coincide (within their coordinate error boxes) with WR stars.

A similar analysis for all longitudes yielded the following results. A total of 152 WR stars and 21 unidentified gamma-ray sources were located within the strip $|b| \le 2^\circ$. WR stars are located within the coordinate error boxes of 12 of these 21 sources, with the mean "coincidence area" being equal to $s = \Delta l \Delta b = 0.2$ square degrees, where Δl and Δb are the differences in the coordinates of the corresponding objects. The mean area per WR star is $S_{\rm WR} = 6.3$ square degrees. Consequently, the probability that the coincidences occur by chance should be $P_{\rm chance} = s/S_{\rm WR} = 0.032$, and only one of the 21 unidentified gamma-ray sources should be so close to a WR star. The resulting probability of coincidence was $\mathrm{P} = 12/21 = 0.57 \pm 0.11$, which exceeds $\mathrm{P}_{\rm chance}$ (the probability of random fluctuations $\approx 10^{-6}$) by five standard deviations.

An additional test of the identifications obtained is the dependence of the measured gamma-ray flux F for the gamma-ray sources on the distance r of the proposed corresponding WR star. Figure 6 shows the

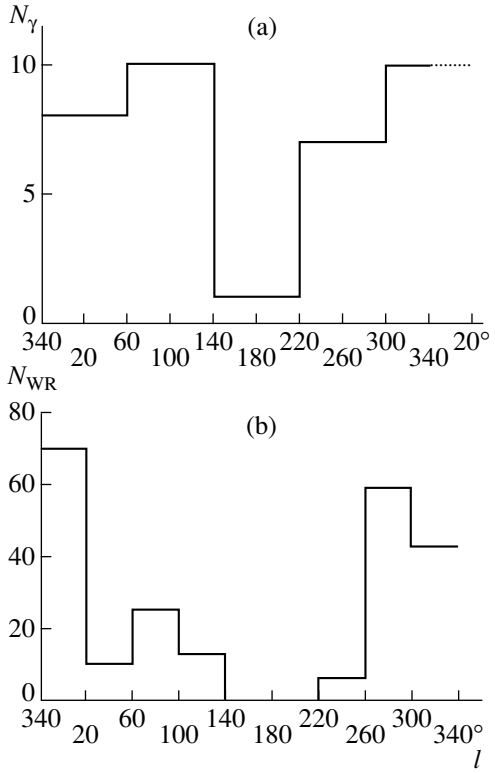


Fig. 4. Galactic-longitude distributions of (a) bright unidentified gamma-ray sources and (b) WR stars.

resulting relation $F(r)$. In spite of the large errors in the measured fluxes, it is clear that, overall, $F(r)$ falls off with distance, as is expected if the identifications are correct.

We noted two peculiarities of the coincident WR stars:

(1) Multiple (binary, triple, multiple) systems were found among them somewhat more often than is usual: $42 \pm 14\%$ of the stars were multiple, whereas 20% of all WR stars are multiple (an enhancement by 1.5 standard deviations).

(2) The highest number of coincidences was shown by stars in the WC (WC4–WC9) subclass. These made up $69\% \pm 13\%$ of the coincidences, in good consistency with their relatively large numbers near the Sun ($r \le 2.5$ kpc), whereas these stars constitute 40% of WR stars overall [4] (an enhancement by more than two standard deviations).

These peculiarities are not highly statistically significant, and should be interpreted only as possible trends.

Of course, the identifications we have carried out must be further verified. In addition to the accumulation of new information about the unidentified gamma-ray sources (refinement of their coordinates, fluxes, and so forth), it is important to turn our attention to new means of studying these sources. It

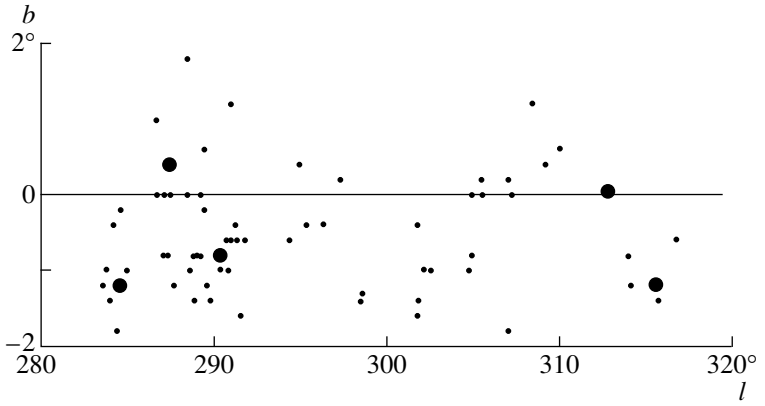


Fig. 5. Map of a section of the Galactic plane ($|b| < 2°$, $l = 280–320°$) showing the positions of WR stars (points) and unidentified gamma-ray sources (circles).

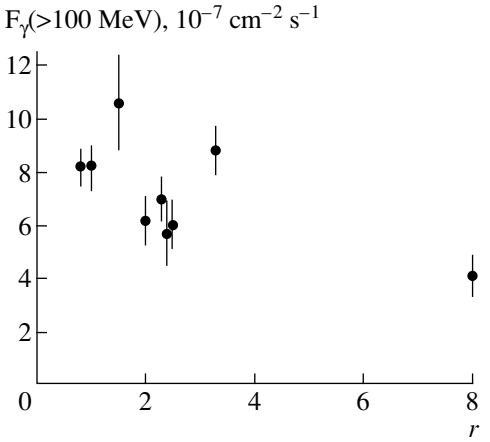


Fig. 6. Dependence of the fluxes of coincident unidentified gamma-ray sources on distance (for the WC–WO subclasses of the WR stars).

is quite probable that the gamma-ray emission of WR stars is not strictly steady, since it is associated with flows of accelerated particles generated during flare processes. The gamma-ray fluxes that have been obtained thus far are the result of averaging over long time intervals (weeks, months). It may be that variability will appear when it is possible to obtain measurements on shorter time scales (days, hours).

Finally, having more penetrating power than optical radiation, the gamma-ray radiation may provide a means to search for new WR stars within the coordinate error boxes of unidentified gamma-ray sources. A preliminary list of these locations is presented in [3].

ACKNOWLEDGMENTS

The authors thank A. Leonov for his help with this work. This work was supported by a contract from the Minister of Industrial Science of the Russian Federation (no. 40022111106).

REFERENCES

1. D. J. Macomb and N. Gehrels, Astrophys. J., Suppl. Ser. **120**, 335 (1999).
2. D. J. Thompson *et al.*, Astrophys. J., Suppl. Ser. **102**, 259 (1995).
3. A. M. Gal'per and B. I. Luchkov, Pis'ma Astron. Zh. (2004, in press).
4. A. K. Van der Hacht, New Astron. Rev. **45**, 135 (2001).
5. A. M. Cherepashchuk, Astron. Zh. **38**, 145 (2001) [Astron. Rep. **45**, 120 (2001)].

Translated by D. Gabuzda

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 39–44. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 43–49.

Massive Close Binary Stars and Gamma-ray Bursts

A. V. Tutukov and A. M. Cherepashchuk

Institute of Astronomy, Russian Academy of Sciences, ul. Pyatnitskaya 48, Moscow, 119017 Russia
Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119992 Russia

Received March 29, 2003; in final form, May 8, 2003

Abstract—We analyze the observed parameters of massive extremely close binaries containing Wolf–Rayet stars and black holes, and identify those systems whose supernova outbursts lead to the formation of rapidly rotating Kerr black holes. It is proposed that the formation of such a black hole is accompanied by a strong gamma-ray burst. Several types of observed systems satisfy the conditions necessary for the formation of a Kerr black hole: BH+WR, BH+OB, WR+O, and BH+K,M. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recently, evidence has been accumulated that gamma-ray bursts (GRBs), at least long-duration ones, are associated with the core collapse of rapidly rotating massive stars, including Wolf–Rayet (WR) stars [1–5]. One possible mechanism for long-duration GRBs is the formation of rapidly rotating, so-called Kerr, black holes possessing the highest angular momenta [6, 7]. The rotation of the cores of single supernova precursors or the components of wide binaries is probably slowed in the course of their evolution, to velocities low enough that the products of their collapse (neutron stars or black holes) have rotational rates far from the highest possible values. As was noted in [8], the situation is different for the components of massive close binaries with short orbital periods. The presence of a close companion to the compact helium presupernova limits the rotational angular velocity of its core—a future black hole—to the orbital angular velocity of the binary system. Thus, the creation of Kerr black holes and the generation of GRBs associated with them are most probable in the closest massive binary systems.

In this paper, based on an analysis of observational data on close binary systems containing black holes or Wolf–Rayet stars [8], we identify systems whose components' evolution can result in or has likely already resulted in the formation of Kerr black holes and, hence, possibly of GRBs.

2. MAIN PROPERTIES OF THE OBSERVED CANDIDATES FOR KERR BLACK HOLES AND GAMMA-RAY BURSTS

The main characteristics of x-ray binaries with black holes were recently reviewed in [9, 10], as well as in [8]. These data can be used to select systems with parameters favorable for the formation of Kerr black holes during their components' explosions as type Ib,c supernovae.

2.1. Systems That Will Form Kerr Black Holes in the Near Future

The systems that are evolutionarily closest to the formation of Kerr black holes are close WR+BH(NS) binaries consisting of a "second-generation" WR star and a black hole (neutron star). In such systems, the secondary mass transfer is over (or nearly over) and the system becomes a WR star with a relativistic companion—a black hole.

Table 1 presents the characteristics of three such systems. The Cyg X-3 system is a typical WR+C binary, and contains a WR star and an accreting x-ray object [11]. The binary's orbital period is very short: $P = 4.8$ hours. The relativistic object was originally thought to be a black whole with a mass of $\sim 17M_\odot$ [12, 13]. However, the new spectroscopic results of [14] have yielded a much lower mass function for the optical star in the Cyg X-3 system—$f_v(m) = 0.027M_\odot$—allowing the presence of either a black hole or a neutron star in the system. Note, however, that no x-ray pulsar or type I x-ray burster is observed, so that there is no direct evidence for the presence of a neutron star.

The SS 433 system possesses a long orbital period, $P \cong 13.1$ days, and consists of an optical star—an A7Ib supergiant with a mass of $19 \pm 7M_\odot$—and a black hole whose mass is $m_x = (11 \pm 5)M_\odot$ [15]. In this system, we observe an optically bright supercritical accretion disk around the black hole, which precesses with a period of 162.5 days [16]. The system is in the final stage of the second mass transfer [16–18]. The recent discovery of an enhanced

Table 1. WR+C systems, possible producers of Kerr black holes in the near future

System	Spectral type	Period, days	m_x, $M_\odot$	m_v, $M_\odot$	$f_v(m)$, $M_\odot$	$f_x(m)$, $M_\odot$
Cyg X-3	WN3-7+C	0.2			0.027	–
SS 433	A7Ib+BH	13.1	11 ± 5	19 ± 7	~ 1.3	~ 7
HD 197406	WN8h+?	4.3			0.28	–

Table 2. OB+BH systems, possible future producers of Kerr black holes

System	Spectral type	Period, days	m_x, $M_\odot$	m_v, $M_\odot$	$f_v(m)$, $M_\odot$
Cyg X-1	O9.7Iab+BH	5.6	16	33	0.24
LMC X-3	B3Ve+BH	1.7	9	6	2.3
LMC X-1	O(7–9)III+BH	4.2	7	22	0.14

helium abundance in the material flowing from the optical star to the relativistic object [19] confirms this model of SS 433.

The system HD 197406 (WN8h+?, $P \cong 4.3$ days) is either a WR+C binary system with a black hole with a low x-ray luminosity [20], or contains a normal star as its invisible companion—a hot B3IV subgiant [20].

2.2. Systems Capable of Forming Kerr Black Holes in the Future

Table 2 presents the parameters of binaries of this type. The Cyg X-1 system ($P \cong 5.6$ days, O9.7Iab+BH) is a typical massive, quasi-steady-state x-ray binary with a black hole. According to [8], it was formed of a massive O+O binary during the common-envelope stage in its evolution. Its second mass transfer will occur in $\sim 10^6$ years. If the second mass transfer occurs in a common-envelope regime (and, based on the mass of the optical component [8], this is most likely), a WR+C binary with a very short period will form after the transfer, so that the system's observed characteristics will resemble those of Cyg X-3. The LMC X-3 system ($P \cong 1.7$ days, B3Ve+BH) has the shortest orbital period among massive x-ray binaries. It is quite probable that it will be associated with GRBs in the future and that it may have been in the past as well.

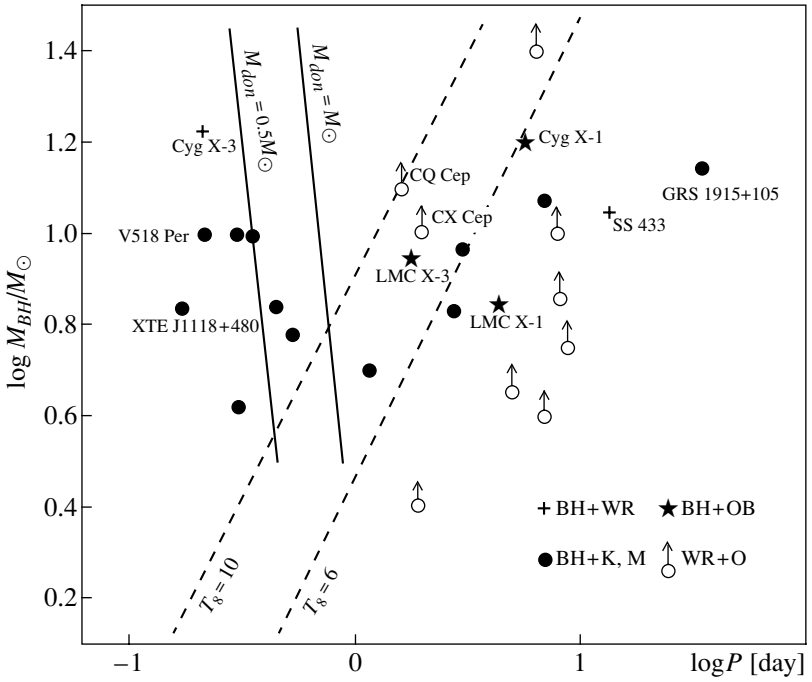
The features of LMC X-1 ($P = 4.2$ days, O(7–9)III+BH) are very similar to those of Cyg X-1. This is also a massive, quasi-steady-state x-ray binary with a black hole, formed from a massive O+O system as a result of common-envelope evolution [8]. It will probably become an analog of Cyg X-3 (WR+C) in about 10^6 years.

2.3. Systems That Probably Formed a Kerr Black Hole in the Past

Table 3 displays parameters of several systems belonging to this category. These binaries can be subdivided into two types: low-mass x-ray binaries (x-ray novae) with black holes possessing very short orbital periods, $P < 0.5$ days (eight systems), and long-period x-ray novae (GRS 1915+105, $P = 33.5$ days, and GRO J1655-40, $P = 2.6$ days) showing clear evidence for rapid rotation of the black hole. This evidence includes an enhanced thermal component of the x-ray spectrum, showing that the inner edge of the accretion disk is less than three gravitational radii from the black hole, as well as the occurrence of sporadic polar jets with velocities $\sim 0.9c$. It has recently become clear [21] that clear evidence for relativistic jets is shown only by x-ray binaries with low-mass companions (x-ray novae) and long orbital periods (4 of 14 known systems: XTE J1550-564, GRO J1655-40, V4641 Sgr, and GRS 1915+105). We should bear in mind, however, that rapid rotation of the black holes in such long-period systems can also result from the accretion of a considerable fraction of the donor's mass, whereas rapid rotation of the black holes in short-period systems has a genetic origin, and is associated with the formation of Kerr black holes in type Ib,c supernova outbursts.

3. POSITIONS OF OBSERVED EVOLVED MASSIVE BINARIES IN THE M_{BH}–P PLANE

The figure can be used to search for possible precursors of GRBs, as well as for systems that have already evolved through this phase. We plotted the positions of the observed short-period massive binaries from Tables 1 and 2 in [8] and for Cygnus X-3 from Koch-Miramond *et al.* [24] in a diagram with



Close binaries in the black-hole mass—orbital period plane. The plotted periods and masses of the binary components were taken from Tables 1 and 2 of [8]. The position of GU Mus corresponds to the parameters presented by Khruzina *et al*. [22]. For the WR+O systems, we plot lower limits for the masses of their future black holes, assuming $M_{BH} \geq 0.5\ M_{WR}$. The $T_8 = 6$ and $T_8 = 10$ boundary lines (where $T_8 = 10^8$ K) are plotted in accordance with (4). The solid lines indicate the longest initial orbital periods of BH+K,M systems when their black holes are formed, taking into account the magnetic stellar wind from donors with initial masses of 0.5 $M_\odot$ and 1 $M_\odot$, in accordance with [23].

axes corresponding to the black hole mass M_{BH} and the system's orbital period P.

The systems shown in the figure represent the four described types of close binaries in evolutionary stages either prior to or following the appearance of a GRB. We list these types below.

First of all, we present positions for two BH+WR systems, in which black holes are companions of Wolf–Rayet stars: Cygnus X-3 and SS 433. The past formation of the black hole in the closest of these systems may have been accompanied by a GRB, and the core collapse of the currently observed WR star will be accompanied by a GRB. The number of such systems known (two) is small due to their low formation rate ($<10^{-3}$/year [8]) and the short lifetimes of Wolf–Rayet stars ($\sim 10^5$ years), as well as the difficulty of finding even short-period systems of this type due to the broad spectral lines of WR stars.

The second type of system plotted is massive BH+OB binaries, such as LMC X-3 and others. The formation of a black hole in these binaries could also be accompanied by a GRB if the system is close enough to bring about rapid rotation of the black-hole precursor.

The third group includes BH+K,M systems such as XTE J1118+480 ($P \approx 4$ hours), in which the black hole's companion is a K,M dwarf with a current mass of several tenths of a solar mass. If the axial rotation of the helium black-hole precursor in such a system is synchronized with the orbital motion, the formation of the black hole could be accompanied by a GRB. An important conclusion that we can draw based on the very existence of systems having such large current component-mass ratios is the unavoidable collapse to a black hole of more than half of the mass of its helium WR precursor. Otherwise, such systems would simply have been disrupted by the time of the supernova explosion accompanying the black hole's formation.

We used this last condition to plot the closest WR+O systems on the Figure [8, Table 1]. To find the positions of such systems, we assumed that the mass of the black hole resulting from the evolution of a WR star should exceed half the star's current mass.

Among close binaries responsible for GRBs, the figure also shows systems consisting of accreting black holes and low-mass K,M dwarfs (0.3–1 $M_\odot$). When comparing their positions with the critical-period line, we must take into account the fact that, due to the loss of angular momentum via the magnetic stellar wind from the donor, their current orbital periods are shorter than their orbital periods at the time of formation of the black hole responsible for

Table 3. Short-period x-ray novae with black holes, producers of Kerr black holes in the past

System	Spectral type	Period, days	m_x, $M_\odot$	m_v, $M_\odot$	$f_v(m)$, $M_\odot$
A0620-00	K5V	0.3	10	0.6	2.91
GRS 1124-68	K2V	0.4	6	0.8	3.01
GS 2000+25	K5V	0.3	10	0.5	4.97
GRO JO422+32	M2V	0.2	10	0.4	1.13
H1705-250	K5V	0.5	6	0.4	4.86
GRS 1009-45	(K6-M0)V	0.3	4	0.6	3.17
XTE J1118+480	(K7-M0)V	0.17	7	0.4	6.1
XTE J1859+226	$\sim$ M	0.38	10	~ 0.4	7.4

Table 4. WR+O systems, possible future producers of Kerr black holes

System	Spectral type	Period, days	$m_{\rm WR}$, $M_\odot$	m_O, $M_\odot$
CQ Cep	WN6+O9II-Ib	1.64	24	30
CX Cep	WN4+O5V	2.13	20	28

the GRB [23]. To constrain these systems' orbital periods at the time of black-hole formation, we should compare the time scale for the loss of orbital angular momentum via the magnetic stellar wind,

$$t_{\rm MSW} \cong 3\times 10^6 \left(\frac{a}{R_\odot}\right)^5 \frac{M_{BH}}{M_\odot} \times \frac{M_\odot^6}{M_{don}(M_{BH}+M_{don})^2}\ \text{years},$$

with the donor's nuclear evolution time for donors with initial masses M_{don} in excess of $\sim 0.9\ M_\odot$ and with the Hubble time for lower-mass donors. The resulting limits for the maximum initial orbital periods required to obtain the observed close semi-detached x-ray binaries with low-mass donors are plotted in the figure as solid lines for two initial donor masses: $M_{don} = M_\odot$ and $M_{don} = 0.5M_\odot$.

Now let us locate in the figure the limit corresponding to the formation of Kerr black holes during the core collapse of WR stars. An important condition when searching for such a limit is the assumption that the axial rotation of the collapsing presupernova core is synchronized with the system's orbital motion. The components of the systems we are considering are close to one another, suggesting this should be true. Naturally, core rotation at rates higher than in the case of synchronism facilitates the formation of Kerr black holes. The formation of a Kerr black hole requires that

$$\omega^2 R_{BH}^3 > GM_{BH}, \qquad (1)$$

where ω is the rotational angular velocity, R_{BH} the black hole's radius, M_{BH} the black hole's mass, and G the gravitational constant. $R_{BH} = 2GM_{BH}c^{-2}$, where c is the velocity of light. As a result, the orbital period P when the Kerr black hole is formed in the system must satisfy the condition

$$P < \frac{2\pi R_0^2}{\sqrt{GM_{BH}R_{BH}}}, \qquad (2)$$

where R_0 is the initial core radius of the presupernova that collapses to form the black hole. We assume conservation of the core's angular momentum during the collapse. To estimate the initial radius of the collapsing core, we assume it is in hydrostatic equilibrium prior to the onset of the collapse: $R_0 = GM_{BH}\mu/kT$, where μ is the molecular weight of the core material, k is the Boltzmann constant, and T is the temperature of the collapsing part of the core. The resulting estimate of the limiting orbital period is

$$P \lesssim \frac{\sqrt{2}\pi c G\mu^2 M_{BH}}{k^2T^2}. \qquad (3)$$

For the presupernova's core material $\mu = 2m_p$ (m_p is the proton mass). Thus, we obtain a condition for the formation of a black hole in a close binary:

$$P \lesssim \frac{12^{\rm d}}{T_8^2}\frac{M_{BH}}{M_\odot}, \qquad (4)$$

where T_8 is the temperature of the collapsing part of the core in units of 10^8 K. We adopt for the core temperature the temperature of carbon burning in the nondegenerate cores of massive stars [25, Fig. 22], $T_8 \cong 6$, since the nuclear energy of the products of carbon burning (O, Ne, Mg) is insufficient to compensate for the gravitational binding energy of the

material in the presupernova core. To demonstrate the role of the core temperature, which is also dependent on the core mass, we also plotted the limit for P when $T_8 = 10$. The resulting lines form the right-hand boundary of the area containing systems in which black hole formation is probably accompanied by a GRB. We showed in our previous paper [8] that the GRB rate corrected for collimation of the gamma rays is close to the of formation rate of Kerr black holes.

Now consider the main results following from the figure. Several short-period close binaries are found in the area that is "promising" from the point of view of the formation of Kerr black holes, and hence the appearance of GRBs. These are, first and foremost, x-ray binaries with black hole accretors and low-mass donors, such as V518 Per. They are distinguished by their long lifetimes ($\sim 10^9$ years [24]) and large x-ray fluxes. For all these systems, GRBs were produced in the remote past.

The position of the Cyg X-3 system is unique. It may contain a black hole, indicating a probable GRB in the past, whereas the presence of a WR star suggests that a GRB may occur in the future. The existence of systems consisting of black holes and WR stars is a natural result of the evolution of massive close binaries [26]. SS 433 may be a similar system, though it is located outside the GRB zone. Detecting such systems is difficult due to their short lifetimes and the complex radial-velocity variability in the spectra of WR stars.

The third type of system, BH+OB, includes only one close binary, LMC X-3, whose black-hole formation may have been associated with a GRB in the past. It is interesting that the further evolution of this system could lead to its transformation into a system similar to Cyg X-1 (BH+WR), with a subsequent GRB. Finally, it is interesting that candidates for systems giving rise to GRBs within the next several hundred thousand years are also encountered among WR+OB systems. One is CQ Cep, whose WR component's mass, $\sim 24\ M_\odot$ [8], is sufficient for it to become a black hole, and whose orbital period is short enough for a Kerr black hole to be formed. The list of GRB progenitors may also include CX Cep ($M_{WR} \cong 20\ M_\odot$, Table 4), but the position of the boundary between the progenitors of Kerr black holes and systems not capable of producing them is not sufficiently certain to draw firm conclusions about this.

It is interesting that the correlation between black-hole masses and orbital periods that has sometimes been claimed (see figure) may be due to an obvious observational selection effect [8]. It is also not ruled out that rotation of the presupernova core itself "participates" in "determining" the mass of the black hole. Namely, cores of rapidly rotating stars in short-period binaries produce low-mass black holes, whereas, in the absence of effective centrifugal forces, cores of slowly rotating stars in long-period binaries produce massive black holes [27]. Future observations, together with a theory of the advanced evolutionary stages of rotating stars based on adequate models, such as those suggested by Brunn *et al.* [28], should help clarify this point.

4. CONCLUSIONS

This paper continues the study of the parameters and evolution of extremely close binaries commenced in our earlier paper [8]. Here, we have aimed to identify systems in which explosions of type Ib,c supernovae result in the formation of rapidly rotating Kerr black holes, and possibly in the appearance of gamma-ray bursts. The most probable predecessors include the Cyg X-3 (BH+WR) and CQ Cep (WR+O) systems. The figure also shows the presence of very close systems with black holes and K,M companions, such as XTE J1118+480. The formation of black holes in such close binaries could also produce GRBs if the axial rotation of their Wolf–Rayet components prior to the supernova outburst was synchronized with the orbital motion. Thus, the existence of such systems, along with the presence of ultrabright x-ray sources ($L_x \cong 10^{39}$–10^{40} erg/s [29]), demonstrates the possibility of the formation of Kerr black holes during the core collapse of the WR components in these extremely close binaries. It turn, the formation of the black hole may be accompanied by a powerful burst of gamma-ray emission—a GRB [1–8]. Note the recent discovery [30] of an ultrabright x-ray source in the galaxy M 51 ($L_x = 10^{39}$–10^{41} erg/s with a period of ~ 2.1 hours, which may be an x-ray binary with a very short orbital period containing a black hole whose mass exceeds $10\ M_\odot$. This object can also be considered a producer of a Kerr black hole and GRB.

Finally, it follows from spectropolarimetric observations [31] that some WR stars are rapidly rotating objects. The fraction of rapid rotators among known WR stars is ~ 15–20% [31]. It is quite possible that the evolution of some of these possibly single, rapidly rotating WR stars likewise ends with the formation of a Kerr black hole accompanied by a GRB. It is interesting that the most massive ($\gtrsim 50\ M_\odot$) main-sequence stars do not pass through a red supergiant phase [8]. Thus, their helium cores (WR stars) can remain rapidly rotating until the formation of a black hole, possibly accompanied by a GRB. The question of the synchronization of the axial rotation and orbital motion of WR stars remains inadequately studies, both theoretically and observationally. Note, however, that WR+OB and OB+OB systems with periods

shorter than several days have zero orbital eccentricities [32]. The systems considered in this paper have shorter periods. This provides a basis to suppose that their rotation and orbital motions are synchronized.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 02-02-17524 and 03-02-16254), the Program of Support for Leading Scientific Schools of Russia, the State Scientific and Technology Program "Astronomy," and the program "Unstable Processes in Astronomy" of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. B. Paczynski, Astrophys. J. **494**, L45 (1998).
2. S. E. Woosley, Astrophys. J. **405**, 273 (1993).
3. A. Mac Fadyen and S. Woosley, Astrophys. J. **524**, 262 (1999).
4. S. S. Gershteĭn, Pis'ma Astron. Zh. **26**, 848 (2000) [Astron. Lett. **26**, 730 (2000)].
5. K. A. Postnov and A. M. Cherepashchuk, Astron. Zh. **78**, 602 (2001) [Astron. Rep. **45**, 517 (2001)].
6. W. Zhang and S. Woosley, astro-ph/0207436.
7. A. Heder and S. Woosley, astro-ph/0206005 (2002).
8. A. V. Tutukov and A. M. Cherepashchuk, Astron. Zh. **80**, 419 (2003) [Astron. Rep. **47**, 386 (2003)].
9. A. M. Cherepashchuk, Usp. Fiz. Nauk **173**, 345 (2003).
10. J. A. Orosz, in *IAU Symp. No. 212: Massive Star Odyssey, from Main Sequence to Supernova,* Ed. by K. A. Van der Hucht, A. Herrero, and C. Esteban (2002).
11. M. N. Van Kerkwik, P. Charles, T. R. Geballe, *et al.*, Nature **355**, 703 (1992).
12. W. Schmutz, T. R. Geballe, and H. Schild, Astron. Astrophys. **111**, L25 (1996).
13. A. M. Cherepashchuk and A. F. J. Moffat, Astrophys. J. **424**, L53 (1994).
14. M. M. Hanson, M. D. Still, and R. P. Fender, Astrophys. J. **541**, 308 (2000).
15. D. R. Gies, W. Huang, and M. V. McSwain, Astrophys. J. **578**, L67 (2002).
16. A. M. Cherepashchuk, Space Sci. Rev. **102**, 23 (2002).
17. A. M. Cherepashchuk, Mon. Not. R. Astron. Soc. **194**, 761 (1981).
18. A. M. Cherepashchuk, Pis'ma Astron. Zh. **7**, 201 (1981) [Sov. Astron. Lett. **7**, 111 (1981)].
19. Y. Fuchs, L. Koch-Miramond, and P. Abraham, astro-ph/0208432 (2002).
20. K. A. Van der Hucht, New Astron. Rev. **45**, 135 (2001).
21. M. R. Garcia, J. M. Miller, J. E. McClintock, *et al.*, astro-ph/0302230 (2003).
22. T. S. Khruzina, A. M. Cherepashchuk, D. V. Bisikalo, *et al.*, Astron. Zh. (2003, in press).
23. A. V. Tutukov and A. V. Fedorova, Astron. Zh. **79**, 847 (2002) [Astron. Rep. **46**, 765 (2002)].
24. L. Koch-Miramond *et al.*, Astron. Astrophys. **396**, 877 (2002).
25. A. G. Masevich and A. V. Tutukov, *Stellar Evolution: Theory and Observations* (Nauka, Moscow, 1988) [in Russian].
26. A. V. Tutukov and L. R. Yungel'son, Nauchn. Inform. Astron. Soveta Akad. Nauk SSSR **27**, 57 (1973).
27. A. V. Tutukov and A. M. Cherepashchuk, Astron. Zh. **62**, 643 (1985) [Sov. Astron. **29**, 654 (1985)].
28. A. Brunn, M. Browning, and J. Toomre, astro-ph/0302598 (2003).
29. A. V. Tutukov, Astron. Zh. (2003, in press).
30. J.-F. Liu *et al.*, Astrophys. J. **581**, L93 (2002).
31. T. J. Harries, D. J. Hillier, and I. D. Howarth, Mon. Not. R. Astron. Soc. **296**, 1072 (1998).
32. A. M. Cherepashchuk and V. G. Karetnikov, Astron. Zh. **80**, 42 (2003) [Astron. Rep. **47**, 38 (2003)].

Translated by N. Samus'

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 45–51. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 50–57.

Dynamical Evolution of Multiple Stars: Influence of the Initial Parameters of the System

A. V. Rubinov

Sobolev Astronomical Institute, St. Petersburg State University, Bibliotechnaya pl. 2, Petrodvorets, 198904 Russia

Received March 20, 2003; in final form, August 8, 2003

Abstract—We have modeled the dynamical evolution of small stellar groups with $N = 6$ components in the framework of the gravitational N-body problem, taking into account possible mergers of stars and ejection of single and binary stars. We study the influence of the initial global parameters of the systems (the mass spectrum, average size, virial factor) on their dynamical evolution. The distribution over states is analyzed for a time equal to 300 initial crossing times of the system. The parameters of binary and stable triple systems that form are studied, as well as the properties of ejected single and binary stars. The rate of dynamical evolution in both expanding and contracting groups is higher than in systems in a state of virial equilibrium. The dynamical evolution is more intense in the case of unequal masses than when the system initially consists of equal-mass stars. In most cases, the evolution of a group ends with the formation of a binary or stable triple system. The semimajor axes of the binaries range from several hundredths to several times the initial size of the system. The distribution of the eccentricities of the binaries formed is consistent with an $f(e) = 2e$ law. When the initial size of the group is small, the number of final binaries with large eccentricities, and also of stable triple systems with elongated inner-binary orbits, decreases due to merging. As a rule, stable triple systems are substantially hierarchical (the average ratio of the semimajor axes of the inner and outer binaries is 1 : 20). On average, the eccentricities of the inner binaries exceed those of the outer binaries: they are equal to $\overline{e_{in}} \approx 0.7$ and $\overline{e_{ex}} \approx 0.5$, respectively. The velocities of ejected stars are from several to several tens of km/s, and tend to increase as the initial size of the system, and hence its virial coefficient, decreases. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Observations of star-formation regions and young T Tauri stars suggest that stars mostly originate in small groups (see, for example, [1] and references therein). This conclusion is also supported by modern simulations of the fragmentation of molecular clouds, which indicate that fragmentation may result in the formation of nonhierarchical stellar systems with various numbers of components (see, for example, [2, 3]). The gravitational interactions between the members of a group can result in its decay. According to the hypothesis of van Albada [4], the decay of young, multiple, nonhierarchical stellar systems can lead to the formation of observed wide binary and multiple stars.

The global characteristics of these small groups remain unclear. This can be understood as a consequence of the fact that few simulations of fragmentation have been carried out with sufficient resolution right to the end of the accretion stage [5, 6]. The very complicated behavior of fragments in the accretion stage make the orbital parameters of forming systems unpredictable. In addition, the number of parameters describing the initial state of a cloud is large, so that nonhierarchical multiple systems originating due to fragmentation cannot currently be adequately described.

Models of the dynamical evolution of stellar systems with numbers of stars $N > 3$ are comparatively scarce. The first studies were carried out in the late 1960s and early 1970s by van Albada [4] and Harrington [7] for a limited number of sets of initial conditions. In the middle 1990s, the distribution over final states, and also the parameters of binaries originating due to the decay of nonhierarchical multiple systems of $N = 3$, 4, and 5 components, were studied for several initial mass spectra [8]. Several studies considered the parameters of single stars ejected from a system in the course of its evolution [9, 10], within a broad range of initial conditions in nonhierarchical multiple systems.

In addition, in [11], the orbital inclinations of inner and outer binaries, ratios of their periods, and component-mass ratios in stable triple systems derived from modeling using a broad range of initial parameters were compared with observed stable triple systems.

In [12], the dynamical evolution of nonhierarchical multiple stellar systems was considered taking into

1063-7729/04/4801-0045$26.00

account the merging of stars for various initial numbers of components N. Modeling was carried out for a specified initial size R and virial factor k of a system, for two mass spectra. The parameters of ejected single stars and of binary and stable triple systems that formed were studied. Here, we study the influence of the parameters R and k on the dynamical evolution of nonhierarchical multiple systems for three different mass spectra.

2. MODELING

We modeled the dynamical evolution of nonhierarchical multiple systems in the framework of the gravitational N-body problem with point masses. The simulation technique is described in detail in [12]. Here, we will only note the most important features of the method.

The modeling was based on the numerical integration of regularized equations of motion. We used the method of chain regularization suggested in [13]. The possible merging of closely approaching stars was taken into account, as well as ejections of single and binary stars that became sufficiently distant from the center of mass of the remaining bodies and thereby isolated from them.

The dynamical evolution was followed over 300 T_{cr}, where

$$T_{cr} = \frac{G}{(2|E|)^{\frac{3}{2}}} \sum_{i<j} m_i m_j \sqrt{\sum_{k=1}^{N} m_k} \qquad (1)$$

is the initial mean crossing time of the system. Here, G is the gravitational constant, E the total energy of the system, and m_i the mass of star i. When an initial system decayed to a binary, the integration was ceased.

We solved the equations of motion numerically using a fourth-order Runge-Kutta integrator with automatic step selection. The accuracy of the calculations was checked using the integrals of motion. The relative errors of the integrals of the energy and area over the computation time did not exceed 10^{-5}. The integrals of motion for the center of mass were maintained with even better accuracy.

The initial conditions were specified as follows. The modeling was carried out for systems initially consisting of $N = 6$ stars. It was assumed that the stars were initially distributed uniformly and randomly within a sphere with radius R. The velocities of the stars were chosen so that the ratio of the kinetic energy and the absolute value of the potential energy of the system was equal to the adopted virial factor k. The velocity distribution was assumed to be isotropic. We considered the parameter values $R = 3$, 10, 100, 1000 AU; and $k = 0.001$, 0.1, 0.5, 0.9.

We studied the dependence of the results of the dynamical evolution on the initial mass spectrum of the system. We considered three different mass spectra: equal masses for all stars (each star had the solar mass), a Salpeter mass spectrum [14]

$$f(m) \sim m^{-2.35}, \quad m \in [0.4; 10] M_\odot, \qquad (2)$$

and a power-law spectrum with index -1.5:

$$f(m) \sim m^{-1.5}, \quad m \in [0.9; 10] M_\odot. \qquad (3)$$

We chose to consider this third type of initial mass spectrum because its index is intermediate between the Salpeter and equal-mass cases. In addition, the cores of molecular clouds from which the stars originate can display tattered fractal structure. In this case, the mass distribution of the fragments inside a cloud can be described by a power law with an index from -2.0 to -1.5 (see, for example, [15]).

Five hundred cases of initial conditions were considered for each set of parameters.

3. RESULTS OF THE MODELING

3.1. Distribution over States

To understand what systems can originate due to the dynamical decay of nonhierarchical multiple stars, we must determine the distribution over states when the integration ceases. The following states were distinguished: binaries with positive total energy (two single stars), binaries with negative total energy, stable triple systems, unstable triple systems, and systems with higher multiplicity. We divided the triple systems into stable and unstable systems, in accordance with the analytical criterion suggested in [16]. When implementing this criterion, stability must be understood as stability with regard to the substitution of a component of the inner binary with a distant component; i.e., with regard to maintaining the hierarchy.

The table presents the distribution over states for various initial virial factors and initial mass spectra for an initial size of the system of 100 AU.

Let us distinguish those characteristics of dynamical systems that are stable against our selection of initial conditions. It is obvious that, as a rule, the dynamical evolution results in the formation of a final binary. The probability of forming a stable triple system is also fairly high (5–20%). The last column of the table illustrates the fact that not all nonhierarchical systems complete their dynamical evolution over 300 initial crossing times.

The dependence of the results on the initial virial factor expresses itself in the acceleration of the dynamical evolution in nonequilibrium (from the point of view of the virial theorem) systems. This can be

Distribution over states as a function of the initial virial factor at time $300T_{cr}$ for $R = 100$ AU

k	Mass spectrum *	Final binaries	Two single stars	Stable triples	Unstable triples	Higher multiplicity
0.001	EM	0.60	0.05	0.13	0.10	0.12
	CM	0.72	0.01	0.07	0.14	0.06
	SM	0.65	0.01	0.11	0.17	0.06
0.1	EM	0.60	0.04	0.14	0.08	0.14
	CM	0.60	0.01	0.11	0.20	0.08
	SM	0.60	0.01	0.11	0.17	0.10
0.5	EM	0.57	0.07	0.06	0.15	0.15
	CM	0.57	0.02	0.19	0.11	0.11
	SM	0.55	0.01	0.18	0.18	0.08
0.9	EM	0.68	0.08	0.15	0.07	0.02
	CM	0.76	0.05	0.08	0.10	0.01
	SM	0.73	0.04	0.10	0.11	0.02

* EM denotes equal masses, CM the mass spectrum (3), and SM a Salpeter mass spectrum.

seen by comparing the number of final binary and stable triple systems (i.e., whose dynamical evolution has finished) with the number of unstable triple and higher-multiplicity systems (whose dynamical evolution has not finished). In addition, the rate of dynamical evolution for the case of unequal masses exceeds that for the case of equal masses. Increasing the initial size of a system also results in a slight increase in the rate of its dynamical evolution, although this tendency is weaker.

We stress that an increase in the rate of dynamical evolution implies a decrease in the time interval over which a system decays, in units of the mean crossing time rather than in physical units. It is obvious that the mean time needed for the decay of the system (in physical units) also increases with the size of the system.

3.2. Final Binaries

In more than half of the cases, the dynamical evolution of a nonhierarchical multiple system ends with the formation of a final binary.

Figure 1 presents the dependence of the median semimajor axis of the final binary on the initial size of a system in the case of virial equilibrium. We can see that, as a rule, the semimajor axes of the binaries are several tenths of the initial size of a system. When the initial size of a nonhierarchical system increases, the final binary becomes wider. Increasing the initial virial factor also results in the formation of a wider final binary. In addition, we can see that the closest binaries are formed in the case of equal-mass the stars and the widest in the case of the mass spectrum (3).

Figure 2 presents the distribution of the eccentricities of the final binaries in the equal-mass case as a function of the initial size of the system. The solid curve shows the $f(e) = 2e$ law, first derived by Ambartsumian [17] for the equilibrium distribution of

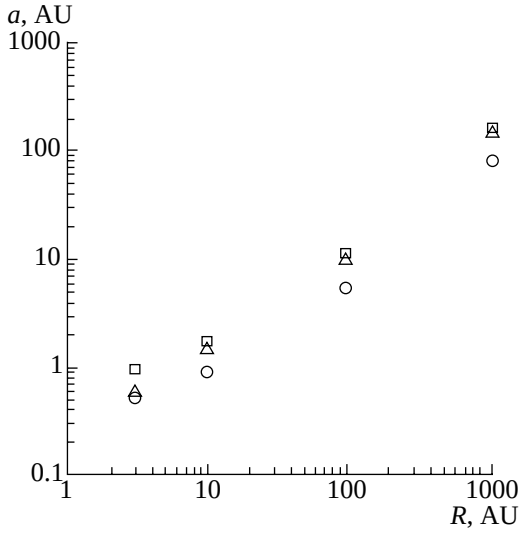


Fig. 1. Median semimajor axis of the final binary as a function of the initial size of the system R. Squares denote the mass spectrum (3), circles the case of equal masses, and triangles a Salpeter mass spectrum.

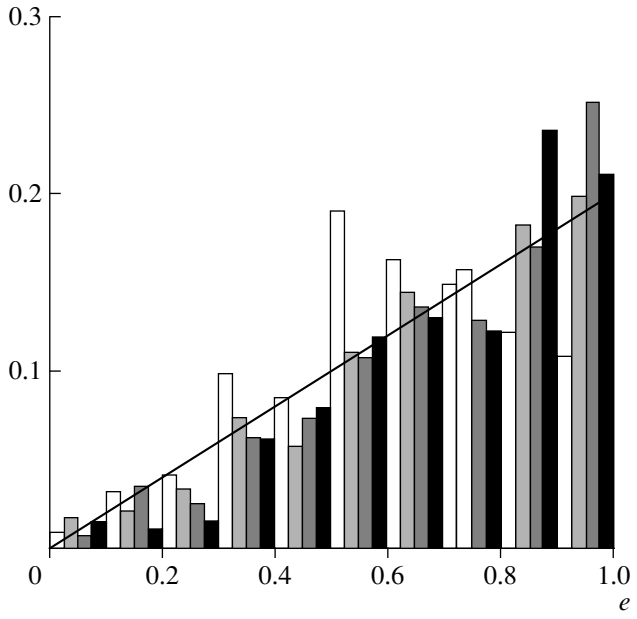


Fig. 2. Distribution of the eccentricities of the final binaries as a function of the initial size of the system R in the equal-mass case. The unfilled bars of the histogram correspond to $R = 3$ AU, the light gray bars to $R = 10$ AU, the dark gray bars to $R = 100$ AU, and the black bars to $R = 1000$ AU. The solid curve represents the distribution $f(e) = 2e$.

binaries in a stellar field. A similar law was derived in [18], where the decay of triple systems was considered. For moderately close, nonhierarchical multiple systems, the distribution of final-binary eccentricities is consistent with the $f(e) = 2e$ law, independent of the initial mass spectrum or virial factor. However, when the initial size of the system decreases, a deficiency of binaries with very elongated orbits is observed. This is apparently due to the increase in the probability of merging for elongated close binaries, particularly at the orbit pericenter.

3.3. Stable Triple Systems

The distribution over states indicates that the probability of forming stable triple systems is fairly high. We will now consider the influence of the initial conditions on their parameters. A triple system can be described as a superposition of two binaries, inner and outer. The inner binary is formed of the two closest stars of the system, while the outer is formed of an object with the mass of the inner binary located at its center of mass and the distant component of the system.

Our modeling indicates that, as a rule, the semimajor axes of the inner binaries are one to two orders of magnitude smaller than the initial sizes of the systems, whereas the semimajor axes of the outer binaries are of the same order of magnitude as the initial size.

When the initial size of a system is varied, the behavior of the median semimajor axes of both the inner and outer binaries is similar to that of the semimajor axes of the final binaries (Fig. 1): they increase with parameter R. Increasing the initial virial factor also results in an increase of the semimajor axes of both the inner and outer binaries. The dependence of the semimajor axes of the inner binaries on the initial conditions can be obtained by shifting the analogous dependence for the outer binaries, since the shapes of the two curves are similar. As for the final binaries, the closest stable triple systems are formed from equal-mass stars.

Figure 3a presents the distributions of the eccentricities of the inner binaries for the initial mass spectrum (3) as a function of the initial size of the system. The solid curve represents the $f(e) = 2e$ law. The eccentricity distributions for the model systems are in poor agreement with the theoretical law due to a deficiency of systems with highly elongated orbits. This discrepancy is enhanced for small systems, probably due to the increase in the probability of merging in close elongated inner binaries. Variation of the remaining parameters considered has little effect on the shape of the distributions.

Figure 3b presents the eccentricity distributions for the outer binaries for the case of the initial mass spectrum (3) as a function of the initial size of the system. We can see that the eccentricities of the outer binaries tend to be lower than those of the inner binaries (the corresponding mean values are $\overline{e_{ex}} \approx 0.5$ and $\overline{e_{in}} \approx 0.7$). This result is preserved for various initial conditions. In addition, the dependence of the shape of the distributions on the initial size is weaker than in the case of the inner binaries, and no stable trends can be distinguished. The shape of the distribution remains essentially the same when the remaining parameters considered are varied.

The stable triple systems are strongly hierarchical. The ratio of the semimajor axes of the inner and outer binaries is stable against the choice of initial conditions and is, on average, 1 : 20.

Figure 4 presents the distributions of the angles between the orbital angular-momentum vectors of the inner and outer binaries in stable triple systems as a function of the initial mass spectrum (virial equilibrium, $R = 100$ AU). The solid curve shows the theoretical law $f(i) = \frac{1}{2}\sin i$, which corresponds to random orientations of the orbital angular-momentum vectors. We can see that all the distributions have a similar shape (one-peaked); however, the distributions of the model systems are asymmetrical in favor

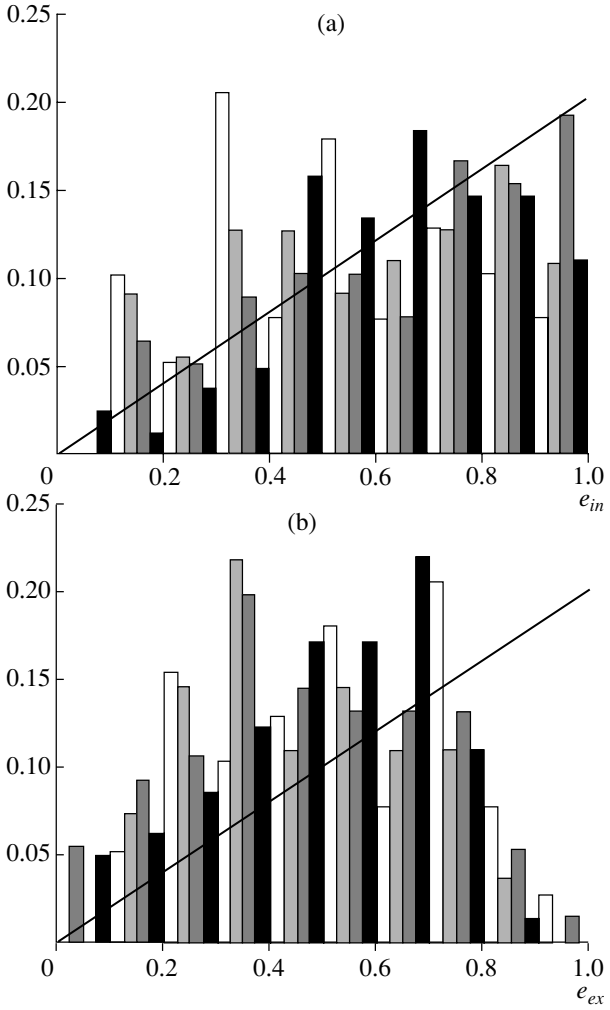


Fig. 3. Same as Fig. 2 for (a) inner and (b) outer binaries in stable triple systems.

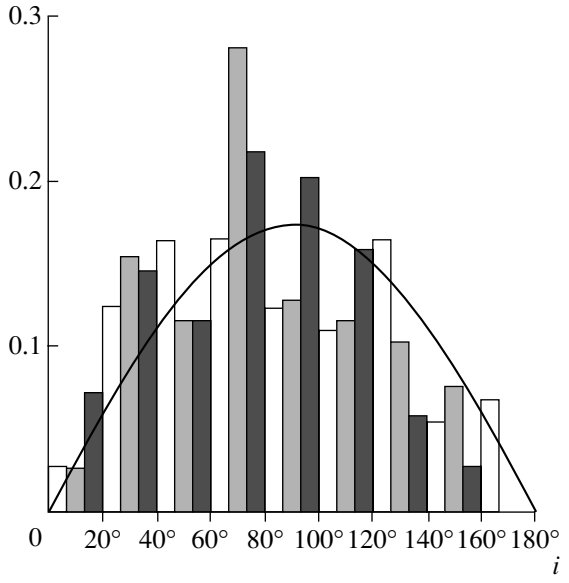


Fig. 4. Distribution of angles between the orbital angular-momentum vectors of the outer and inner binaries as a function of the initial mass spectrum. The unfilled bars of the histogram correspond to the equal-mass case, the gray bars to the mass spectrum (3), and the black bars to a Salpeter mass spectrum. The solid curve represents the law $f(i) = \frac{1}{2}\sin i$.

of direct motions. This is apparently due to the fact that, according to the criterion of Golubev, the stability of triple systems with reverse motions requires higher isolation of the components.

The angles between the orbital angular-momentum vectors tend to be larger in the case of equal-mass stars than in the case of stars with unequal masses, with the asymmetry being only weakly pronounced. No other stable tendencies can be distinguished when the other initial parameters are varied.

3.4. Ejected Single Stars

The dynamical evolution of nonhierarchical multiple stars is accompanied by ejections of single and binary stars from the system.

Figure 5 presents the velocity distributions for ejected single stars as a function of the initial virial factor for systems with the mass spectrum (3). We can see that, in the case of small virial factors, the distributions reach their maximum for stars with velocities of 5–10 km/s. When the virial factor increases, the fraction of stars with high velocities decreases, and the maximum of the distribution shifts towards small velocities.

As the parameter R increases, the mean and median velocities of the stars decrease, whereas the fraction of the kinetic energy carried away by the ejected star remains essentially constant.

The influence of the initial mass spectrum of the system on the velocities of ejected single stars is substantially weaker than that of the initial size of the system. The highest velocities are reached for the initial mass spectrum (3), while stars ejected from systems with equal-mass stars possess the lowest velocities. As a rule, the ejected stars have low masses.

The above dependences of the stellar velocities on the initial size and virial factor can be understood as follows. As either the initial size or virial factor decreases, the stars' approaches become closer. Therefore, the energy transfer between the stars becomes more intense, which in turn can increase the velocities of the ejected stars.

The velocities of ejected binaries display similar trends; on average, they are a factor of 1.5–2 smaller than those of single ejected stars.

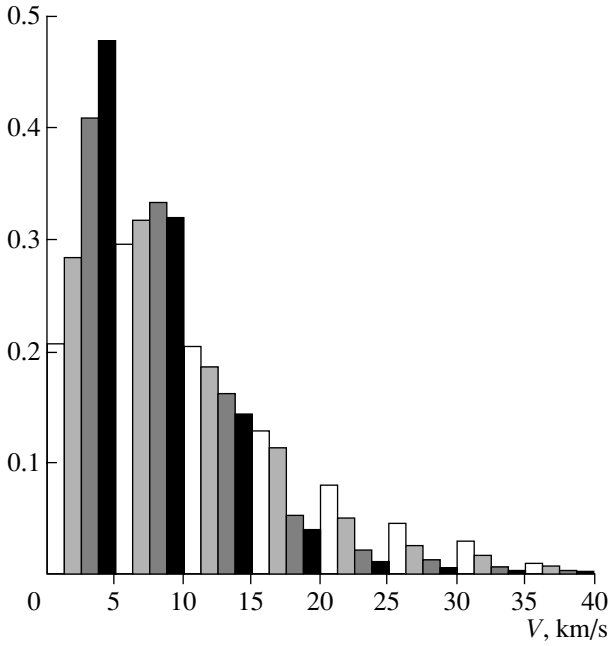


Fig. 5. Velocity distribution for ejected single stars as a function of the initial virial factor of the system k for the mass spectrum (3). The unfilled bars of the histogram correspond to $k = 0.001$, the light gray bars to $k = 0.1$, the dark gray bars to $k = 0.5$, and the black bars to $k = 0.9$.

Despite the fact that the distributions of ejection velocities display their maximum near 2–4 km/s and 1–2 km/s for single and binary stars, respectively, single stars fairly frequently leave the system with larger velocities, reaching several tens of km/s (Fig. 5; see also, for example, [9, 10]). It is possible that young T Tauri stars located at large distances from star-formation regions (see, for example, [19–21]) were formed as a result of decays of young nonhierarchical systems (see, for example, [9, 10]).

4. CONCLUSIONS

Observations of star-formation regions and simulations of star formation indicate that a rather large number of stars may form in groups with low multiplicity (see, for example, [1, 22]). The evolution of these systems may include a stage in which the gravitational interactions between the stars critically affect the subsequent evolution of the system. Such systems can be studied by numerically integrating the equations for the gravitational N-body problem taking into account the possible merging of stars in close approaches.

Here, we have studied how the results of such numerical integration depend on the initial mass spectrum, size, and virial factor of a system. In the dynamical evolution of nonhierarchical multiple systems, we can recognize the following characteristic features that are stable against variations of the initial conditions.

(1) As a rule, the evolution of a system results in the formation of a binary. The fraction of stable triple systems is high, 5–20%.

(2) The eccentricity distributions for final and ejected binaries can be described by the law $f(e) = 2e$.

(3) In stable triple systems, the eccentricities of the outer binaries tend to be smaller than those of the inner binaries ($\overline{e_{in}} \approx 0.7$, $\overline{e_{ex}} \approx 0.5$).

(4) Stable triple systems display pronounced hierarchy (the average ratio of the semimajor axes of the outer and inner binaries is 20 : 1).

(5) Triple systems with direct motions dominate.

When the initial mass spectrum, virial factor, and size of nonhierarchical multiple systems are varied, the following trends are observed.

(1) The dynamical evolution of both contracting and expanding systems is faster than in systems in a state of virial equilibrium.

(2) The final binary and triple systems for initial systems with the initial mass spectrum (3) tend to be wider than those with the other mass spectra considered.

(3) A deficiency of final binaries with large eccentricities and of stable triple systems with elongated orbits for their inner binaries is observed in systems with small initial sizes, probably due to merging of their components.

(4) When either the initial size of a system or its virial factor decreases, the velocities of ejected single and binary stars tend to increase.

ACKNOWLEDGMENTS

This work was supported by the Program of Support for Leading Scientific Schools of the Russian Federation, (grant 00-15-96775), the Russian Foundation for Basic Research (project no. 02-02-17516), and the program "Universities of Russia" of the Ministry of Education of the Russian Federation (grant UR.02.01.027).

REFERENCES

1. R. B. Larson, *IAU Symp. No. 200: The Formation of Binary Stars*, Ed. by H. Zinnecker and R. D. Mathieu (2001), p. 93.
2. A. P. Boss, Astrophys. J. **410**, 157 (1993).
3. J. K. Truelove, R. I. Klein, C. F. McKee, *et al.*, Astrophys. J. **495**, 821 (1998).
4. T. van Albada, Bull. Astron. Inst. Netherlands **19**, 479 (1968).
5. A. Burkert and P. Bodenheimer, Mon. Not. R. Astron. Soc. **280**, 1190 (1996).

6. R. S. Klessen, A. Burkert, and M. R. Bate, Astrophys. J. **501**, L205 (1998).
7. R. S. Harrington, Astron. J. **80**, 1081 (1975)
8. M. F. Sterzik and R. H. Durisen, Astron. Astrophys. **339**, 95 (1998).
9. M. F. Sterzik and R. H. Durisen, Astron. Astrophys. **304**, L9 (1995).
10. L. Kiseleva, J. Colin, B. Dauphole, and P. Eggleton, Mon. Not. R. Astron. Soc. **301**, 759 (1998).
11. M. F. Sterzik and A. A. Tokovinin, Astron. Astrophys. **384**, 1030 (2002).
12. A. V. Rubinov, A. V. Petrova, and V. V. Orlov, Astron. Zh. **79**, 1044 (2002) [Astron. Rep. **46**, 942 (2002)].
13. S. Mikkola and S. J. Aarseth, Celest. Mech. Dyn. Astron. **57**, 439 (1993).
14. E. E. Salpeter, Astrophys. J. **121**, 161 (1955).
15. B. G. Elmegreen and E. Falgarone, Astrophys. J. **471**, 816 (1996).
16. V. G. Golubev, Dokl. Akad. Nauk SSSR **174**, 767 (1967).
17. V. A. Ambartsumyan, Astron. Zh. **14**, 207 (1937).
18. J. J. Monaghan, Mon. Not. R. Astron. Soc. **176**, 63 (1976).
19. R. Neuhäuser, M. F. Sterzik, G. Torres, and E. L. Martin, Astron. Astrophys. **299**, L13 (1995).
20. M. F. Sterzik, J. M. Alcala, R. Neuhäuser, and J. H. M. M. Schmitt, Astron. Astrophys. **297**, 418 (1995).
21. M. F. Sterzik and J. H. M. M. Schmitt, Astron. J. **114**, 1673 (1997).
22. F. C. Adams and P. C. Myers, Astrophys. J. **553**, 744 (2001).

Translated by K. Maslennikov

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 52–57. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 58–63.

The Nature of Sectorial–Hemispheric Asymmetry in the Distribution of Photospheric Structures on the Quiet Sun

V. P. Mikhaĭlutsa

Kislovodsk Institute, Academy of Defensive Industries of the Russian Federation, Kislovodsk, Russia

Received April 22, 2003; in final form, July 26, 2003

Abstract—The poorly studied, spatially coherent large-scale structures on the solar surface characterized by sectorial–hemispheric asymmetry may be produced by ultralow-frequency nonlinear oscillations of the Sun. This requires that (i) the wavevector of the oscillation be normal to the rotational axis of the Sun, (ii) the radial component of the wavevector be about a factor of five greater than the component tangential to the spherical surface, (iii) the Brunt–Väisälä frequency in the oscillation-trapping region correspond to a period of ≈ 22 yr, and (iv) the sectorial number of the spatially periodic component of the oscillation be equal to three ($m = 3$). © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

During solar-activity minima, the photosphere mapped in the Bartels coordinates appears as a pattern of steady, spatially coherent structures in the distribution of bright points of sizes $1–3''$ or so. These structures cover the entire solar surface, and the pattern of their coherent distribution is known as the sectorial–hemispheric asymmetry [1]. Figure 1 shows this distribution for the minimum of 1964 as an example. The solar surface is depicted in cylindrical coordinates, and the contours of the distribution of photospheric bright points delineate regions of increased density. The sectorial–hemispheric asymmetry of such regions can clearly be seen in both equatorial and polar regions of the Sun. The structures in the equatorial region and the diametrically opposed polar regions are linked like the fingers of left and right hands. Moreover, we can trace some correspondence in the distribution of structures between the northern polar and the southern equatorial region, and between the southern polar and the northern equatorial region [2]. A 22-yr recurrence cycle was detected in the Bartels-longitude distribution of these structures, based on data for four successive solar minima (1964–1996) [1]. No physical interpretation has been found for the sectorial–hemispheric asymmetry in the distribution of such features over the solar surface.

Here, we suggest an explanation for this asymmetry based on the possibility that an ultralow-frequency nonradial oscillation with a specific direction for its wavevector is present on the Sun. The modeled pattern of the distribution of surface elements includes the above features of the sectorial–hemispheric asymmetry.

2. NONRADIAL OSCILLATIONS ON THE SUN

The physics of nonradial oscillations is described by Unno [3]. Such oscillations can be detected on a stellar surface if they have the form of standing waves; i.e., there is a region of trapped nonradial oscillations on the stellar atmosphere. The Coriolis force, gravitational force, and pressure play the major roles in exciting such waves.

The presence of such waves on the Sun remains an open question [4, 5], although some experimental evidence for their existence has recently been found [6, 7].

The dispersion relation for a low-frequency nonradial oscillation has the form [3]

$$\omega^2 \approx \frac{N^2 k_H^2 + (2\mathbf{\Omega}\,\mathbf{k})^2}{k^2}, \qquad (1)$$

where N is the Brunt–Väisälä frequency ($N^2 = \dfrac{\nabla\rho\left(\nabla\ln\rho - \dfrac{1}{\Gamma_1}\nabla\ln p\right)}{\rho}$; $\Gamma_1 = \left(\dfrac{\partial\ln p}{\partial\ln\rho}\right)_S$), $\mathbf{\Omega}$ the rotational angular velocity of the star, $\mathbf{k}$ the wavevector, k_H the horizontal (tangential to the surface) component of the wavevector, p the gas pressure, and ρ the gas density. It follows from (1) that the oscillation frequencies are generally "tied" to the rotational angular velocity of the star (the second term in the numerator).

When applied to the Sun, relationship (1) leads to two postulates.

(1) If the sectorial–hemispheric asymmetry in the distribution of surface structures is a manifestation of trapped nonradial oscillations, the second term in

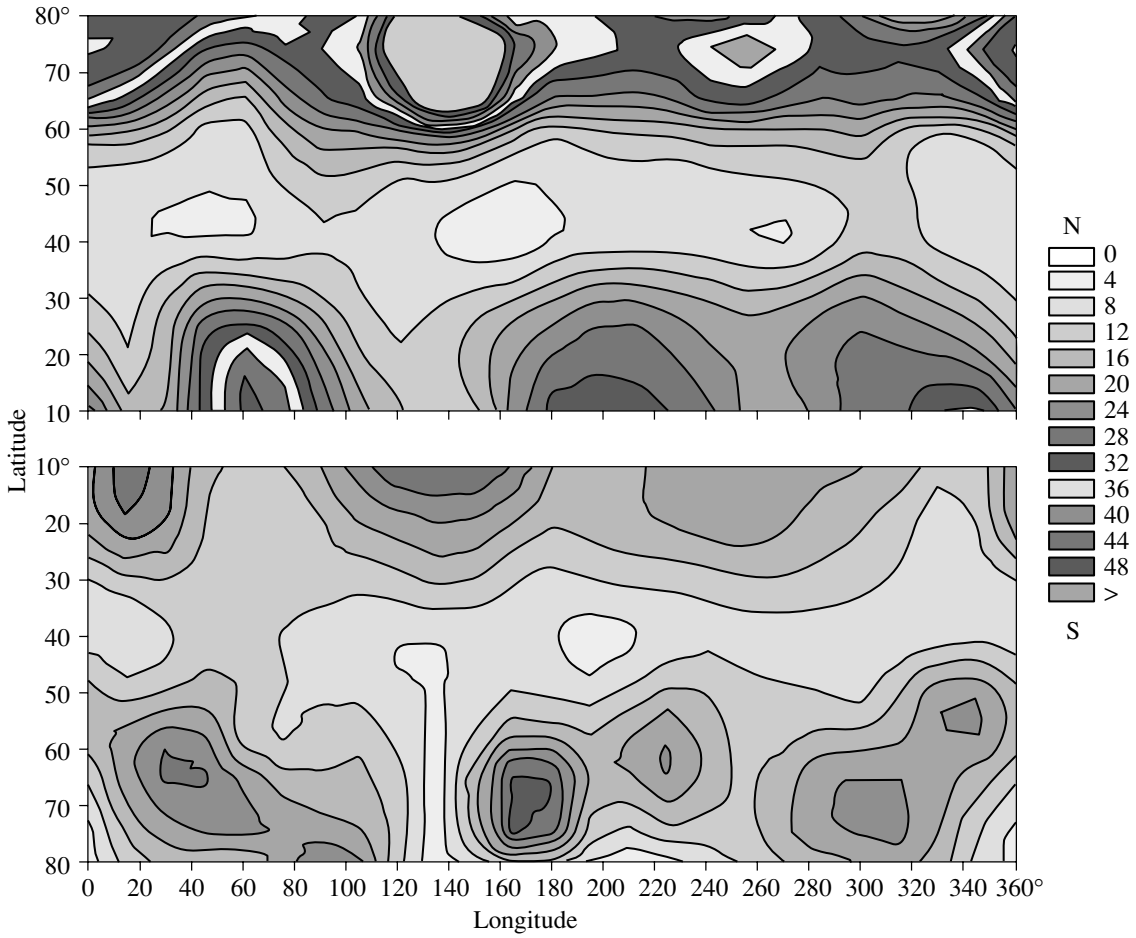


Fig. 1. Distribution of the density of photospheric bright points over the solar surface during the activity minimum of 1964. A cylindrical projection of the northern and southern hemispheres is shown. The density scale in relative units is indicated to the right.

(1) should vanish due to their 22-yr periodicity, i.e., $\mathbf{k} \perp \mathbf{\Omega}$, or $k_r \cos\theta - k_\theta \sin\theta = 0$.

(2) According to postulate (1), the frequency of the trapped waves will be determined by the Brunt–Väisälä frequency, and the factor $\frac{k_H}{k}$ should be constant throughout the trapping region ($\frac{k_H}{k} = \frac{1}{\alpha}$).

Postulates (1) and (2), together with the fact that a trapped standing wave must form an integer number of periods (m) along any latitude circle on the spherical surface, enable the derivation of the following expressions for the wavevector components of the trapped nonradial oscillation:

$$k_r = \sqrt{\frac{\alpha-1}{\alpha}}k; \quad k_\theta = \sqrt{\frac{\alpha-1}{\alpha}}k\cot\theta; \qquad (2)$$

$$k_\varphi = \sqrt{\frac{1-(\alpha-1)\cot^2\theta}{\alpha}}k; \quad k_\varphi = \frac{m}{r\sin\theta};$$

$$(kr) = \frac{m}{\sin\theta}\sqrt{\frac{\alpha}{1-(\alpha-1)\cot^2\theta}}.$$

Saio and Lee [8, 9] proposed the introduction of the displacement vector for an element of the medium, $\boldsymbol{\xi}(dr, rd\theta, r\sin\theta d\varphi)$, to describe the effect of nonradial oscillations on the medium in which they propagate, so that the linearized equations of motion for the elementary volume assume the form

$$-\rho\omega^2\xi_r - 2i\rho\omega\Omega\sin\theta\xi_\varphi = -\frac{\partial p'}{\partial r} - \rho'\frac{\partial\Psi}{\partial r},$$

$$-\rho\omega^2\xi_\theta - 2i\rho\omega\Omega\cos\theta\xi_\varphi = -\frac{1}{r}\frac{\partial p'}{\partial\theta}, \qquad (3)$$

$$-\rho\omega^2\xi_\varphi + 2i\rho\omega\Omega\cos\theta\xi_\theta + 2i\rho\omega\Omega\sin\theta\xi_r = -\frac{1}{r\sin\theta}\frac{\partial p'}{\partial\varphi},$$

where Ψ is the gravitational potential. The displacement vector $\boldsymbol{\xi}$ is very convenient for detecting regions with concentrations of surface elements on the Sun.

The magnitude and direction of this vector directly indicate the locations of such regions. In Eqs. (3), the time dependence of the displacement vector has the form $e^{i\omega t}$, as it should for a standing wave, while the spatial orientation is determined by the components ξ_r, ξ_θ, and ξ_φ.

Since the trapped nonradial oscillation is a spatially periodic wave, it is reasonable to assume that the perturbed pressure and density of the medium, p' and ρ', are also spatially periodic. We represent these perturbations in the form

$$\begin{pmatrix} p' \\ \rho' \end{pmatrix} = \begin{pmatrix} p'_0 \\ \rho'_0 \end{pmatrix} e^{\pm \mathbf{i}(\mathbf{k}\cdot\mathbf{r})} = \begin{pmatrix} p'_0 \\ \rho'_0 \end{pmatrix} e^{\pm i(k_r r)}. \quad (4)$$

We use $k_r r$ from (2) to find that the gradient of the pressure perturbation in (3) has the components

$$\frac{\partial p'}{\partial r} = 0; \quad \frac{\partial p'}{\partial \varphi} = 0; \quad (5)$$

$$\frac{\partial p'}{\partial \theta} = p'_0 im\sqrt{(\alpha-1)}\,\frac{(\alpha-1)\cos 2\theta - \sin 2\theta}{[2\sin^2\theta - (\alpha-1)\sin 2\theta]}.$$

To locate regions with concentrations of surface elements on the Sun, it is sufficient to determine the vector component ξ_θ from the system (3) using (5):

$$\xi_\theta \begin{Bmatrix} \mathrm{N} \\ \downarrow \\ \mathrm{S} \end{Bmatrix} = \left\{ \frac{1}{\rho\omega^2}\left[\frac{\partial\Psi}{\partial r}\rho'_0 \cos\left(m\sqrt{\frac{2(\alpha-1)}{2\sin^2\theta - (\alpha-1)\sin 2\theta}} \right) \sin 2\theta \right.\right.$$

$$\left.\left. - p'_0 \frac{m\sqrt{\alpha-1}}{r}\,\frac{(\alpha-1)\cos 2\theta - \sin 2\theta}{[2\sin^2\theta - (\alpha-1)\sin 2\theta]^{3/2}} \sin\left(m\sqrt{\frac{2(\alpha-1)}{2\sin^2\theta - (\alpha-1)\sin 2\theta}} \right) \right] \right\}, \quad (6)$$

$$\xi_\theta \begin{Bmatrix} \mathrm{N} \\ \uparrow \\ \mathrm{S} \end{Bmatrix} = \left\{ \frac{1}{\rho\omega^2}\left[-\frac{\partial\Psi}{\partial r}\rho'_0 \cos\left(m\sqrt{\frac{2(\alpha-1)}{2\sin^2\theta + (\alpha-1)\sin 2\theta}} \right) \sin 2\theta \right.\right.$$

$$\left.\left. + p'_0 \frac{m\sqrt{\alpha-1}}{r}\,\frac{(\alpha-1)\cos 2\theta + \sin 2\theta}{[2\sin^2\theta + (\alpha-1)\sin 2\theta]^{3/2}} \sin\left(m\sqrt{\frac{2(\alpha-1)}{2\sin^2\theta + (\alpha-1)\sin 2\theta}} \right) \right] \right\}.$$

The first and second formulas in (6) represent the meridional components of the displacement vectors in waves whose wavevectors are directed from north to south and vice versa, respectively. The azimuthal component of the displacement vector ξ_φ does not play a determining role in this case, since the number of regions with concentrations of surface elements along a latitude circle will ultimately be determined by the sectorial number m. In contrast, the positions and number of these regions along a meridional direction are important. Formulas (6) indicate that the meridional component of the displacement vector is composed of two parts. One results from the gradient of the gravitational field and the other from the gradient of the pressure field. The ratio of these parts is not known, and we will assume this ratio to be a free parameter. For this reason, we will measure the displacement on an absolute, rather than relative, scale.

Thus, we obtain a two-parameter model in which one parameter is the quantity α specifying the degree to which the wavevector of the nonradial oscillation is radial, and the other is the ratio of the contributions of the gravitational and pressure fields to the displacement vector. The displacements of the surface elements in two adjacent longitudinal sectors on the solar surface are determined by the quantities $\xi_\theta(\mathrm{N} \to \mathrm{S})$ and $\xi_\theta(\mathrm{N} \leftarrow \mathrm{S})$ in (6). Thus, the model describing the degree of concentration of the surface elements is qualitative.

3. RESULTS OF THE MODELING

We first determine the azimuthal number m from the observed pattern (Fig. 1). There are three maxima in the distribution of the density of photospheric bright points in both equatorial and polar regions; i.e., $m = 3$. The parameter α can be estimated from the

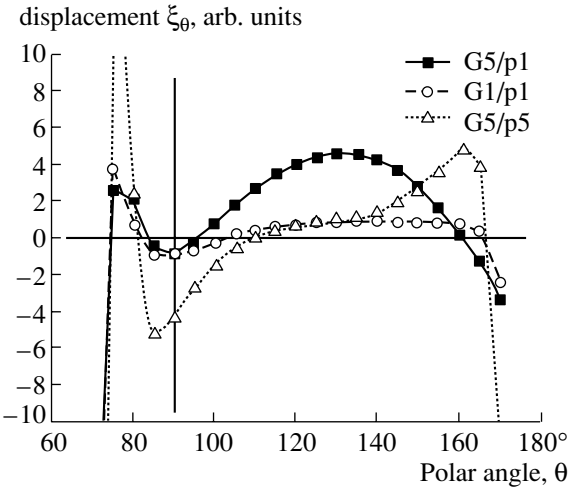


Fig. 2. The meridional displacement $\xi_\theta(\mathrm{N} \to \mathrm{S})$ for $m = 3$ and $\alpha = 5$ as a function of the ratio of the contributions of the gravitational (G) and pressure (p) fields. Shown are the cases when the contribution of gravity is a factor of five greater than that of the pressure (squares), the two contributions are equal (circles), and the contribution of the pressure is a factor of five greater than that of gravity (triangles).

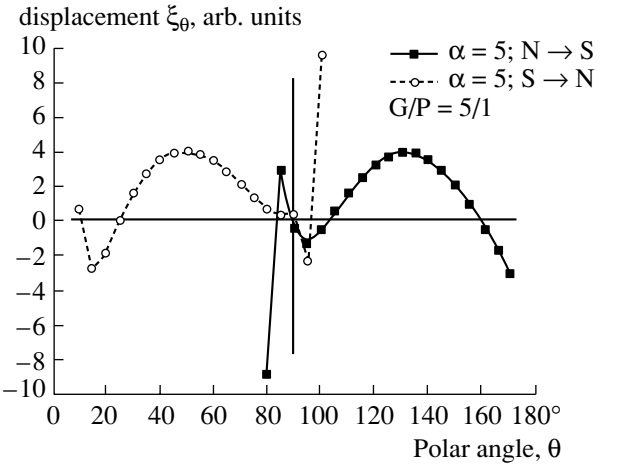


Fig. 3. Meridional displacement of surface elements ξ_θ as a function of the polar angle for waves propagating in opposite directions along the meridian. The parameters of the waves are indicated at the top right.

details of the pattern of hemispheric asymmetry in the distribution of regions with high concentrations of points. If $1 < \alpha < 2$, formulas (6) indicate that ξ_θ is distributed nearly symmetrically about the equator, irrespective of the relative contributions of the gravitational and pressure fields. In turn, this results in a nearly symmetrical distribution of the regions with concentrations of surface elements. Indeed, if the component ξ_θ is, for example, positive in the northern and negative in the southern hemisphere, the surface elements will move toward the equator in both hemispheres. Otherwise, the displacement will be poleward in both hemispheres.

However, the observations reveal a different pattern: the displacement appears to be toward the pole in the northern hemisphere and toward the equator in the southern hemisphere within a given sector, and vice versa in a contiguous sector (Fig. 1). In view of this, preliminary modeling yielded the estimate 5 ± 1 for α. The hemispheric asymmetry of the displacements is well defined starting from $\alpha = 3$.

The relative contributions of the gravitational and pressure fields to ξ_θ can be judged from Fig. 2, where the meridional displacement $\xi_\theta(\mathrm{N} \to \mathrm{S})$ is shown as a function of the polar angle for three relative contributions of these two fields. The squares correspond to the case when the contribution of the gravitational force to the displacement is a factor of five larger than that of the pressure, the triangles to the opposite situation, and the circles to equal contributions of gravity and pressure.

We can see that a predominance of the gravitational force mainly results in a "depletion" at middle latitudes. The surface elements are carried from there to the polar region 20° from the pole. The predominance of the pressure mainly depletes the equatorial belt. The surface elements drift from the equatorial belt to regions located 15° from the equator. Figure 1 indicates a definite depletion at middle latitudes. We take this to indicate the predominance of the gravitational force in the generation of ultralow-frequency nonradial oscillations.

A graphical summary of the relative values of the meridional component of the displacement vector as a function of the polar angle for waves traveling over the Sun from both north to south and in the opposite direction is given in Fig. 3 for certain parameter values. We can see that the wave $\mathrm{N} \to \mathrm{S}$ originates in the equatorial region of the northern hemisphere, intersects the equator, and reaches the southern pole. The wave $(\mathrm{N} \leftarrow \mathrm{S})$ emerges in the equatorial region of the southern hemisphere, intersects the equator, and reaches the northern pole. Each wave should propagate within its own longitudinal sector. This links the surface structures in a polar region of the Sun to structures in the equatorial region of the opposite hemisphere. Such a coherence of structures can clearly be seen in Fig. 1. The wave $\mathrm{N} \to \mathrm{S}$ forms regions with concentrations of surface elements in the equatorial region of the northern hemisphere (where the sign of the displacement ξ_θ changes from positive to negative) and in the polar region of the southern hemisphere. The wave $\mathrm{N} \leftarrow \mathrm{S}$, which propagates in the adjacent sector, forms high-concentration regions in the equatorial region of the southern hemi-

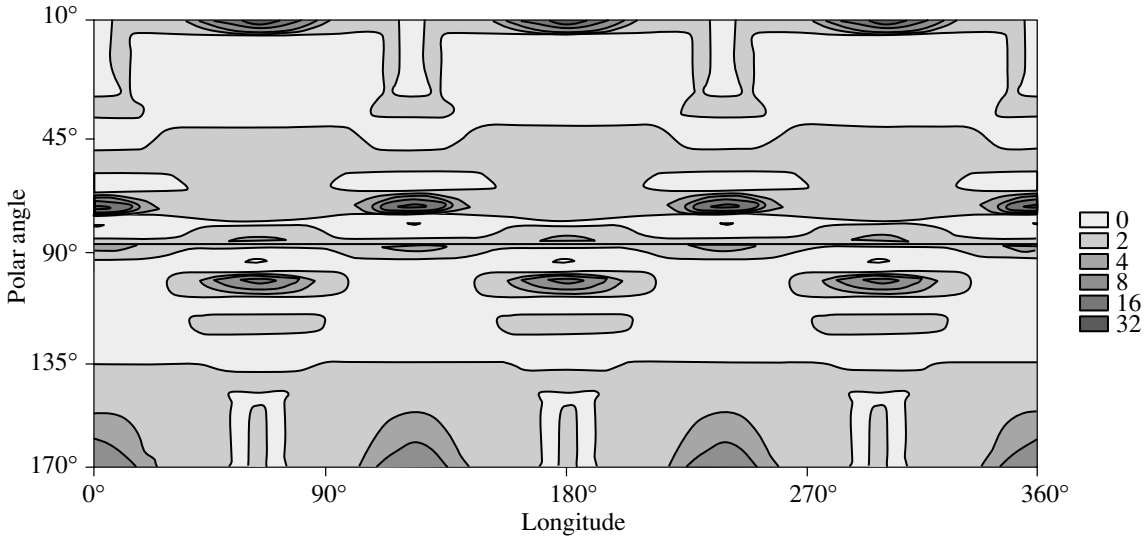


Fig. 4. Modeled distribution of the regions with concentrations of surface elements produced by ultralow-frequency nonradial oscillations in cylindrical coordinates. The sectorial number is $m = 3$, the wavevector $\mathbf{k}$ is mainly radial ($k/k_H = 5$), and the gravitational force dominates over the pressure (G/p = 5). The scale for the concentration (in relative units) is shown at the bottom right.

sphere and the polar region of the northern hemisphere.

We used the dependence presented in Fig. 3 to construct a two-dimensional map of the distribution of regions with concentrations of surface elements. To this end, we computed the derivative $\dfrac{\Delta\xi_\theta}{\Delta\theta}$. We can easily see that negative values of this derivative mark regions with concentrations of surface elements, while positive values mark regions with a depletion of surface elements. The modeled distribution of high-concentration regions is shown in Fig. 4.

4. DISCUSSION

Under the assumptions made, the modeled pattern of the concentration of surface elements clearly demonstrates the main features of sectorial–hemispheric asymmetry. Thus, we have shown that ultralow-frequency nonradial oscillations can, in principle, produce regions with concentrations of surface elements on the Sun, and the distribution of these regions exhibits a spatially coherent asymmetry. However, some details of the observed distributions remain unaccounted for.

In particular, our model relates the numbers of equatorial and polar regions with concentrations of bright photospheric points in the two hemispheres. The numbers of such points should be equal and, in a true realization of hemispheric asymmetry, odd. The distribution observed in 1964 (Fig. 1) truly satisfies this requirement. However, the distributions for other epochs (1975, 1985–1986, 1996) [1] are not so consistent. Distributions with $m = 3$ can be detected in the equatorial region of the Sun at these epochs, whereas up to five high-concentration zones can be found in the polar region. The reason for this remains unknown.

Another peculiarity is related to the overall pattern of the distributions near the equator. They are sometimes well-defined (1964, 1985–1986), and sometimes smeared to some extent (1975, 1996). These features suggest that the real situation is complicated by some effects that have not been taken into account.

The solar surface rotates differentially. This was not included in the model for two reasons. First, the coherent asymmetry in the distribution of regions with concentrations of surface structures was manifest in a rigidly rotating system of longitudinal coordinates with a rotational period of 27.0^{d}. Second, physical arguments suggest that the region of trapping of the traveling waves of pulsations cannot rotate differentially, and should be located in rigidly rotating layers of the Sun.

To conclude, our modeling has demonstrated that low-frequency nonradial oscillations can exert an organizing effect on spatially coherent surface structures on the Sun. This is possible if the wavevector of the oscillation is essentially normal to the rotational axis of the Sun. The radial component of the wavevector must dominate in terms of its magnitude. The sectorial number in the spatially periodic component of the wave must be equal to three. The Brunt–Väisälä frequency in the trapping region should correspond to the observed period for variation of the spatial coherence (≈ 22 yr).

The presence of ultralow-frequency nonradial oscillations enriches the spectrum of models describing the cyclic magnetic activity of the Sun. In particular, we cannot rule out the possibility that the cyclic formation of sunspots is controlled by these oscillations.

REFERENCES

1. V. P. Mikhailutsa, Solar Phys. **199**, 13 (2001).
2. V. P. Mikhailutsa and V. V. Makarova, Astron. Zh. **77**, 554 (2000) [Astron. Rep. **44**, 487 (2000)].
3. W. Unno *et al.*, *Nonradial Oscillations of Stars* (Univ. Tokyo, Tokyo, 1989).
4. C. I. Wolf and J. B. Blizard, Solar Phys. **105**, 1 (1986).
5. M. Woodard, Astrophys. J. **347**, 1176 (1989).
6. V. P. Mikhailutsa, A. D. Wittmann, and M. Bianda, Astron. Astrophys. Trans. **19**, 79 (2000).
7. V. P. Mikhailutsa and V. V. Makarova, Izv. Ross. Akad. Nauk, Ser. Fiz. **63**, 2152 (1999).
8. H. Saio, Astrophys. J. **256**, 717 (1982).
9. U. Lee and H. Saio, Astrophys. J. **491**, 839 (1997).

Translated by A. Getling

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 58–64. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 64–71.

The Pattern of Supergranular in a Solar Active Region and the Formation of Filaments

V. M. Grigoryev, L. V. Ermakova, and A. I. Khlystova

Institute for Solar–Terrestrial Physics, P.O. Box 4026, Irkutsk, 664033 Russia

Received April 3, 2003; in final form, August 8, 2003

Abstract—The formation of filaments in solar bipolar active regions is investigated, giving particular attention to the relationship between this process and the pattern of supergranular convection. SOHO MDI and Kitt Peak magnetograms and Hα filtergrams are used. The large decaying active region NOAA 8525 is considered over the period May 4–7, 1999. The boundaries of supergranules are identified as concentrations of the line-of-sight photospheric field in magnetograms. Filaments in the central part of the active region are studied; as a whole, they are aligned with the supergranule boundaries. Variations in the magnetic field in this period were manifest primarily in the form of "cancellations" and spatial-redistribution processes consistent with the pattern of developing supergranules. These factors created the conditions necessary for the formation of a filament stretched across the entire active region; i.e., the straightening of the polarity-inversion line and reduction of the horizontal gradients of the magnetic field. One possible explanation of the results is that the magnetic-field component along the filament axis is associated with the vortical structure of horizontal flows in the supergranulation cells. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Quiescent solar filaments lie over polarity-inversion lines (PILs) of the large-scale magnetic field. Studies of the structure of filaments and the surrounding regions in Hα (see [1–3] and many other papers) suggest that the magnetic skeleton of a quiescent filament forms a system of arches hooked together, with their ends protruding from the body of the filament and going down into the photosphere (the so-called "feet"). These are clearly visible in large filaments. The feet issue from most filaments to the right in the northern hemisphere and to the left in the southern hemisphere, irrespective of the side from which the filament is viewed. The first type of filaments are called dextral and second type sinistral [2]. These two types of filaments are associated with two directions of the twisting of the filament's magnetic field—right-handed and left-handed. The structure of filaments suggests that their magnetic fields should have an appreciable axial component. The structure of the chromosphere indicates that the magnetic field in the immediate vicinity of a filament is directed along the filament [1]. Direct measurements of the photospheric magnetic field with moderate spatial resolution indicate that the transverse component of the magnetic field near a filament is mainly directed along the filament, irrespective of whether the filament is quiescent [4, 5] or located in an active region [6]. The formation and stability of such systems can be provided by shear motions.

There is a relationship between the position of the filament and the pattern of the supergranular convection. Based on his analysis of the distances between the feet of filaments, Sykora [7] suggested that the feet are anchored between supergranules, and Plocienia and Rompolt [8] confirmed this result. These studies gave no information about whether the filaments were completely aligned with the boundaries of supergranules or whether they crossed the supergranular cells while being supported in the areas between them. Grigoryev and Ermakova [4, 5] reconstructed the supergranular pattern near a quiescent filament from the line-of-sight velocity field measured with the vector magnetograph of the Sayan observatory, and concluded that the filaments are mainly located over supergranular boundaries.

Investigations of the physical conditions in the solar atmosphere necessary for the formation of quiescent filaments can be considerably aided by studies of filaments in active regions, since they develop in a medium with a strong, rapidly varying magnetic field, making the factors affecting the formation and stability of filaments more clearly manifest. If an active region is large, it will be associated with a pair of cells with positive and negative polarities in maps of the large-scale magnetic field. The filament formed in an active region during the "spotted" stage of its development can be observed over several solar rotations at the PIL of the large-scale magnetic field after the disappearance of the spots.

1063-7729/04/4801-0058$26.00

Maksimov and Ermakova [9, 10] showed that a filament can originate and remain stable in an active region if the PIL is sufficiently long and has a simple shape, and if the magnitudes of the horizontal magnetic-field gradients transverse to the boundary lie within a certain range (Wang and Li [6] came to a similar conclusion). These conditions are satisfied for the PIL separating the leading and following parts of an active region only in the decay stage of the active region. Isolated filamentary fragments can first appear at segments of the PIL where the above conditions are already satisfied, with a full filament forming as the bipolar active region ages. As the active region decays, the mean magnetic-flux density decreases and the PIL straightens. The evolution of the distribution of the line-of-sight magnetic field that is not associated with spots in a bipolar active region can be described analytically over a fairly long period, using some evolutionary parameter that depends on the horizontal gradient of the line-of-sight field averaged over the active region [11]. Therefore, there is a threshold value for the development phase of a bipolar active region, starting from which the horizontal gradients of the line-of-sight magnetic field near the PIL lie within the range favorable for the formation of a filament.

Grigoryev and Ermakova [12] found that a filament in an active region appears after the total spot area has been reduced by about 40%. As the active region evolves, not only the ratio of the magnetic fields of different polarities but also the spatial distribution of the magnetic-field concentration changes. Annular structures delineating the supergranular convection cells are visible in magnetograms of the line-of-sight field in old active regions. Based on their analysis of the line-of-sight velocity field, Grigoryev and Selivanov [13] demonstrated that the supergranular convection is preserved during the formation of a bipolar active region. As a rule, the PIL has a complex shape at this stage, and should intersect supergranules. It is possible that the PIL becomes aligned with the boundaries of convection cells during the decay of the active region and the simplification of the magnetic-field configuration, and that this is one of the necessary conditions for the formation of filaments.

Here, we continue our studies of the development of filaments in bipolar active regions, paying particular attention to elucidating the relationship between this process and the pattern of supergranular convection.

2. OBSERVATIONAL MATERIAL AND THE OBJECT OF OUR STUDIES

The observational material used consists of SOHO MDI and Kitt Peak magnetograms in FITS format and Hα filtergrams from various observatories in GIF format. We considered the active region NOAA 8525, which appeared from beyond the limb on April 29, 1999. According to *Solar Geophysical Data,* the group reached its maximum area (more than 300 mph) on May 2 or 3; its extent was about 20° and its magnetic configuration bipolar. The group consisted of a large leading spot of positive polarity and small spots of both polarities. Subsequently, the area of the group decreases, with the size of the leading spot varying only slightly. During the next rotation, a small spot of positive polarity was situated in the location of the active region, and had decayed by June 6.

This active region had a heliographic position convenient for investigations of the supergranular structure and magnetic field on May 4–6, when its center was in the longitudinal interval E15–W13 at a latitude of N20. We used full-disk Kitt Peak magnetograms for May 4–7 taken once a day and SOHO MDI magnetograms for May 4, 5, and 7 taken at 96-min intervals (the data for May 6 on the website are erroneous). The Kitt Peak magnetograms have a spatial resolution of 1″ and a sensitivity of 5 G; 55 minutes were required to obtain each magnetogram. The spatial resolution of the MDI magnetograms was about 2″, and their sensitivity 30 G. Daily Big Bear Solar Observatory Hα images available on the website and the Kitt Peak magnetograms were obtained at similar times. We estimate that the quality of the best filtergrams enables location of the ends of filaments to within 3–6″. We considered only filaments lying on the PIL separating the leading and following portions of the active region. The high sensitivity of the Kitt Peak spectromagnetographs ensured high accuracy for the location of the PIL. The boundaries of supergranules were identified in the MDI magnetograms as concentrations of the line-of-sight field in the photosphere. We used the IDL software package to rotate the magnetic images of the Sun to the position of the central meridian in order to directly superpose the images and eliminate the effect of foreshortening. Further, we cut out a $400'' \times 320''$ area that included the entire active region and its neighborhood. The resulting magnetograms were viewed in a slide-show regime. The supergranules are not always clearly defined in the magnetograms; however, the entire time sequence of magnetograms makes it possible to delineate the supergranules fairly confidently, since their lifetimes are substantially longer than the interval between successive MDI magnetograms, and the magnetic fields move along the supergranular boundaries in the course of the evolution.

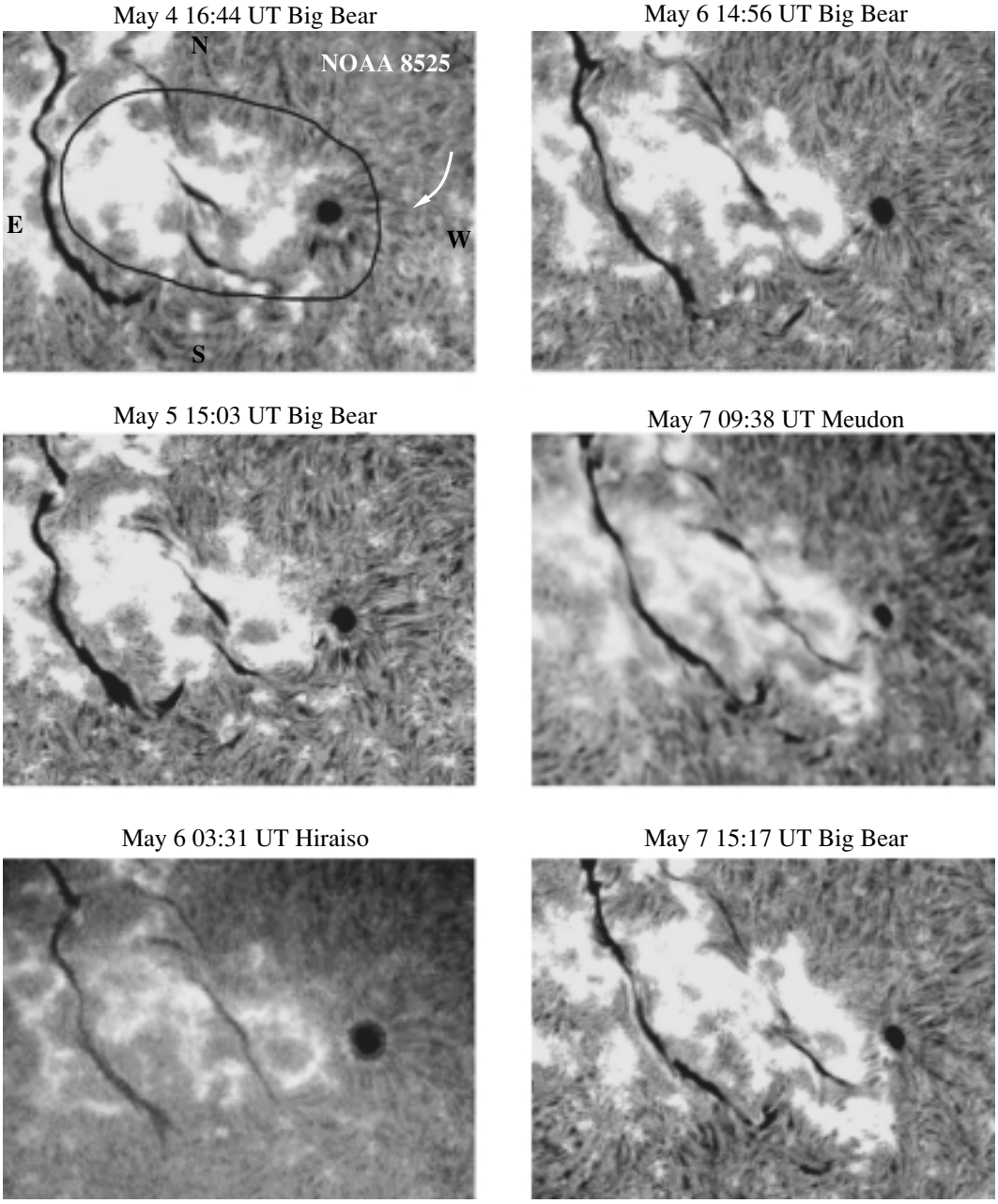


Fig. 1. Hα filtergrams.

3. RESULTS AND DISCUSSION

Figure 1 shows fragments of Hα filtergrams that include the active region NOAA 8525, while Fig. 2 presents Kitt Peak magnetograms of the line-of-sight field. The filtergrams were obtained with various instruments, making the series appear nonuniform. We consider only filaments located within the active region. These could be observed without difficulty from May 3 to 9; later, they were close to the limb and difficult to identify. The total extent of the filaments was no less than half that of the PIL, and the lengths of the filaments varied. The PIL separating the leading and following parts of the active region was wavy until May 6, and two filaments that formed near two parallel segments of the PIL deviated substantially from the meridional direction. Later, the PIL was straight and formed a small angle with the meridian. On May 6 at 03:31 UT, a single filament stretched over the entire active region, but subsequently fragmented. Thus, a filamentary channel crossing the entire active region had formed by May 6. By that time, the total spot area in the group was reduced by at least 45%, in agreement with the previously obtained result of [12].

The supergranule contours are superposed on

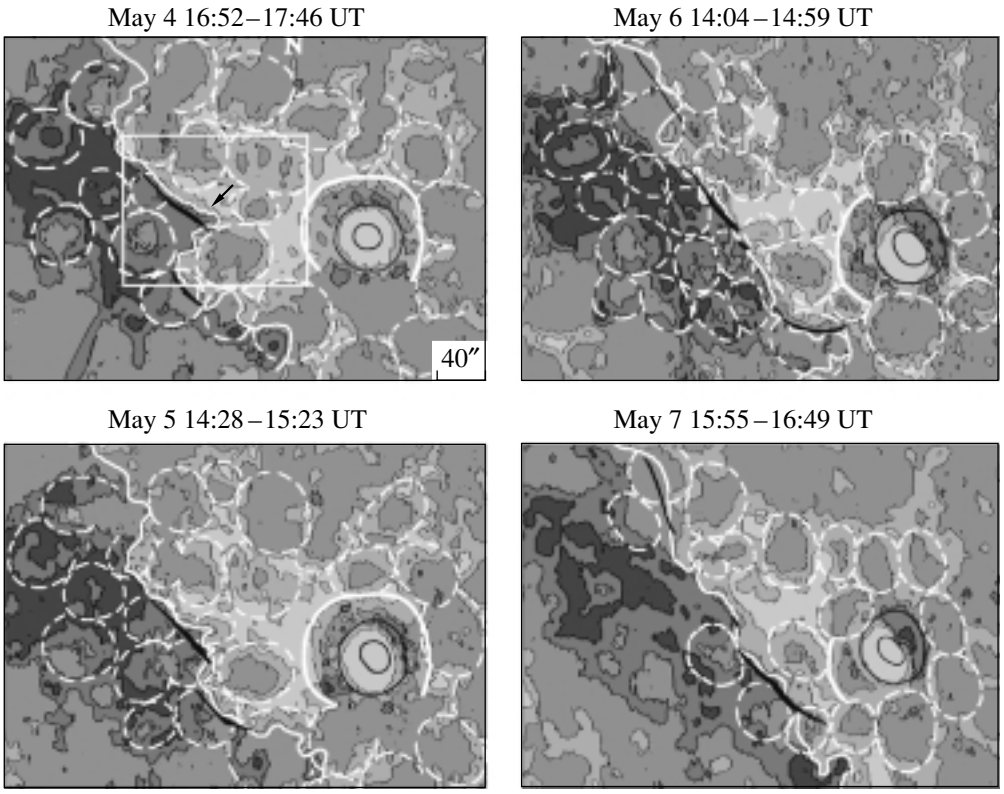


Fig. 2. Kitt Peak maps of the line-of-sight magnetic field with contours of 20 and 180 G. Filaments located using the Big Bear filtergrams and the contours of the umbra and penumbra of the leading spot are shown in black. The polarity-inversion lines (light solid), supergranule contours (dashed), and contours corresponding to the annular convection cell ("moat") around the leading spot (heavy solid) are shown in white. For other notation, see text.

maps of the line-of-sight magnetic field in Fig. 2. We can see that the filaments in the active region lie near supergranular boundaries and end at places where they join. The lengths of the filaments are initially one to two supergranular diameters and reach three diameters at later times. The available series of MDI magnetograms enables us to analyze the behavior of the supergranular pattern on May 4–5, when the PIL was straightening at the center of the active region.

Figure 3 shows fragments of the magnetograms of the central part of the active region. We can see that three supergranules formed during this period—two with the field of the following polarity and one with the field of the leading polarity. These are labeled 1–3 in order of their appearance. New supergranules formed at the places where the old ones joined in the immediate vicinity of the filament, which remained at its position near the supergranular boundary. Supergranule 2 originated in a high-density magnetic field of the leading polarity, and the pattern of field variations was more pronounced in this case. The first signatures appeared on May 4 at 22:24 UT, as a concavity of both of the magnetic-field contours, and a significant portion of the contour became well defined within three hours. Subsequent changes were manifest as an extension of the area free of magnetic field within the supergranule and the propagation of the leading-polarity magnetic field along the contour of the supergranule toward the PIL. The diameter increased simultaneously. The leading-polarity field outlined the entire supergranular contour within a day. In their studies of the relationship between the photospheric velocity field and the magnetic-field distribution over the solar surface, Simon *et al.* [14] noted that small magnetic elements on the quiet Sun move toward supergranular boundaries, then drift slowly along these boundaries. Supergranules 1 and 3 originated in the area of the following polarity. According to Fig. 3, the following-polarity magnetic field also drifts along the southern boundary of supergranule 3 toward the PIL. We suppose that the same was true for supergranule 1 at magnetic-field levels below 30 G. Thus, the formation of supergranules near the PIL results in the "convergence" of magnetic fields of opposite polarities, which, according to [15], is a necessary condition for the formation of a filament.

At the same time, the straightening of the PIL is associated with a weakening of the line-of-sight magnetic field and a decrease in the horizontal gradients of the field (Figs. 2, 3). This is due to numer-

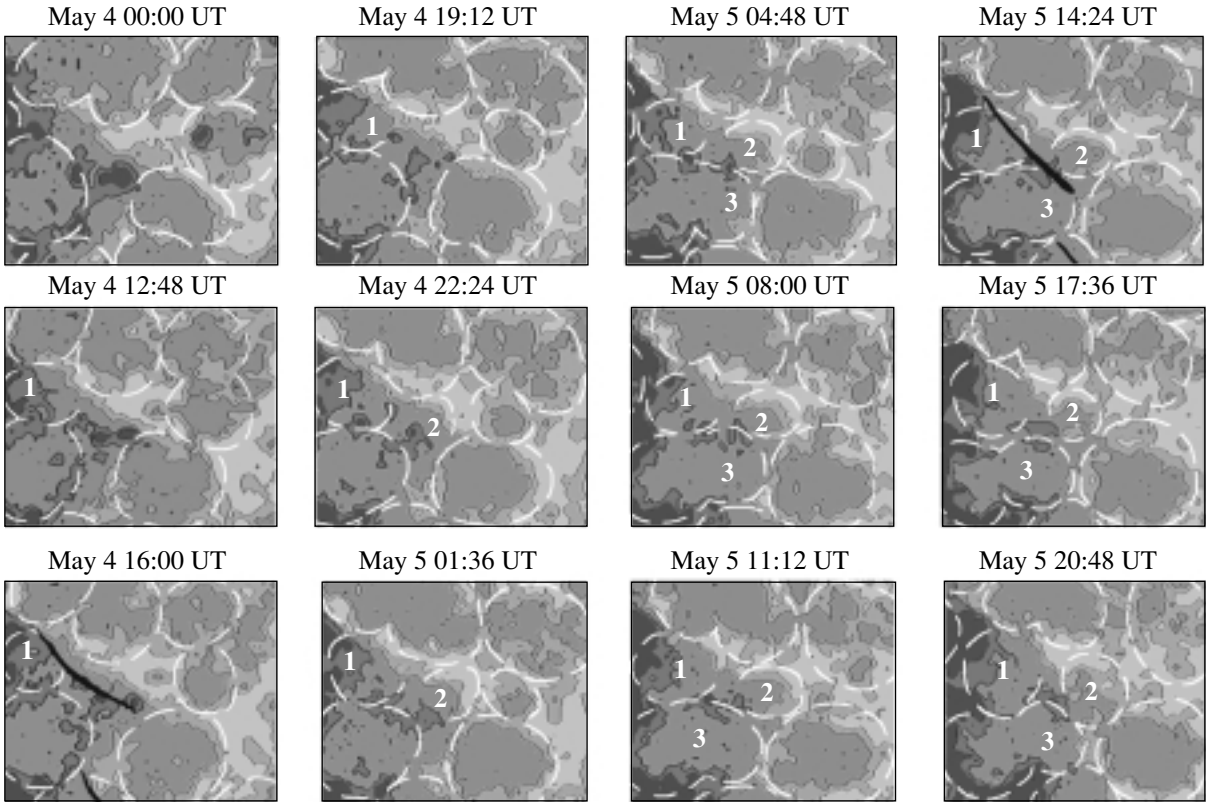


Fig. 3. MDI magnetograms of the central part of the active region marked in Fig. 2. Contours of 30 and 150 G are shown. The notation is the same as in Fig. 2.

ous magnetic-field "cancellation" events. The figures show the presence of field cancellation on May 4 and 5 along the entire PIL, with the cancellation being especially violent near the northern filament, where the magnetic-flux density was higher. One such case is illustrated in Fig. 4; this took place on May 4 to the north of the western part of the northern filament (this region in the active region is shown by an arrow in Fig. 2). The negative pole that underwent cancellation was local, while the opposite pole could not be identified with certainty, except near 16:00 UT; this is typical of cancellation events in the center of an active region [16]. Otherwise, the pattern was quite regular and included a convergence of the poles, increases in the magnetic-field gradient at the site where they came into contact, and the gradual disappearance of the negative pole. The center of supergranule 2 was observed at this location on the following day. North of this filament, cancellation events were observed during the intervals 01:36–12:48 and 17:36–22:24 UT. On May 5, this region corresponded to the center of the new supergranule 3. After the formation of new supergranules, cancellation events occurred when opposite poles moving along the supergranular boundaries toward the PIL collided. Both the very existence of filaments and their instability are associated with cancellation [16, 17]. They may also play a role in generating the helical structure of the magnetic fields of filaments.

The changes in the convection pattern near the leading spot associated with the evolution of the active region are of interest (Fig. 2). During the initial period, on May 4 and 5, a "moat" with an angular extent of about 270° can be seen around the leading spot in the line-of-sight magnetograms. This is a region of reduced line-of-sight field crossed by small magnetic elements of both polarities moving from the spot toward the surrounding fields [18, 19]. This feature is an annular convection cell that forms around a well-developed spot and transfers excess heat outward from the base of the spot, thereby stabilizing it [20, 21]. The moat around the leading spot broke down gradually over the next two days; on May 7, the arc of the moat was about 60°. The supergranular convection in the immediate vicinity of the spot was simultaneously reestablished (Fig. 2). No substantial changes in the shape and area of the leading spot occurred during that period. Moats are primarily associated with decaying spots; however, many reports of moats around growing spots can also be found in the literature. The formation of a moat at the end of the growth stage of main spots whose area exceeded

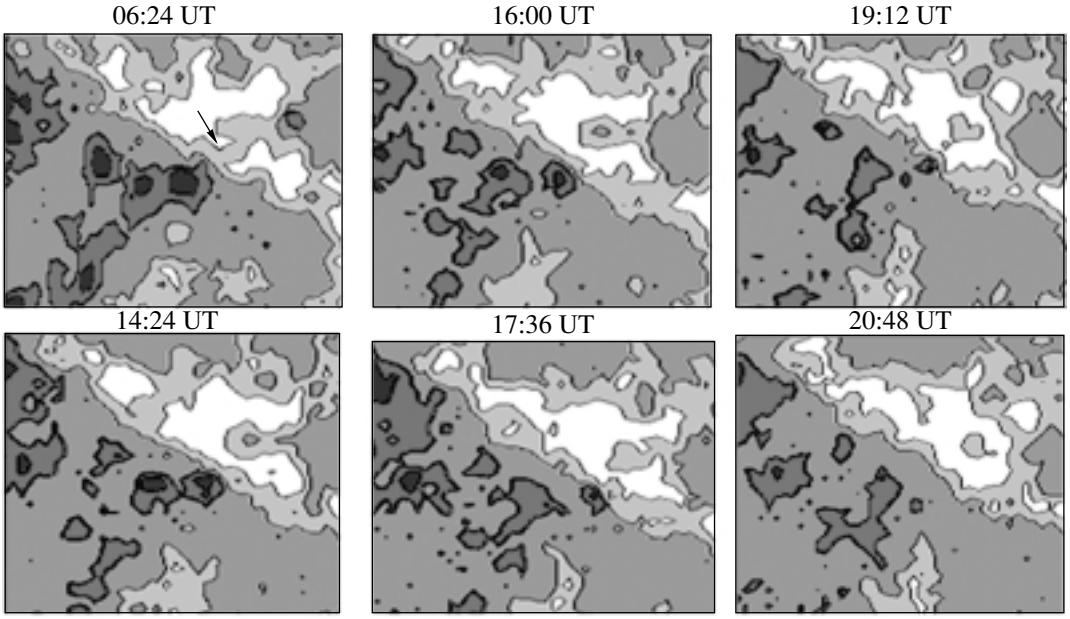


Fig. 4. MDI magnetograms (contours of 30 and 150 G are shown) for May 4 illustrating the cancellation of magnetic flux.

that of a supergranule is described in [13, 22]. It is likely that the development of a large spot ends with the formation of such a cell. In the case considered here, the moat was first observed on May 3, near the maximum development of the active region. This suggests that the formation and decay of a moat not only determine a spot state, but also characterize the intensity and state of the entire flux tube forming the active region; i.e., they represent a characteristic phase in the growth of the active region.

Thus, in the decay stage of a bipolar active region, magnetic flux is lost due to cancellation in the central part of the region at the line where the magnetic fields of opposite polarities come into contact; simultaneously, motions in the newly forming supergranules result in a spatial redistribution of the flux. These factors straighten the PIL and reduce the horizontal gradients of the magnetic field. An extended filamentary channel forms near the PIL. The developing filaments are aligned with supergranular boundaries.

A number of observational conditions required for the formation and existence of a filament are known. We emphasize here three of these that are directly related to the structure of the magnetic field in the filament.

(1) Filaments are located along the line separating different magnetic polarities.

(2) A magnetic-field component associated with filaments and aligned with the PIL is present; this is indicated by the structure of the fibrils in chromospheric rosettes near filaments and filamentary channels [2], as well as by direct measurements of the photospheric magnetic fields [4–6].

(3) The supergranular cells near a filament form an almost regular pattern, with supergranules arranged along the PIL [4, 5] (this can also be seen from our results).

To explain these observational facts, we suggest that the formation of the component of the magnetic

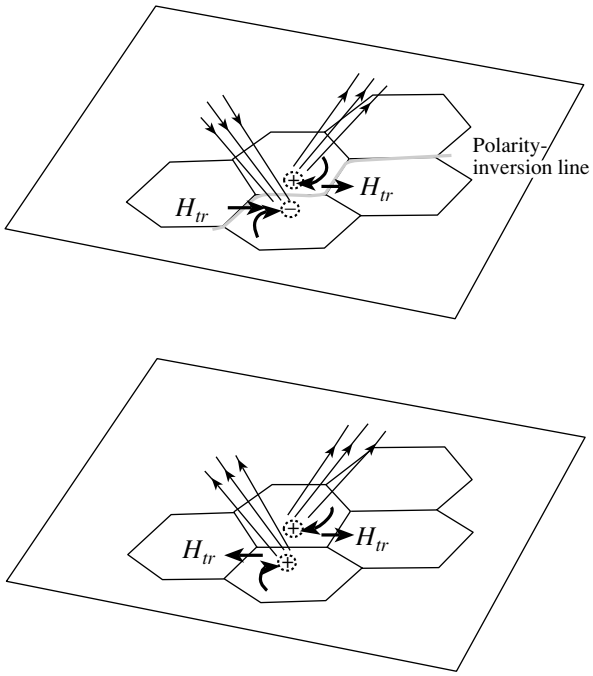


Fig. 5. Schematic of the formation of the axial component of the magnetic field under the action of the vortical component of the motions in supergranules.

field aligned with the PIL is related to the vortical structure of the horizontal flows in supergranular cells (Fig. 5). The vortical structure in a supergranule results from the action of the Coriolis force on the horizontal divergence of the flow. If supergranules are arranged along a PIL, the vortical flows from neighboring supergranules converging to this line squeeze together the magnetic fields of opposite polarities, leading to the disappearance of the vertical component of the field; the observed magnetic-flux cancellation events provide evidence for this effect. This reduces the magnetic-field gradient at the PIL (so that the necessary condition for the formation of filaments is satisfied [9]). At the same time, the magnetic field component aligned with the PIL is increased, since the vortical component of the flows in the neighboring supergranules has opposite directions on either side of the PIL. This ensures the presence of a filament-aligned component of the field. This hypothesis is supported not only by our results concerning the structure of supergranular cells, but also by a theoretical consideration of cellular convection in a rotating fluid [23]. The first evidence for the rotation of cells was presented by Kubičela [24]. Using modern techniques, Duvall and Gizon [25] recently confirmed the effect of the Coriolis force on supergranular flows. They estimated the quantity curl/div and found that it varies with latitude from 0 at the equator to ± 0.07 at latitudes of $\pm 60°$. In addition, studies of motions in supergranules based on the advection of mesogranules [26] yield horizontal velocities of 500–1000 m/s, which provides an energetically more convincing argument for the possibility that the flows affect the magnetic field, compared to the velocities of 350–500 m/s determined earlier.

Our hypothesis requires further development. However, it can also provide a natural explanation for certain additional features, such as the disappearance of magnetic knots at the polarity-inversion line, the lack of filaments in the equatorial zone, and the presence of intense filaments at high latitudes.

ACKNOWLEDGMENTS

This work was supported by the program "Leading Scientific Schools of Russia" (grant NSh-733.2003.2) and the Federal Science and Technology program "Astronomy."

REFERENCES

1. P. Foukal, Solar Phys. **19**, 59 (1971).
2. S. F. Martin, R. Bilimora, and P. W. Tracadas, *Solar Surface Magnetism,* Ed. by R. J. Rutten and C. J. Schrijver (Kluwer, 1994), p. 303.
3. S. F. Martin, Solar Phys. **182**, 107 (1998).
4. V. M. Grigoryev and L. V. Ermakova, *Current Problems in Solar Cyclicity,* Ed. by V. I. Makarov, V. N. Obridko, *et al.* (GAO RAN, St. Petersburg, 1997), p. 317 [in Russian].
5. V. M. Grigoryev and L. V. Ermakova, Astron. Astrophys. Trans. **17**, 355 (1999).
6. Wang and W. Li, *New Perspectives of Solar Prominences,* Ed. by D. Webb, D. Rust, and B. Schmieder; ASP Conf. Ser. **150**, 98 (1998).
7. J. Sykora, Bull. Astron. Inst. Czechosl. **19**, 37 (1968).
8. S. Plocienia and B. Rompolt, Solar Phys. **29**, 399 (1973).
9. V. P. Maksimov and L. V. Ermakova, Astron. Zh. **62**, 558 (1985) [Sov. Astron. **29**, 323 (1985)].
10. V. P. Maksimov and L. V. Ermakova, Astron. Zh. **64**, 841 (1987) [Sov. Astron. **31**, 438 (1987)].
11. L. V. Ermakova, Solar Phys. **191**, 161 (2000).
12. V. M. Grigoryev and L. V. Ermakova, *New Solar Activity Cycle: Observational and Theoretical Aspects,* Ed. by V. I. Makarov, V. N. Obridko, *et al.* (GAO RAN, St. Petersburg, 1998), p. 233 [in Russian].
13. V. M. Grigoryev and V. L. Selivanov, Contrib. Astron. Observ. Skalnate Pleso **15**, 87 (1986).
14. G. W. Simon, A. M. Title, K. P. Topka, *et al.*, Astrophys. J. **327**, 964 (1988).
15. S. Martin, *Dynamics of Quiescent Prominences,* Ed. by Ruzdjak and E. Tandberg-Hanssen (Springer-Verlag, 1990); Lect. Notes Phys. **363**, 1 (1990).
16. S. F. Martin, S. H. B. Livi, and J. Wang, Austral. J. Phys. **38**, 929 (1985).
17. J. Wang, Z. Shi, and S. F. Martin, Astron. Astrophys. **316**, 201 (1996).
18. D. Vrabec, *IAU Symp. No. 56: Chromospheric Fine Structure,* Ed. by R. G. Athey (Reidel, Dordrecht, 1974), p. 201.
19. K. Harvey and J. Harvey, Solar Phys. **28**, 61 (1973).
20. F. Meyer, H. U. Scmidt, N. D. Weise, *et al.*, Mon. Not. R. Astron. Soc. **169**, 35 (1974).
21. A. Nye, D. Brunning, and B. J. Labonte, Solar Phys. **115**, 25 (1988).
22. L. V. Ermakova, *The Magnetic and Velocity Fields of Solar Active Regions,* Ed. by H. Zirin, G. Ai, and H. Wang; ASP Conf. Ser. **46**, 75 (1993).
23. G. Veronis, J. Fluid Mech. **5**, 401 (1959).
24. A. Kubicela, *Solar Activity and Related Interplanetary and Terrestrial Phenomena* (Proc. 1st European Astron. Meeting), Ed. by J. Xanthakis (Springer, Berlin, 1973), p. 123.
25. T. L. Duvall, Jr. and L. Gizon, Solar Phys. **192**, 177 (2000).
26. R. A. Shine, G. W. Simon, and N. E. Hurlburt, Solar Phys. **193**, 313 (2000).

Translated by A. Getling

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 65–79. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 72–87.

Cyclotron Instability in Solar Flares

G. D. Fleĭshman

Purple Mountain Observatory, 2 West Beijing Rd, Nanjing, 210008 China
Ioffe Physicotechnical Institute, Russian Academy of Sciences, ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia

Received May 20, 2003; in final form, August 8, 2003

Abstract—The properties of the cyclotron-maser radiation generated by anisotropic distributions of fast electrons with power-law momentum distributions are studied. The momentum and pitch-angle distributions of the electrons are chosen with regard to available observational constraints. The results obtained for these model distributions are compared with those in the literature for other distributions. The dependence of the properties of the cyclotron instability on the parameters of the distribution of fast electrons is analyzed. These properties are in good quantitative agreement with observational data on millisecond solar radio spikes. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

It is well known that, under certain conditions, anisotropic distributions of particles in a plasma are unstable to the generation of plasma or electromagnetic modes. The idea that the cyclotron instability (i.e., the direct amplification of electromagnetic waves near the gyrofrequency and its harmonics) might arise in magnetoactive plasmas was proposed long ago [1–3], though comprehension of the importance of this mechanism under cosmic conditions came much later, following the critical work of Wu and Lee [4], which demonstrated the fundamental importance of relativistic effects in calculations of cyclotron-instability growth rates. It was also found that the properties of cyclotron-maser radiation depend strongly on the details of the distribution function for the suprathermal electrons, as well as the ratio of the plasma frequency to the gyrofrequency in the source.

The cyclotron-maser mechanism was first successfully applied to explain the kilometer radio emission from the polar regions of the Earth's magnetosphere (so-called auroral kilometer radiation). In this case, the success of the theory [5] was provided by the ability to use realistic values for the magnetic field, plasma density, and distribution function of fast suprathermal particles obtained from direct spacecraft measurements.

At the same time, the cyclotron instability has come to be considered a probable mechanism for the formation of the superfine (millisecond) structure of the radio emission of solar flares [6, 7]. The cyclotron-maser mechanism was developed primarily in connection with millisecond solar radio spikes [8–17], though it was also invoked to explain the millisecond quasi-periodic pulsations of the radio emission [18] and the superfine temporal structure of the decameter solar radio emission [19]. The recent analysis [20] rigorously demonstrated that the millisecond radio pulsations are generated by a plasma mechanism; thus, the radio spikes probably represent the only type of decimeter solar radio emission that appears to be associated with cyclotron instability.

Much indirect evidence in favor of the cyclotron-maser mechanism for the generation of solar radio spikes has been accumulated. In particular, in the recent review [21], the observed characteristics of radio spikes (spectral, temporal, spatial, etc.) are compared in detail with the predictions of various models considered in the literature. The entire collection of observational data can be satisfactorily explained by a model with cyclotron-maser radiation generated by fast electrons with a power-law momentum distribution and loss-cone anisotropy. The source of a cluster of spikes is a magnetic loop containing more or less pronounced magnetic-field inhomogeneities that form local traps, each of which can be a spike source.

This model is able to explain many detailed properties of radio spikes. In particular, the quasi-linear saturation of cyclotron-maser radiation [22] can naturally reproduce the observed time profiles of spikes [23, 24], which consist of a Gaussian phase of growth of the radiation and a subsequent exponential decay. It was later shown that spatial nonuniformity of the spike source (a local trap) can result in splitting of the radiation line profile [25]. In this case, the ratio of the frequencies of the spectral components of the produced doublet is determined by the relative inhomogeneity of the magnetic field $\delta B/B$ in the local trap. Since the value of $\delta B/B$ changes (from event to event) continuously, the broad distribution of the

1063-7729/04/4801-0065$26.00

harmonic ratios of radio spikes [26] from 1.06 to 1.5 becomes understandable.

The most direct evidence that radio spikes are produced by the cyclotron-maser mechanism was obtained recently in a study of correlations between the properties of spike clusters and the accompanying gyrosynchrotron (microwave) continuum emission [27]. It turned out that the ratio of the gyrofrequency to the plasma frequency ω_{Be}/ω_{pe} is systematically higher in microwave sources accompanying spikes (we will call the corresponding radio bursts spike events) than in all remaining microwave bursts; this is a condition favorable for the realization of the cyclotron-maser mechanism.

The optically thin gyrosynchrotron radiation in a spike event demonstrates an anomalously high degree of polarization ($\sim$90% in the quasi-longitudinal case and $\sim$50% in the quasi-transverse case). According to the calculations of [28], this provides evidence for anisotropy of the distribution of trapped electrons, such as a loss cone, and rules out beam-like distributions, for which the degree of polarization is considerably lower. The properties of spike events depend substantially on the angle at which the gyrosynchrotron source is viewed with respect to the magnetic field. In the quasi-transverse case, the spectra of the optically thin gyrosynchrotron radiation are harder than for all other microwave bursts. The harder the microwave spectrum, the more intense the average radiation flux of the radio spikes. This means that radio spikes are generated more efficiently the harder the spectrum of fast electrons captured in the trap. This conclusion is consistent with the results obtained in studies of the correlations between radio spikes and the hard x-ray radiation of flares [29].

However, in the quasi-longitudinal case, the situation is fundamentally different. First of all, the optically thin gyrosynchrotron spectra are much softer than the average spectra for all microwave bursts. Moreover, the softer this spectrum, the more intense the spike radiation. This behavior can be readily explained if we allow for pitch-angle anisotropy of the radiating fast electrons. Indeed, Fleishman and Melnikov [28] recently showed that, in the case of cone anisotropy of the fast electrons captured in a trap, a much softer gyrosynchrotron spectrum is radiated in quasi-longitudinal directions than in the transverse direction.

Therefore, radio spikes are generated more efficiently in the presence of stronger cone anisotropy of the trapped electrons and a harder energy spectrum. It is well known [21] that the same conditions (together with relatively low values of $Y = \omega_{pe}/\omega_{Be}$ in spike events) are necessary and sufficient conditions for the efficient operation of the cyclotron-maser mechanism. However, the properties of the cyclotron-maser radiation generated by fast electrons with power-law energy distribution have been considered in only a few papers [14, 30, 31] and remain poorly studied. A fairly rarefied plasma ($Y \leq 0.9$) was considered in [14, 30], with the angular part of the fast-electron distribution function being modeled by a Gaussian dependence over the entire range of polar angles:

$$f(\mu) \propto \exp(-\mu^2/\mu_0^2), \tag{1}$$

where $\mu = \cos\theta$ and θ is the pitch angle of the electrons.

The choice of this angular distribution was determined solely by its simplicity and convenience. A power-law electron-energy distribution and idealized angular distribution with a loss cone were considered in [31]:

$$f_2(\mu) \propto \begin{cases} 1, & \mu < \cos\theta_c \\ \dfrac{\cos(\theta_c - \Delta\theta_c) - \mu}{\cos(\theta_c - \Delta\theta_c) - \cos\theta_c}, & \cos\theta_c < \mu < \cos(\theta_c - \Delta\theta_c) \\ 0, & \mu > \cos(\theta_c - \Delta\theta_c), \end{cases} \tag{2}$$

i.e., the angular function is constant at large angles, zero at small angles, and displays a rapid linear decay near the critical angle. Accordingly, the derivative of this function is nonzero (and constant) in a narrow interval of angles near the critical angle and is zero outside this interval. The values of Y were varied up to 1.5. The question of the observability of generated eigenmodes was analyzed in [31], but the dependence of the results on the parameters of the electron distribution function was not discussed.

Using the above simple distribution functions in the model calculations seemed quite justified at the time, since observational data on the properties of the angular distributions of the electrons accelerated in flares were lacking. However, the situation has radically changed. The analysis of the spectral evolution

of microwave radiation carried out in [32, 33] suggests anisotropy of the injection of fast electrons into the trap. The presence of a radio brightness peak at the loop top at high frequencies (17 and 34 GHz) represents clear evidence for an anisotropic (loss-cone) distribution of the accelerated electrons captured in the trap [34].

Recent detailed calculations of gyrosynchrotron radiation by fast electrons with an anisotropic pitch-angle distribution [28] enable us to proceed to a quantitative analysis of the angular distributions of the particles realized in solar flares. In particular, the electron pitch-angle distributions in a spike event must satisfy a number of criteria (following from observations). They must provide the cyclotron instability over the entire range covered by the energy spectra of fast electrons observed in a spike event (derived from data on hard x-ray and/or microwave radiation). Moreover, the efficiency of the cyclotron-maser radiation must decrease as the electron-energy spectral index increases (i.e., for softer distributions).

Thus, there is no doubt that the analysis of cyclotron-maser radiation generated by trapped fast electrons with maximally "realistic" distribution functions based on observations is a topical problem.

Though cyclotron-maser radiation has been studied in detail in the literature, many of the results obtained are model dependent and have a restricted range of application. The extrapolation of such results beyond their range of applicability can lead to serious errors (and in any case should be carried out with caution). For instance, Melrose and Dulk [35] made many simplifying assumptions when obtaining their estimates of cyclotron-maser radiation: they assumed semirelativistic expansions, replaced Bessel functions by their asymptotic expansions for small arguments, etc. As a result, their maximum amplification takes place in a narrow interval of angles relative to the magnetic field, 70–84°. However, the subsequent papers [9, 13, 14, 18, 30] showed that amplification can take place in any direction with respect to the external magnetic field (from quasi-transverse to quasi-longitudinal), though the amplification of waves propagating along the external magnetic field is, indeed, usually less efficient.

Further, a kinematic estimate was presented for the spectral bandwidth of the amplification, $\Delta\omega/\omega \sim (v/c)^2$, where v is the characteristic velocity of the fast electrons. In fact, this estimate may have nothing to do with the real bandwidths of cyclotron-maser radiation. The reason is that the change of the sign of the absorption coefficient within some interval of frequencies and angles (i.e., the appearance of amplification) requires the fulfillment of a number of fairly severe constraints. Therefore, as a rule, amplification arises in a rather restricted range of frequencies, so that $\Delta\omega/\omega \ll (v/c)^2$; i.e., even relativistic electrons yield quite narrow emission lines. This has been clearly shown, for example, in [31]: a very broad energy distribution (10 keV–1 MeV) produces extremely narrow cyclotron-maser spectral lines (with a relative width on the order of 1%).

We emphasize that, in general, assuming semirelativistic expansions [4] in a quantitative analysis of the cyclotron-maser radiation of solar flares is not correct and can result in serious errors; to avoid these, the exact relativistic relationships must be used.

Finally, let us turn our attention to the characteristic features of the quasi-linear saturation of cyclotron-maser radiation. Beginning with the early papers [8, 9], the idea arose that only the dominating mode (i.e., with the greatest growth rate) can be appreciably amplified, its back action on the distribution of fast electrons will cause their isotropization, and "slower" modes are virtually unamplified. These concepts are based on one-dimensional analyses of the relaxation of unstable distributions of fast electrons in a plasma, when the quasi-linear approximation results in a plateau in the (one-dimensional) particle distribution function. The generation of cyclotron-maser radiation by a conical distribution of electrons is a fundamentally two-dimensional problem. It is well known in plasma physics [36] that, in this case, quasi-linear relaxation proceeds in a substantially different manner than in the one-dimensional case. In particular, relaxation in a faster mode does not provide full stabilization of an initially unstable particle distribution, and thus does not result in saturation of the radiation of waves with smaller growth rates. This general conclusion is confirmed by detailed numerical studies of the saturation of cyclotron-maser radiation under the conditions of solar flares [22].

Since the fastest growing waves are often extraordinary or ordinary waves near the first (fundamental) harmonic of the gyrofrequency and, as a rule, the absorption of these waves by thermal electrons in the second gyrofrequency layer is very strong, the concept of a "dominating mode" has essentially led to the conclusion that cyclotron-maser radiation cannot be observed on the Earth, even if it is, in fact, generated in a radio source (because this radiation would be completely absorbed in the gyrofrequency layers). In reality, modes without the very highest growth rates can also be amplified considerably. Thus, their growth rates should nevertheless be fairly high to provide appreciable amplification of waves on scales on the order of or less than the source size.

In this paper, we pay special attention to our choice of model momentum and pitch-angle distribution functions for the fast electrons taking into account available observational constraints (Section 2).

In the third section, we study the dependences of the spatial growth rates of the cyclotron instability on the parameters of this distribution. In the fourth section, we discuss these results.

2. FORMULATION OF THE PROBLEM

We consider here the dependence of the spatial growth rates (negative absorption coefficients) of cyclotron-maser radiation on the plasma parameters and distribution function of fast electrons. As is known [37], the exact relativistic formula for the gyrosynchrotron absorption coefficient is

$$\kappa_\sigma(\omega,\vartheta) = -\frac{\pi^2\omega_{pe}^2}{\omega|\eta|v_\sigma N(1+K_\sigma^2+\Gamma_\sigma^2)\left(n_\sigma^2\frac{\partial(\omega n_\sigma)}{\partial\omega}\right)} \times\sum_{n=1}^{\infty}\int\frac{p^2}{\gamma_e\beta}dp(1-\mu^2) \times\left[\frac{-\Gamma_\sigma\sqrt{1-\eta^2}+K_\sigma(\eta-\beta\mu n_\sigma)}{n_\sigma\beta\sqrt{(1-\mu^2)(1-\eta^2)}}J_n(z) + J_n'(z)\right]^2\left[p\frac{\partial}{\partial p}-(\mu-n_\sigma\beta\eta)\frac{\partial}{\partial\mu}\right]f(\mathbf{p})|_{\mu=\mu_*}, \quad (3)$$

where $\mu=\cos\theta$ is the cosine of the pitch angle of the electrons,

$$\mu_* = \frac{\gamma_e\omega - n\omega_{Be}}{\gamma_e\beta n_\sigma\eta\omega}, \quad |\mu_*|\le 1, \quad (4)$$

$\eta=\cos\vartheta$ is the cosine of the radiation angle, v_σ and n_σ are the group velocity and refraction index of the wave σ, ω_{pe} is the electron plasma frequency, K_σ and Γ_σ are the components of the polarization vector of the wave σ, $\beta=v/c$ is the dimensionless velocity of the particle, $\gamma_e=(1-\beta^2)^{-1/2}$ is its Lorentz factor, $J_n(z)$ and $J_n'(z)$ are the corresponding Bessel function and its first derivative with respect to $z=(\omega/\omega_{Be})\gamma_e n_\sigma\beta\sqrt{(1-\mu^2)(1-\eta^2)}$, and N is the number density of thermal electrons in the plasma.

The relationships for the index of refraction and the polarization vectors are well known, and are given, for example, in the Appendix to [28]. The value and sign of the absorption coefficient are determined by the fast-electron distribution function, $f(\mathbf{p})$, normalized to the total number of fast electrons:

$$\int f(\mathbf{p})p^2dpd\mu d\varphi = N_e. \quad (5)$$

We will restrict our consideration to model distribution functions of the form

$$f(\mathbf{p}) = \frac{N_e}{2\pi}f_1(p)f_2(\mu), \quad (6)$$

so that

$$\int f_1(p)p^2dp = 1, \qquad \int_{-1}^{1} f_2(\mu)d\mu = 1. \quad (7)$$

Since the properties of the instability are quite sensitive to details of the fast-electron distribution function, the choice of the initial distribution function requires a detailed discussion.

Let us consider various models for $f_1(p)$ and $f_2(\mu)$ in the literature and the constraints imposed on them by available observations. We shall begin with the distribution of the momentum modulus (or equivalently of the velocity or energy). The majority of studies of cyclotron-maser radiation have used "hot" electron models (see, e.g., [13]), when the fast electrons obey a Maxwellian velocity distribution corresponding to a temperature that is higher than the background plasma temperature. These models are probably not consistent with the tight correlation between radio spikes, on the one hand, and nonthermal hard x-ray and microwave continuum radiation, on the other hand, since the latter have extended power-law spectra implying a power-law distribution for the nonthermal electrons. Proceeding from this, we consider a power-law momentum distribution for the particles

$$f_1(p) = \frac{(\gamma-3)p_0^{\gamma-3}}{p^\gamma}, \qquad p_0 < p < p_{br} \quad (8)$$

from 10 keV to several MeV. At still higher energies, we have considered three possibilities: the spectrum continues with the same index γ, there is a break and transition to a power-law segment with a higher index γ_1, or there is an exponential decay of the spectrum. In our calculations, we found no substantial differences between these three model functions (for the particular angular distribution chosen for the calculations).

Let us now discuss the angular part of the distribution function $f_2(\mu)$. Various distributions have been considered in the literature: beams [16, 17], oblique beams [5], and loss cones [13, 14, 31, 38]. Since the degree of polarization of the optically thin gyrosynchrotron radiation of the microwave continuum bursts accompanying spike clusters (and tightly correlated with them) is anomalously high [27], the average pitch-angle distributions of the fast electrons must have loss cones [28]. Therefore, it is natural to assume that the angular distributions of the particles in spike sources are also loss-cone distributions (note that the estimated sizes of spike sources are from several tens to several hundreds of kilometers [39], much smaller than the size of the gyrosynchrotron burst source). For example, in the local-trap model [21], the anisotropy of the electron distribution in the spike source is formed as a result of amplification of the "mean" anisotropy by a locally nonuniform magnetic

field. On the other hand, there is no observational evidence of the realization of beam-like distributions in spike sources. Thus, a loss-cone angular distribution for the fast electrons currently seems to be the most probable for spike sources.

The available observations enable us to narrow the class of admissible angular distributions even further. The angular distribution must [29, 27] (i) provide the cyclotron instability for all values of the spectral index $5 < \gamma < 12$ in the fast-electron momentum distribution for which radio spikes are observed and (ii) be such that the efficiency of the instability in this case decreases with growth in γ.

Three kinds of functions to model the loss cone have been considered in the literature. In addition to the two functions (1) and (2) mentioned above, we have considered a function that decreases as the Nth power of the sine of the pitch angle (which we will call a $\sin^N$ function):

$$f_2(\mu) \propto \begin{cases} \sin^N\left(\frac{\pi}{2}\frac{\theta}{\theta_c}\right), & 0 < \theta < \theta_c \\ 1, & \theta_c < \theta < \pi - \theta_c \\ \sin^N\left[\frac{\pi(\theta-\pi)}{2\theta_c}\right], & \pi - \theta_c < \theta < \pi. \end{cases} \quad (9)$$

It is interesting that these functions do not match the listed set of conditions. For example, the $\sin^N$ function (Fig. 1) is a smooth function that is continuous, together with its first derivative. Its derivative is also large in the angular interval where the function is already very small (marked with an arrow in Fig. 1). Moreover, the ratio of these quantities diverges at small angles. Therefore, the positive contribution to the gain coefficient (3) due to the derivative with respect to the angle turns out to be much greater than the negative contribution due to the derivative with respect to the modulus of the momentum. As a result, the growth rates for the amplification of waves demonstrate a very weak dependence on the hardness of the fast-electron spectrum. The same (with the caveat that the derivative is not a continuous function) is true for the idealized loss cone [31].

In contrast, the Gaussian distribution [14, 30] readily provides a decreasing efficiency of the cyclotron-maser radiation with increasing γ. However, its angular derivatives are very small (Fig. 1), and do not result in instability at $\gamma \gtrsim 8$ [14, 30]. In addition, the instability can be strongly suppressed for extended (to $p/mc > 1$) power-law electron-momentum distributions [14, 30].

It is obvious that we will obtain an appreciable spectral-index dependence in the electron-momentum distribution if the derivatives of the distribution functions for the momentum modulus and the cosine of the pitch angle are comparable.

Accordingly, we consider here a new trial function that is constant outside the loss cone and exponentially decreases inside it (Fig. 1):

$$f_2(\mu) \propto \begin{cases} \exp\left(-\frac{\mu - \cos\theta_c}{\Delta\mu}\right), & \mu > \cos\theta_c \\ 1, & -\cos\theta_c < \mu < \cos\theta_c \\ \exp\left(\frac{\mu + \cos\theta_c}{\Delta\mu}\right), & \mu < -\cos\theta_c, \end{cases} \quad (10)$$

where θ_c is the loss-cone angle and $\Delta\mu$ describes the rate of decrease of the distribution function in the loss cone. Advantages of this distribution function include the constant ratio of the angular derivative to the function itself. Accordingly, no artifact regions with anomalously high angular gradients arise, which enables us to study the properties of the instability in the weakly above-threshold regime.

The momentum-modulus distribution function is chosen to be

$$f_1(p) \quad (11)$$
$$\propto \begin{cases} \frac{(\gamma-3)p_0^{\gamma-3}}{p^\gamma}, & p_0 < p < p_{br} \\ \frac{(\gamma-3)p_0^{\gamma-3}}{p_{br}^\gamma}\exp\left(-\gamma\frac{p-p_{br}}{p_{br}}\right), & p > p_{br}. \end{cases}$$

The computation of the normalization of the distribution function $f_1(p)$ approximately assumes an extended decreasing spectrum with $\gamma > 3$. As we have already pointed out above, the form of this function when $p > p_{br}$ only weakly affects the results of the calculations for the angular distribution (10), though this effect is present and quite important in the case of the Gaussian angular distribution (1) [14, 30].

We calculated dimensionless spatial gain coefficients (i.e., the familiar absorption coefficients taken with the opposite sign). The dimensionless quantities (given in the graphs) can be translated to dimensioned quantities using the formula

$$\kappa_\sigma(\omega, \vartheta) = \frac{\pi\omega_{Be}}{2c}\frac{N_e}{N}\kappa_\sigma^{num}(\omega, \vartheta). \quad (12)$$

Each value of $\kappa_\sigma(\omega, \vartheta)$, defined by (3), was calculated using the integration method of Gauss and summing a series over the necessary number of harmonics. We used the exact (built-in) Bessel functions, and found their derivative by means of recurrence relationships. We then determined for each set

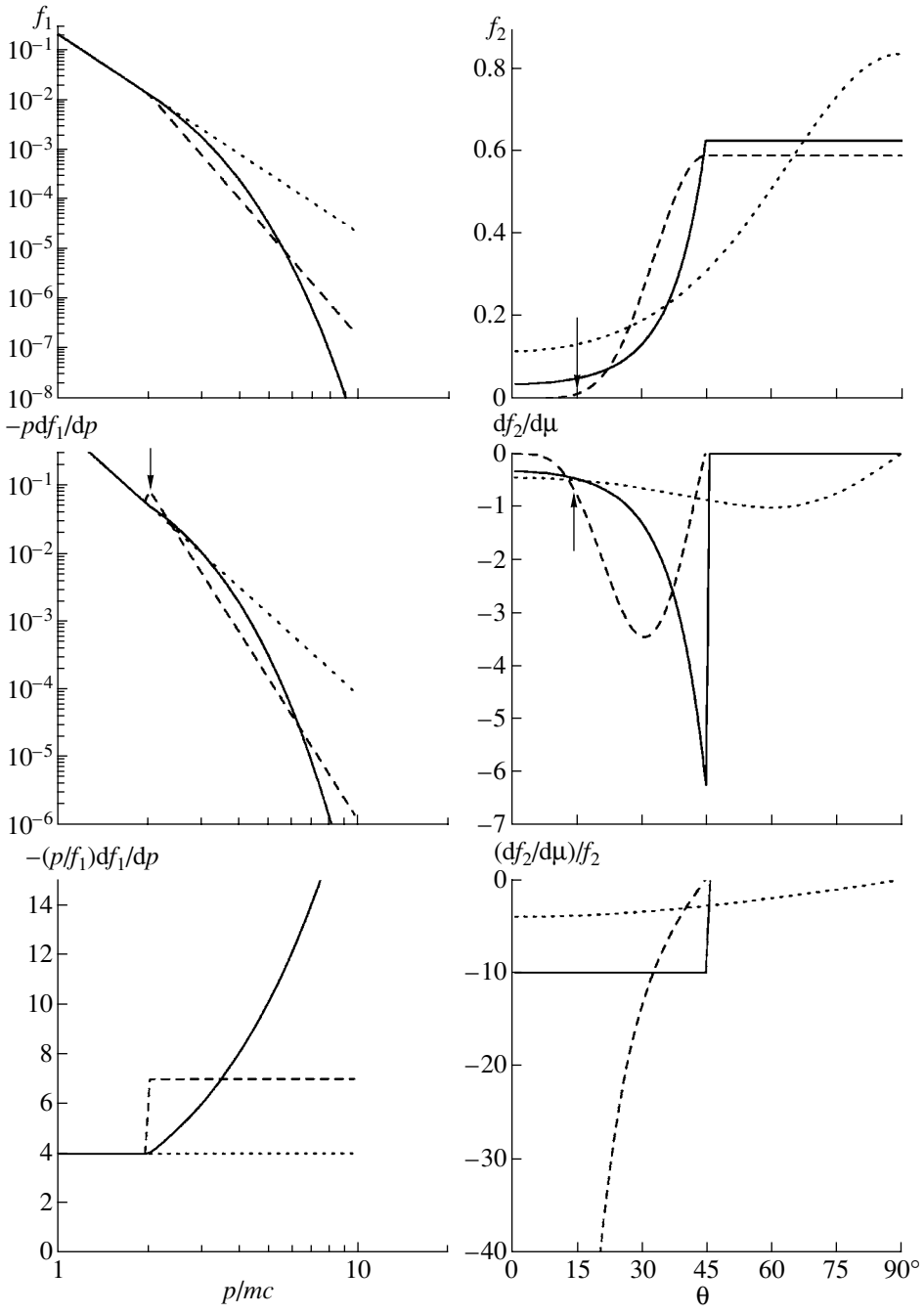


Fig. 1. Various models describing the momentum (left) and angular (right) distribution functions of the fast electrons. A single power law (dotted), double power law (dashed), and power law with exponential cutoff (solid) are shown for the momentum distributions. A Gaussian (dotted), $\sin^N$ function (dashed), and exponential decay in the loss cone (solid) are shown for the angular distributions. The upper panels show the functions, the middle panels their derivatives, and the lower panels the ratios of the derivatives and original functions.

of input parameters the maximum spatial growth rate (as a function of frequency and radiation angle); its dependences on various parameters of the problem are shown in the graphs.

3. PROPERTIES OF THE CYCLOTRON INSTABILITY

The hierarchy of the dominating modes as a function of the plasma parameter Y has been considered in a number of previous studies. Figure 2 presents

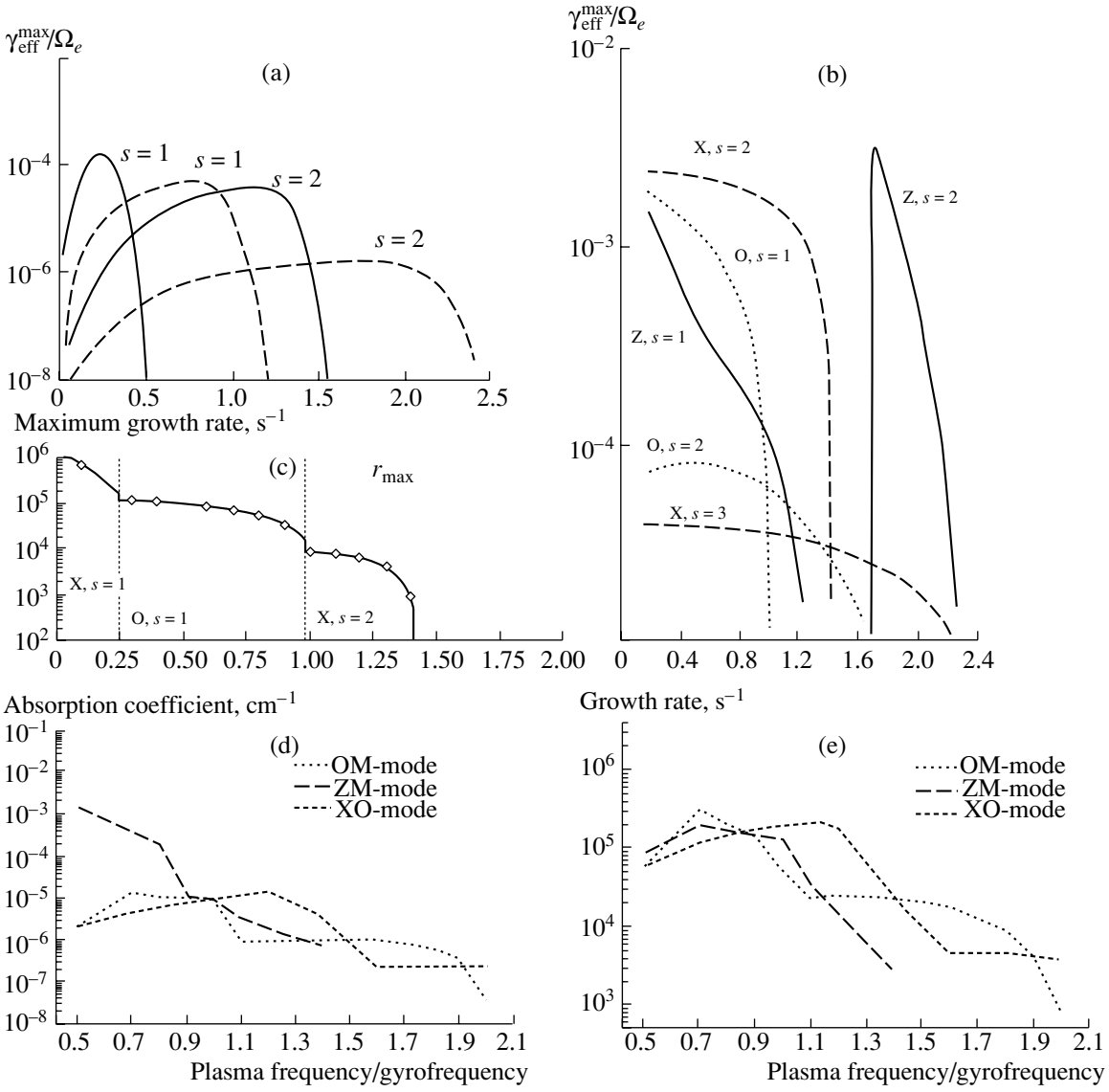


Fig. 2. The hierarchy of dominating modes according to (a) [9], (b) [10], (c) [13], (d, e) [31].

the fast-electron distribution functions from several of these. Note that the mode hierarchy depends on both Y and the form of the fast-electron distribution. Therefore, we will first consider the dependence of the eigenmode gain factors on the background plasma density (more precisely, on the parameter Y, the ratio of the plasma frequency and gyrofrequency) for our model fast-electron distribution function. In Fig. 3, Y varies from 0.1 to 2.5 in steps of 0.02; growth rates were calculated near the three lowest harmonics of the gyrofrequency. The maximum spatial growth rates for the ordinary and extraordinary modes are shown in the upper panels. The middle panels indicate the frequencies (if $s \approx 1$, we are dealing with emission at the first harmonic, if $s \approx 2$, with emission at the second harmonic, etc.).The lower panels present the cosine of the angle at which the corresponding growth rate peaks. Note that the preferential direction of the radiation can be from 0° to 90° as Y is varied for each mode and at any harmonic [9, 13, 14, 18, 30]. Thus, the quasi-transverse directions of the radiation are not selected out or preferred.

The hierarchy of the"dominating" modes (i.e., those with the greatest growth rate for a given set of parameters) for this distribution function is consistent with data in the literature for other distributions [9, 10, 13, 31]. For small values of Y, <0.25–0.35, the extraordinary mode near the first harmonic (X1) dominates; as Y increases, the ordinary mode at the first harmonic (O1), extraordinary mode at the second harmonic (X2), ordinary mode at the second harmonic (O2), and extraordinary mode at the third

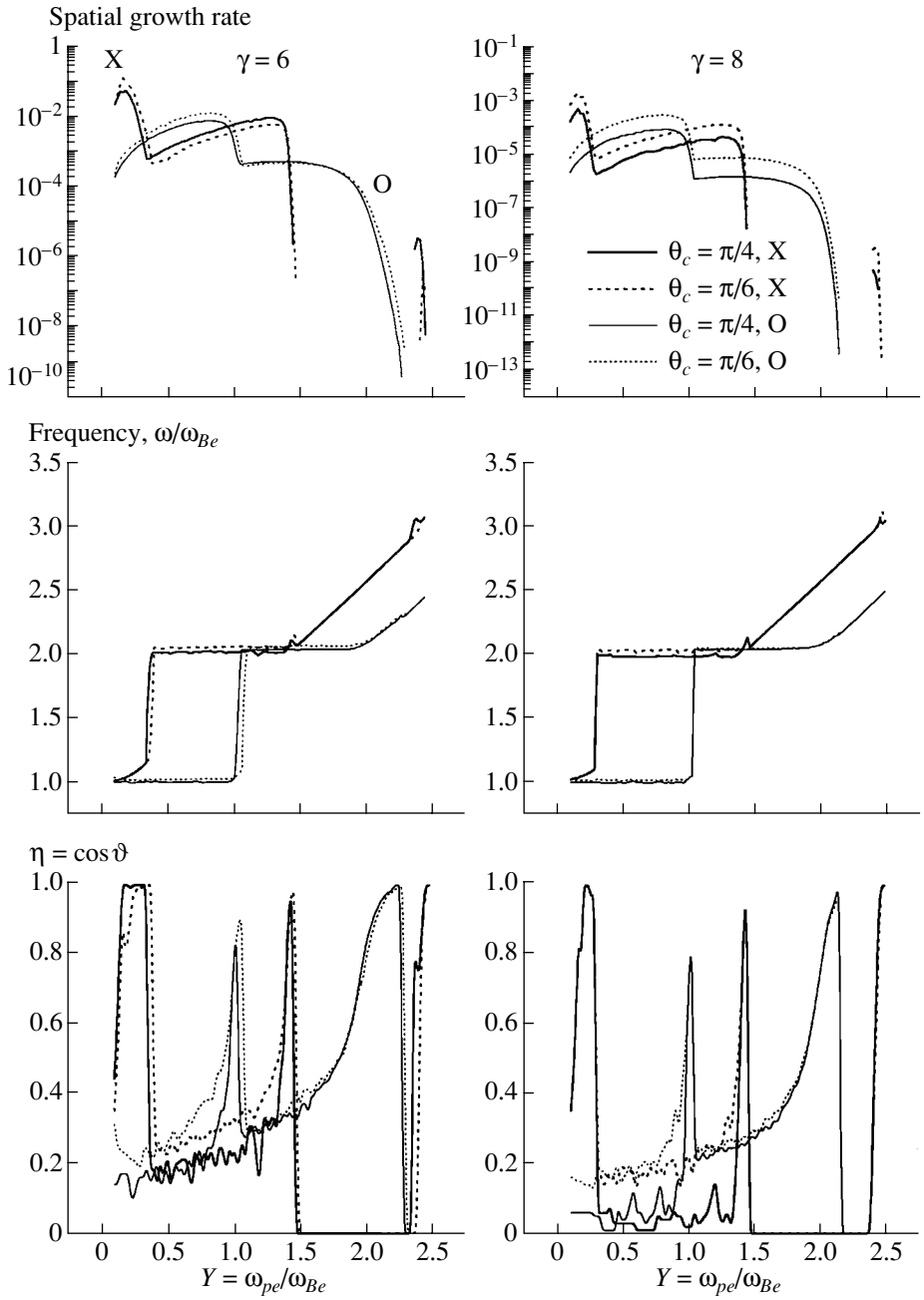


Fig. 3. Maximum growth rates of the extraordinary (X, bold curves) and ordinary (O, thin curves) as functions of the ratio of the plasma frequency and gyrofrequency $Y = \omega_{pe}/\omega_{Be}$ (upper panels). The middle and lower panels show the frequency and cosine of the emission angle for which the maximum growth rate is achieved. The parameters of the electron distribution function are $p_0/mc = 0.2$, $p_{br}/mc = 3$, $\Delta\mu = 0.1$, $\theta_c = \pi/4$ (solid curves) and $\theta_c = \pi/6$ (dotted curves), and $\gamma = 6$ (left plots) and $\gamma = 8$ (right plots).

harmonic (X3) become dominating in turn, although the growth rates of the X3 modes for this fast-electron distribution function are too small to expect appreciable generation in these modes. The exact values of Y at which the dominating mode changes depend on the parameters of the fast-electron distribution function—in particular, the hardness of the electron spectrum. For example, when $\gamma = 6$, the X1 mode dominates up to $Y = 0.35$–0.4 (depending on the loss-cone angle θ_c) and the O1 mode dominates for $Y < 1$. However, when $\gamma = 8$, the lower boundary shifts to 0.26 (Fig. 3). These values are very close to those obtained in studies of the growth rates produced by hot (Maxwellian) electrons with a loss cone and $T = 10^8$ K (0.24 and 1) [13]. This agreement is not accidental: when the fast-electron spectrum is soft

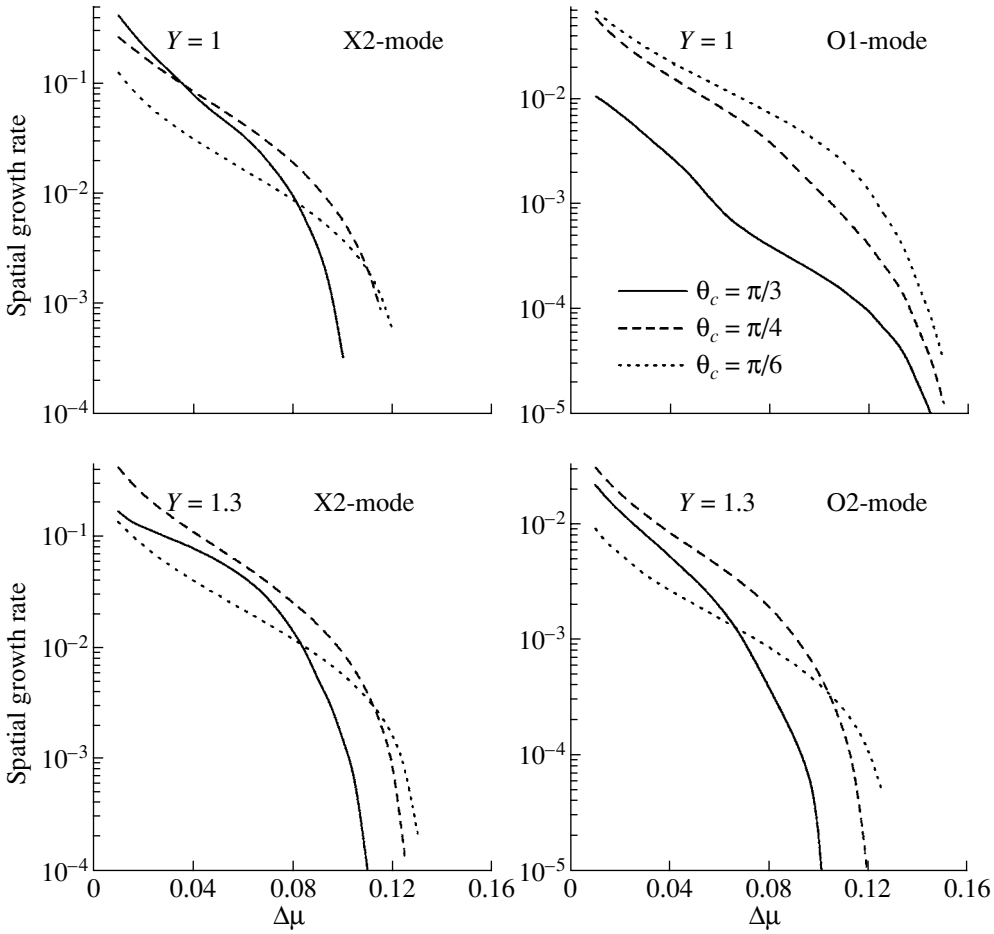


Fig. 4. Maximum growth rates of the extraordinary (X2) and ordinary (O1, O2) waves as functions of the angular gradient $\Delta\mu$. The values of Y and θ_c are given in the figure; the remaining parameters of the electron distribution function are $p_0/mc = 0.2$, $p_{br}/mc = 3$, and $\gamma = 6$.

($\gamma = 8$), the growth rates are determined mainly by the contribution of low-energy electrons, whereas the role of high-energy electrons increases in the case of harder spectra ($\gamma = 6$).

Analysis of the Y dependence of the gain factors enables us to draw important conclusions about which eigenmode (ordinary or extraordinary) generates the solar spike radiation and at which harmonics. In such analyses, we must take into account the pronounced dependence (close to inverse proportionality) of the characteristic duration of the spikes on the frequency at which they are observed [23].

The extraordinary mode at the first harmonic demonstrates a very strong Y dependence of the growth rate, so that it should not make a significant contribution to the spike generation. It is not ruled out that the cyclotron-maser radiation generated at the first harmonic of the gyrofrequency is completely absorbed in the second gyrofrequency layer.

The growth rates of the extraordinary waves at the second harmonic and of the ordinary mode at the first and second harmonics demonstrate very smooth Y dependences, consistent with the observed frequency dependence of the spike durations.

It was recently found [24] that the spike durations at 1.42 GHz and especially at 2.695 GHz exceed the values following from the dependence of [23], derived from observations at $f \lesssim 1$ GHz. Even larger deviations from this law have been found at 2.6–3.8 GHz [40]. This means that either the frequency dependence of the spike duration is not described by a single power law, or that the index in this law must be decreased. Refining this dependence is extremely important for identifying the mode responsible for the formation of the radio spikes.

Let us now discuss the dependence of the gain factors on the parameters of the distribution function. We initially chose a set of "base" parameters, then consecutively varied each of them. The base distribution for the momentum modulus is a power law with index $\gamma = 6$ in the interval from $p_0/mc = 0.2$ to $p_{br}/mc = 3$, with an exponential decay at higher

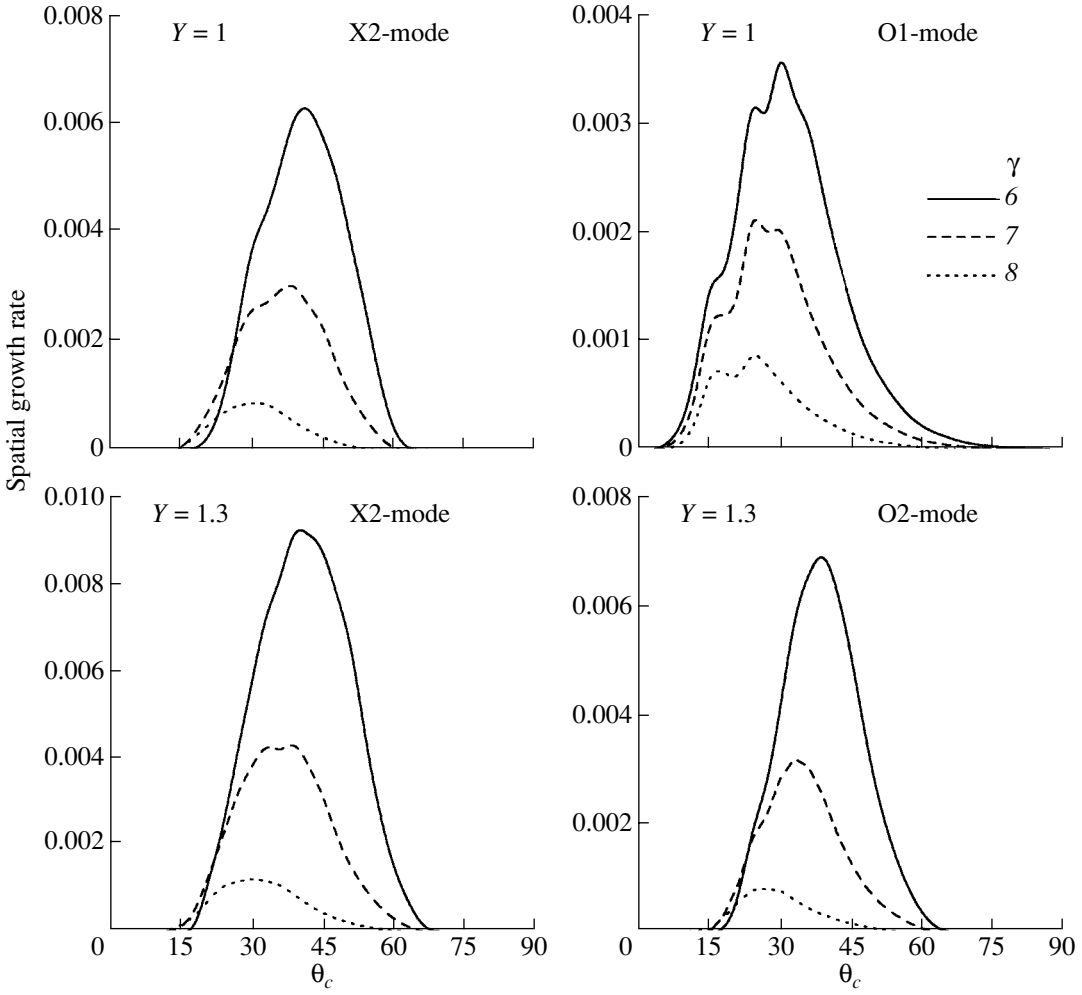


Fig. 5. Maximum growth rates of the extraordinary (X2) and ordinary (O1, O2) waves as functions of the loss-cone angle θ_c for various values of γ (shown in the figure). The remaining parameters of the electron distribution function are $p_0/mc = 0.2$, $p_{br}/mc = 3$, and $\Delta\mu = 0.1$.

momenta. Calculations were carried out for the three loss-cone angles, $\theta_c = \pi/3$, $\pi/4$, and $\pi/6$, with the distribution width inside the loss cone chosen to be $\Delta\mu = 0.1$. The graphs of the growth rates as functions of p_0 are an exception; here $\Delta\mu = 0.05$ since the cyclotron instability threshold depends considerably on p_0, and the instability occurs only in a restricted interval of p_0 when $\Delta\mu = 0.1$.

The growth rates as functions of the parameters of the fast-electron distribution function are studied for the two values $Y = 1$ and 1.3. In the former case, we consider the extraordinary mode at the second harmonic and the ordinary mode at the first harmonic, and, in the latter, the extraordinary and ordinary mode at the second harmonic.

Figure 4 shows the spatial growth rates of these waves as functions of $\Delta\mu$, which describes the angular gradient of the distribution inside the loss cone. It is natural that the growth rates increase with increasing angular gradient (decreasing $\Delta\mu$). Note that the growth rates decrease sharply at $\Delta\mu \sim 0.1$, and the instability completely vanishes with further increase in $\Delta\mu$; i.e., when the electron distribution is still strongly anisotropic.

This behavior is due to the threshold character of the cyclotron instability: the positive contribution of the angular gradient must exceed the negative contribution of $p\,\partial f/\partial p = -\gamma f(p)$, the derivative with respect to the momentum modulus.

The dependences of the growth rates on the loss-cone angle θ_c (Fig. 5) have broad maxima, and the cyclotron instability can be realized for virtually any θ_c from 20° to 80°, though the interval 30–60° is preferred.

It was shown in [14, 30] that, with a Gaussian pitch-angle distribution for the fast electrons, the cyclotron instability is suppressed if the momentum-modulus distribution is fairly extended. Figure 6,

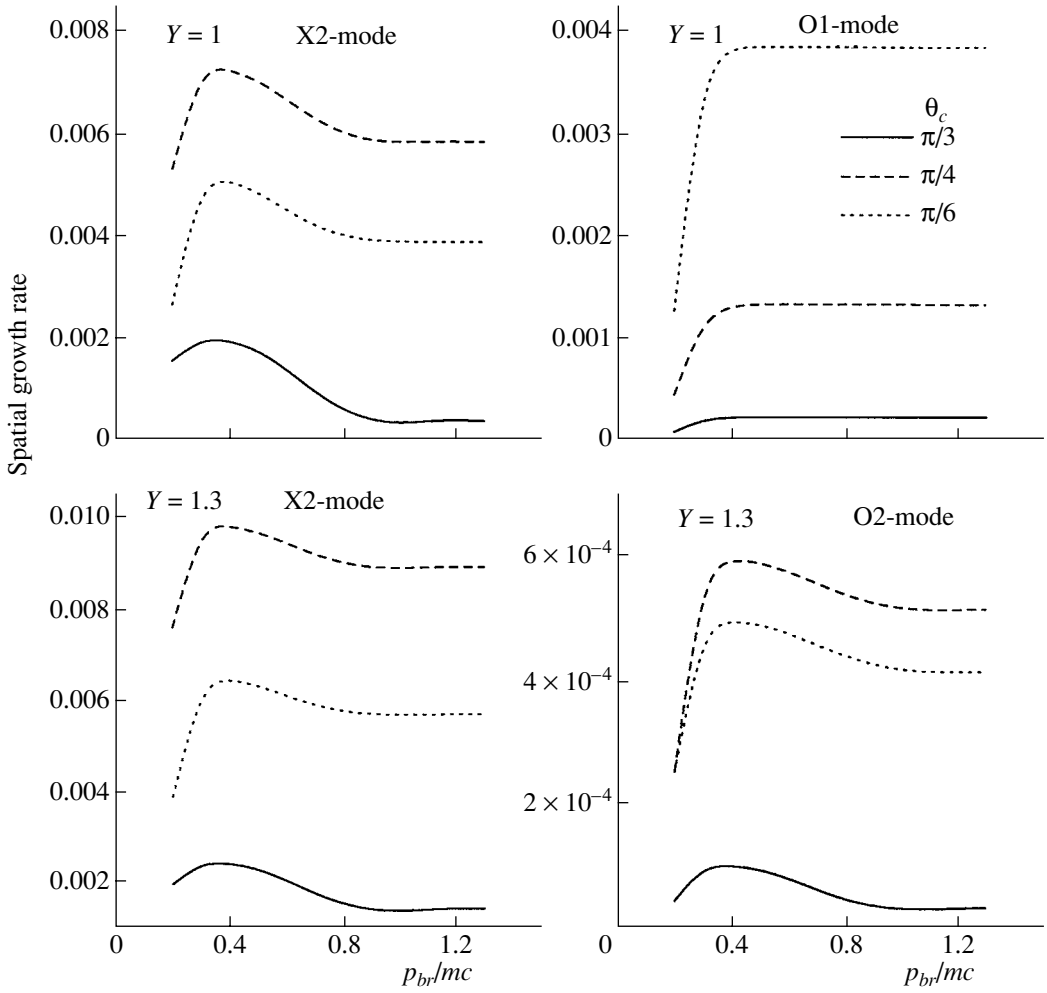


Fig. 6. Maximum growth rates of the extraordinary (X2) and ordinary (O1, O2) waves as functions of the break momentum p_{br}/mc for various values of θ_c (shown in the figure). The remaining parameters of the electron distribution function are $p_0/mc = 0.2$, $\gamma = 6$, and $\Delta\mu = 0.1$.

which presents the growth rates as functions of the break momentum p_{br} at which the exponential cutoff in the electron distribution begins, shows that there is no suppression for the specified angular distribution. This distinction is associated with the smallness of the angular gradients produced by the Gaussian pitch-angle distribution compared to the distribution used in our analysis. Note that there is nevertheless a dependence of the growth rates on p_{br}, with its character differing for different modes. This is due to the fact that different groups of fast electrons are responsible for the generation of different eigenmodes.

This is also confirmed by the dependences of the growth rates on the minimum momentum of the fast electrons p_0 (Fig. 7); namely, for each of the three modes considered, there is a range of p_0 values that is preferential for its generation: $p_0/mc = 0.3$–0.5 for the O1 mode, $p_0/mc = 0.15$–0.4 for the O2 mode, and $p_0/mc < 0.2$ for the X2 mode. The instability of the X2 waves vanishes for higher values of p_0/mc (0.5–0.6 ($E_0 \sim 50$ keV)) remaining fairly efficient for lower values.

Figure 8 shows the spatial growth rates as functions of the hardness of the fast-electron spectrum (the spectral index γ), which are bell-shaped curves. The position of the maximum and the detailed shape of the wave vary with the parameters of the pitch-angle distribution. The decrease in the growth rates for harder spectra (small γ) is due to the destructive contribution of the energetic electrons (at high harmonics), while the decrease for soft spectra (large γ) is due to the increase (in absolute value) of the derivative with respect to the momentum modulus ($p\partial f/\partial p = -\gamma f(p)$). As a result of the interplay of these two factors, there arises an optimum (for a given angular distribution) value of the hardness of the electron spectrum that provides the most efficient amplification of the waves. The greatest amplification

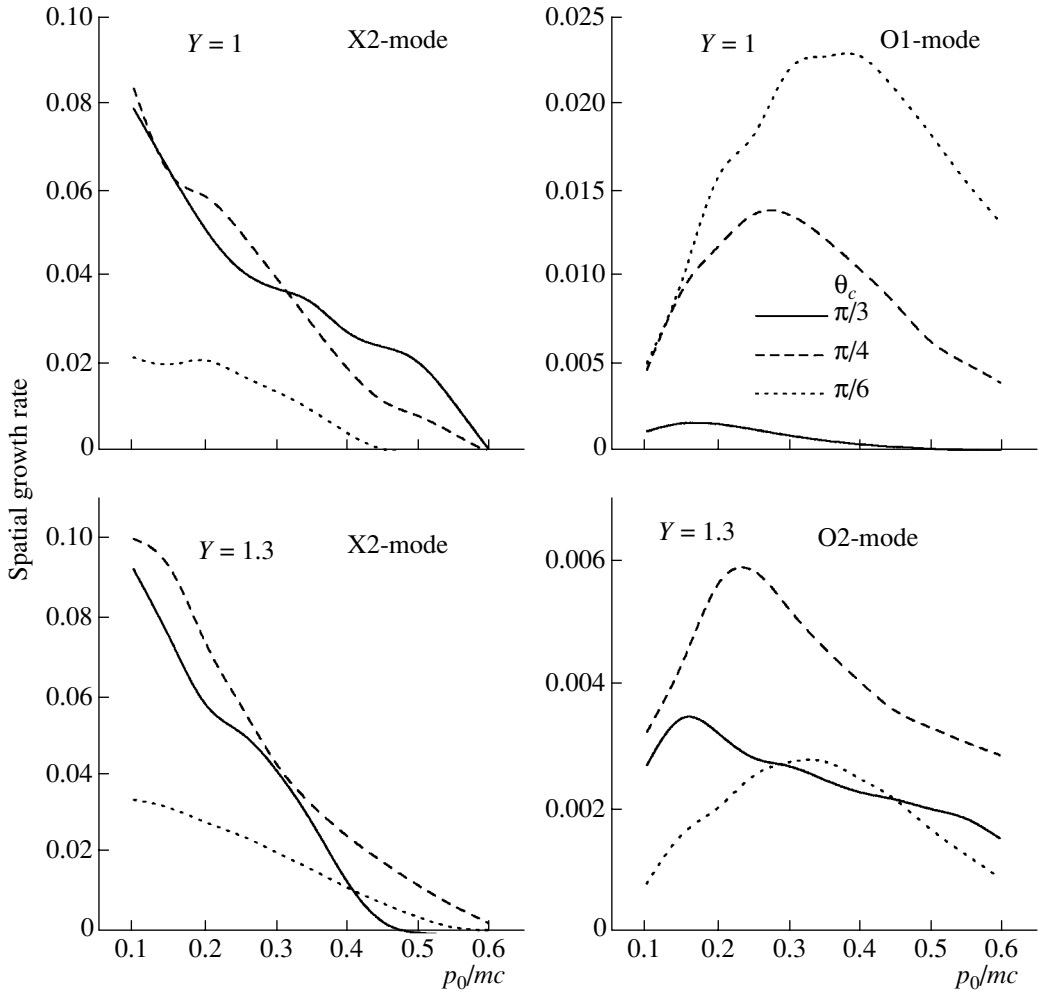


Fig. 7. Maximum growth rates of the extraordinary (X2) and ordinary (O1, O2) waves as functions of the initial momentum p_0/mc for various values of θ_c (shown in the figure). The remaining parameters of the electron distribution function are $p_{br}/mc = 3$, $\gamma = 6$, and $\Delta\mu = 0.05$.

usually occurs at $\gamma = 5$–7, though wave amplification can take place at least to $\gamma \approx 12$ (for admissible values of the angular gradient). It stands to reason that the stronger the angle gradients (the smaller $\Delta\mu$) are, the larger are the values of γ for which the cyclotron instability is possible.

4. DISCUSSION

We can ask ourselves whether the relationships obtained for the growth rates of the cyclotron-maser radiation are in agreement with the observational data on decimeter solar radio spikes. Before discussing this question, we note the following. According to the estimates of [39], the characteristic source sizes of individual spikes do not exceed 200 km, approximately two orders of magnitude smaller than the sizes of typical magnetic traps in the solar corona (and, accordingly, of the sources of microwave continuum bursts). However, the very tight correlation between the average fluxes of radio spikes and the gyrosynchrotron continuum radiation suggests that their sources are not independent, and that these two kinds of radiation are generated by the same distributions of fast electrons.

In particular, the pitch-angle distribution of the electrons generating the gyrosynchrotron continuum must be anisotropic, but stable to the generation of transverse waves in a large fraction of the magnetic trap's volume. If the distribution were unstable everywhere, this would result in very powerful radiation covering a continuum of frequencies (tracking the regular change of the magnetic field in the loop). Instead, clusters of spikes—isolated emission peaks in the frequency–time plane—are observed. This can occur if the angular part of the fast-electron distribution function in the trap fluctuates (i.e., there is some scatter of the $\Delta\mu$ values, in the framework of the model pitch-angle distribution (10)). It stands to

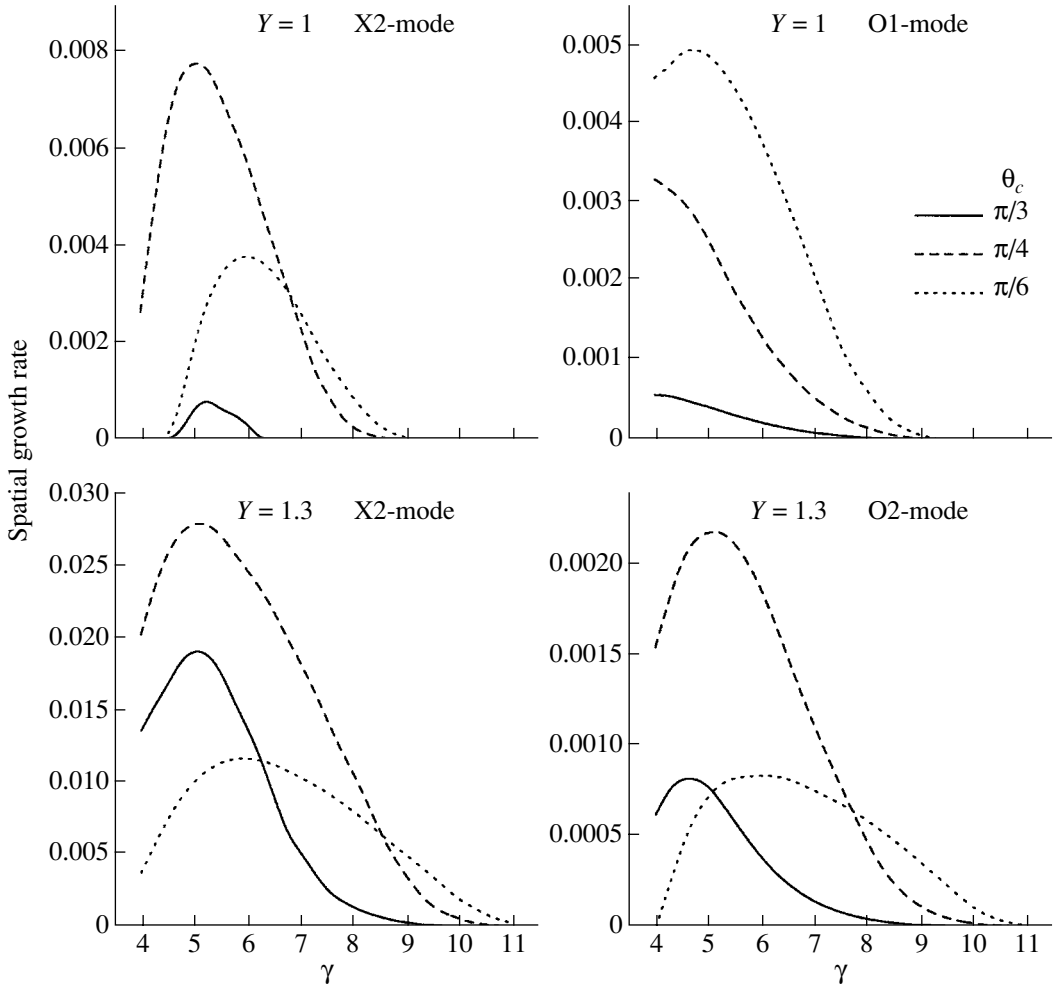


Fig. 8. Maximum growth rates of the extraordinary (X2) and ordinary (O1, O2) waves as functions of the spectral index γ for various values of θ_c (shown in the figure). The remaining parameters of the electron distribution function are $p_0/mc = 0.2$, $p_{br}/mc = 3$, and $\Delta\mu = 0.1$.

reason that the loss-cone angle θ_c changes (depending on the position in the trap) due to the regularly nonuniform magnetic field and its fluctuations.

It is easy to see that the dependences in Fig. 4 allow such a model. The presence of sharp cutoffs in the growth rates at $\Delta\mu = 0.10 - 0.12$ demonstrates that a strongly anisotropic but stable fast-electron distribution (with $\Delta\mu > 0.12$) can exist in the trap. However, if a slightly greater anisotropy ($\Delta\mu \lesssim 0.10$) is realized in some part of the trap, efficient cyclotron-maser radiation arises. It is natural that the stronger the "mean" anisotropy of the pitch-angle distribution in the trap (i.e., the closer to the corresponding threshold) is, the greater is the number of locations in the trap at which the instability conditions will be fulfilled. Therefore, there will be more spikes and their radio flux will increase, as observed [27].

Figure 8 demonstrates that, for a given value of the angle gradient $\Delta\mu$ (and changing values of θ_c), the efficiency of the cyclotron-maser radiation decreases for softer electron spectra when $\gamma > 5$. This behavior is confirmed by observations: the gyrosynchrotron radiation observed in the quasi-transverse direction is accompanied by the strongest spike emission when the hardness of the optically thin spectrum is highest (in this case, the spectral index of the gyrosynchrotron radiation varies from 1–1.5 to 4). This is also consistent with the correlation between radio spikes and hard x-ray radiation [29].

Let us now ask whether we can draw any conclusions about the electromagnetic mode and the number of harmonics providing the majority of the observed radio spikes. Analysis of the observational data [41, 27] testifies to the extraordinary polarization of the spikes. However, the very strong Y dependence of the growth rates of the extraordinary waves on the first harmonic of the gyrofrequency (Fig. 3) is inconsistent with the observed frequency dependence of the

spike durations [23, 24]. Therefore, the radio spikes can only be due to the emission of extraordinary waves at the second harmonic.

It is important that the efficiency of the generation of X2 waves drops sharply when $Y > 1.4$. This means that the frequency ($2\omega_{Be}$) at which the decimeter radio spikes are generated is at least 30% higher than the plasma frequency (ω_{pe}) in the source and, accordingly, than the frequency of any radiation arising near the plasma frequency.

Note that the amplification factor for the O1 and O2 waves can be of the same order as for the X2 waves. Therefore, at favorable values of Y, ordinary-wave cyclotron radiation should be observed at the second harmonic (and also at the first, if this is not precluded by absorption in the second gyrofrequency layer). The maser radiation of the O2 mode differs from that of the X2 mode (apart from, obviously, its opposite polarization) in the proximity of the radiated frequency to the local plasma frequency (within 30%); for example, for $Y = 1.8$, we have $2\omega_{Be} = 1.1\omega_{pe}$. Proceeding from this, we can suppose that the ordinary-wave maser radiation results, in particular, in the generation of the meter-wavelength spikes observed near the high-frequency edges of many type III bursts.

The above analysis suggests that appropriate choice of the energy and angular distribution functions of the fast electrons (captured in the magnetic trap) provides properties of the cyclotron-maser radiation that agree well with observations of the coherent radiation of solar radio spikes. The (averaged) distribution of the electrons in the trap can be stable to the generation of transverse waves and at the same time be strongly anisotropic, as is observed in spike events [27]. In this case, a very small increase in the anisotropy is enough for the efficient development of the cyclotron instability. This fact enables us to link the generation of radio spikes with fluctuations of the electron pitch-angle distribution, and the sources of individual spikes with parts of the magnetic trap in which the nonuniform magnetic field results in a stronger-than-average pitch-angle anisotropy. The closer the "mean" anisotropy is to the threshold, the greater is the number of locations (due to fluctuations) at which the instability conditions will be fulfilled, and the more spikes per unit time will be emitted. A similar dependence on the hardness of the energy spectrum arises: the harder the spectrum (under the condition, however, that $\gamma > 5$), the lower the cyclotron instability threshold; therefore, hard spectra are preferential for the generation of spikes. All these properties are indeed observed in spike events with a very high degree of confidence [27].

Typical values of the dimensionless growth rates of the X2 mode (presented in the figures) are $\sim 10^{-3}$–10^{-2}. With the adopted normalization, this corresponds to the dimensioned gain factor

$$\kappa_\sigma \sim \frac{\pi\omega_{Be}}{2c}\frac{N_e}{N} \times (10^{-3}\text{–}10^{-2}). \qquad (13)$$

For emission at $f = 2$ GHz ($\omega_{Be} = 2\pi \times (1$ GHz)), we have for $n_b/n_0 \sim 3 \times 10^{-3}$:

$$\kappa_\sigma \sim 10^{-6}\text{–}10^{-5}\ \text{cm}^{-1}; \qquad (14)$$

i.e., the e-folding length for the amplification of this emission is on the order of 1–10 km, consistent with the estimated sizes of spike sources.

Thus, our development of the local-magnetic-trap model proposed in [21] has led to further quantitative agreement between the observations and theory. Comparing the model with observations, we can obtain new information about the properties of the pitch-angle distributions of the fast electrons in magnetic traps.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project nos. 02-02-39005 and 03-02-17218), the Chinese NFS foundation (project 10273025), and Chinese Program 973 (G2000078403).

REFERENCES

1. R. Q. Twiss, Austral. J. Phys. **11**, 564 (1958).
2. E. G. Harris, Phys. Rev. Lett. **2**, 34 (1959).
3. R. Z. Sagdeev and V. Safranov, Zh. Éksp. Teor. Fiz. **39**, 181 (1960) [Sov. Phys. JETP **12**, 130 (1960)].
4. C. S. Wu and L. C. Lee, Astrophys. J. **230**, 621 (1979).
5. C. S. Wu, Space Sci. Rev. **41**, 215 (1985).
6. A. V. Stepanov, Pis'ma Astron. Zh. **4**, 193 (1978) [Sov. Astron. Lett. **4**, 103 (1978)].
7. G. D. Holman, D. Eichler, and M. R. Kundu, in *Radiophysics of the Sun,* Ed. by M. R. Kundu and T. E. Gergely (Reidel, Dordrecht, 1980), p. 457.
8. R. R. Sharma, L. Vlahos, and K. Papadopoulos, Astron. Astrophys. **112**, 377 (1982).
9. R. R. Sharma and L. Vlahos, Astrophys. J. **280**, 405 (1984).
10. R. R. Winglee and G. A. Dulk, Solar Phys. **104**, 93 (1986).
11. H. W. Li, Solar Phys. **104**, 131 (1986).
12. P. Louarn, D. Le Queau, and A. Roux, Solar Phys. **111**, 201 (1987).
13. M. J. Aschwanden, Astron. Astrophys., Suppl. Ser. **85**, 1141 (1990).
14. G. D. Fleishman and S. G. Yastrebov, Solar Phys. **153**, 389 (1994).
15. G. D. Fleishman and S. G. Yastrebov, Solar Phys. **154**, 361 (1994).
16. V. G. Ledenev, Solar Phys. **179**, 405 (1998).

17. V. G. Vlasov, A. A. Kuznetsov, and A. T. Altyntsev, Astron. Astrophys. **382**, 361 (2002).
18. M. J. Aschwanden and A. O. Benz, Astrophys. J. **332**, 447 (1988).
19. C. H. Barrow, P. Zarka, and M. G. Aubier, Astron. Astrophys. **286**, 597 (1994).
20. G. D. Fleishman, Q. J. Fu, M. Wang, *et al.*, Phys. Rev. Lett. **88**, 251101 (2002).
21. G. D. Fleishman and V. F. Mel'nikov, Usp. Fiz. Nauk **168**, 1265 (1998).
22. G. D. Fleishman and K. Arzner, Astron. Astrophys. **358**, 776 (2000).
23. M. Güdel and A. O. Benz, Astron. Astrophys. **231**, 202 (1990).
24. H. Meszarosova, A. Veronig, P. Zlobec, and M. Karlicky, in *Proc. of the 10th European Solar Physics Meeting: Solar Variability: From Core to Outer Frontiers, Prague, Czech Rep., 2002* (ESA SP, 2002), Vol. 1, p. 347.
25. K. Yu. Platonov and G. D. Fleishman, Astron. Zh. **78**, 238 (2001) [Astron. Rep. **45**, 203 (2001)].
26. S. Krucker and A. O. Benz, Astron. Astrophys. **285**, 1038 (1994).
27. G. D. Fleishman, D. E. Gary, and G. M. Nita, Astrophys. J. **593**, 571 (2003).
28. G. D. Fleishman and V. F. Melnikov, Astrophys. J. **587**, 823 (2003).
29. M. J. Aschwanden and M. Güdel, Astrophys. J. **401**, 736 (1992).
30. G. D. Fleishman and S. T. Yastrebov, Astron. Zh. **71**, 531 (1994) [Astron. Rep. **38**, 468 (1994)].
31. A. Stupp, Mon. Not. R. Astron. Soc. **311**, 251 (2000).
32. J. Lee and D. E. Gary, Astrophys. J. **543**, 457 (2000).
33. J. Lee, D. E. Gary, and K. Shibasaki, Astrophys. J. **531**, 1109 (2000).
34. V. F. Melnikov, K. Shibasaki, and V. E. Reznikova, Astrophys. J. **580**, L185 (2002).
35. D. B. Melrose and G. A. Dulk, Astrophys. J. **259**, 844 (1982).
36. *Plasma Electrodynamics*, Ed. by A. I. Akhiezer (Nauka, Moscow, 1974) [in Russian].
37. R. Ramaty, Astrophys. J. **158**, 753 (1969).
38. D. B. Melrose, K. G. Rönnmark, and R. G. Hewitt, J. Geophys. Res. **87**, 5140 (1982).
39. A. O. Benz, Solar Phys. **104**, 99 (1986).
40. S. J. Wang, Y. H. Yan, and Q. J. Fu, Solar Phys. **209**, 185 (2002).
41. M. Güdel and P. Zlobec, Astron. Astrophys. **245**, 299 (1991).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 48, No. 1, 2004, pp. 9–20. Translated from Astronomicheskiĭ Zhurnal, Vol. 81, No. 1, 2004, pp. 11–23.

Mixing of Metals during Stripping of Galactic Gaseous Halos

S. Yu. Dedikov and Yu. A. Shchekinov

Space Research Department, Rostov State University, ul. Sorge 5, Rostov-on-Don, 344090 Russia

Received May 8, 2003; in final form, August 8, 2003

Abstract—The mixing of metals in the intergalactic gas when a galaxy with a metal-rich envelope moves through the intergalactic medium is analyzed. Two simple models for the initial distribution of metals are considered. In the first case, the metals are concentrated in a fairly thin envelope with thickness $\Delta R_s = 1$ kpc, outer radius $R_s = 31$ kpc, and metallicity $Z = 10^{-3}$. In the second case, material with the same metallicity uniformly fills an entire spherical region of radius R_s. After 2.85 Gyr, the metals are distributed over a fairly extended volume with a typical size of $\simeq 200$ kpc in the direction of the motion of the intergalactic gas, with a mean metallicity of $\simeq 4.6 \times 10^{-4}$ in metal-enriched regions. However, the distribution of metals remains extremely nonuniform, so that the main contribution to the overall metallicity is provided by metal-rich islands $Z \gtrsim 6 \times 10^{-4}$ that occupy only $\sim 10\%$ of the total mixing volume. Moreover, metal-free regions remain in this volume. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

In the last decade, compelling evidence has emerged suggesting that intergalactic medium (IGM) at high redshifts ($z = 2$–5.5) has been enriched in heavy elements [1–4]. The metallicity of this gas $Z \sim 10^{-3} Z_\odot$ is close to that of the oldest Population-II stars, and depends only slightly on redshift out to $z \simeq 5.5$ [3], beyond which it shows a tendency to decrease with increasing z. Interestingly, metals are observed in clouds of the Lyα forest, which have extremely low column hydrogen densities, $N(\mathrm{HI}) \simeq 10^{14}$ cm^{-2}, in agreement with modern concepts about regions of IGM in cosmic voids with sizes ~ 1 Mpc. The metal enrichment of IGM is obviously due to the activity of massive stars, including explosive processes, such as supernovae or hypernovae. At the same time, many details of these processes remain unclear: When (at what redshifts) did the active enrichment of IGM begin? Were supernovae and hypernovae capable of ejecting the enriched matter from galaxies? What must have been the nature of the accompanying gasdynamic processes in order to bring about mixing of metals on scale lengths of the order of 1 Mpc?

The active star formation in young galaxies that was responsible for the onset of the metal enrichment of the intergalactic medium led to reionization of the IGM. The epoch when this reionization began can be determined from observations of the Gunn–Peterson effect—a dip in the continuous spectrum of quasars to the blue of the resonance emission line due to extinction in neutral intergalactic gas [5]. Modern observations of the Gunn–Peterson effect at redshifts $z \simeq 5$–6 show that the reionization started in the region $z > 5.8$ [6]; there is only one data point at $z = 6.05$, where the mean Lyα transparency and the ionizaton parameter $\Gamma/(\alpha_r n)$ (ratio of the photoionization and recombination rates) decreases abruptly (by approximately an order of magnitude compared to the value at $z = 5.8$). This may mean that IGM zones begin to overlap at $z \simeq 6$, while gas is transparent in the Lyα line at lower z, although conclusive results can be obtained only after further observations of the Gunn–Peterson effect in this range of redshifts. The first stage of stellar nucleosynthesis apparently took place during the same period. Observations of damped Lyα systems at redshifts $z < 4.5$ show that their metal abundances vary little with redshift, whereas the metallicity exhibits a well-defined decrease at $z > 4.5$, albeit represented by only two observational data points [6]. In any case, $z \simeq 6$ appears to be the lower redshift boundary for the onset of stellar nucleosynthesis and the possibly related enrichment of the IGM in heavy elements. The upper boundary—$z \sim 20$–30—is less definite, and is currently determined by theoretical models incorporating cold dark matter and the vacuum energy (so-called CDM and ΛCDM models).

The enrichment of the IGM in metals and the subsequent mixing of these metals is a complex gasdynamical process, which includes the combined effect of collective supernovae, the pressure of stellar radiation on dust particles, tidal interactions between galaxies, and stripping of metal-enriched gaseous galactic halos as the galaxies move through the intergalactic medium. There is currently no doubt that the observed galaxies and their protogalactic progenitors are capable of producing the required amount

1063-7729/04/4801-0009$26.00

of heavy elements. Moreover, if the "specific" (i.e., per unit mass, including dark mass) stellar activity of each galaxy is proportional to that of the Milky Way, the amount of metals ejected by supernovae outside their parent galaxies (i.e., out to galactocentric distances 10–30 kpc) must be comparable to what we observe in the intergalactic medium [7]. However, the pressure of intergalactic gas, which is $p_{\rm igm} \simeq 0.24\Omega_b h^2(1+z)^3$ K cm^{-3} after reionization, prevents the penetration of shock waves from collective supernovae, so that the maximum filling factor of such metal-enriched regions does not exceed $q_v \sim 10^{-4}$ [7]. Rather extreme assumptions concerning the efficiency of stellar activity in low-mass, early-type galaxies ($M \lesssim 10^9\ M_\odot$, with dark matter being the main contributor) must be adopted if the metal-enriched regions around galaxies are to overlap [8]. Blowout of dust particles from their parent galaxies by stellar-radiation pressure, with the subsequent (possibly partial) destruction of these particles in the intergalactic medium, may be an efficient mechanism for the enrichment of the IGM in heavy elements [9]. However, a possible limiting factor in this case may be the magnetic fields of galaxies, which confine charged dust particles in cyclotron orbits; this remains an open question (see discussion in [10]).

In all cases, the distribution of metals in the intergalactic medium must be extremely nonuniform immediately after the ejection of another portion of heavy elements from the galaxy. In cases where the ejection is due to collective supernovae, the metallicity of the gas in the enriched region is inversely proportional to the total mass of the galaxy: $Z \propto M^{-1}$ [11]. Therefore, the observed nearly uniform heavy-element distribution (which depends only slightly on redshift) implies the existence of fairly efficient mixing mechanisms, which should also operate on scale lengths of the order of a megaparsec, corresponding to regions of low hydrogen column density in the Lyα forest occupying cosmic voids. Ferrara *et al.* [7] suggested that one such mechanism may be multiple tidal interactions, which are capable of redistributing metals over much greater regions owing to the velocity perturbations they produce in the already metal-enriched limited gas volume around the galaxies. Another possible efficient mixing mechanism is the stripping of metal-enriched gas of the galactic halo as the galaxy moves through the intergalactic medium. Unlike the tidal mechanism, which operates only episodically, stripping of halo gas as the galaxy moves through the IGM is a prolonged process. Moreover, it can amplify the tidal mechanism, since the associated extended tidal tails occupy a much greater volume than do unperturbed spherical halos, and are thus more susceptible to tidal forces produced by passing nearby galaxies. In this paper, we use a simple model to analyze the characteristic features of this very process—metal enrichment of the IGM via stripping of the gas of galactic halos. We will numerically analyze other mixing mechanisms in future papers.

The main characteristic of the mixing of metals in diffuse media such as the interstellar or intergalactic medium is that the time scale for turbulent or convective motions is the same order of magnitude as the time scales for the processes determining the physical state of the media. In the case of the intergalactic gas, with characteristic scale lengths of $L \sim 30$–100 kpc and velocities of $v \sim 100$ km/s, the time scale for the transport of inhomogeneities in chemical composition, $t \sim 0.3$–1 Gyr, is only a little shorter than the time scale for galaxy formation at redshifts $z \sim 3$–5. Similarly, in the interstellar gas, the time scale corresponding to scale lengths $L \sim 1$ kpc and velocities $v \sim 10$ km/s is $t \sim 10^8$ yr, which is comparable to the rotation period of the galaxy and to the Jeans time scale for a gas with density $n \sim 1$ cm^{-3}. Roy and Kunth [12] were the first to discuss the problem of mixing in the diffuse interstellar medium in detail. Numerical simulations of mixing in galactic disks driven by chaotic motions associated with shock waves from multiple supernovae [13] have shown that the mixing time scale may be even longer, $t \sim (2$–$3) \times 10^8$ yr. Mixing of metals on large scales may also be due to the ejection of enriched gas to large heights by collective supernovae and its subsequent radial flow to the periphery of the galaxy, as suggested in [14]. However, powerful, chaotic, small-scale motions are required to smooth small-scale metallicity inhomogeneities. Clearly, chaotic quasi-diffusion motions cannot completely "wash out" traces of the initial inhomogeneous distribution of the chemical composition due to the local nature of the enrichment process, and the observed distribution would be rather nonuniform. We analyze here mixing due to instabilities, primarily due to shear flows that occur as the galaxy moves through the intergalactic medium. Recently, Brüggen [15] and Marcolini *et al.* [16] discussed mixing in the intergalactic medium along similar lines. We argue that mixing is most efficient during initial stages, $t \sim L/v$, whereas the distribution of chemical composition shows signs of "freezing-out" on intermediate time scales $L/v < t \lesssim 10L/v$, in the sense that the metallicity distribution varies very slowly.

Section 2 describes the model, Section 3 discusses the results obtained, including the characteristic features of mixing in our model, and Section 4 presents our conclusions.

2. DESCRIPTION OF THE MODEL

We adopt the following simple model: metal-enriched gas has been ejected from the galaxy by an explosive mechanism to form a fairly thin, dense spherical envelope. The rarefied cavity may have the same metallicity as the envelope. We consider two versions: in the first, the envelope alone has a high metallicity, while, in the second, the tenuous gas in the cavity has the same metallicity. We ignore the effect of the parent galaxy on the envelope, and thus analyze the case where the envelope ejected by explosive processes has received sufficient energy to allow it to escape from the galaxy. At the same time, to simplify the analysis, we assume that the envelope is static (i.e., not expanding) at the initial time. This corresponds to the case when the expansion of the envelope is decelerated and stopped by the pressure of the intergalactic gas (see the discussion in [7, 8]). It is evident that taking into account the gravity of the galaxy would result in the attraction of the envelope and the entire perturbed region of the IGM toward the galaxy (i.e., toward the gravitating center), decreasing the volume of the mixing region compared to the case without the galaxy's gravity; in this sense, the ejection of metals into the IGM and their mixing will be suppressed.

We assume that the envelope is initially in dynamical equilibrium with the ambient gas and with the gas filling the cavity: $p^0_{igm} = p^0_s = p^0_h$. We assume that the envelope has an outer radius of $R_s = 31$ kpc and a thickness of $\Delta R_s = 1$ kpc. The gas density is $n_{igm} = 4.4 \times 10^{-6}$ cm^{-3} in the external region, $n_s = 4.4 \times 10^{-4}$ cm^{-3} in the envelope, and $n_h = 4.4 \times 10^{-6}$ cm^{-3} in the cavity; the corresponding temperatures are $T_{igm} = 10^6$ K, $T_s = 10^4$ K, and $T_h = 10^6$ K. We analyzed two distributions of metals in the envelope and cavity: (1) with the metals concentrated exclusively in the envelope, $Z_s = 10^{-3}$, $Z_h = 0$, and (2) with the mass concentrations of metals the same in the envelope and cavity, $Z_s = Z_h = 10^{-3}$. The total mass of metals injected into the system is $M_Z = 10^5\ M_\odot$, equivalent to about $N_{sn} = 5 \times 10^4$ supernovae. The external intergalactic gas moves relative to the cavity in the ζ direction at a velocity of $v_{igm} = 100$ km/s. We specified the initial density profile at the trailing boundary of the cloud so that the velocity was zero at the boundary of the cloud and equal to the unperturbed velocity of the external flow a small distance $\delta z = 0.3$ kpc from it. This introduces certain perturbations into the initial dynamics of the flow, however, the flow becomes steady over a relatively short time, $\Delta t \sim \delta R_s / v_{igm} = 10^7$ yr. We consider a more realistic flow model in which the flow dynamics are analyzed from the time of the injection of energy at the galactic center in a separate paper.

We obtained a numerical solution using a two-dimensional code operating on a cylindrical grid (r, ζ). We subdivided the $\Delta r \times \Delta\zeta = 125 \times 300$ kpc^2 domain under study into 1000×2400 cells. The center of the envelope was located along the ζ axis at the point $r = 0$, $\zeta = 100$ kpc. The computations were carried out using an explicit, multistep, finite-difference code with splitting operators (ZEUS-2D), described in [17]. We used a fixed grid, and accordingly stopped the computations at times $t \sim 3$ Gyr when some of the envelope fragments began to escape from the computational domain. We neglect effects of the Hubble expansion. For the adopted model parameters and the time interval $t \leq 9 \times 10^{16}$ s that is of interest to us, this assumption corresponds to a redshift constraint $z \leq 1.7h^{-2/3} - 1$, where $h = H_0/100$ is the Hubble constant, provided that the galaxy is not a member of a group, so that the surrounding field objects take part in the Hubble expansion. We have $z \leq 1.3$ for $h = 0.65$. Note, however, that, at redshifts $z \simeq 5$–6, large-scale structure is well separated from the mean uniform background, and most galaxies are located in denser regions corresponding to walls of the large-scale structure, where the density decreases more slowly than the isotropic Hubble expansion [18]. In this sense, our results may qualitatively describe the mixing pattern in such regions (walls of cosmic voids) at $z < 5$–6 and in clusters of galaxies.

We restrict our analysis to an adiabatic model, neglecting radiative cooling. In reality, radiative cooling may play an important role, since regions with higher metal abundances cool faster, and so rapidly contract, hindering their participation in mixing. We will demonstrate new mixing phenomena that may result due to radiative cooling in a future paper.

The metallicity Z of the gas is described by the following transport equation for a passive scalar variable

$$\frac{\partial Z}{\partial t} + \mathbf{u}\nabla Z = 0, \qquad (1)$$

written on a cylindrical grid in a second-order approximation. Here, $Z = \rho_Z/\rho$, ρ_Z is the density of metals, and ρ is the density of the gas.

The transport of metals into the intergalactic medium and their mixing is determined by the dynamical pressure $\rho_{igm} v^2_{igm}$ of the gas falling onto the envelope accompanied by the development of Kelvin–Helmholtz (KH) instability due to shear flows at the envelope boundary. In the general case, the problem is characterized by two time scales [19, 20]:

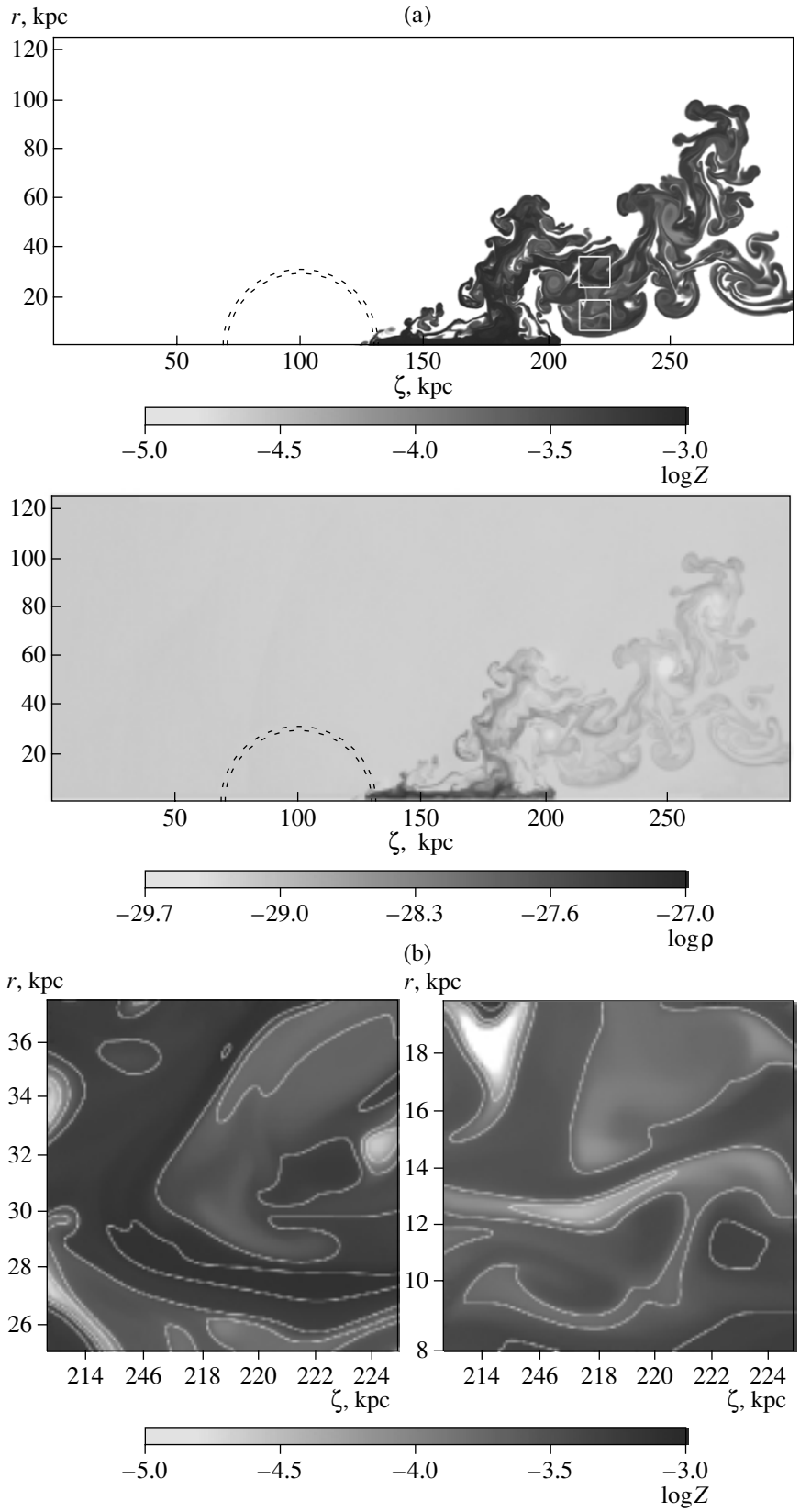


Fig. 1. (a) Overall pattern of mixing dynamics in model 1: the upper and lower panels show the distributions of metallicity and density, respectively, in the perturbed region at $t = 3$ Gyr ($\simeq 9t_{cs}$). The dashed curve shows the initial position of the envelope. (b) Spatial distribution of the metallicity in two fragments shown in Fig. 1a at time $\simeq 9t_{cs}$.

the time scale for the flow of intergalactic gas along the envelope

$$t_{\rm is} = \frac{2R_s}{v_{igm}}, \qquad (2)$$

and the time scale for the propagation of a shock across the envelope (identified with the time scale for the destruction of the latter)

$$t_{cs} \simeq 4\frac{\chi^{1/2}\Delta R_s}{v_{igm}}. \qquad (3)$$

Here, $\chi = \rho_s/\rho_{igm}$ is the density contrast between the envelope and IGM. In general, t_{cs} includes the durations of two processes: the time for penetration of a shock from the outer boundary of the envelope to its inner radius, equal to $\sqrt{\rho_s}\Delta R_s/\sqrt{\rho_{igm}}v_{igm}$, and the time it takes a shock that has passed through the envelope to cross the cavity, which is on the order of $\sqrt{\rho_h}R_s/\sqrt{\rho_{igm}}v_{igm}$. We have for the adopted parameters of the envelope and cavity $R_s\sqrt{\rho_h/\rho_{igm}} = 3\Delta R_s\sqrt{\chi}$, and this explains the factor of four in (3). In our case, $t_{cs} = 1.2 \times 10^{16}$ s.

The Kelvin–Helmholtz instability develops on a time scale [21, 22]

$$t_{\rm KH} \sim \frac{\chi^{1/2}}{kv_r}, \qquad (4)$$

where k is the wavenumber of the perturbation and v_r is the relative velocity of the intergalactic gas and envelope in the initial stages, $v_r \sim v_{igm}$. We thus have

$$t_{\rm KH} = \frac{1}{4}\frac{t_{cs}}{k\,\Delta R_s}. \qquad (5)$$

We can easily see that the instability increment increases for perturbations of the short-wavelength range, $k\ \Delta R_s \gg 1$. However, the problem has a preferred scale length, R_s, so that the major contributors to the development of vortex-type flows, and consequently to mixing, are long-wavelength perturbations, $kR_s \sim 1$, and the subsequent cascade process of the fragmentation of vortex-like fragments. Klein *et al.* [20] were the first to point out this circumstance. Under these conditions, the metal-rich envelope can be expected to break up comparatively easily into increasingly smaller fragments, but the exchange of mass between the fragments and ambient diffuse gas is hindered. The corresponding pattern for the spatial distribution of metals should include multiple metal-rich islands surrounded by tenuous regions with lower concentrations of metals.

3. RESULTS

Figure 1a shows the steady-state distributions of the metallicity (upper panel) and density (lower panel) that are established during the destruction of the envelope by the time $t = 3$ Gyr in model 1 (where metals initially fill only the envelope). The overall metal distribution in model 2 established by the same time virtually coincides with the pattern in Fig 1a; the spatial distribution of metals in model 1 retains some traces of the initial envelope structure only at times $t \lesssim 2t_{cs}$. We can clearly see that the distribution of metals remains extremely nonuniform, and that this nonuniformity is conserved on smaller scales; Fig. 1b shows separately the Z distributions in two fragments indicated in Fig. 1a. The metallicity remains nearly equal to its initial value in an appreciable part of the volume in regions close to the symmetry axis, and the density in this region remains significantly lower than the initial density of the envelope. This corresponds to the situation where, following the passage of a shock and subsequent reloading, some portions of the envelope deform and expand without taking part in vortex-like motions and mixing. Note that the total masses of gas and metals are conserved to high accuracy, so that the total mass decreases by only 3% by $t = 3$ Gyr, mostly (2%) due to the escape of some of the dense envelope gas from the computational domain. The same factor is the main reason for the $\simeq 4\%$ decrease of the total mass of metals during the same time.

Mixing can be represented as a hierarchical fragmentation of the envelope driven by the Kelvin–Helmholtz instability. This instability develops most rapidly at short wavelengths, however, Klein *et al.* [20] have pointed out that long-wavelength perturbations with wavelengths of the order of the radius of the sphere are most destructive when the gas flow interacts with the spherical structure. (Matvienko and Shchekinov [23] showed that this is due to the fact that, in such cases, the initial amplitude of the perturbation produced by the spherical structure itself, which has a wavelength equal to the radius of the sphere, is maximum, while the amplitudes of all other perturbations are comparable to the level of errors introduced by the difference representation.) In this scheme, the fragmentation process can be represented as the successive division of spherical fragments into smaller ones with half the initial radii. The KH instability for perturbations with wavelengths $\lambda \ll \sqrt{\chi}R$ develops on time scales that are substantially shorter than the time over which the cloud acquires the velocity of the external flow due to friction. We can therefore assume that the newly-formed fragments continue to move in the ζ direction at a velocity that is substantially lower than v_r. The fragments move in the radial direction at a velocity

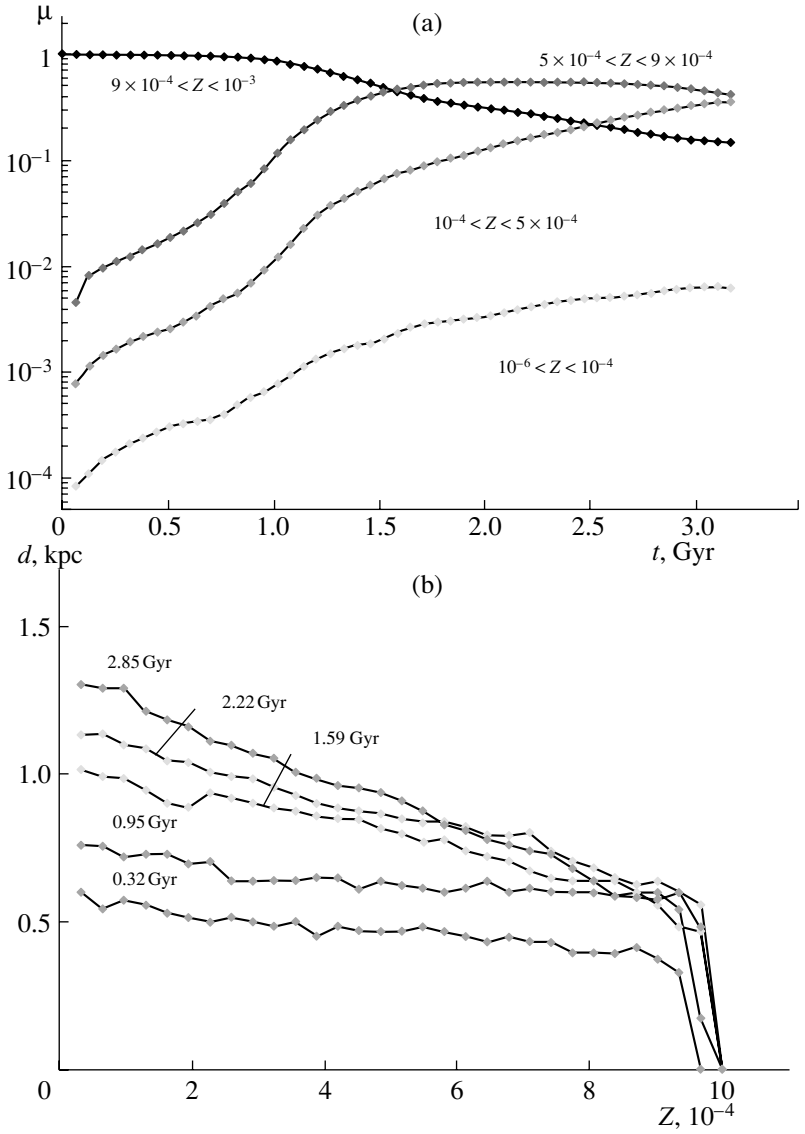


Fig. 2. (a) Mass fraction of metals (μ) contained in a given metallicity interval as a function of the time elapsed since the onset of mixing. (b) Dependence of the characteristic size d of enriched regions on their metallicity at various times.

of u_r, whose mean value $\langle u_r \rangle$ is determined by the mean density in the volume occupied by the envelope, $\langle \rho \rangle = 3\rho_s \Delta R_s / R_s + \rho_h \simeq 10\rho_{igm}$, and the velocity of the ambient intergalactic gas. After the passage of the shock through the envelope and the cavity, the mean density in the envelope increases to approximately double the unperturbed value [20]. We can therefore assume that the mean density after the passage of the shock is on the order of $\langle \rho \rangle \simeq 20\rho_{igm}$. The characteristic velocity variations in the perturbed region can be estimated from $\rho_{igm} v_r^2 \sim \langle \rho \rangle \langle u^2 \rangle$, which yields $\sqrt{\langle u^2 \rangle} \sim v_r / \sqrt{20}$. Here, we can set $\langle u_r \rangle \sim \sqrt{\langle u^2 \rangle}$— the resulting estimate coincides with the mean radial velocity in the numerical model, $\simeq 20$ km/s.

At each stage of the fragmentation, the forming fragments are shifted by $\sqrt{\chi} a u_r / v_r$ in the radial direction, where a is the size of the fragment. Therefore, if the fragments maintained the initial density contrast χ, the total displacement by the end of fragmentation would be $\Delta r = \sqrt{\chi} R_s u_r \sum_1^\infty 2^{-i} / v_r$, assuming that $|u_r| \ll v_r$ and $a_i = a_{i-1}/2$ at each stage of the fragmentation, so that radial motions contribute insignificantly to the development of the KH instability. However, the envelope material is first compressed, then expands after the decrease in the pressure, so that the density contrast decreases in the process of fragmentation. As a consequence, the fragment deceleration time decreases, with the result that fragments are carried away more rapidly by the ambient flow and, in contrast, spread more slowly in the radial

direction. After the first passage of the shock through the envelope, the temperature inside it increases to $T'_s = 3\mu v_r/16k$ and the gas pressure becomes $p'_s = 3\rho_s v_r^2/4$. During the subsequent adiabatic expansion, the compressed envelope or its fragments reach equilibrium with the ambient gas and the density inside the fragments decreases, so that the density contrast becomes $\chi' \simeq 7\chi^{3/5} M_e^{-6/5}$ by the end of the expansion, where $M_e = v_r/c_{igm}$ is the Mach number in the ambient gas. For the adopted parameters of the IGM and flow velocity we have $M_e \lesssim 1$, and we can therefore assume that $\chi' \simeq 7\chi^{3/5} \simeq 110$ in our estimates.

Subsequent interaction of fragments with the ambient flow leads to the development of shocks inside the fragments, but the temperature of the gas in these fragments is fairly high at these stages, $T_f \simeq T'_s(\chi'/\chi)^{2/3} \simeq 7^{2/3}\chi^{-4/15} \simeq T'_s$, and the additional heating by the already weak shocks is insignificant, so we can treat χ' as a steady-state value in the case considered. Thus, during cloud fragmentation driven by KH instability, the newly-formed fragments are shifted in the ζ direction by [23]

$$\Delta\zeta_i = v_{\zeta,i} t^i_{\rm KH} \simeq \frac{v_{\zeta,i-1} t^i_{\rm KH}}{1 + 1.5 t^i_{\rm KH} v_{\zeta,i-1}/(\chi' a_{i-1})} \quad (6)$$

between each two successive acts of fragmentation. Here, $t^i_{\rm KH}$ is the time scale for the development of the KH instability at the ith fragmentation step and $v_{\zeta,i} and\ a_i$ are the corresponding velocity and size of the fragment, respectively. Since $t^i_{\rm KH} \sim t^{i-1}_{\rm KH}$ and $t^{i-1}_{\rm KH} v_{\zeta,i-1} \simeq \sqrt{\chi'} a_{i-1}$, we obtain for the total displacement in the ζ direction

$$\Delta\zeta \sim \sqrt{\chi'}(1 + 1.5/\sqrt{\chi'}) R_s \simeq 260\ {\rm kpc}; \quad (7)$$

in the numerical computations, $\Delta\zeta \simeq 200$ kpc. Similarly, the radial displacement is

$$\Delta r \sim \sqrt{\chi'} \frac{\langle u_r \rangle R_s}{v_r} \simeq 75\ {\rm kpc}. \quad (8)$$

This estimate essentially gives us the maximum region of mixing under the adopted assumptions. Because of the drift in the ζ direction, the perturbed region has the shape of an irregular cone with a characteristic height and base given by (7) and (8), respectively.

We can easily see from Fig. 1 that the numerical computations are in agreement with the above estimate. We can infer from this the mean metallicity over the cylindrical volume corresponding to the entire perturbed region filled by the envelope fragments $Z = M_Z/\pi\rho'\Delta r^2\Delta\zeta \simeq Z_0/200$, where Z_0 is the initial metallicity in the envelope. However, the real metallicity distribution differs substantially from the mean distribution, since the process is highly nonlinear.

In the hierarchical fragmentation scheme presented above (described in detail in [20]), each newly-formed fragment is treated as an isolated object that continues to interact with the unperturbed external flow. In reality, after each successive fragmentation of the envelope, the resulting fragments interact not only with the ambient flow, which is perturbed, but also with each other; in particular, they shield each other, thereby decreasing the velocities of shear flows. As a result, most of the envelope remains concentrated in large and comparatively dense, weakly mixed fragments. This is evident from Fig. 2a, which shows the mass fraction μ of metals (in units of their total mass) inside regions within a given metallicity interval. The metal mass is dominated by the contribution of fragments with the maximum metallicity up to times $t \simeq 5t_{cs}$: $9 \times 10^{-4} < Z < 10^{-3}$. Only when $t > 5t_{cs}$ do fragments with $5 \times 10^{-4} < Z < 9 \times 10^{-4}$ become the main contributors to the mass of enriched regions, while metal-poor regions $Z < 5 \times 10^{-4}$, which occupy most of the total volume, contain less than 30% of the total mass.

The properties of the mixing are also clearly visible in the dependence of the characteristic size of regions on the metallicity of the gas they contain. Figure 2b shows this dependence for several times. The horizontal axis gives the size of the regions in the (r, ζ) plane calculated as the square root of the characteristic (i.e., averaged over the entire perturbed zone) area of those regions whose metallicity exceeds a specified level. It is clear that the characteristic size of regions increases with decreasing metallicity. Note that the first fairly efficient mixing takes place during early stages when $t < t_{cs}$. Subsequently, when $t \gtrsim t_{cs}$, the metal distribution remains virtually unchanged—the sizes of enriched regions increase proportionally, apparently as a result of expansion. The most metal-rich regions begin to expand more slowly than metal-poor regions only after $t \sim 4t_{cs} \sim$ 1.5 Gyr, partially because of fragmentation. However, the sizes of the fragments with the maximum metallicity (equal to its initial value) always remain substantially larger than the grid size and only a factor of a few smaller than the maximum size of the (most metal-poor) fragments; this difference is less than a factor of four over long times. This behavior indicates the existence of regions that play only a very small role in the mixing. This conclusion is supported by the fact that the time dependence of the characteristic size of regions with metallicities $Z = (8\text{–}9.5) \times 10^{-4}$ varies nonmonotonically when $t \geq 0.95$ Gyr. It is evident from the distribution in Fig. 2b that, over long times, the characteristic cross section of metal-rich regions

($Z > 8 \times 10^{-4}$) is about a factor of four smaller than that of metal-poor regions ($Z < 2 \times 10^{-4}$), implying that the probability of finding such regions along the line of sight could be smaller by the same factor. As a result, determinations of the metallicity of the intergalactic gas based on quasar absorption spectra may be affected by observational selection effects, which result in substantial underestimation of the contribution of high-metallicity regions.

The mixing is affected substantially by the deformation of the fragments, which, in turn, amplifies short-wavelength perturbations via the KH instability, which further deform the fragment surfaces. This results in a lower efficiency for the fragmentation of the envelope and a substantial increase of its surface area and thereby the deceleration force, which carries the fragment with the flow and decreases the relative velocity of the flow with respect to the surface of the cloud. As a result, some of the fragments remain undestroyed and, due to the decrease of the relative shear velocity, actually become inaccessible for subsequent mixing—"pockets" of nearly constant or slowly varying heavy-element abundances develop. Thus, the distribution of metals during the asymptotically intermediate stages that we analyze here displays features typical of intermittent distributions [24]. This is testified to, in particular, by the fact that relatively small regions occupying $\sim 10\%$ of the total perturbed volume and having high metallicities ($\gtrsim 6 \times 10^{-4}$) are the main contributors to the mean metallicity in the entire volume of the perturbed motions covered by our computations.

Another characteristic feature of intermittent systems—sharp peaks in the distributions of random variables—can be seen in the metallicity histograms in Fig. 3a, i.e., the number $N(Z)$ of grid cells with metallicities in a given interval. It is evident that, although an appreciable fraction of the metal mass spreads throughout the envelope over an ever increasing number of cells, a considerable number of cells also retain the initial metallicity, as is shown in the four panels corresponding to various times. This suggests that, on small scales, the metals are concentrated in fairly isolated islands with limited mass exchange. In our case, intermittency in the mixing of the metals is a natural consequence of the nonlinear nature of the initial metallicity distribution and transport equation (1), of which function $Z[\mathbf{r} - \mathbf{v}(t)t, 0]$ is a solution. Here, $Z[\mathbf{r}, 0]$ is the initial metal distribution, which we specify in the form of a Θ function. It can easily be seen (see [24]) that, even in the case of a Gaussian velocity field $\mathbf{v}(t)$, the final, or asymptotic, distribution of Z is characterized by the presence of regions with metallicities deviating strongly from the mean value. We expect that including radiative losses for regions containing heavy elements should enhance the intermittency, since the most metal-rich regions should then lose thermal energy faster and contract to higher densities, thereby decreasing substantially the increment of the Kelvin-Helmholtz instability $\Gamma \propto \chi^{-1/2}$.

The distribution of the metallicity over the number of grid cells shown in Fig. 3a correctly reproduces the time evolution of the spatial distribution of the metals, but it conveys a wrong impression about the volume filling factors for regions with various metallicities, and consequently about the masses of metals contained in these regions. This is due to the cylindrical geometry of our model: the volume of grid cells increases proportional to the distance r from the axis of the cylinder.

Figure 3b (solid) shows the volume distribution of the metals $V(Z)$. The substantially smaller amplitudes of the peaks in the distribution in Fig. 3b compared to the peaks in Fig 3a is obvious, and is due to the fact that the negligible numbers of particles with certain metallicities Z corresponding to sharp minima of $N(Z)$ are located at fairly large distances and occupy large volumes. However, the behavior of the $V(Z)$ distribution shows that the mixing in our model clearly differs from simple diffusion: even on short time scales, $t \sim 8t_{cs}$, the volume occupied by regions with metallicities $Z \simeq 5 \times 10^{-4}$ is only a third the volume occupied by gas with metallicities that are two and a half orders of magnitude lower, $Z \sim 2 \times 10^{-6}$. In addition, the $V(Z)$ distribution shows well-defined extrema on time scales $t > 6t_{cs}$, in particular, near $Z \sim 10^{-4}$.

The redistribution of metals from metal-rich regions over a large volume appears to take place mostly during the very first stages of the interaction between the intergalactic flow and the envelope surrounding the galaxy. This is clear from the (ρ, Z) diagram relating the gas density and metallicity in a cell (Fig. 4): in the very first stages of interaction ($t \sim 2t_{cs}$), some metals are ejected together with fragments into the tenuous ambient gas where they are mixed (upper panel), whereas subsequent and even rather late ($t \sim 9t_{cs}$) stages correspond mostly to the expansion of dense (most metal-rich) regions ($6 \times 10^{-29} < \rho < 8 \times 10^{-28}$ g cm^{-3}) and weak mixing in tenuous regions ($5 \times 10^{-30} < \rho < 6 \times 10^{-29}$ g cm^{-3}).

The differences of the properties of mixing due to the KH instability from the properties of ordinary diffusion mixing become especially clear when the metallicity histograms $V(Z)$ obtained in our model (Fig. 3b, solid lines) are compared with the metallicity distribution for pure diffusion. The dashed lines in Fig. 3b show the metallicity distribution produced by

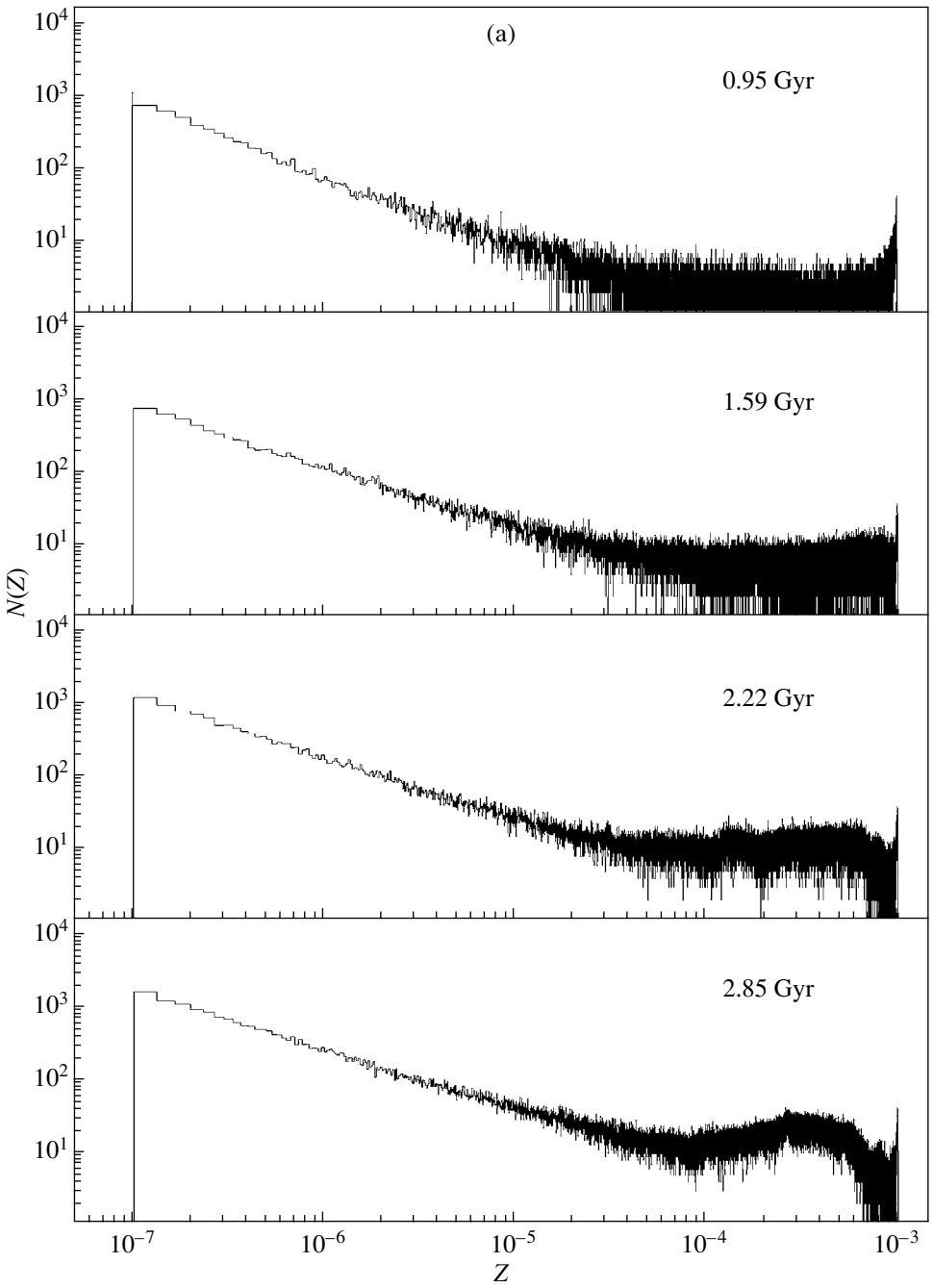


Fig. 3. (a) Distribution of metallicity over numbers of grid cells $N(Z)$ for various times (indicated in the top right corners of the panels). (b) Volume distribution of the metallicity $V(Z)$ in units of the initial volume of the envelope for various times (indicated in the top right corners of the panels); the dotted curve shows the distribution $V(Z)$ expected for diffusion mixing with diffusion coefficient $\kappa = 10^{29}$ cm^2/s.

ordinary diffusion, described by the linear diffusion equation in the spherical case:

$$\frac{\partial Z}{\partial t} = \frac{\kappa}{r^2}\frac{\partial}{\partial r}\left(\frac{1}{r^2}\frac{\partial Z}{\partial r}\right), \qquad (9)$$

with the diffusion coefficient $\kappa = \langle u\rangle R_s/3 \sim 10^{29}$ cm^2/s, where we have taken the minimum size of a turbulent cell associated with KH instability to be R_s, $\langle u\rangle \sim \langle u_r\rangle \simeq 20$ km/s.

We solved (9) on the same grid as for our hydrodynamical model, with the initial distribution of metals being identical to that adopted for the enriched spherical envelope. The distribution obtained in these numerical computations clearly differs fundamentally

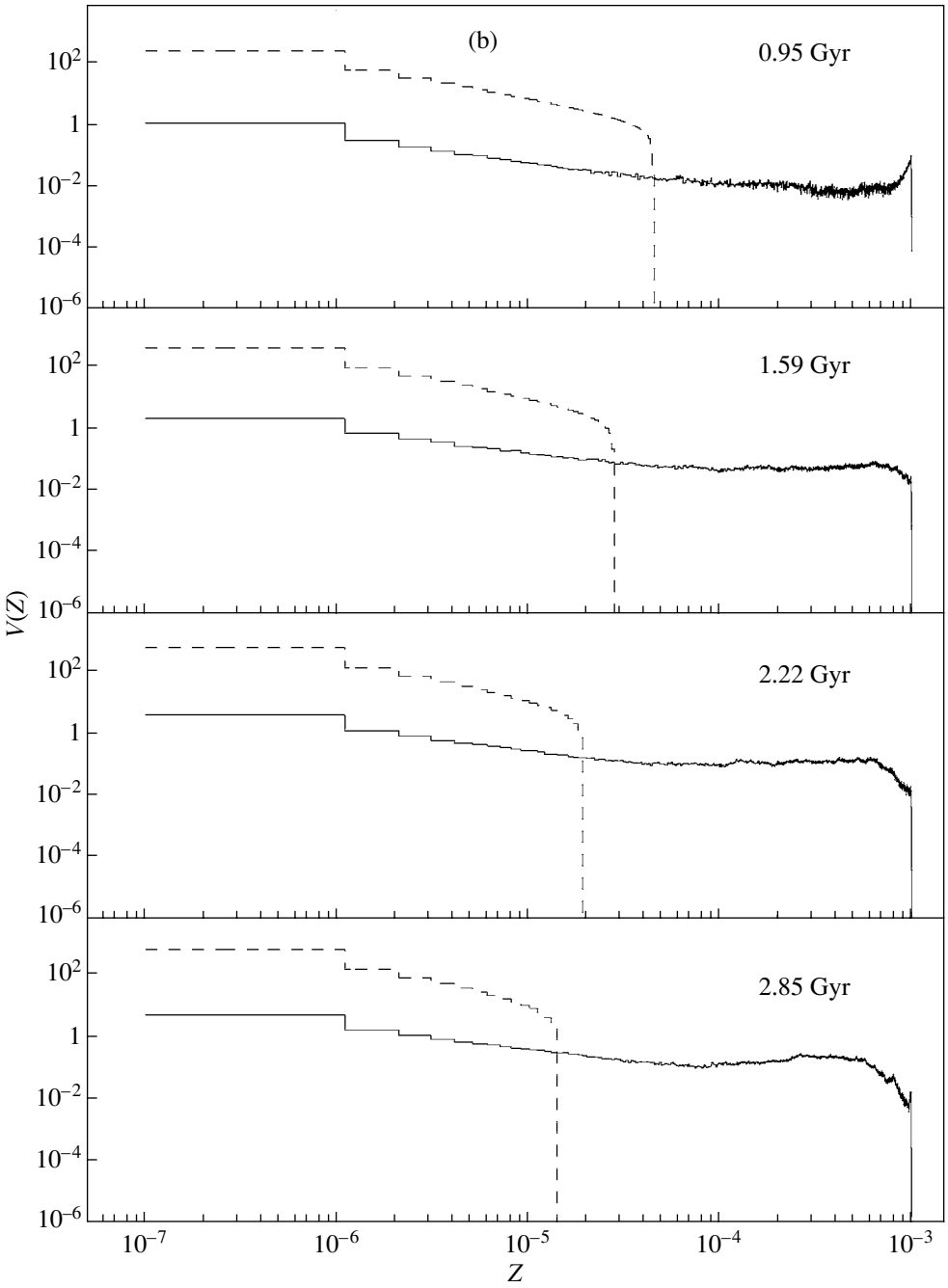


Fig. 3. (Contd.)

from the distribution for a diffusion process, and displays certain features of intermittency. This is especially clearly manifested in the fact that the mixing in the numerical model leaves many grid cells with high metallicities, whereas ordinary diffusion is characterized by some maximum metallicity (which decreases with time). Note that mixing can be described as an efficient diffusion process (with an exponentially growing diffusion coefficient) in the model of [13] for mixing in the interstellar medium driven by multiple supernova shocks. This is apparently due to the fact that each element of gas is repeatedly subjected to shocks, which lead to the repeated development of the KH instability, and consequently to the progressive fragmentation of vortex cells on ever decreasing scales. In our case, each newly formed vortex cell de-

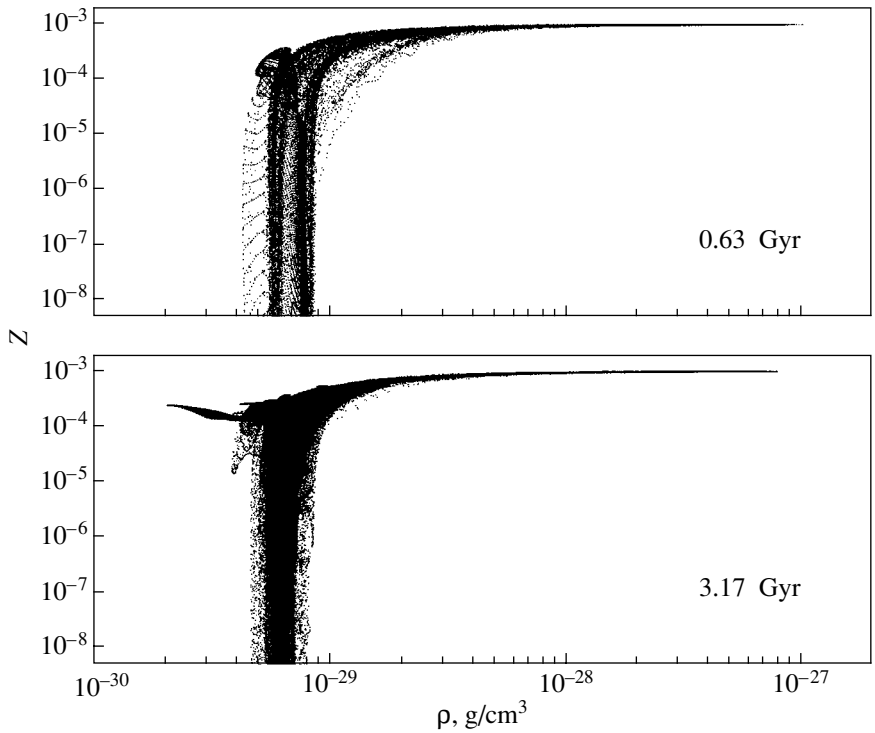


Fig. 4. Density–metallicity diagrams for $t = 1.7t_{cs} = 6.34 \times 10^8$ and $t = 3.17 \times 10^9$ yr.

creases its velocity relative to the perturbing flow, so that the instability reaches a "saturated" mode—the formation of new vortices slows, as does the mixing associated with this process.

4. DISCUSSION AND CONCLUSIONS

(1) We have analyzed a numerical model for the mixing of metals in the intergalactic medium as a result of the stripping of metal-rich circumgalactic envelopes by the ram pressure of the incident intergalactic gas. If radiative losses can be neglected, the metals are mixed in a conical volume with a length of $\sim 8R_s$ (equal to 200 kpc in the numerical computations) and a maximum transverse size of $\sim \sqrt{6}R_s \sim$ 75 kpc.

(2) The distribution of metals remains highly nonuniform and is characterized by intermittency of regions of high and very low metallicities, so that the main contributors to the mean metallicity of the flow volume are isolated metal-rich "islands" with a total volume of $\sim 10\%$ of the entire volume of the perturbed region. At the same time, fairly extended metal-free regions remain in the volume encompassed by strong perturbed flows. We expect that the inclusion of radiative losses would result in an even more nonuniform Z distribution.

(3) The characteristic sizes of the metal-rich regions are a few times smaller and their geometric cross sections about a factor of four smaller. If the metal distribution is the same in the intergalactic medium, this could lead to observational selection effects that could result in underestimation of the metallicity of the intergalactic gas.

ACKNOWLEDGMENTS

We are grateful to B.M. Shustov for critical comments. This work was supported by the Russian Foundation for Basic Research (project no. 00-02-17689) and the Federal Science and Technology Program "Astronomy" (project no. 1.3.1.1).

REFERENCES

1. D. Tytler, X.-M. Fan, S. Burles, *et al.*, in *QSO Absorption Lines*, Ed. by G. Meylan (Springer, Berlin, 1995), p. 289.
2. L. L. Cowie, A. Songaila, T.-S. Kim, and E. Hu, Astron. J. **109**, 1522 (1995).
3. A. Songaila, Astrophys. J. **561**, L153 (2001).
4. A. Songaila and L. L. Cowie, Astron. J. **112**, 335 (1996).
5. I. S. Shklovskiĭ, Astron. Zh. **41**, 801 (1964) [Sov. Astron. **8**, 638 (1964)].
6. A. Songaila and L. L. Cowie, Astron. J. **123**, 2183 (2002); astro-ph/0202165.
7. A. Ferrara, M. Pettini, and Yu. A. Shchekinov, Mon. Not. R. Astron. Soc. **319**, 539 (2000).

8. P. Madau, A. Ferrara, and M. Rees, Astrophys. J. **555**, 92 (2001).
9. B. M. Shustov and D. Z. Vibe, Astron. Zh. **72**, 650 (1995) [Astron. Rep. **39**, 578 (1995)].
10. R.-J. Dettmar and Yu. A. Shchekinov, Astron. Astrophys. (2003, in press).
11. Yu. A. Shchekinov, Astron. Astrophys. Trans. **21**, 131 (2002).
12. J.-R. Roy and D. Kunth, Astron. Astrophys. **294**, 432 (1994).
13. M. A. de Avillez and M.-M. Mac Low, Astrophys. J. **581**, 1047 (2002).
14. G. Tenorio-Tagle, Astron. J. **111**, 1641 (1996).
15. M. Brüggen, Astrophys. J. **571**, L13 (2002).
16. A. Marcolini, F. Brighenti, and A. D'Ercole, astro-ph/0212440 (2002).
17. J. M. Stone and M. L. Norman, Astrophys. J., Suppl. Ser. **80**, 753 (1992).
18. Y. Zhang, A. Meiksin, P. Anninos, and M. L. Norman, Astrophys. J. **495**, 63 (1998).
19. K. V. Bychkov and S. B. Pikel'ner, Pis'ma Astron. Zh. **1**, 29 (1975) [Sov. Astron. Lett. **1**, 14 (1975)].
20. R. I. Klein, C. F. McKee, and P. Colella, Astrophys. J. **420**, 213 (1994).
21. L. D. Landau and E. M. Lifshitz, *Hydrodynamics* (Nauka, Moscow, 1986) [in Russian].
22. S. Chandrasekhar, *Hydrodynamic and Hydromagnetic Stability* (Dover, New York, 1961).
23. E. E. Matvienko and Yu. A. Shchekinov, Astron. Zh. (2003, in press).
24. Ya. B. Zel'dovich, S. A. Molchanov, A. A. Ruzmaikin, and D. D. Sokolov, Usp. Fiz. Nauk **152**, 3 (1987) [Sov. Phys. Usp. **30**, 353 (1987)].

Translated by A. Dambis