Astronomy Reports, Vol. 45, No. 3, 2001, pp. 163–172. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 195–204.

Current Status of Models with Hot and Cold Dark Matter

E. V. Mikheeva, V. N. Lukash, N. A. Arkhipova, and A. M. Malinovskiĭ

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

Received January 20, 2000

Abstract—An analysis of a five-parameter family of cosmological models in a spatially flat Friedmann Universe with a zero Λ term is presented. The five parameters are (1) σ_8, the dispersion of the mass fluctuations in a sphere with radius $8h^{-1}$ Mpc, where $h = H_0/100$ km s^{-1} Mpc^{-1} and H_0 is the Hubble constant; (2) n, the slope of the density-perturbation spectrum; (3) Ω_ν, the normalized energy density of hot dark matter; (4) Ω_b, the baryon density; and (5) h, the normalized Hubble constant. The density of cold dark matter is determined from the condition $\Omega_{cdm} = 1 - \Omega_\nu - \Omega_b$. Analysis of the models is based on comparison of computational results with observational data for: (1) the number density and mass function of galaxy clusters (a so-called Press–Schechter formalism) and (2) the cosmic microwave background anisotropy. The first method enabled us to determine the value $\sigma_8 = 0.52 \pm 0.01$ with high accuracy. Using the resulting normalization of the density-perturbation spectrum, we calculated a model for the anisotropy of the cosmic microwave background radiation on large scales ($l \simeq 10$, where l is the harmonic number) and the required contribution of cosmological gravitational waves, characterized by the parameter T/S. The restrictions on T/S become weaker as Ω_ν increases. Nevertheless, even when $\Omega_\nu \le 0.4$, models with $h + n \ge 1.5$ require a considerable contribution from gravitational waves: T/S $\gtrsim 0.3$. On the other hand, in models with $\Omega_\nu \le 0.4$ and a scale-invariant density-perturbation spectrum ($n = 1$), we find T/S $\gtrsim 10(h - 0.47)$. The minimization of T/S is possible only for the family of models with red spectra ($n < 1$) and small h (<0.6). The value of Ω_ν is determined most accurately by the data on $\Delta T/T$ near the first acoustic peak ($l \simeq 200$). By imposing a general restriction on the amplitude of gravitational waves T/S $\in [0, 3]$ and taking into account the available observational data on the amplitude of the acoustic peak of Sakharov oscillations, ranges of possible values n and Ω_ν are derived. If the baryon number is constrained by nucleosynthesis data, the models under consideration can have both moderately red and blue power spectra $n \in [0.9, 1.2]$ with a rather high concentration of hot particles $\Omega_\nu \in [0.2, 0.4]$. The conditions that $n < 0.9$ and/or $\Omega_\nu < 0.2$ decrease the relative amplitude of the acoustic peak by over 30% compared to its value in the standard cold-dark-matter (CDM) model normalized using COBE data. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The vigorous development of the observational basis for cosmological studies in the 1990s has made it possible to progress from theoretical investigations of cosmological models to direct testing of them. The most important result has been the rejection of the standard CDM model,[1] following the discovery of large-scale anisotropy of the cosmic microwave background radiation (CMBR) in 1992 [2–4], whose amplitude on scales of 10° turned out to be appreciably greater than predicted by theory.

Since the simplest cosmological models could not give an adequate description of the large-scale structure of the Universe, it was clear these models needed to be modified. Obviously, this could be done in several ways: by changing the model for the dark matter (i.e., using mixed instead of cold dark matter) and/or introducing a non-zero Λ term ([5] and references therein; [6]), by abandoning spatially flat models of the Universe in favor of open models [7], by modifying the primordial spectrum of density perturbations and cosmological gravitational waves [8–17], etc. (see the review [19]).

The most successful modern theory of the early Universe—inflation—pays special attention to cosmological models whose total energy density is equal to the critical density; i.e., models in which the comoving three-dimensional space is Euclidean. In the simplest versions of this type of model, the cosmological constant is zero. We will investigate this type of model, using the following quantities as free parameters:

(1) σ_8, the dispersion of mass fluctuations in a sphere with radius $8h^{-1}$ Mpc ($\sigma_8 \in [0.47, 0.61]$, 15 models with step 0.01), which is linearly related to the amplitude of the density-perturbation spectrum;

[1] The standard CDM (cold dark matter) model has energy density of ordinary matter in the Universe in units of the critical density $W_m = W_{cdm} + W_b = 1$, where W_{cdm} and W_b are the normalized energy densities of dark matter and baryons respectively, $W_b = 0.05$ [1], $h = H_0/100$ km s^{-1} Mpc^{-1} = 0.5, H_0 is the Hubble constant, the slope of the density-perturbation spectrum is $n = 1$, the amplitude of cosmological gravitational waves is zero, and the spectrum of density perturbations is normalized by the COBE data. Note also that, with this normalization, $\sigma_8 > 1$; where σ_8 is the dispersion of the mass fluctuations in a sphere with radius $8h^{-1}$ Mpc.

(2) n, the slope of the density-perturbation spectrum ($n \in [0.8, 1.4]$, seven models with step 0.1);

(3) Ω_ν, the energy density of hot dark matter in units of the critical density ($\Omega_\nu \in [0, 0.4]$, five models with step 0.1);

(4) Ω_b, the baryon energy density in the Universe in the same units ($\Omega_b \in [0.01, 0.11]$, six models with step 0.02); and

(5) the Hubble constant $H_0 = 100h$ km s^{-1} Mpc^{-1} ($h \in [0.45, 0.70]$, six models with step 0.05).

Combining the various values of these parameters, we obtain 18900 distinct cosmological models to be investigated.

The amount of cold matter is not a free parameter, and is determined by the condition $\Omega_{cdm} = 1 - \Omega_\nu - \Omega_b$. The amplitude of the tensor mode is expressed by the parameter T/S (representing the relative contribution of gravitational waves to the large-scale anisotropy of the CMBR), which is calculated for models that reproduce the observed mass function of galaxy clusters by comparing them with COBE data. We restrict the range of allowed values of T/S to the interval $0 \le T/S \le 3$.

The main aim of our study is to put useful limits on these five model parameters, allowing for a cosmological background of gravitational waves, based on the observed mass function of galaxy clusters and the CMBR anisotropy on both large ($l \sim 10$) and small ($l \sim 200$) angular scales. (Here, l is the harmonic number in an expansion of the CMBR fluctuations $\Delta T/T$ in spherical functions.) We shall restrict our consideration to these most important tests, and address the question of the adequacy and sensitivity of the model parameters to the observational data. The basic problem is to determine the importance of hot matter in cosmological models with gravitational waves (without a Λ term). The results will be used to optimize the step choice and the allowed intervals of the parameters in more complex models (in particular, with cosmological constant $\Lambda > 0$), which will require a more careful analysis, taking into account data on the Ly_α forest, the power spectrum of galaxy clusters, etc.

The first parameter—σ_8—gives a more accurate normalization of the density-perturbation spectrum than the COBE data, because of the possible contribution of primordial gravitational waves to the large-scale CMBR anisotropy.

The second parameter—n—generalizes a flat density-perturbation spectrum (Harrison–Zeldovich spectrum) in the most natural and simple way. Power-law spectra are predicted by many inflationary models, for example, by chaotic inflation based on scalar field theory with a power-law potential [19]. The spectra of other inflationary models can be both "red" ($n < 1$; i.e., the spectral power at large wavelengths is increased over smaller scales) [20] and "blue" ($n > 1$), and non-power-law spectra are also possible [8–17, 21, 22].

The introduction of hot dark matter is partially associated with the recent discovery of atmospheric neutrino oscillations [23–26], which imply a non-zero rest mass for at least one kind of neutrino. The masses of all kinds of neutrinos remain unknown, so that we can consider Ω_ν a free parameter of the models (we shall assume later that only one kind of neutrino has a mass).

The fourth parameter—the number of baryons in the Universe—affects the evolution of density perturbations only slightly. It is evident that neither the mass function of galaxy clusters nor the large-scale CMBR anisotropy can be used to estimate this parameter with high statistical significance. Nevertheless, the value of Ω_b is extremely important for the amplitude of the CMBR anisotropy on moderate ($\theta \sim 1^\circ$) and small ($\theta \ll 1^\circ$) angular scales (Sakharov oscillations). The range of Ω_b chosen in our study is in accordance with modern data on primordial nucleosynthesis [1].

Finally, the fifth free parameter is the Hubble constant. There is currently a discrepancy of about ~20% in estimates of h obtained using different methods, based on observations of type Ia supernovae, Cepheid variables, etc. (see, for example, [27–30]). It is clear that h cannot be less than 0.45. On the other hand, if h were too high, we would be forced to introduce a positive Ω_Λ, in accordance with the ages of old globular clusters. Precisely these reasons determined our choice of the range of this parameter in our studies.

Our subsequent presentation will be organized as follows. In Section 2, based on observational data on numbers of galaxy clusters, we calculate a χ^2 for each model and determine the most probable values of the parameters. The highest level of statistical confidence can be attained for σ_8. We consider the anisotropy of the CMBR in Section 3. By normalizing the density-perturbation spectra by the obtained value of σ_8, we calculate T/S using data on large-scale CMBR anisotropy. By imposing the constraint $T/S \in [0, 3]$, we separate out models satisfying this criterion. Further, we select preferred models based on observational data in the region of the first acoustic peak. Our final results are formulated and discussed in Sections 4 and 5.

2. THE MASS FUNCTION OF GALAXY CLUSTERS

The Press–Schechter method [31] enables derivation in explicit form of the mass function of gravitationally bound, virialized objects. This method is based on two main assumptions:

(1) the field of density fluctuations $\delta \equiv \delta\rho/\rho$ in a linear approximation can be described by a Gaussian distribution;

(2) the gravitational collapse of regions of enhanced density isolated from the Hubble expansion (massive halos) can be described using a spherically symmetric approximation.

Although the second assumption is a rather crude idealization of the real evolution of density peaks, stud-

ies carried out over 20 years have shown that the main result of such an approach—the integral mass function—is stable and in good agreement with numerical simulations of the *N*-body problem ([32] and references therein). Unfortunately, the basic formalism does not contain a mechanism for the merging of several objects, resulting in the formation of a more massive object. Therefore, as a rule, the Press–Schechter method is applied to analyses of spatial distributions of objects with large masses (such as galaxy clusters), when the effects of merging are negligible.

In accordance with the Press–Schechter formalism, the spatial density of virialized halos with masses exceeding some value M is determined by the expression

$$N(>M) = \int_M^\infty \frac{dn}{dM'} dM', \qquad (1)$$

where

$$\frac{dn}{dM} = \sqrt{\frac{2}{\pi}} \frac{\rho_0 \delta_c}{M} \frac{1}{\sigma_R^2} \left|\frac{d\sigma_R}{dM}\right| \exp\left(-\frac{\delta_c^2}{2\sigma_R^2}\right) \qquad (2)$$

is the differential mass function, $M = \frac{4}{3}\rho_0 R^3$, ρ_0 is the matter density in the Universe, δ_c the threshold density contrast for the formation of a halo (δ_c =1.686 [33]), and σ_R the dispersion of linear mass perturbations in a sphere with radius R (a "top-hat" filter), which is an integral function of the power spectrum of the density perturbations:

$$\sigma_R^2 = \int_0^\infty \Delta_k^2 W^2(kR) \frac{dk}{k}. \qquad (3)$$

Here, $\Delta_k^2 = \frac{1}{2\pi^2} P(k) T^2(k)$ is the dimensionless power spectrum; $P(k) = Ak^{n+3}$ the primordial density-perturbation spectrum; A a normalization constant; $T(k)$ the transition function, which describes the evolution of density perturbations in the Universe and depends on Ω_ν, Ω_b, and h; and $W(kR)$ the Fourier transform of a "top-hat" smoothing function, $W(x) = \frac{3}{x^3}(\sin x - x\cos x)$.

We can see from Eqs. (1) and (2) that the integral curve is determined primarily by the normalization of the perturbation spectrum, and consequently can be used to calculate this normalization. The dependence of $N(>M)$ on σ_8 for fixed values of the other parameters is illustrated in Fig. 1, and the dependence on Ω_ν is shown in Fig. 2. The ten observational points were taken from [34] and are listed in the table. Similar plots can be drawn for the other model parameters, however, the dependence of $N(>M)$ on n, Ω_b, and h is less pronounced. For example, when n is varied, the change in the shape of the curves has an opposite character than in the case of Ω_ν. The curves have a constant amplitude of $M \simeq 10^{14.8} M_\odot$; their slopes vary, but the discrepancy between the curves for the limiting values $n = 0.8$ and 1.4 is approximately half that for the corresponding limiting values $\Omega_\nu = 0.4$ and 0 in Fig. 2. The functions $N(>M)$ for various values of Ω_b and h virtually coincide.

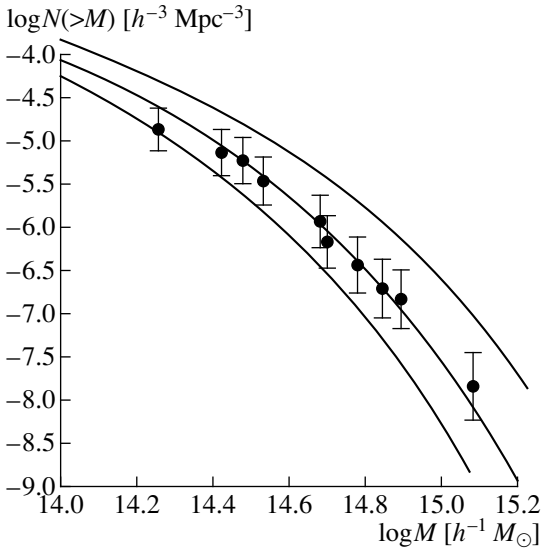


Fig. 1. Dependence of $N(>M)$ on σ_8 ($n = 1$, $\Omega_\nu = 0$, $\Omega_b = 0.05$, $h = 0.5$). The upper curve corresponds to the model with $\sigma_8 = 0.61$, the middle curve to $\sigma_8 = 0.52$, and the lower curve to $\sigma_8 = 0.47$. The observational data are marked by dots.

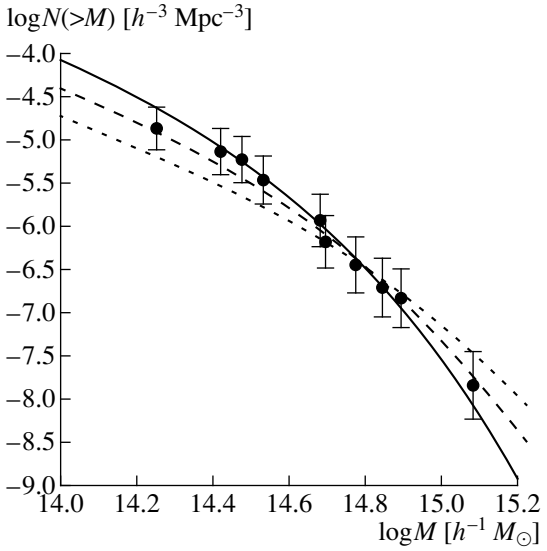


Fig. 2. Dependence of $N(>M)$ on Ω_ν ($\sigma_8 = 0.52$, $n = 1$, $\Omega_b = 0.05$, $h = 0.5$). The solid curve corresponds to the model with $\Omega_\nu = 0$, the dashed curve to $\Omega_\nu = 0.2$, and the dotted curve to $\Omega_\nu = 0.4$. The observational data are marked by dots.

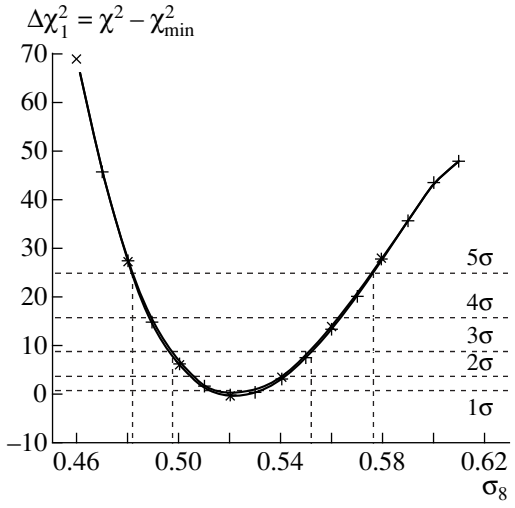


Fig. 3. $\Delta\chi_1^2 \equiv \chi^2 - \chi^2_{\min}$ as a function of σ_8. The curve marked by right-angle crosses corresponds to the approximation for the transition function from [36], while the curve marked by oblique crosses corresponds to the approximation from [35]. The confidence levels are shown by horizontal lines.

We used a χ^2 criterion for our statistical analysis. This distribution describes the sum of the squares of independent quantities characterized by a normal distribution with zero mathematical expectation and unit dispersion. The model with minimum χ^2 ($\chi^2_{\min} = 1.43$) has $\sigma_8 = 0.52$, $n = 1.3$, $\Omega_\nu = 0.3$, $\Omega_b = 0.01$, and $h = 0.7$. Yet, what is the statistical significance of this result?

A statistical analysis of the cosmological models under investigation is illustrated in Fig. 3, which shows the distribution function $\Delta\chi_1^2(a) = \chi_a^2 - \chi^2_{\min}$, where $a = \sigma_8$ and χ_a^2 is the minimum χ^2 at the hypersurface (a = const) in the five-parameter model space. The 1σ, 2σ, and 3σ levels correspond to the values $\Delta\chi_1^2 = 1$, 4, and 9, respectively. We can see that a high level of statistical significance (3σ or more) was achieved in the determination of σ_8. Using only the Press–Schechter method does not enable the specification of the values of the other model parameters, even with a lower statistical significance (for example, 1σ), and any values of these parameters within the intervals under consideration are allowed.

Observational data on the number density of galaxy clusters [16]: $N(>M) = A \times 10^{-B}$ $(h / \mathrm{Mpc})^3$, $M = C \times 10^{15}\, M_\odot/h$

A	*B*	*C*
13.5	6 ± 0.18	0.18
7.5	6 ± 0.23	0.264
6.0	6 ± 0.18	0.30
3.5	6 ± 0.23	0.34
1.2	6 ± 0.18	0.48
6.9	6 ± 0.23	0.50
3.7	6 ± 0.23	0.60
2.0	6 ± 0.23	0.70
1.5	7 ± 0.3	0.78
1.5	8 ± 0.3	1.20

Note: Statistical errors are 1σ.

To estimate deviations associated with the accuracy of our approximation of the transition function, we calculated cosmological models based on two different approximations. The first (the curve is marked by oblique crosses in Fig. 3) was taken from [35], whereas the second (marked by right-angle crosses) was taken from [36]. Figure 3 shows that the two analytical approximations do not contradict each other, and lead to similar results for σ_8. Our conclusion about the large statistical uncertainty of the other model parameters remains valid, and does not depend on the chosen approximation.

Thus, the Press–Schechter method enables us to determine a range of allowed values only for the single parameter σ_8, but with very high accuracy: the 1σ statistical error is 0.01. Taking into account systematic experimental errors and the accuracy of the Press–Schechter approximation increases somewhat the total error in σ_8 (see, for example, [37, 38], where the total error was estimated to be 0.04).

3. ANISOTROPY OF THE COSMIC MICROWAVE BACKGROUND RADIATION

3.1. Large-Scale CMBR Anisotropy

Large-scale CMBR anisotropy $\Delta T/T$ is intrinsically associated with metric perturbations due to the Sachs–Wolfe effect [39]:

$$\frac{\Delta T}{T}(\mathbf{e}) = \frac{1}{2}\int_R^E \frac{\partial h_{ik}}{\partial \eta} e^i e^k d\eta, \qquad (4)$$

where E and R are the times of the emission and reception of a photon, h_{ik} is the metric perturbation tensor, $\partial/\partial\eta$ is the derivative with respect to conformal time, the integration is carried out along the line of sight, and $e^i = (1, \mathbf{e})$.

As is known, perturbations of the metric can be separated into three classes: scalar (responsible for density perturbations and the potential velocity field), vector (associated with the vortex velocity field), and tensor (describing gravitational waves) [40]. In inflationary theory, based on the existence of a scalar field in the early Universe, only perturbations of the first and third types are created. Inflation does not discriminate against any type of perturbation, but the absence of vector and spinor fields in the linear approximation leads

to the absence of a vortex mode in first-order perturbation theory.

The relation between the amplitudes of the scalar and tensor modes depends on the shape of the inflaton potential; the relative amplitude of gravitational waves in some inflationary models is small, whereas in other models it approaches the amplitude of the scalar mode [15–17, 21, 22, 41–44]. As a rule, gravitational waves are intensively created if the density-perturbation spectrum produced in the course of inflation differs considerably from a flat spectrum, or is not a power-law at all (as in the cases of power-law [20] or Λ inflation [21, 22], respectively). The energetic scale of such inflation should be of the order of the Grand Unification energy ($H_{\rm inf} \gtrsim 10^{13}$ GeV).

Despite the large amount of observational data that has been obtained, the available information remains insufficient to draw unambiguous conclusions about the amplitude of the gravitational-wave mode of the metric perturbations. This will become possible only in several years, when high-precision measurements of the polarization properties of the CMBR anisotropy and the spatial distribution of galaxies on scales to ~1000 Mpc will be carried out. Such polarization measurements are scheduled for the largest Russian radio telescope RATAN-600 and other ground-based telescopes, as well as satellites associated with future space missions; measurements of the spatial distribution of galaxies will be conducted as part of the SDSS (Sloan Digital Sky Survey) and other projects. Thus, due to the absence of data on the amplitude of gravitational waves, we will treat this as an extra parameter to be calculated in the model.

From a quantitative point of view, the amplitude of gravitational waves can conveniently be expressed in terms of the relative contribution of gravitational waves to the large-scale CMBR anisotropy T/S:

$$\left\langle \left(\frac{\Delta T}{T}\right)^2 \right\rangle_{10^\circ} = \mathrm{S} + \mathrm{T} = \mathrm{S}\left(1 + \frac{\mathrm{T}}{\mathrm{S}}\right) \simeq 1.1 \times 10^{-10}. \quad (5)$$

In all reasonable models, the value of T/S does not exceed several units. Our subsequent analysis will be restricted to models with $\mathrm{T/S} \in [0, 3]$.

Using an expansion of $\Delta T/T$ in spherical functions and several well known relations, we can write the contribution of density perturbations in the form

$$\mathrm{S} = \sum_{l=2}^{\infty} S_l W_l, \quad (6)$$

where

$$S_l = \frac{2l+1}{64\pi} A H_0^{3+n} \frac{\Gamma(3-n)\Gamma\left(l + \frac{n-1}{2}\right)}{\Gamma^2\left(2 - \frac{n}{2}\right)\Gamma\left(l + \frac{5-n}{2}\right)}. \quad (7)$$

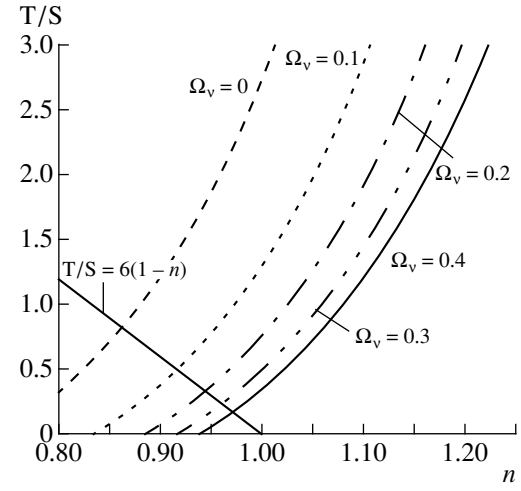


Fig. 4. The required contribution from cosmological gravitational waves to the CMBR anisotropy T/S as a function of n and Ω_ν ($\sigma_8 = 0.52$, $\Omega_b = 0.05$) when $h = 0.5$. The dashed curve corresponds to the model with $\Omega_\nu = 0$, the dotted curve to $\Omega_\nu = 0.1$, the dot–dash curve to $\Omega_\nu = 0.2$, the double-dot–dash curve to $\Omega_\nu = 0.3$, and the solid curve to $\Omega_\nu = 0.4$. The thick line shows the relation $\mathrm{T/S} = (1 - n)$.

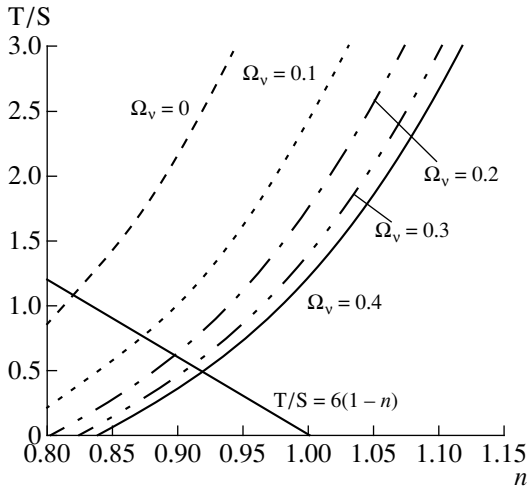


Fig. 5. Same dependences as in Fig. 4 for $h = 0.6$.

Here,

$$W_l = \exp\left[-\left(\frac{2l+1}{27}\right)^2\right] \quad (8)$$

is the instrumental function of DMR COBE. The quantities S_l were calculated in a large-scale approximation ($T(k) = 1$), whose accuracy is somewhat different for different harmonics, but is always $\lesssim 3\%$. The main contribution to S is obviously produced by harmonics with $l \lesssim 10$. The results of calculating T/S for $\Omega_b = 0.05$ and $h = 0.5$ and 0.6 are shown in Figs. 4 and 5, respectively.

We can see in these figures that the CDM model normalized to $\sigma_8 = 0.52$ is inconsistent with the observa-

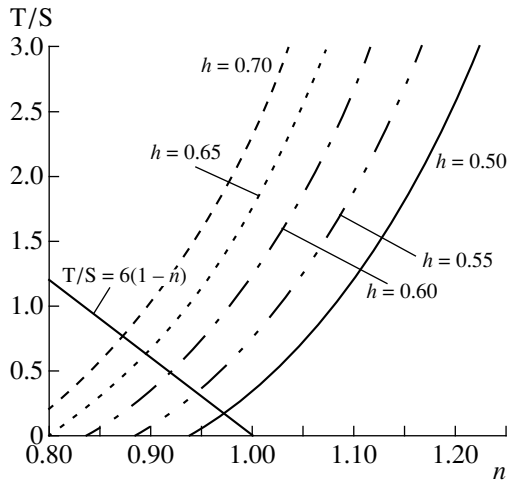


Fig. 6. The required contribution from cosmological gravitational waves to the CMBR anisotropy T/S as a function of n and h ($\sigma_8 = 0.52$, $\Omega_b = 0.05$) when $\Omega_\nu = 0.4$. The dashed curve corresponds to the model with $h = 0.70$, the dotted curve to $h = 0.65$, the dot–dash curve to $h = 0.60$, the double-dot–dash curve to $h = 0.55$, and the solid curve to $h = 0.50$. The thick line shows the relation $T/S = 6(1 - n)$.

tions. Imposing a restriction on the contribution of gravitational waves ($T/S < 0.5$) and requiring that deviations from a flat spectrum be small ($0.92 \le n \le 1.02$), we obtain a lower limit for the density of hot particles $\Omega_\nu > 0.1$ and an upper limit for the Hubble constant $h < 0.6$. If we consider a stronger restriction $T/S \le 0.3$, only models with $h \simeq 0.5$ and $n < 1$ are consistent with the observational data. Therefore, small values of T/S are allowed only in the region of red spectra ($n < 1$) and small h (<0.6). The violation of these conditions leads to large values of T/S ($\gtrsim 1$).

As Ω_ν increases, the constraints on T/S become weaker; however, even when $\Omega_\nu \le 0.4$, cosmological models with $h + n \ge 1.5$ require a considerable contribution from gravitational waves:

$$\frac{T}{S} \gtrsim 0.3. \qquad (9)$$

On the other hand, if the scale-invariant spectrum ($n = 1$) is fixed, and T/S varies over a wide range of values, the observational data are consistent with any value of Ω_ν when $h \lesssim 0.6$.

Models satisfying the relation $T/S = 6(1 - n)$, which is well known for power-law inflation (see, for example, [41–44]), are denoted in Figs. 4 and 5 by the thick straight lines. If the density-perturbation spectrum was formed in the course of power-law inflation, the allowed region of parameter values is the segment of this line delimited by the dashed ($\Omega_\nu = 0$) and solid ($\Omega_\nu = 0.4$) curves. In this case, for various values of Ω_ν and h, the slope of the density-perturbation spectrum can vary from 0.87 to 0.97 (and T/S, respectively, from 0.8 to 0.2) when $h = 0.5$ and from 0.82 to 0.92 (and T/S from 1.1 to 0.5) when $h = 0.6$.

In general, T/S increases linearly with h (with the other parameters fixed), as we can see in Fig. 6. Therefore, the T/S curves with the highest $\Omega_\nu = 0.4$ can be used to find the minimum allowed values of T/S. In particular, we have for the models with $n = 1$ the following estimate of the concentration of primordial gravitational waves:

$$T/S \gtrsim 10(h - 0.47). \qquad (10)$$

For the family of models with $\Omega_b = 0.05$ and $h = 0.5$, we were able to obtain a simple approximation of T/S as a function of n and Ω_ν, which reproduces $\Delta T/T$ to better than 11% in the interval $0.1 \le T/S \le 3$:

$$\frac{T}{S} = \frac{30(n - 0.7)^2}{10\Omega_\nu + 1} + 10\Omega_\nu(n^{3/2} - 1.06). \qquad (11)$$

3.2. The Acoustic Peak in $\Delta T/T$

The position and amplitude of the first acoustic (Doppler) peak of Sakharov oscillations in the CMBR anisotropy represent a sensitive test of many key cosmological parameters. Although the current measurement accuracy does not enable us to draw firm conclusions about allowed cosmological models, the data from numerous measurements of $\Delta T/T$ near the acoustic peak indicate that its characteristics are approximately the same as in the standard CDM model without gravitational waves, normalized to the amplitude of the large-scale CMBR anisotropy [45–47]. Despite the inexact nature of this conclusion, associated with the appreciable systematic and statistical errors of the observations at $1 \sim l \sim 200$, it is reasonable to ask the following question: Can any of the models described above imitate the corresponding behavior of the coefficients S_l of the standard CDM model normalized to the COBE data? We list below the models possessing precisely this property—a rather powerful acoustic peak; however, we delay attempts to estimate the parameters of this peak more exactly until data from future experiments become available.

The number density of galaxy clusters and the large-scale CMBR anisotropy impose a certain relation on possible values of cosmological parameters. After excluding from consideration models that do not satisfy the above two tests, let us turn to the amplitude of the acoustic peak in the allowed models. It is obvious that the position of the acoustic peak (which depends primrily on the curvature of the Universe) will be approximately the same in all the models under consideration (i.e., $l_p \simeq 200$, corresponding to a three-dimensional scale $\sim 100h^{-1}$ Mpc). We computed S_l using the CMBFAST numerical code [48], whose accuracy is sufficient for our estimates.

The computational results obtained with the CMBFAST code can be conveniently presented in the form of the functions $\Re_l \equiv l(l + 1)S_l/(l + 0.5)$ and the corresponding quantities for the primordial gravitational waves. Since the sum of the contributions from both

types of perturbations (scalar and tensor) was already normalized to the COBE signal, and the contribution from the gravitational-wave mode is negligible on the scale of the first Sakharov-oscillation peak, we will call $\Re \equiv \Re_{l=200}/1.1 \times 10^{-10}$ the relative amplitude of the acoustic peak. In the standard model, $\Re = 5.1$. We will use this number as a reference point in our subsequent analysis of the models (it corresponds to a first acoustic peak with amplitude ~70 μK).

Let us now estimate factors affecting the value of $\Re$ in the selected models. (Recall that these models correspond to $\sigma_8 = 0.52$ and $T/S \in [0, 3]$.) An increase in T/S (with other model parameters fixed) obviously leads to a decrease in $\Re$. If $\Omega_\nu = 0$, and n, Ω_b, and h are near their values for the standard model, the relative amplitude of the acoustic peak decreases by a factor of $\simeq T/S + 1 \simeq 4$; i.e., this peak is virtually absent. To increase the peak amplitude to its standard value, we need a very large baryon density $\Omega_b \simeq 0.35$, which contradicts data on primordial nucleosynthesis (see also [49]).

Another way to increase the acoustic peak is to consider a red density-perturbation spectrum and/or increase the fraction of hot dark matter, since both these factors will flatten the dimensionless power spectrum in the range $10–100h^{-1}$ Mpc. On the other hand, when Ω_ν is appreciably increased, blue spectra begin to play an important role, since red spectra become incompatible with the condition $T/S \geq 0$. We can judge the influence of a slope in the density-perturbation spectrum with $n > 1$ only through numerical analysis.

Our computations show that the relative amplitude of the acoustic peak $\Re$ when $\Omega_b = 0.05$ and $h = 0.5$ for models with $n \geq 0.9$ and $\Omega_\nu \geq 0.2$ is slightly smaller than the value of $\Re$ in the standard model (Figs. 7 and 8). Either of these conditions—$n < 0.9$ and/or $\Omega_\nu < 0.2$—decrease the relative amplitude of the acoustic peak by over 30% (i.e., $\Re < 3.5$). In CDM models ($\Omega_\nu = 0$), the acoustic peak is virtually absent.

Such a strong relation between the amplitude of the acoustic peak and the value of Ω_ν can be explained by the fact that all the models considered here were normalized to the observed mass function of galaxy clusters (i.e., to $\sigma_8 = 0.52$; see Section 2). Therefore, the height of the acoustic peak, which is determined by the power spectrum on scales $\sim 100h^{-1}$ Mpc, depends directly on Ω_ν (since it is responsible for the shape of the transition function). When Ω_b is increased to 0.1 (see the right-hand vertical axis in Figs. 7 and 8), the amplitude of the acoustic peak coincides with its standard value ($\Re = 5.1$) to better than 10% in all the models with blue spectra ($n \in [1, 1.2]$) and $\Omega_\nu \geq 0.3$, as well as in models with moderately red spectra ($n \in [0.9, 1]$) and $\Omega_\nu \geq 0.2$.

Thus, the presence of a high acoustic peak in $\Delta T/T$ (at the level of $\Re \gtrsim 4$) can be satisfied in the group of standard power-law spectra ($n \in [0.9, 1.2]$), but requires a considerable concentration of hot dark matter ($\Omega_\nu \in [0.2, 0.4]$) in the class of cosmological models studied here.

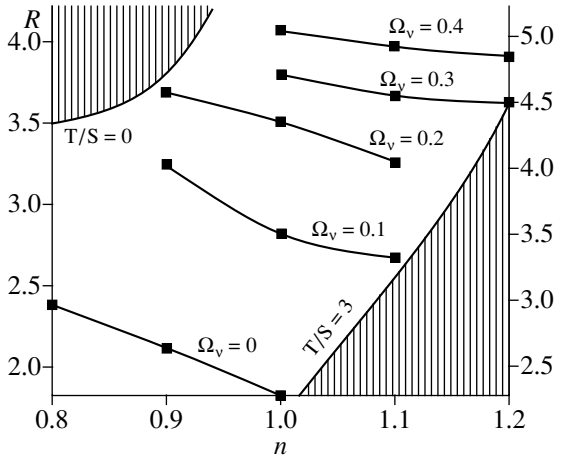


Fig. 7. Dependence of the relative height of the Doppler peak $\Re$ on n for five values of Ω_ν (0.4 to 0, from top to bottom, respectively), $h = 0.5$, and $\sigma_8 = 0.52$. The numbers along the left-hand vertical axis correspond to $\Omega_b = 0.05$, and those along the right-hand vertical axis to $\Omega_b = 0.1$. The unshaded region corresponds to $T/S \in [0, 3]$.

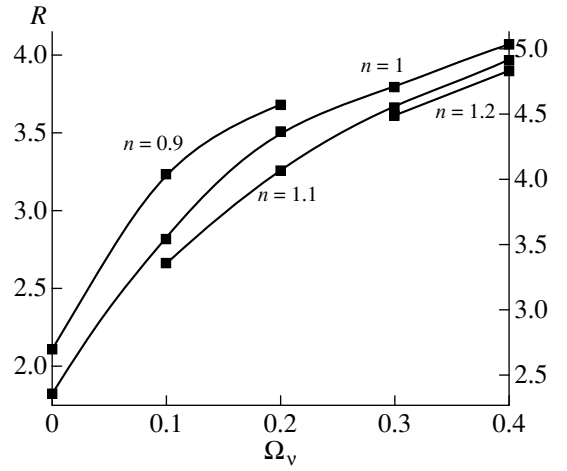


Fig. 8. Dependence of the relative height of the Doppler peak $\Re$ on Ω_ν for four values of n (0.9 to 1.2, from top to bottom and from left to right), $h = 0.5$, and $\sigma_8 = 0.52$. The numbers along the left-hand vertical axis correspond to $\Omega_b = 0.05$, and those along the right-hand vertical axis to $\Omega_b = 0.1$.

4. DISCUSSION

As observational data on the large-scale structure of the Universe are accumulated, the requirements for theoretical cosmology grow. One manifestation of this tendency is the increase in the number of parameters required in models of the Universe.

There are several paths for the development of the theory, depending on the adopted assumptions about the nature of dark matter in the Universe. In particular, cosmological models with stable particles (of both cold and hot dark matter) include both spatially flat models with a Λ term and the parameter T/S and open models ($\kappa = -1$). The recent BOOMERanG and MAXIMA experiments testify that the Universe is flat ($\kappa \simeq 0$). Modern observations of supernovae at large distances demonstrate that models with $\Lambda > 0$ are promising; however, it will not be possible to draw final conclusions for quite some time [50].

To formulate the problem of estimating the parameters of cosmological models most correctly, all these quantities should be initially treated as free parameters. Next, based on observational data, a range of allowed values can be specified for each of the parameters. (These intervals will obviously depend on the amount and quality of observational information available.) Of course, this formulation is idealized and far from realization, primarily because the accuracy and uniqueness of the observational data are insufficient for the solution of multi-parameter problems.

In the present paper, we have studied cosmological models with mixed dark matter and primordial gravitational waves, and shown that there is a family of models satisfying three structural tests: the large-scale CMBR anisotropy (on scales $\sim 1000h^{-1}$ Mpc), the first acoustic peak in $\Delta T/T$ (on scales $\sim 100h^{-1}$ Mpc), and the number density of galaxy clusters (on scales $\sim 10h^{-1}$ Mpc). Our analysis has revealed no substantial constraints on either the slope of the density-perturbation spectrum or the baryon density in the Universe ($n \in [0.9, 1.2]$, $\Omega_b \in [0.3, 0.11]$). Larger values of the Hubble constant require larger values of T/S, but moderate values of h (≤ 0.6) are compatible with fairly small values of T/S (≤ 1). We have derived significant constraints only on the parameter $\Omega_\nu \in [0.2, 0.4]$: generation of the observed acoustic peak requires a considerable contribution to the total density of the Universe from hot dark matter. Thus, inclusion of hot dark matter solves the problem of matter-dominated models with a moderate number of baryons (i.e., consistent with the observations and standard nucleosynthesis theory).

The requirement that there be an appreciable contribution from hot dark matter could be decisive for the class of models considered here if we take into account other observational information (Ly_α clouds; early galaxies; quasars; the power spectra of spatial distributions of galaxies and their groups and clusters; the large-scale velocity field), as well as the results of future experiments on determination of the Hubble constant, the age of the Universe, the deceleration parameter (using type Ia supernovae), and neutrino rest masses. On the other hand, the introduction of a non-zero Λ term in cosmological models could lead to less restrictive constraints on Ω_ν. Therefore, in the next stage of investigations, it seems reasonable to use additional tests already available to "fine tune" more general models (i.e., incorporating such parameters as Λ, κ, and so on).

5. CONCLUSIONS

Our analysis of a five-parameter family of cosmological models (with $\Lambda = \kappa = 0$) has led to the following results.

(1) Using the Press–Schechter method has enabled us to calculate the dispersion of the mass fluctuations on scales $8h^{-1}$ Mpc to high accuracy. After analyzing the observational data [34], we obtain $\sigma_8 = 0.52 \pm 0.01$.

(2) Using only a Press–Schechter formalism and data on the number density of galaxy clusters does not enable us to draw conclusions about other parameters of the cosmological models (such as n, Ω_ν, Ω_b, and h) with sufficiently high statistical significance.

(3) None of the models with mixed dark matter, $n = 1$, and T/S = 0 satisfies the observational data on the number density of galaxy clusters and the amplitude of the large-scale CMBR anisotropy. Therefore, in the framework of the family of cosmological models considered, we must either abandon a flat density-perturbation spectrum or introduce a non-zero amplitude for the primordial gravitational waves (or both simultaneously).

(4) Small values of T/S are allowed only in the region of red spectra ($n < 1$) and small h (<0.6). Violation of these conditions leads to large values of T/S ($\gtrsim 1$).

(5) As Ω_ν increases, constraints on T/S become less restrictive. However, even when $\Omega_\nu \leq 0.4$, the cosmological models with $h + n \geq 1.5$ require a considerable contribution from gravitational waves: T/S $\gtrsim 0.3$.

(6) We estimate for the concentration of primordial gravitational waves in the models with $n = 1$ and $\Omega_\nu \leq 0.4$, $T/S \geq 10(h - 0.47)$.

(7) An analytic approximation for T/S as a function of n and Ω_ν was derived for the models with $\Omega_b = 0.05$ and $h = 0.5$ [see formula (11)].

(8) Due to the requirement that both normalizations ($\sigma_8 = 0.52$ and $(\Delta T/T)^2_{10^\circ} = 1.1 \times 10^{-10}$) be satisfied simultaneously via the introduction of the additional parameter T/S, the relative height of the acoustic peak $\Re$ turns out to be lower than in the standard model. $\Re$ is decreased to < 30% of the standard value ($\Re = 5.1$) in models with a large concentration of hot particles $\Omega_\nu \in [0.2, 0.4]$ for a wide range of power spectra—from moderately red to blue ($n \in [0.9, 1.2]$). Either of the conditions $\Omega_\nu < 0.2$ or $n < 0.9$ decreases the relative height of the acoustic peak by over 30% (i.e., $\Re < 3.5$ in the models with $\Omega_b = 0.05$ and $h = 0.5$). The acoustic peak is virtually absent in CDM models.

(9) The discrepancy in $\Re$ for the models considered here and for the standard model decreases when the baryon density is increased. When Ω_b is increased to

0.1, the amplitude of the acoustic peak coincides with its standard value ($\Re$ = 5.1) to better than 10% in all models with blue spectra ($n \in [1, 1.2]$) and $\Omega_\nu \geq 0.3$, as well as in models with moderately red spectra ($n \in [0.9, 1]$) and $\Omega_\nu \geq 0.2$.

(10) Thus, increasing Ω_ν to values in the interval [0.2, 0.4] is an effective way to solve the problem of the first acoustic peak in the CMBR anisotropy for spatially flat models with a zero Λ term, since the corresponding baryon density remains within the constraints imposed by data on primordial nucleosynthesis.

Therefore, if we introduce a cosmological gravitational-wave background that contributes to the large-scale CMBR anisotropy, models dominated by mixed dark matter are quite reasonable. However, reproduction of the high acoustic peak in $\Delta T/T$ on angular scales $\theta \simeq 1^\circ$ observed in the BOOMERanG, MAXIMA, VSA, MAP, and other experiments requires a considerable fraction of hot dark matter in these models ($\Omega_\nu \gtrsim 0.2$). This makes observational tests on galactic scales and consideration of more general models even more urgent.

ACKNOWLEDGMENTS

This work has been partially supported by the Foundation for the Development and Support of a Radio Astronomy Scientific–Educational Center (project 315) and the State Scientific Technological Program in Astronomy (project "Kosmomikrofizika"). E.V. Mikheeva and V.N. Lukash are also grateful to the Swiss National Scientific Foundation (project SNSF 7IP 050163.96/1) and INTAS (project no. 97-1192) for financial support.

REFERENCES

1. T. P. Walker, G. Steigman, H. Kang, *et al.*, Astrophys. J. **376**, 51 (1991).
2. I. A. Strukov, A. A. Brukhanov, P. P. Skulachev, *et al.*, Mon. Not. R. Astron. Soc. **258**, 37 (1992).
3. G. F. Smoot, C. L. Benett, A. Hogut, *et al.*, Astrophys. J. Lett. **396**, L1 (1992).
4. C. L. Bennet, A. J. Banday, K. M. Górski, *et al.*, Astrophys. J. Lett. **464**, L1 (1996).
5. R. Valdarnini, T. Kahniashvili, and B. Novosyadlyj, Astron. Astrophys. **336**, 11 (1998).
6. J. R. Primack and M. A. Gross, astro-ph/9810204 (1998).
7. S. Borgani, L. Moscardini, M. Plionis, *et al.*, New Astron. **1**, 321 (1997).
8. S. Gottlober and J. P. Mücket, Astron. Astrophys. **272**, 1 (1993).
9. S. Mollerach, S. Matarrese, and F. Lucchin, Phys. Rev. D **50**, 4835 (1994).
10. B. J. Carr and J. H. Gilbert, Phys. Rev. D **50**, 4853 (1994).
11. J. Gilbert, Phys. Rev. D **52**, 5486 (1995).
12. F. Lucchin, S. Colafrancesco, G. de Gasperis, *et al.*, Astrophys. J. **459**, 455 (1996).
13. L. V. Semig and V. Müller, Astron. Astrophys. **308**, 697 (1996).
14. J. Lesgourgues, D. Polarski, and A. A. Starobinsky, astro-ph/9711139 (1997).
15. V. N. Lukash and E. V. Mikheeva, Gravit. Cosmol. **2**, 247 (1996).
16. J. Lesgourgues, D. Polarski, and A. A. Starobinsky, astro-ph/9807019 (1998).
17. A. Melchiori, M. V. Sazhin, V. V. Shulga, *et al.*, Astrophys. J. **518**, 562 (1999).
18. J. R. Primack, *Critical Dialogues in Cosmology* (World Scientific, Singapore, 1996).
19. A. D. Linde, Phys. Lett. B **129**, 177 (1983).
20. F. Lucchin and S. Matarrese, Phys. Rev. D **32**, 1316 (1985).
21. V. N. Lukash and E. V. Mikheeva, astro-ph/9910135 (1999).
22. V. N. Lukash, E. V. Mikheeva, V. Müller, *et al.*, Mon. Not. R. Astron. Soc. **317**, 795 (2000).
23. Y. Fukuda *et al.*, Phys. Rev. B **335**, 237 (1994).
24. C. Athanassopoulos, L. B. Auerbach, D. A. Bauer, *et al.*, Phys. Rev. Lett. **75**, 2650 (1995).
25. C. Athanassopoulos, Phys. Rev. B **81**, 1562 (1998).
26. Y. Fukuda, T. Hayakawa, E. Ychihara, *et al.*, Phys. Rev. Lett. **81**, 1562 (1998).
27. A. Saha, A. Sandage, G. A. Tammann, *et al.*, Astrophys. J. **522**, 802 (1999).
28. S. Jha, P. Garnavich, R. Kirshner, *et al.*, Astrophys. J., Suppl. Ser. **125**, 73 (1999).
29. L. Ferrarese, B. Gibson, D. Kelson, *et al.*, astro-ph/9909134 (1999); G. A. Tammann, A. Sandage, and B. Reindl, astro-ph/9904360 (1999).
30. B. F. Madore, W. L. Freedman, N. Silberman, *et al.*, astro-ph/9812157 (1998).
31. W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).
32. J. Lee and S. F. Shandarin, astro-ph/9709200 (1997).
33. J. E. Gunn and J. R. Gott, Astrophys. J. **176**, 1 (1972).
34. N. A. Bahcall and R. Cen, Astrophys. J. Lett. **407**, L49 (1993).
35. D. J. Eisenstein and W. Hu, Astrophys. J. **511**, 5 (1999).
36. B. Novosyadlyj, R. Durrer, and V. N. Lukash, Astron. Astrophys. **347**, 799 (1999).

37. V. R. Eke, S. Cole, and C. S. Frenk, Mon. Not. R. Astron. Soc. **282**, 263 (1996).
38. S. Borgani, M. Girardi, R. G. Carlberg, *et al.*, astro-ph/9907323 (1999).
39. P. K. Sachs and A. M. Wolfe, Astrophys. J. **147**, 73 (1967).
40. E. M. Lifshits, Zh. Éksp. Teor. Fiz. **16**, 587 (1946).
41. J. E. Lidsey and P. Coles, Mon. Not. R. Astron. Soc. **258**, 57 (1992).
42. R. L. Davis, H. M. Hodges, G. F. Smoot, *et al.*, Phys. Rev. Lett. **69**, 1856 (1992).
43. F. Lucchin, S. Matarrese, and S. Mollerach, Astrophys. J. Lett. **401**, L49 (1992).
44. R. Crittenden, J. R. Bond, and R. L. Davis, Phys. Rev. Lett. **71**, 324 (1993).
45. C. H. Lineweaver and D. Barbosa, astro-ph/9706077 (1997).
46. A. de Oliveira-Costa, M. Devlin, T. Herbig, *et al.*, Astrophys. J. Lett. **509**, L77 (1998).
47. K. Coble, M. Dragovan, J. Kovac, *et al.*, astro-ph/9902195 (1999).
48. U. Seljak and M. Zaldarriaga, Astrophys. J. **469**, 437 (1996).
49. D. J. Eisenstein, W. Hu, J. Silk, *et al.*, Astrophys. J. Lett. **494**, L1 (1998).
50. E. I. Sorokina, S. I. Blinnikov, and O. S. Bartunov, astro-ph/9906494 (1999).

Translated by Yu. Dumin

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 173–179. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 205–212.

Peculiarities of the Radio Emission of Very Rich Abell Clusters

V. I. Zhuravlev and T. S. Fetisova

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

Received June 30, 1999

Abstract—The mutual arrangement of very rich Abell clusters of galaxies influences the radio properties of these clusters, making it possible to explain certain peculiarities of their radio correlation functions. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

According to our current understanding, the large-scale structure of the Universe arose during an early stage in the development of nonlinear perturbations. Quantitative estimates of the parameters of this large-scale structure have relied primarily on statistical analyses of catalogs of galaxies and clusters of galaxies. Characteristic scales for the distribution of matter in the Universe have been identified using widespread correlation-analysis methods. The largest scales have been derived from correlation-function calculations for samples of galactic clusters. Analysis of the two-point correlation function for all clusters in the catalog of Abell [1] reveals a characteristic scale $r_{0,1} \sim 20h^{-1}$ Mpc [2–4], where $h = H_0/100$ km s^{-1} Mpc^{-1} and H_0 is the Hubble constant. Together with this scale, another, larger scale was identified, $r_{0,2} \sim 125h^{-1}$ Mpc, associated with the distribution of the richest clusters. In 1974, Sokolov and Shvartsman [5] proposed to use the richest Abell clusters as tracers of large-scale structure. The scale $r_{0,2}$ was first derived in 1984 using the 6-m telescope of the Special Astrophysical Observatory as part of the "Northern Cone" program [6], based on optical observations of Abell clusters with richness $R_A \geq 2$, listed as compact by Zwicky *et al.* [7]. Kopylov *et al.* [6] considered clusters within a conical surface with opening angle 60° and its axis oriented toward the Northern Galactic pole.

There are several independent pieces of evidence for the existence of inhomogeneities in the distribution of emitting matter on scales of the order of 100–150 h^{-1} Mpc. These include studies of one-dimensional slices to redshifts of $z = 0.3$ [8]; the Great Attractor [9]; clustering of galactic clusters, such as Shapley concentrations [10]; and the tendency for groups and clusters of galaxies to be concentrated in comparatively thin surfaces, filaments, and nodes of the large-scale structure that are bounded by large, relatively empty regions (voids) [11]. Note that, based on upper limits to the temperature fluctuations $\Delta T/T$ on the corresponding angular scales ~1° [12, 13], the inhomogeneity in the distribution of matter on these scales (which is apparently primarily dark; i.e., non-emitting) must be small, with mass fluctuations $\delta M/M < 0.05$. Pariĭskiĭ [14] identifies excess blackbody noise detected at 7.6 cm in the "Kholod" experiment and the "Large Trio" Project (based on the RATAN-600 telescope, VLA, and 6-m SAO telescope) with Sakharov oscillations arising in the multi-component (photon, electron, baryon) plasma long before recombination.

The results of correlation analyses distinguish large-scale structure in the spatial distribution of the richest clusters—the most massive and concentrated formations (in terms of both emitting and dark matter)—while the poorest clusters display a Poisson distribution on these scales. On the other hand, it is known that radio sources tend to be located in regions with enhanced galaxy densities. For example, Shaver and Pierre [15] and Zhuravlev and Fetisova [16] note anisotropy in the distribution of radio sources on the sky in the direction of the supergalactic plane. The data of Ledlow and Owen [17] indicate that clusters with bright optical galaxies (see the classifications of Bautz and Morgan [18] and Rood and Sastry [19]) contain radio sources appreciably more often than do clusters without dominant bright members. These radio sources are usually identified with giant cD galaxies, which possess massive halos and often have multiple nuclei. Recall that, according to the classification of Fanaroff and Riley [20], extragalactic radio sources can be divided into two types, called FRI and FRII, based on their radio power and the morphology of their extended radio structures. Note that any relation between radio luminosity P_r and radio size l_r is likely subject to cosmological evolution. This is probably why a study of 540 radio sources from the 3CR and 4C catalogs [21] did not confirm the existence at large z of two sequences—a "main sequence" with $P_r \sim l_r^{+1.5}$, corresponding to FRI objects, and a "giant sequence" with $P_r \sim l_r^{-4.8}$, corresponding to FRII objects—as proposed earlier by Shkovskiĭ based on data for about three dozen nearby radio galaxies.

1063-7729/01/4503-0173$21.00 © 2001 MAIK "Nauka/Interperiodica"

The evolution of a radio source in the dense medium of a cluster could differ appreciably from the evolution of a source in a low-density medium. It is noted in [23] that the probability of a galaxy becoming a radio source increases if it has a nearby neighbor. Isolated close pairs have a higher probability of containing radio sources than do wide pairs. Recently, the characteristics of radio sources have been investigated based on much richer statistical material from a deep survey by the RATAN-600 radio telescope [24]. According to [25], the P_r values for galaxies in clusters lies in a substantially narrower interval than do those for field galaxies. Ledlow and Owen [17, 26] found that the radio sources associated with 747 Abell clusters were usually located within 0.1° of the cluster centers, with the number of radio sources growing with the richness of the cluster.

If radio sources are associated with rich clusters, we should expect the distribution of radio sources to have a scale equal to the clustering scale characteristic of rich clusters. However, according to analyses of the spatial distribution on the sky of 329 radio sources with flux densities S (1.4 GHz) > 0.5 Jy, redshifts $0.01 < z < 0.1$, and Galactic latitudes $|b| > 15°$ [27], and also of 21490 radio sources with S (4.85 GHz) ≥ 35 mJy, $20° < \delta < 74°$, $0^h < \alpha < 24^h$, and $|b| \geq 10°$ from the 87GB survey [28], the scale $r_{0,2}$ characterizing the richest clusters is not displayed by these samples of radio sources.

In addition, Kooiman *et al.* [28] note certain peculiarities in the behavior of the correlation function as a function of flux on modest angular scales. Figure 1 of [28] presents two-point correlation functions for four samples of radio sources with limiting fluxes S (4.85 GHz) > 300, 60, 45, and 35 mJy. This figure shows that the amplitude of the correlation function becomes negative for angular distances $\theta = 0.^\circ2–0.^\circ4$ for the sample with S (4.85 GHz) ≥ 300 mJy and for $\theta = 0.^\circ8–1.^\circ0$ for all four samples. Does this occur by chance, or is it the result of some factors acting on the correlation between radio sources?

In order to elucidate the possible factors capable of masking intrinsic correlation scales, we considered the radio emission from VLA surveys associated with the richest Abell clusters. We posed the question: how does a quasi-structure with a characteristic scale for inhomogeneity influence the radio emission of the brightest members of galactic clusters? The current paper analyzes the radio emission of very rich Abell clusters for various distances to the nearest massive cluster and various redshifts.

2. OBSERVATIONAL DATA

We investigated a sample of the richest Abell clusters with the most powerful radio emission. We selected clusters with richness $N_A \geq 70$ according to the definition of Abell *et al.* [29]. The radio data for the nearest Abell clusters were taken from the VLA survey [30], which includes clusters with Abell distance classes to $D \leq 3$. We supplemented this sample with more distant objects from the surveys [31, 32]. The observations for these surveys were carried out at 1.4 GHz in the A, B, C, and D configurations of the VLA [33]; the angular resolution achieved varies from 1.″2 in the A configuration to 50″ in the D configuration. The sample is complete to $S \geq 10$ mJy for $z < 0.09$ and $S \geq 200$ mJy for more distant objects ($0.09 < z \sim 0.25$) [17].

The final sample contains 54 rich Abell clusters, listed in the Table. The first six columns present (1) the Abell cluster number, (2)–(3) the right ascension α and declination δ of the cluster centers at epoch B1950, (4) the cluster richness N_A, (5) the redshift z (the redshifts for the three clusters A1234, A1650, and A1763 were measured using the 6-m SAO telescope as part of the "Northern Cone" program [6]), and (6) the distance r in Mpc to the nearest neighbor of the same richness $N_A \geq 70$. For close pairs with small differences in redshift between the ith cluster, which contains a radio source, and the jth neighbor ($|z_i - z_j| \ll 1$), the distance i was determined from the condition

$$r = \min_{j \neq i} |\mathbf{d}_i - \mathbf{d}_j|,$$

where

$$d = \frac{cz}{H_0(1+z)}\left(1 + \frac{z(1-q_0)}{1 + q_0 z + \sqrt{2q_0 z + 1}}\right)$$

is the distance to the cluster in a comoving coordinate system. All data in the table are presented for H_0 = 50 km s^{-1} Mpc^{-1} and $q_0 = 0.5$. Column 7 gives the angular distance θ to the nearest neighbor in degrees, and column 8 gives the logarithm of the 1.4-GHz radio luminosity P_r in W/Hz, where

$$P_r = 4\pi d_i^2 S (1+z)^2,$$

and S is the flux density. Column 9 presents the logarithm of the X-ray luminosity L_X in erg/s. We took the X-ray data from the complete ROSAT sample [34], which includes all 242 X-ray-bright Abell clusters over the entire sky at $|b| \geq 20°$. The sample of X-ray clusters was limited to those with $|b| \geq 20°$ nad fluxes exceeding 5.0×10^{-12} erg cm^{-2} s^{-1} in the 0.1–2.4 keV band.

To estimate the completeness of the sample of radio sources in the table, let us consider the sample of rich Abell clusters. In any catalog, there are selection effects due to distance; i.e., objects begin to be lost starting at some distance z. Rich galactic clusters are not numerous, so that it is very important not to lose any; a guarantee of completeness can be obtained by limiting the region under study. For this reason, we must establish the distance to which selection effects can be avoided as accurately as possible, in order not to lose distant objects from the sample.

We analyzed the completeness of the sample using the procedure applied in the "Northern Cone" program [35], and investigated the dependence depicted in Fig. 1.

Rich Abell Clusters

No. Abell	α(1950)	δ(1950)	N_A	z	r, Mpc	θ	$\log P_r$ [W/Hz]	$\log L_X$ [erg/s]
1	2	3	4	5	6	7	8	9
24	$0^h19\overset{m}{.}9$	23° 1′	127	0.1338	69.2	$5\overset{\circ}{.}26$	25.23	
69	0 35.0	18 4	106	0.1448	41.8	0.93	25.30	
84	0 39.2	21 8	76	0.1030	18.6	1.41	25.51	44.45
98	0 43.8	20 13	185	0.1053	18.6	1.41	25.45	
115	0 53.3	26 3	174	0.1971	154.7	8.61	26.38	45.16
224	1 35.8	−7 12	75	0.1617	133.1	3.23	25.35	
286	1 55.9	−2 1	93	0.0791	22.8	2.87	23.81	
401	2 56.2	13 23	90	0.0748	112.5	9.86	25.18	44.99
423	3 8.9	−12 18	89	0.0796	143.2	16.01	23.73	
426	3 15.3	41 20	88	0.0179	179.9	101.17	25.20	45.10
478	4 10.6	10 21	104	0.0881	102.5	11.82	24.59	45.11
514	4 45.5	−20 31	78	0.0734	183.2	22.72	24.71	44.16
526	4 57.2	5 22	71	0.0541	162.9	4.37	24.76	
562	6 46.5	69 20	70	0.1100	86.1	6.11	25.64	
618	8 4.1	67 42	76	0.1193	93.1	7.26	25.49	
629	8 10.3	66 35	78	0.1380	94.9	1.27	25.37	
754	9 6.4	−9 26	92	0.0534	190.6	25.98	23.82	44.90
1190	11 9.0	41 7	87	0.0794	5.4	0.69	25.41	44.17
1203	11 11.2	40 34	75	0.0795	5.4	0.69	24.65	
1234	11 19.8	21 40	88	0.1663	54.8	3.50	25.84	
1307	11 30.2	14 48	71	0.0834	39.2	4.15	24.72	
1356	11 39.9	10 43	77	0.0698	81.3	4.72	24.44	
1367	11 41.9	20 7	117	0.0214	44.4	19.00	25.04	44.21
1412	11 53.1	73 45	86	0.0836	61.2	5.38	24.30	
1446	11 59.3	58 18	85	0.1035	76.3	7.49	25.57	
1455	12 1.3	28 16	91	0.1390	37.9	2.45	25.28	
1552	12 27.3	12 1	75	0.0843	116.8	14.15	25.82	
1562	12 31.8	41 27	77	0.1910	152.9	8.63	26.07	
1650	12 56.2	−1 29	114	0.0845	20.3	2.44	24.60	44.89
1656	12 57.4	28 15	106	0.0231	44.4	19.00	24.19	44.86
1667	13 0.9	32 5	98	0.1648	26.6	0.70	25.61	
1674	13 1.7	67 46	165	0.1055	28.8	1.68	25.18	
1731	13 20.8	58 26	92	0.1932	133.2	4.63	25.76	
1761	13 30.9	57 54	114	0.2272	149.2	7.09	25.73	
1763	13 33.1	41 13	152	0.2279	95.5	3.35	26.29	45.17
1775	13 39.6	26 37	92	0.0724	56.5	1.61	24.88	44.46
1790	13 44.4	54 17	73	0.1210	61.9	4.65	25.22	
1795	13 46.7	26 50	115	0.0622	56.5	1.61	25.20	45.05
1940	14 33.9	55 22	130	0.1396	6.9	0.36	25.60	
1942	14 36.1	3 53	138	0.2240	236.4	9.95	26.28	
2029	15 8.5	5 57	82	0.0767	95.6	1.75	25.11	45.19
2061	15 19.2	30 50	71	0.0777	35.7	3.63	24.28	44.59
2089	15 30.6	28 11	70	0.0734	17.2	2.22	24.93	
2141	15 55.9	35 36	74	0.1579	73.8	4.82	25.40	
2142	15 56.2	27 22	89	0.0896	96.1	8.79	24.67	45.32
2198	16 26.5	43 56	85	0.0798	136.1	11.86	23.93	
2219	16 38.9	46 47	159	0.2070	287.1	7.17	25.63	45.30
2255	17 12.2	64 9	102	0.0808	112.7	10.95	25.25	44.68
2256	17 6.6	78 47	88	0.0581	133.1	15.07	24.63	44.85
2345	21 24.4	−12 21	107	0.1760	137.6	7.59	25.56	45.00
2365	21 40.2	−18 55	94	0.1873	137.6	7.59	25.74	
2538	23 6.0	−20 9	72	0.0829	53.3	6.28	24.79	
2617	23 30.8	9 12	95	0.1623	98.0	3.89	25.42	
2672	23 52.6	26 10	70	0.2404	329.3	13.62	25.93	

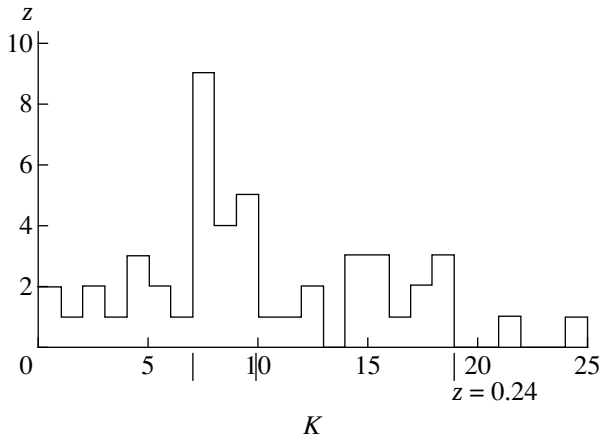


Fig. 1. Number density of clusters as a function of distance. When constructing this relation, the volume of a cone was divided by concentric spherical surfaces into layers with equal volume. The number of clusters falling within each of these layers of volume 26×10^6 Mpc^3 was calculated. The horizontal axis shows the number of equal volumes K.

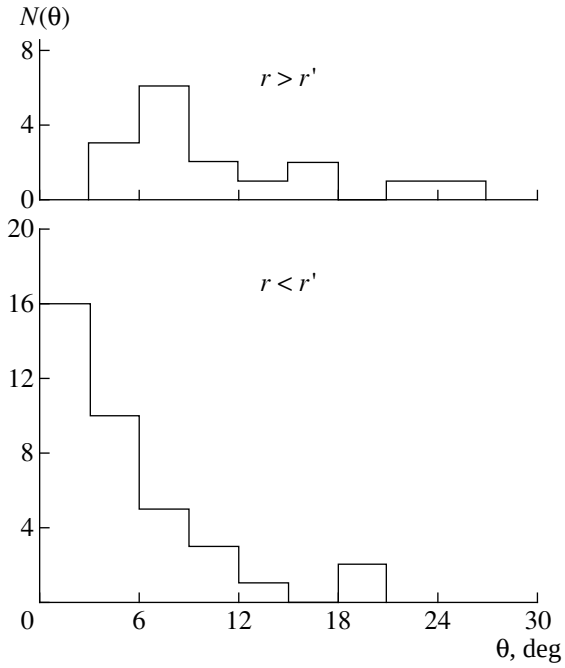


Fig. 2. Distribution of angular distances θ to the nearest neighbor as a function of r.

As a first step, we chose the outer boundary for the sample to be obviously further than the distance at which the luminosity function begins to fall off. We then divided the sample into "nearby" and "distant" parts and calculated their mean fluxes. We used a χ^2 criterion to test for the statistical equivalence of the two subsamples. If the two sample fluxes were found to be significantly different, we moved the boundary slightly closer and repeated the analysis. After several trials, the "distant" sample was found to be free of distance selection effects, and the corresponding boundary was adopted as that for the sample as a whole. By excluding distance selection effects in this way, we ensure that we do not lose any clusters from the sample studied. Our data indicate that the richest Abell clusters ($R_A \geq 2$) are complete to $z = 0.24$ with probability $P = 0.95$.

The resulting sample can thus be considered complete to $z = 0.24$. We adopted a spherical surface with this radius as the outer boundary of the region to be studied. In order to avoid possible distance selection effects influencing the classification of cluster richness and compactness, we also introduced an inner boundary to the study region, $z = 0.015$. We can see that all clusters from the Table with the exception of A2672 ($z = 0.2404$) fall within these two spherical surfaces; since A2672 lies very near the outer boundary, we have also included it in our analysis.

3. DISCUSSION OF RESULTS

Let us consider the radio luminosity P_r on the two scales $r_{0,1}$ and $r_{0,2}$, having initially determined two subsamples with z greater than and less than $z' = 0.09$. The first subsample, with $z > z'$, contains 27 clusters. The flux densities of the radio sources in this subsample are limited to values $S \geq 200$ mJy, or radio luminosities $P_r \geq 10^{24.98}$ W/Hz. The second subsample, with $z < z'$, also contains 27 clusters. Its radio sources have $S \geq 10$ mJy and $P_r \geq 10^{23.55}$ W/Hz.

Further, to elucidate the dependence of P_r on the scales $r_{0,1}$ and $r_{0,2}$, each of the two subsamples was divided into two groups according to the distance r to the nearest neighbor larger and smaller than r'. We chose the boundary between these subsamples to be $r' = 120$ Mpc, between $r_{0,1} = 40$ Mpc and $r_{0,2} = 250$ Mpc. This made it possible to distinguish clusters according to the two clustering scales $r_{0,2}$ and $r_{0,1}$. The first subgroup contained clusters with relatively distant neighbors ($r > r'$) from both z subsamples, and included rich clusters responsible for clustering on the scale $r_{0,2}$. The second subgroup contained clusters with closer neighbors $r < r'$ (with clustering on the scale $r_{0,1}$). The subgroups obtained in this way contain the following numbers of clusters N_{cl}:

z	r	N_{cl}
$>z'$	$>r'$	10
$>z'$	$<r'$	17
$<z'$	$>r'$	7
$<z'$	$<r'$	20

Figure 2 shows the distribution of angular distances to the nearest neighbors for 16 wide pairs (upper), excluding the large angular distance $\theta = 101.^\circ 7$ for the wide pair with $r = 179.9$ Mpc, which includes the nearest cluster, in Perseus A426. We constructed the same dependence for 37 close pairs (lower). The histograms representing these angular distributions on the sky are

very different. The maximum for wide pairs is observed at $\theta = 6°–9°$, while the maximum for close pairs is shifted toward zero, $\theta = 0°–3°$. Fewer than 30% of the close pairs have an angular distance larger than 6°, and separations exceeding 20° are absent. The probability that these clusters are distributed randomly according to a χ^2 criterion is less than 0.05.

To further analyze the luminosities of the radio sources in rich clusters, we compare in Fig. 3 $N(P_r)$ for the four subgroups: the subsamples of distant ($z > z'$) and nearby ($z < z'$) clusters and of wide ($r > r'$) and close ($r < r'$) pairs. We can see that, among the wide pairs ($r > r'$), sources with the highest radio luminosities are concentrated in the distant clusters ($z > z'$), while wide pairs with $z < z'$ have weaker radio sources (Figs. 3a and 3d). The distant ($z > z'$) and nearby ($z < z'$) subsamples for the close pairs ($r < r'$) have less different radio luminosities (Figs. 3b and 3c). Thus, it appears that the influence of the surrounding medium on the evolution of radio sources that has been discussed in the literature (see the Introduction) extends to the characteristic scales for the distribution of rich clusters. We do not exclude the possibility of gravitational action of the rich clusters on the evolution of their associated radio sources. This suggestion is based on the sizes of gravitationally bound systems indicated by model computations. The estimates of [36] indicate that Abell clusters gravitationally interact to distances $r = 50$ Mpc ($\sim r_{0,1}$), while gravitational interactions are virtually absent at larger distances $r \geq 100$ Mpc ($\sim r_{0,2}$). The different tidal perturbations exerted by nearby and distant clusters apparently influence the evolution of a radio source differently. For our sample, the mean radio luminosities $\bar{P}_r$ for clusters with nearby neighbors lie in a narrower interval, $\bar{P}_r(r < r') = 10^{24.79}–10^{25.49}$ W/Hz, than do the mean $\bar{P}_r$ for clusters with distant neighbors, $\bar{P}_r(r > r') = 10^{24.40}–10^{25.84}$ W/Hz. The lower ends of the $\bar{P}_r$ intervals correspond to $z < z'$, and the upper ends to $z > z'$.

Together with the variation of P_r in the evolution of clusters, we expect appreciable variations of the corresponding X-ray luminosities L_X. Figure 4 shows a diagram of the luminosity in the $\log P_r$–$\log L_X$ plane. Below are the number of clusters N_X with X-ray fluxes above a threshold of 5.0×10^{-12} erg cm^{-2} s^{-1} and the percent of detected X-ray clusters in the total number of clusters in each subgroup:

z	r	N_X	%
$>z'$	$>r'$	3	30
$>z'$	$<r'$	2	12
$<z'$	$>r'$	4	57
$<z'$	$<r'$	12	60

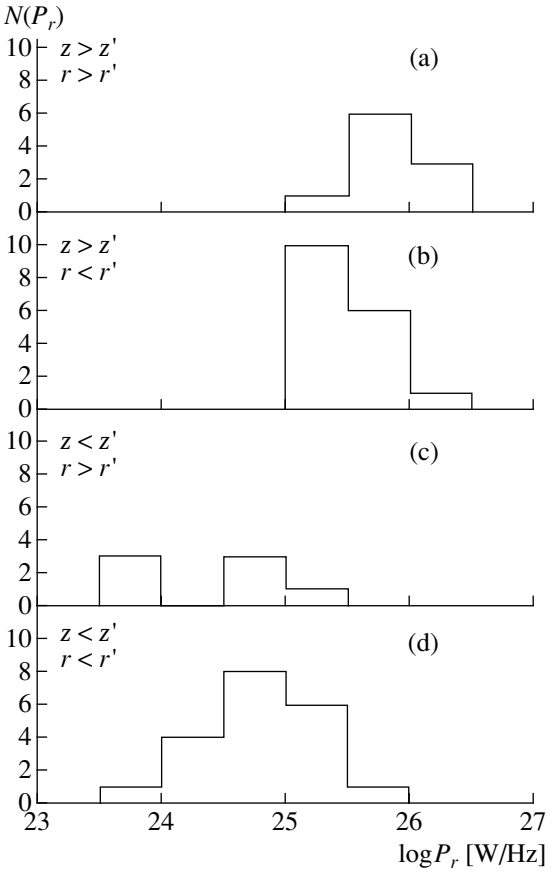


Fig. 3. Comparison of the distributions of radio luminosities P_r for four subgroups of Abell clusters with richness $N_A \geq 70$ as a function of z and r.

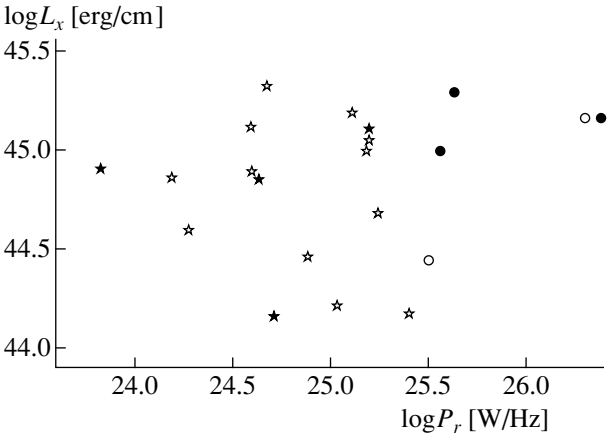


Fig. 4. Luminosity diagram $\log P_r$–$\log L_X$ for Abell clusters: the region $z > z'$, $r > r'$ is shown by fill circles; $z > z'$, $r < r'$ by hollow circles, $z < z'$, $r > r'$ by filled stars, and $z < z'$, $r < r'$ by hollow stars.

We can see from Fig. 4 that there is no strong correlation between P_r and L_X. This is consistent with the results of [37], but we note the following peculiarity. The relative number of rich clusters with X-ray emis-

sion grows in the transition to smaller z, fivefold for close pairs and twofold for wide pairs. This can affect the morphology of the jets in radio sources.

It is interesting that, as a rule, the morphologies of these radio sources classify them as FRI objects, and only strong sources have FRII properties. Such FRI radio sources are usually located in the central regions of rich clusters, where there is much X-ray gas, and where powerful cooling flows directed toward the center form. These cooling flows can act against the development of extended radio lobes by hindering their expansion and possibly even forcing them backward, which should decrease their extent and radio luminosity with time. It would therefore not be surprising if, with time, FRII radio sources could be transformed into FRI sources. This should be accompanied by a decrease in the radio brightness at the source's outer edge and an increase in the radio brightness closer to the center, in the direction of the host galaxy. Taking into account the lower mean radio luminosities of FRI sources, this could explain the different positions of FRI and FRII sources in the $\log P_r - \log l_r$ plane [21, 38].

The difference in the properties of radio sources could also be associated with different rates of release of gravitational energy, or different rotational angular momenta for the central black holes brought about by the accretion of gas onto them [39]. Other differences in the properties of radio sources will come about due to the propagation of relativistic particles into surrounding media with different densities. As pointed out by Gopal-Krishna and Wiita [40], the efficiency of transforming the kinetic energy of the jets emerging from the active galactic nucleus into radio emission is enhanced in a dense medium, due to lower energy losses to adiabatic expansion in the extended radio structures.

We thus can see that the properties of radio sources depend on many factors, but the dependence of P_r on z is probably directly connected to the nearness of neighboring rich clusters. The correlation between P_r and the clustering scale supports this connection.

4. CONCLUSION

The rich galactic clusters whose radio luminosities P_r we have analyzed are exclusively Abell clusters for which redshifts, VLA maps, and X-ray fluxes have been measured. The need to take into account the geometrical characteristics of the Universe, the spatial distribution of the radio sources, and possible evolution of their properties complicated the interpretation of the observational data from the start. The situation is made more difficult by the fact that the radio sources have a range of spectral indices. However, some of the properties of radio sources in relatively nearby Abell clusters may be more general, and be manifest in more distant radio sources as well.

First, the presence of negative correlation-function amplitudes in the two angular-distance intervals $\theta = 0.^\circ2–0.^\circ4$ and $\theta = 0.^\circ8–1.^\circ0$ (see Fig. 1 in [28]) is easy to understand if the dependence of P_r on the clustering scale is preserved on these scales. Therefore, we can expect the appearance of selection effects associated with the restriction of the sample from the side of strong fluxes. The loss of a statistically significant peak in the correlation function on large scales should be manifest at low fluxes due to the wide interval of P_r on these scales. Indeed, as can be seen in Fig. 1 from [28], the negative correlation-function amplitudes on large angular scales $\theta = 0.^\circ8–1.^\circ0$ are observed for low fluxes S (4.85 GHz) = 35 mJy, while negative amplitudes are present only to S (4.85 GHz) = 300 mJy on the smaller angular scales $\theta = 0.^\circ2–0.^\circ4$. For samples with limiting fluxes S (4.85 GHz) $\geq$ 300, 60, 45, and 35 mJy, the amplitudes of the correlation function on scales $\theta = 0.^\circ8–1.^\circ0$ increasingly grow, becoming positive and then exiting onto a plateau at S (1.4 GHz) $\cong$ 5 mJy (from FIRST VLA survey data [41], Fig. 1). This indicates that the amplitude of the correlation function on these scales is not random, and the limiting fluxes at which an excess number of pairs forms on the two angular intervals distinguished reflects variations in P_r as a function of the clustering scale.

Second, modeling the spatial distribution of radio sources with limiting flux S (4.85 GHz) = 35 mJy from the 87GB survey reveals a clustering scale of $25h^{-1}$ Mpc [28], similar to $r_{0,1}$. On the other hand, analysis of the spatial distribution of radio sources with higher limiting flux S (1.4 GHz) = 0.5 Jy indicates a smaller clustering scale bounded by the value $11h^{-1}$ Mpc [27]. Turning again to our hypothesis of a dependence of P_r on clustering scale, we obtain direct evidence for an increase in the fraction of radio sources on the scale $r_{0,1}$ as the limiting flux decreases from S (1.4 GHz) = 0.5 Jy to S (4.85 GHz) = 35 mJy. In deeper surveys, we can expect a decrease in the deficit of radio sources on the larger scale $r_{0,2}$ due to the wider interval of P_r.

The relations we have detected between the evolution of radio sources and the spatial distribution of rich clusters do not enable us to judge unambiguously whether the evolution of radio sources in distant ($z < 5$) clusters is fully analogous to that in more nearby clusters. However, we point out that such an analogy is possible. More accurate estimates require detailed studies of radio luminosity functions and further analysis of radio correlation relations.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project code 98-02-16917).

REFERENCES

1. G. O. Abell, Astrophys. J., Suppl. Ser. **3**, 211 (1958).
2. M. G. Hauser and P. J. E. Peebles, Astrophys. J. **185**, 757 (1973).
3. S. F. Shandarin, A. G. Doroshkevich, and Ya. B. Zel'dovich, Usp. Fiz. Nauk **139**, 83 (1983) [Sov. Phys. Usp. **26**, 46 (1983)].
4. N. A. Bahcall and M. J. West, Astrophys. J. **392**, 419 (1992).
5. D. D. Sokolov and V. F. Shvartsman, Zh. Éksp. Teor. Fiz. **66**, 412 (1974) [Sov. Phys. JETP **39**, 196 (1974)].
6. A. I. Kopylov, D. Yu. Kuznetsov, T. S. Fetisova, and V. F. Shvartsman, Astron. Tsirk., No. 1347, 1 (1984).
7. F. Zwicky, E. Herzog, P. Wild, *et al.*, in *Catalogue of Galaxies and Clusters of Galaxies* (Calif. Inst. Tech., Pasadena, 1961–1968), Vols. 1–6.
8. T. J. Broadhurst, R. S. Ellis, D. C. Koo, and A. S. Szalay, Nature **343**, 726 (1990).
9. A. Dressler and S. M. Faber, Astrophys. J. **354**, 13 (1990).
10. R. B. Tully, R. Scaramella, G. Vettolani, and G. Zamorani, Astrophys. J. **388**, 9 (1992).
11. N. A. Bahcall, M. J. Henriksen, and T. E. Smith, Astrophys. J. Lett. **346**, L45 (1989).
12. Yu. N. Pariĭskiĭ, A. A. Starobinskiĭ, and A. V. Chepurnov, Pis'ma Zh. Éksp. Teor. Fiz. **56**, 561 (1992) [JETP Lett. **56**, 541 (1992)].
13. D. S. Alsop, E. S. Cheng, A. C. Clapp, *et al.*, Astrophys. J. **395**, 317 (1992).
14. Yu. N. Pariĭskiĭ, in *Proceedings of the XXVII Radio Astronomical Conference "Problems in Modern Radio Astronomy," St. Petersburg, 1997*, Vol. 1, p. 73.
15. P. A. Shaver and M. Pierre, Astron. Astrophys. **220**, 35 (1989).
16. V. I. Zhuravlev and T. S. Fetisova, Astron. Astrophys. Trans. **12**, 281 (1997).
17. M. J. Ledlow and F. N. Owen, Astron. J. **109**, 853 (1995).
18. L. P. Bautz and W. W. Morgan, Astrophys. J. Lett. **162**, L149 (1970).
19. H. J. Rood and G. N. Sastry, Publ. Astron. Soc. Pac. **83**, 313 (1971).
20. B. L. Fanaroff and J. M. Riley, Mon. Not. R. Astron. Soc. **167**, 31 (1974).
21. N. S. Kardashev, B. V. Komberg, and V. I. Zhuravlev, Astron. Astrophys. Trans. **18**, 39 (1999).
22. I. S. Shklovskiĭ, Astron. Zh. **39**, 591 (1962) [Sov. Astron. **6**, 465 (1962)].
23. J. T. Stocke, W. G. Tifft, and M. A. Kaftan-Kassim, Astron. J. **83**, 322 (1978).
24. M. G. Larionov, Yu. N. Parijskij, V. I. Zhuravlev, *et al.*, Astron. Astrophys., Suppl. Ser. **106**, 119 (1994).
25. V. I. Zhuravlev, Candidate's Dissertation (Astro Space Center, Lebedev Institute of Physics, Russian Academy of Sciences, 1997).
26. M. J. Ledlow and F. N. Owen, Astron. J. **112**, 9 (1996).
27. J. A. Peacock and D. Nicholson, Mon. Not. R. Astron. Soc. **253**, 307 (1991).
28. B. L. Kooiman, J. O. Burns, and A. A. Klypin, Astrophys. J. **448**, 500 (1995).
29. G. O. Abell, H. G. J. Corwin, and R. P. Olowin, Astrophys. J., Suppl. Ser. **70**, 1 (1989).
30. J. Zhao, J. O. Burns, and F. N. Owen, Astron. J. **98**, 64 (1989).
31. F. N. Owen, R. A. White, and J. O. Burns, Astrophys. J., Suppl. Ser. **80**, 501 (1992).
32. F. N. Owen, R. A. White, and J. P. Ge, Astrophys. J., Suppl. Ser. **87**, 135 (1993).
33. A. R. Thompson, B. G. Clark, C. M. Wade, and P. J. Napier, Astrophys. J., Suppl. Ser. **44**, 151 (1980).
34. H. Ebeling, W. Voges, H. Böhringer, *et al.*, Mon. Not. R. Astron. Soc. **281**, 799 (1996).
35. T. S. Fetisova, D. Yu. Kuznetsov, V. A. Lipovetskiĭ, *et al.*, Pis'ma Astron. Zh. **19**, 508 (1993) [Astron. Lett. **19**, 198 (1993)].
36. P. H. Coleman and W. C. Saslaw, Astrophys. J. **353**, 354 (1990).
37. A. G. Gubanov, Astrofizika **25**, 573 (1986).
38. V. I. Zhuravlev and B. V. Komberg, Astron. Zh. **76**, 163 (1999) [Astron. Rep. **43**, 135 (1999)].
39. S. A. Baum, E. L. Zirbel, and C. P. O'Dea, Astrophys. J. **451**, 88 (1995).
40. Gopal-Krishna and P. J. Wiita, Astrophys. J. **373**, 325 (1991).
41. C. M. Cress, D. J. Helfand, R. H. Becker, *et al.*, Astrophys. J. **473**, 7 (1996).

Translated by D. Gabuzda

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 180–194. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 213–228.

Estimating the Masses of the Spherical and Disk Components of Galaxies via Numerical Simulations

A. V. Khoperskov, A. V. Zasov, and N. V. Tyurina

Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119899 Russia

Received February 28, 2000

Abstract—A series of numerical *N*-body simulations is performed in order to dynamically model the properties of four galaxies (NGC 5603, NGC 3198, NGC 891, and NGC 1566) with known rotation curves, radial disk scales *L*, and velocity dispersions of old disk stars at various galactocentric distances *r*. Each model includes a three-dimensional collisionless disk and rigid spherical components, whose relative mass μ was treated as a free parameter that differed from simulation to simulation. The observed disk stellar velocity dispersions were assumed to be equal to or (in the general case) greater than the corresponding line-of-sight projections of the simulated values for the adopted μ after the initially unstable disk is heated and arrives at a steady state. A comparison of the simulated and observed rotational velocities and velocity dispersions provides evidence for "light" disks with $\mu \geq 2$ in the disk ($r < 4L$). © *2001 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

The analysis of mass distributions in disk galaxies is based on studies of their rotation curves $V(r)$, which are usually derived from measurements of the Doppler gas velocities. Knowledge of the disk rotational velocity enables the rough estimation of the galactic mass within some radius without performing simulations, while the shape of the galactic rotation curve reflects the density distribution of its main components. However, the decomposition of a galactic rotation curve into the contributions due to various components is not unique, necessitating the incorporation of additional information about the radial scales of the individual components, which can be derived from surface photometry of the disk and bulge. This approach could provide full information about the relative masses of the disk and bulge if it weren't for the unseen component (dark halo) of *a priori* unknown mass and poorly known density distribution, which complicates the problem substantially. Interpretations of rotation curves are usually limited either to the so-called maximum-disk solution (i.e., determining the maximum mass of a flat component with known radial density distribution that is consistent with the observed rotation curve) or the best-fit solution (when model parameters are selected to minimize the difference between the computed and observed rotation curves). In the general case, modeling results are not unique, and the shape of the rotation curve can be satisfactorily interpreted in terms of various ratios of the disk and spherical component masses. At the same time, the relative masses of its components are very important parameters determining a galaxy's dynamic and photometric evolution, and their estimation is very important.

Different approaches to estimating the masses of flat and spherical galaxy components can yield contradictory results. According to some studies, the disk masses in luminous galaxies are close to those inferred for maximum-disk solutions, suggesting that the disk dominates inside a large fraction of the optical radius of the galaxy [1–3]. Other studies suggest that the masses of galaxy disks are much smaller than those for maximum-disk solutions, so that the disks are much less massive than the corresponding spherical components inside the optical radius [4–7]. However, given the variety of observed shapes of galactic rotation curves, there is every reason to believe that the relative masses of disks can differ, even among galaxies of similar luminosity. Therefore, the rotation curve of each galaxy must be decomposed individually.

2. USING STELLAR VELOCITY DISPERSIONS TO ESTIMATE DISK SURFACE BRIGHTNESSES

The number of galaxies for which both the rotation curve and velocity dispersion in the old stellar disk are known has increased in recent years. Velocity dispersion analyses can be used to narrow the range of possible galaxy models fitting the observed rotation curve. The principal objective of this paper is to apply an additional constraint when modeling galactic mass distributions, formulated as follows: the observed velocity dispersion of old disk stars at various galactocentric distances cannot be lower than the corresponding computed values that are established in an initially gravitationally unstable model disk after it enters its steady state (the two values should coincide in the case of marginally stable disks).

Attempts to use the velocity dispersions for the old stars of a disk to compute its surface brightness some dis-

1063-7729/01/4503-0180$21.00 © 2001 MAIK "Nauka/Interperiodica"

tance from the galactic center are far from new; combined with rotation-curve data, such analyses enable estimation of the masses of individual components. Such studies have usually assumed that the stellar disks are marginally stable against local gravitational or warping perturbations, or made assumptions about the thickness of the stellar disk at various distances from the center r [7–12]. The criteria for stability against spiral- or bar-forming modes can be applied to this same end [12–14, 25]. Bottema and Gerritsen [5] used numerical simulations varying the disk masses to explain the observed velocity dispersions in the disk NGC 6503. Their approach is very similar to our own treatment of this galaxy, although the computational schemes differ.

The problem with using stellar velocity dispersions is that no well-defined disk stability criteria are known (available analytical criteria have been derived in terms of oversimplified models), and the mechanisms responsible for heating collisionless disks are poorly known, even for numerical models. It nevertheless seems qualitatively clear that sufficiently massive disks must be "heated" to some steady state due to the development of gravitational instability.

Local stability against local radial perturbations in an infinitely thin disk is determined by the criterion of Toomre [15]:

$$Q_T = c_r/c_T \geq 1,$$

where $c_T = 3.36G\sigma/\kappa$, c_r is the radial velocity dispersion, σ the local disk surface brightness, and κ the epicycle frequency, which is a function of $V(r)$ and its radial derivative. It is well known that non-axisymmetric waves in a nonuniformly rotating disk are more unstable, and numerical simulations indicate that a steady-state disk should have $Q_T \approx 2$–3, with this parameter possibly varying with radius (see, e.g., [5, 11, 16–18]). This last circumstance makes it rather difficult to determine the critical parameter in a simple analytical form for any realistic model.

Polyachenko *et al.* [19] derived a stability criterion in the limiting case of extremely non-axisymmetric perturbations in an infinitely thin disk, which corresponds to Toomre parameter $Q_T \approx 3$ and a flat rotation curve $n = -rd\Omega/\Omega dr = 1$ (without allowance for surface-brightness and velocity-dispersion inhomogeneities). The effect of the radial inhomogeneity of the disk density and velocity-dispersion profiles on the gravitational instability of the disk has been analyzed only for weakly non-axisymmetric perturbations [20]. Note that stability criteria have been derived in a local epicyclic approximation $c_r \ll V$, where the disk parameter scales substantially exceed the wavelengths of the most unstable waves. Furthermore, the stability criterion can depend critically on the transition from two- to three-dimensional disk models (see Appendix II of the book of Fridman and Gor'kavyĭ [21]).

In the current paper, we use an approach based not on analytical gravitational-instability criteria, but on numerical simulations of the dynamical evolution of a three-dimensional disk made up of a large number N of gravitationally interacting particles that is initially unstable against gravitational perturbations. Such a disk should evolve into an equilibrium marginally stable state, which we can compare to the dynamical properties of real galaxies. The main simplifying assumptions in this approach are: (1) restricting the analysis to rigid axisymmetric models for the spherical components, (2) assuming an axisymmetric and circular pattern (up to noise fluctuations) for the initial stellar disk rotation, and (3) neglecting the mass of the dynamically cool layer of gas and young stars. The galaxies studied do not include any with pronounced bars, high gas contents, or intensive star formation.

The principal aim of our numerical simulations is to determine the ratio of the masses of the disk and spherical components required to explain the observed rotational velocities and velocity dispersions of the old stars at various galactocentric distances.

The dynamical model constructed yields the spatial distributions of various dynamical parameters of the disk after it enters the steady state. We selected the parameters to fit the model circular-velocity curve to the observed rotation curve of the gaseous disk of the galaxy. At the end of each numerical simulation, the disk is taken to be marginally stable and to have the lowest possible velocity dispersion.

Note that disks can possess excess stability, so that marginal estimates are in effect only limits. For example, when it is sufficiently high to ensure gravitational stability, the stellar velocity dispersion in the disk plane can continue to increase due to a number of processes, such as scattering by massive disk or halo objects, the accretion of companions or star formation in the gas accreted by the disk, and bar disruption or tidal effects in interacting systems. Therefore, verifying whether the observed velocity dispersion of the stars making up most of the disk mass satisfies the marginal gravitational stability criterion is of interest in its own right.[1]

Obviously, the condition $c_{\rm sim} \leqslant c_{\rm obs}$ must be satisfied throughout the disk. Here, $c_{\rm sim}$ and $c_{\rm obs}$ are the simulated and observed velocity dispersions, respectively. We refer to the disk as being "overheated" in regions where $c_{\rm sim} < c_{\rm obs}$, which is possible in the presence of additional factors, ignored by the model, that increase the velocity dispersion.

We considered four galaxies, whose integrated parameters are summarized in Table 1. We adopted velocity-dispersion estimates for a wide range of galactocentric distances r and photometric scales from [11] (see also references therein). The shapes of the rotation curves differ from galaxy to galaxy. The sizes of the rigid-rotation portions in NGC 6503 and NGC 3198 exceed the corresponding photometric scales L, and the rotation curves exhibit a plateau at $r > 2L$ extending to at least $4L$. The rotation curve of NGC 1566 shows a conspicuous

[1] We plan to consider this issue in a future paper.

Table 1. Galaxy parameters and μ estimates

Galaxy	Type	L, kpc	D, Mpc	R_{25}, kpc	μ
NGC 891	Sb	4.6	9.4	18.4	1.8
NGC 1566	Sc	2.25	15.0	18.14	1.7
NGC 3198	Sc	2.6	8.8	10.9	2.1
NGC 6503	Sc	1.16	5.9	6.2	2.0

Note: Galaxy—NGC number; Type—morphological type; L—exponential disk scale according to [11]; D—distance to the galaxy (H_0 = 75 km s^{-1} Mpc^{-1}); R_{25}—isophotal radius according to [38]; μ—optimum ratio of the masses of the spherical subsystem (nucleus + bulge + halo) and disk from this paper.

maximum at $r \simeq L$ and an extended rigid-rotation portion, whereas that of NGC 891 has two maxima and a rigid-rotation portion that does not exceed $0.1L$.

3. NUMERICAL MODEL AND SIMULATION CONDITIONS

The models considered in this paper include the following components.

(1) A stellar disk with a specified initial surface-brightness distribution $\sigma(r)$ and total mass M_d. We adopted the following exponential law as a first approximation:

$$\sigma(r) = \begin{cases} \sigma_0 \exp(-r/L), & \text{when} \quad r \le R \\ 0, & \text{when} \quad r > R, \end{cases} \tag{1}$$

where the radial disk scale is taken to be equal to the photometric scale and $R \simeq 4L$ is the outer disk boundary.

(2) A "rigid" halo with volume density distribution

$$\rho_h = \frac{\rho_{h0}}{1 + (r/a)^2}, \tag{2}$$

core radius a, and total mass M_h within a sphere of constant radius r.

(3) A "rigid" bulge and nucleus (the dense innermost part of the bulge) with density distributions approximated by the formulas

$$\rho_{B,n} = \begin{cases} \dfrac{\rho_{B0,n0}}{(1 + (r/b_{B,n})^2)^{3/2}}, & \text{when} \quad r < r_{B,n} \\ 0, & \text{when} \quad r > r_{B,n}, \end{cases} \tag{3}$$

where B, n refer to the bulge and nucleus, respectively. We determined the parameters of these components for each model by minimizing the residuals between the model circular velocity and the observed rotation curve in the central region of the galaxy. Here and below, we refer to the sum of components (2) and (3) as the spherical subsystem.

In the case of the disk component, we must distinguish between the rotational velocity V and the circular velocity V_c, determined by the equation

$$\frac{V_c^2}{r} = \frac{\partial \Psi}{\partial r}, \tag{4}$$

where Ψ is the total potential produced by all components of the galaxy.

The difference between the stellar-disk rotation curve V and V_c can roughly be estimated from force balance, with allowance for the components of the dispersion of the residual stellar velocities c_r and c_φ:

$$\frac{V^2}{r} - \frac{V_c^2}{r} = \frac{d(c_r^2 \sigma)}{\sigma dr} + \frac{c_r^2 - c_\varphi^2}{r}. \tag{5}$$

In our simulations, we followed the evolution of both the velocity dispersion and the velocities V_c and V.

Our numerical simulations are based on a direct N-body method that provides high spatial resolution (although it is rather computationally intensive), which reduces to solving the equations of motion for a three-dimensional system of N gravitationally interacting particles (the disk). In this approach, all particles interact with each other. The disk is immersed in the external fixed gravitational field of the spherical subsystem (nucleus + bulge + halo), which is determined by the three-dimensional potential $\Psi_S(r, z)$. The procedures we used to construct our equilibrium disks and perform the numerical simulations are standard and described in detail elsewhere [22–24]. Each particle i with mass $m = 1/N$ at a point with coordinates (x_i, y_i, z_i) creates the potential

$$\Psi_i(x, y, z) = -\frac{m}{\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 + r_c^2}}, \tag{6}$$

where the smoothing radius r_c ensures collisionless behavior for a system in which the total force acting on a particle due to all other disk particles exceeds substantially the force of interaction with its closest neighbor. In addition, r_c is small compared to the vertical disk scale height h, as is required for correct simulation of collective processes responsible for the density distribution in the disk.

Our initial data for modeling each galaxy were its observed rotation curve and photometric disk scale. We set the total disk mass to be $M_d = 1$ in all models. The initial radial-velocity dispersion at $t = 0$ corresponded to a dynamically unstable state for the disk ($Q_T \simeq$ 1–1.5), ensuring its rapid dynamical heating. The initial ratio $c_z(r, t = 0)/c_r(r, t = 0)$ was set to a value in the interval 0.3–0.5, and all simulations showed both c_r and c_z to increase with time. The relation between the initial radial

and azimuthal components of the velocity dispersion was specified by the condition $c_\varphi = \frac{\kappa}{2\Omega} c_r$, which is valid in the epicyclic approximation for an axisymmetric, steady-state disk. We determined the rotational velocities of the stellar disk from the equilibrium condition (5). We treated the ratio of the masses of the spherical and disk subsystems $\mu = M_s/M_d$ as a free parameter that was varied when fitting the models for each galaxy.

We specified certain masses and radial scales (in units of the disk mass and radius) for the spherical components, which we took to remain constant in time. In this process, we checked that the total contribution of the above components to the circular velocity was in agreement with the observed rotation curve of the galaxy. We selected the component parameters for a specified μ so as to minimize the deviation of the simulated curve $V_c(r)$ from the observed curve $V(r)$ in the region $r < 4L$.

At time $t = 0$, all the disk mass is contained within $r \le R = 4L$. In the process of evolution, a small fraction of the mass can leave this region, and the simulations continued to include the motions of these particles and their contribution to the gravitational field. However, the disk parameters (density, velocity, dispersions, etc.) were computed only for $r < 4L$.

Figure 1 illustrates the evolution of the macroscopic parameters $r \lesssim R$ (velocity dispersion, surface brightness, rotational velocity, and c_z/c_r ratio) for about ten revolutions of the outer disk boundary using the model for NGC 6503 as an example. In the adopted units, the rotational period of the outer disk boundary is equal to π, and therefore the total time covered by the simulations is $t \approx 32$. The dependences shown in the figure demonstrate that the disk parameters remain virtually unchanged after three to five revolutions. Since the initial state is unstable, the system is heated, which manifests as an increase of the velocity dispersion. In the course of evolution of the system, the parameter $\beta = \frac{\kappa}{2\Omega}\frac{c_r}{c_\varphi}$ was equal to unity with very good accuracy (to several per cent) in all the simulations.

Figure 2 shows the particle distributions in the plane of the disk and in the perpendicular direction. The disk develops spiral waves, which persist for several revolutions; their amplitudes are substantial, and they are responsible for collective heating processes. Sellwood and Carlberg [16] were the first to point out the efficiency of stellar-disk heating by short-lived spirals. The amplitudes of these spirals decrease rapidly with time, and virtually disappear after three to four rotation periods. As expected, the stellar velocity dispersion ceases to increase at the same time. Under the initial conditions used, the spiral waves begin to develop at the center, and this is the disk region that is heated first (see Figs. 2a and 2b and also [16]). Only by the end of the third rotation period (at the outer boundary) do the perturbations begin to affect the peripheral regions. The waves at the disk periphery are more persistent, increasing the time for the heating processes by factors of two to three over that for the center of the system.

The radial profiles of various parameters shown here for a disk that has been heated to a steady state are the result of averaging over the last two to three revolutions, when the azimuthally averaged parameter variations resemble noise rather than a systematic pattern. Note, however, that, as shown below, models with "loose" halos can show a slower approach to the steady state, requiring that the computations be continued for several dozen revolutions.

To make sure that the simulations do not lead to unrealistically strong disk heating due to the limited number of massive particles representing the disk, we compared several models with identical dynamical parameters that differed only in the number of point masses used, which we varied from $N = 10000$ to $N = 80000$. The models with $N \gtrsim 18000$–20000 yielded virtually identical results for both the disk-heating rates and the final parameter values. This leads us to conclude that our simulation results can be used to model stellar disks in flat galaxies. To illustrate this point, we show in Fig. 3 the temporal behavior of c_r as a function of the number of particles for various galactocentric distances, with the other model parameters fixed. Note that the volume densities of the interacting particles at $r = 0.5L$ and $r = 3L$ differ by more than an order of magnitude, whereas the final c_r value for each disk zone depends only slightly on N provided that the number of particles is large enough.

Our testing of the dynamical model indicates that a system of $N \gtrsim 15000$ particles is collisionless when the ratio of the potential smoothing radius to the exponential disk scale is $r_c/L = 0.004/0.25$. This enables an adequate description of collective processes in both the plane of the disk and in the vertical direction.

Due to the instability of the initial state, some simulations involved mass redistribution, with the density at the disk center increasing and that at $r \simeq L$ somewhat decreasing. This led to a surface-density distribution that differed appreciably from an exponential law (see Fig. 1b). We used an iterative approach to ensure that the final radial disk scales corresponded to the observed scales: we increased the initial Q_T in each subsequent simulation by half the total amount by which it changed, so that the system started evolving from a state that was closer to a steady state. In this way, we were able to bring the final density distribution close to the observed (exponential) disk brightness distribution after several iterations.

We varied the final velocity dispersion at the end of the simulation via selection of the parameters of the spherical subsystem, first and foremost the mass of the dark halo. The changes in the nucleus and bulge parameters had a weaker effect on the results when the size of the region where these components dominate the disk gravitationally does not exceed L.

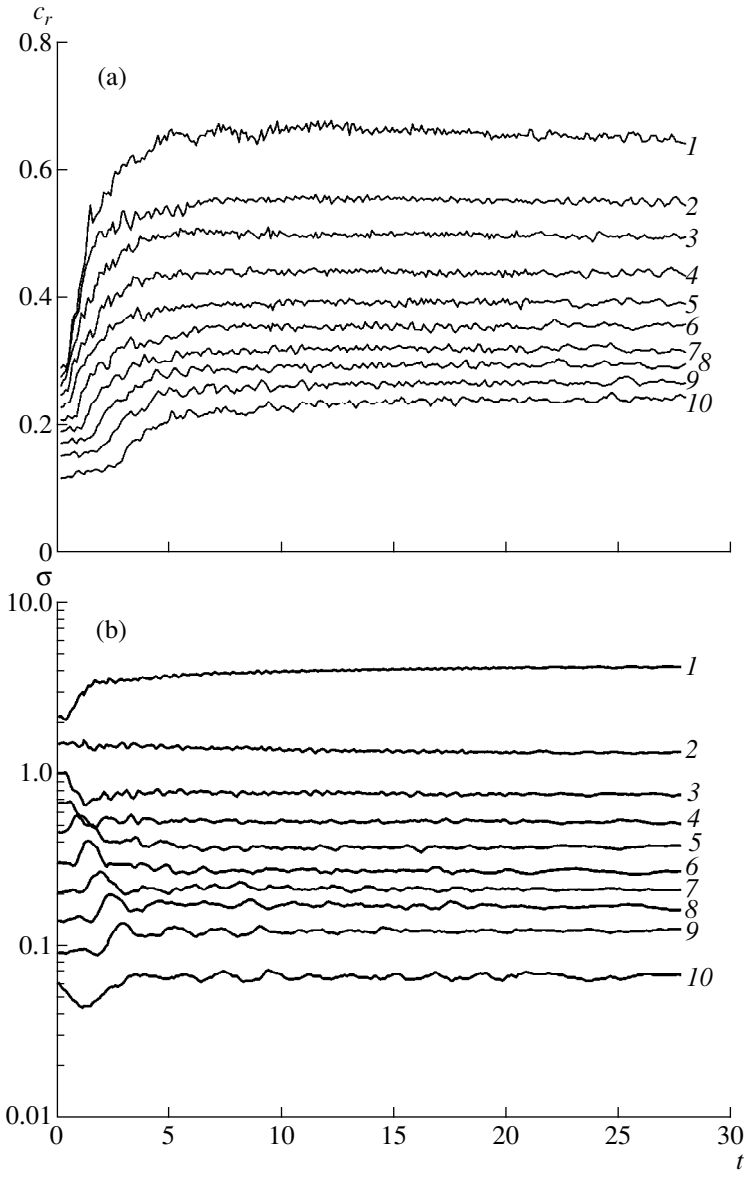


Fig. 1. Dynamical evolution of the stellar disk illustrated by the case of NGC 6503 (model SG2): (a) radial-velocity dispersion $c_r(t)$ at ten different radii ($r_i = 0.1i$, $i = 1, \ldots, 10$), (b) disk surface brightness $\sigma(t)$, (c) disk rotational velocity $V(t)$, and (d) $c_z/c_r(t)$. All quantities are expressed in the units adopted in this paper: $G = 1$, $R = 1$, and $M = 1$. In these units, the period of revolution at the outer disk boundary is $2\pi R/V \sim 3$.

We followed the ratio of the vertical and radial velocity dispersions c_z/c_r in the course of the simulations (Fig. 1d). All the simulations showed an initial increase of c_z, probably due to the instability of small-scale warping modes [36]. An increase of the vertical velocity dispersion increases the disk thickness (see Figs. 2b and 1d). The vertical velocity dispersion $c_z(t)$ increases more slowly than the velocity dispersion in the plane of the disk, explaining the initial decrease of the c_z/c_r ratio (Fig. 1d). That is why, in some cases, the final ratio c_z/c_r was smaller than its initial value at $t = 0$, although both parameters increased during their evolution.

We now turn to a description of the models for individual galaxies. We adopted all observed velocity dispersions and radial scales from [11].

NGC 6503

We now describe in more detail the technique for matching the observed and simulated radial profiles of the rotational velocity and velocity dispersion. This galaxy has a small exponential disk scale ($L = 1.16$ kpc), and the estimates of the velocity dispersions correspond to $r \le 2L$. We took the rotation curve $V(r)$ from [26]. The observed

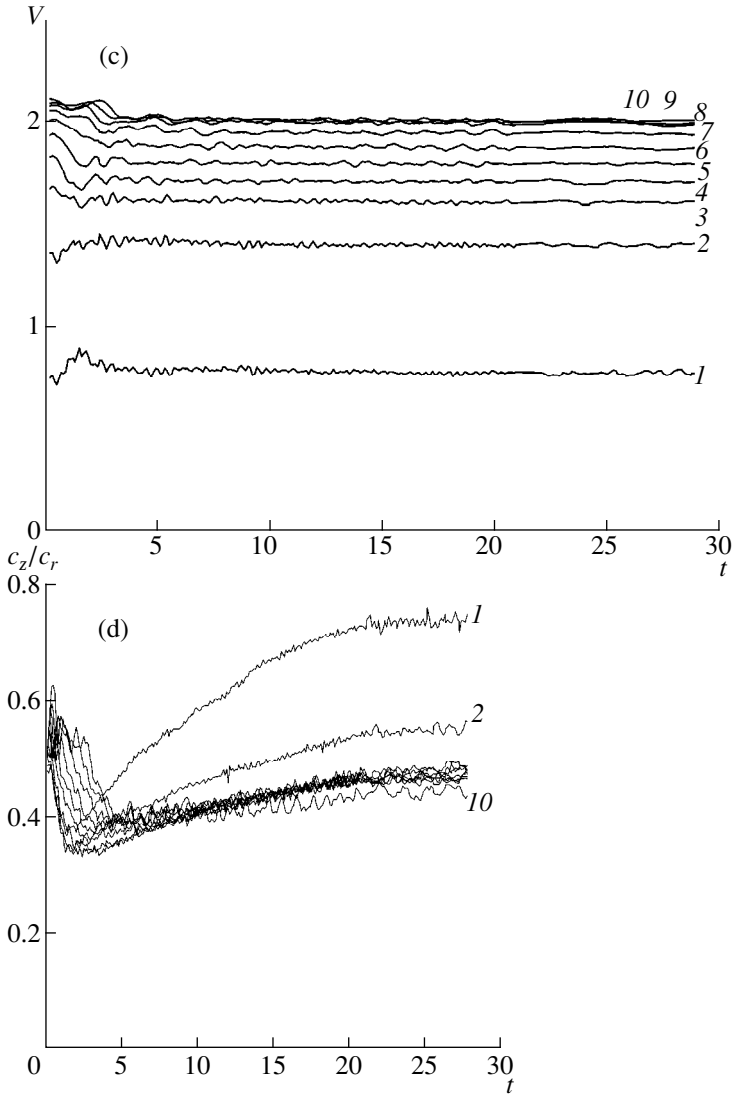


Fig. 1. (Contd.)

$V(r)$ can be interpreted in terms of models with different ratios of the spherical and disk subsystem masses $\mu = M_s/M_d = 0.8$–4.4 (see Table 2 and Fig. 4). We estimated the disk mass within $4L$ of the center (which corresponds to 4.6 kpc in the case of NGC 6503). All models for this galaxy (Table 2) can explain the observed rotation curve satisfactorily without the contribution of the nucleus.

The final radial dependence $c_r(r)$ is determined by the parameters of the spherical subsystem and the exponential disk scale, and is independent of the initial disk velocity dispersion, provided this dispersion corresponds to an unstable state ($Q_T = 1$–1.5). If we wish to compare our results with the observed velocity dispersions, we must take into account the inclination i of the galaxy relative to the plane of the sky.

Taking into account projection effects and the line-of-sight velocity dispersion, we obtain

$$c_{\rm sim} = \sqrt{c_z^2\cos^2(i) + c_\varphi^2\sin^2(i)\cos^2(\alpha) + c_r^2\sin^2(i)\sin^2(\alpha)}, \quad (7)$$

where α is the angle between the slit direction and the major axis projected on the plane of the galaxy. Following [5], we adopt 6503 $i = 74°$ for NGC 6503. We aligned the spectrograph slit along the major axis of the galaxy ($\alpha = 0$) and took the ratios of the c_z, c_φ, and c_r velocity-dispersion components from the simulations.

The system was unstable for models with low-mass spherical subsystems (SF0 and SA2) with $\mu \le 1.0$ and initial radial velocity-dispersion profiles $c_r(r, t = 0)$ that are close to those observed, and there was strong disk heating. Because of this, the final velocity dispersions

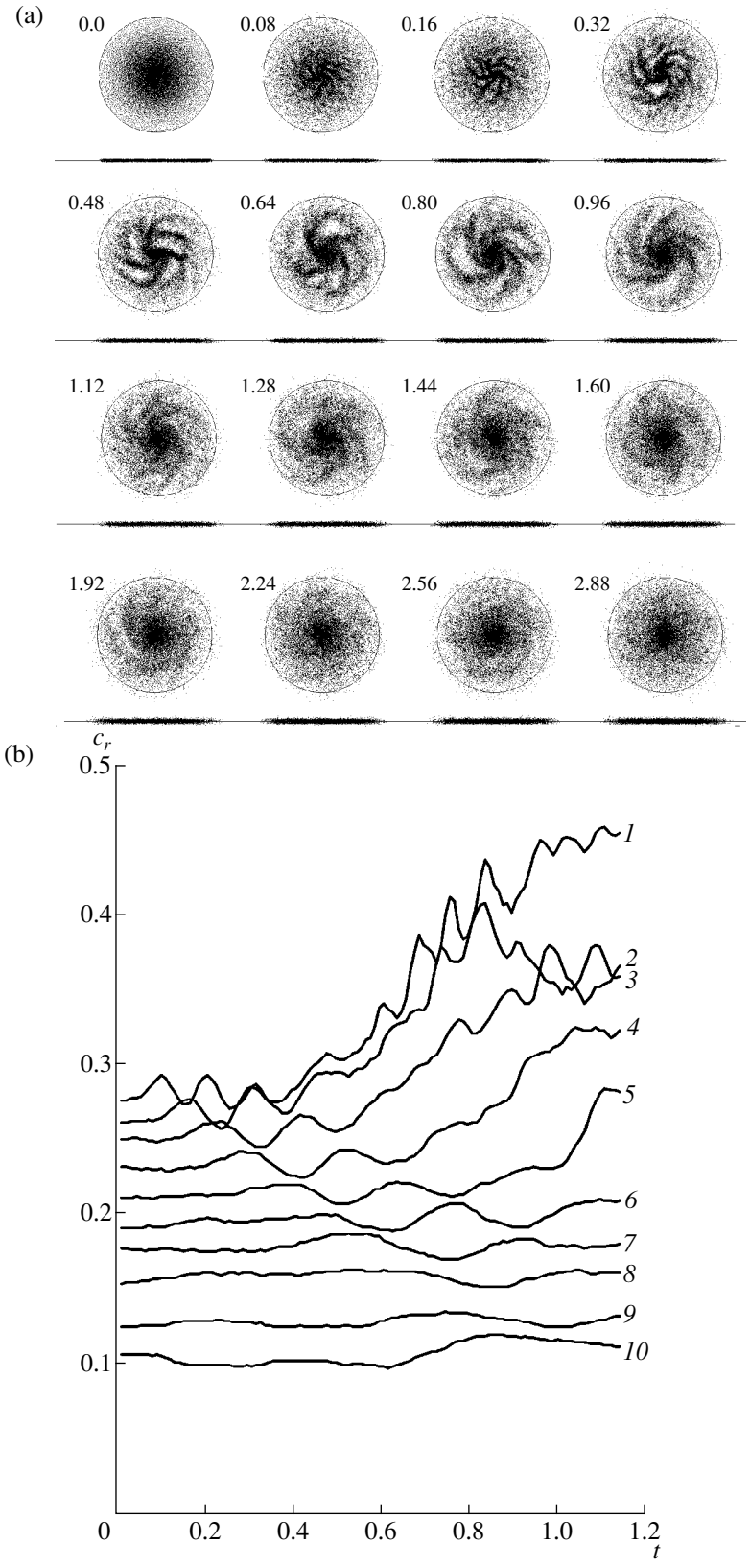


Fig. 2. (a) Evolution of the computed spatial particle distribution for the same model as in Fig. 1 (SG2). Half of all the particles included in the computations are shown. The numbers indicate time in units of the period of revolution at the outer disk boundary. (b) Temporal dependence of the velocity dispersion c_r at the initial stage of instability development in model NN9 with $N = 80000$. The velocity dispersions (in the adopted units) at various r are shown: from the central region (*1*) to the disk periphery (*10*).

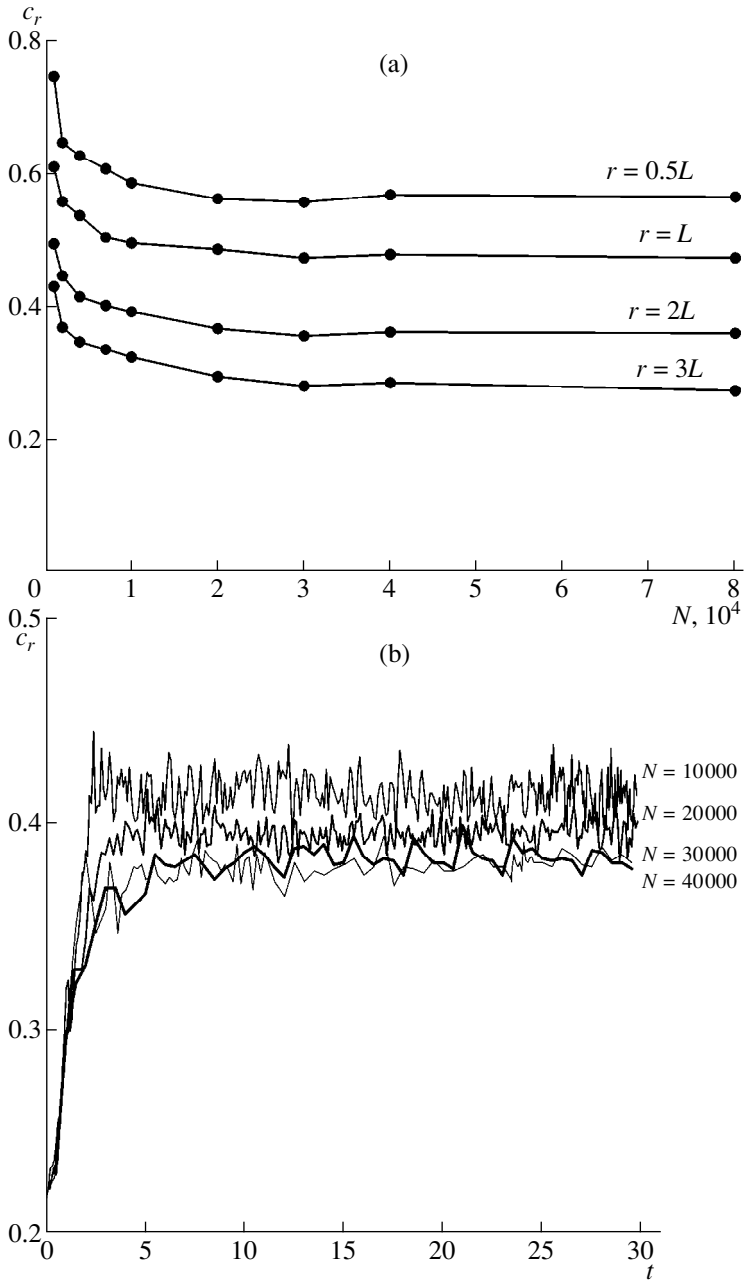


Fig. 3. (a) Radial velocity dispersion c_r (in the adopted units) for various r at the end of the simulation as a function of the number of model particles N. (b) Evolution of the radial velocity dispersion in NGC 6503 (at $r = 2L$) for models with differing numbers of test particles (10, 20, 30, and 40 thousand particles).

exceeded substantially the observed values (Fig. 5a). Model SF0 corresponds to the maximum-disk solution. In the two models considered, the spherical and disk subsystems have similar masses, and the inner part of the disk develops a bar (Fig. 6). This is apparently due to the large core radius of the halo (the so-called "loose halo" case), which implies that the disk mass dominates over the spherical component when $r \lesssim 2L$. As a result, a bar mode has already developed by the end of the second galactic revolution. The final c_r value at the disk periphery at the end of the simulation is about twice the observational estimates (Fig. 5a). The discrepancy is less pronounced in the central region; however, $c_{\rm sim} > c_{\rm obs}$ throughout the entire disk.

In the maximum-disk model (SF0), the observed circular velocity curve can be best reproduced when $M_b = 1.9 \times 10^7 M_\odot$, $M_d = 7.1 \times 10^9 M_\odot$, and $M_h = 5.6 \times 10^9 M_\odot$, which correspond to $\mu = 0.8$. These results indicate that a dynamical model with these parameters

Table 2. Parameters of numerical models

Model	Galaxy	N	μ	M_h	a_h	M_b	a_b	$R_{\max_b}$	M_n	a_n	$R_{\max_n}$
SF6	6503	18000	1.62	1.60	0.25	0.02	0.05	0.17	–	–	–
SF0	6503	18000	0.80	0.80	0.70	0.003	0.02	0.06	–	–	–
SA2	6503	18000	1.03	1.01	0.49	0.02	0.11	0.22	–	–	–
SF8	6503	18000	2.38	2.33	0.25	0.05	0.05	0.17	–	–	–
SG1	6503	40000	3.18	3.00	0.26	0.18	0.11	0.22	–	–	–
SG2	6503	30000	3.18	3.00	0.26	0.18	0.11	0.22	–	–	–
NN9	6503	80000	3.18	3.00	0.26	0.18	0.11	0.22	–	–	–
SF9	6503	18000	4.40	4.18	0.25	0.22	0.11	0.25	–	–	–
JF0	3198	18000	0.78	0.77	0.60	0.01	0.03	0.10	–	–	–
JF3	3198	18000	1.78	1.74	0.28	0.04	0.11	0.25	–	–	–
JF5	3198	18000	3.15	3.01	0.28	0.14	0.11	0.25	–	–	–
JF7	3198	18000	2.15	2.11	0.27	0.04	0.09	0.25	–	–	–
PP1	1566	18000	2.00	0.64	0.74	1.23	0.09	0.31	0.13	0.10	0.28
PP2	1566	18000	2.72	1.08	0.26	1.51	0.09	0.31	0.13	0.10	0.28
PP6	1566	25000	1.33	0.41	0.74	0.57	0.13	0.22	0.35	0.19	0.28
PP7	1566	25000	1.73	0.46	0.74	0.88	0.28	0.13	0.39	0.19	0.28
PF0	1566	18000	1.03	0.59	1.02	0.44	0.10	0.18	–	–	–
PF1	1566	18000	0.84	0.17	1.22	0.67	0.11	0.25	–	–	–
WF4	891	18000	1.77	1.44	0.25	0.23	0.03	0.08	0.10	0.11	0.03
WF0	891	18000	0.16	0.02	0.52	0.14	0.01	0.05	–	–	–
WF5	891	18000	3.21	2.67	0.25	0.19	0.04	0.01	0.35	0.01	0.01
WH9	891	18000	1.77	1.38	0.25	0.28	0.05	0.22	0.11	0.01	0.09
WX6	891	18000	4.66	3.98	0.25	0.24	0.04	0.01	0.44	0.01	0.09

Note: Model—simulation number; Galaxy—NGC number; N—number of particles in the simulation; μ—spherical system-to-disk mass ratio; M_h, M_b, and M_n are the halo, bulge, and nucleus masses; a_h, a_b, a_n, $R_{\max_n}$, and $R_{\max_b}$—parameters describing the density distribution in the halo (h), bulge (b), and nucleus (n) (see text).

should develop a strongly heated disk ($c_r/V \simeq 0.3$), in contradiction with the observations (Fig. 5a). Thus, although the computed rotation curve is in good agreement with the observations, a realistic model requires a more massive spherical subsystem (Fig. 5a).

The simulated circular velocity curve $V_c(r)$ can also be brought into satisfactory agreement with the observed rotation curve $V(r)$ by adopting an even higher mass for the spherical component. However, dynamical models with very massive spherical subsystems ($\mu \geq 4$) yield velocity dispersions for which $c_{\rm sim} < c_{\rm obs}$ throughout the galaxy (Fig. 5a, model SF9). Such a model can be consistent with the observations of NGC 6503 only if the entire stellar disk of the galaxy is strongly overheated and has a large margin of gravitational stability. In this case, in the ring $0.2 < r/L < 0.5$, the stability margin in terms of c_r reaches a factor two or higher. If, on the other hand, the stellar disk is close to the gravitational stability limit, models with $\mu \gtrsim 3.5$ must clearly be rejected. The results of the dynamical simulations are consistent with the observational data if $\mu = 1.6$–3.5. Of all the models considered, SF6 and SF8 are most consistent with observations. They imply a mass ratio for the spherical and disk subsystems of at least 1.6 for NGC 6503. The bulge's contribution to the mass of the spherical subsystem remains negligible.

One distinctive feature of the observed radial velocity-dispersion profile of NGC 6503 is the presence of a maximum at $r \simeq 0.3L$: at the very center of the galaxy, c_r decreases sharply, to as low as 25 km/s. One way to quantitatively fit the observational data in this region is to suppose the presence of a concentrated, kinematically distinct nucleus with a low velocity dispersion, as proposed by Bottema and Gerritsen [5]. However, we cannot rule out the possibility that the decrease in the velocity dispersion noted above could be due to a rapidly rotating disk in the region dominated by the mass of the spherical component.

The conclusion that the mass of the disk in NGC 6503 is small is consistent with the modeling results of Bottema and Gerristen [5] for this galaxy. However, their model velocity dispersions, which are consistent

with the observations for $r < 2.3$ kpc, decrease rapidly at large distances from the center, becoming smaller than 10 km/s at $r > 3$ kpc, which seems unlikely for an old stellar population. We do not find any such low velocities in our dynamical models: $c_r > 14$ km/s throughout the disk.

The ratio c_z/c_r in this galaxy depends only weakly on the radial coordinate, although it increases slightly towards the disk center. The resulting ratios (c_z/c_r = 0.5–0.75) are consistent with data for the Milky Way Galaxy [27] and other galaxies [28].

NGC 3198

The shape of the rotation curve of this galaxy is similar to that of NGC 6503, although the rotational velocity in the flat part of NGC 3198 is about 40 km/s higher than in NGC 6503 [26].

The model corresponding to the maximum-disk solution (JF0) is not able to simultaneously describe the radial profiles of both $c_{\rm obs}(r)$ and $V_c(r)$: the model velocity dispersions are appreciably higher than the observed values (Fig. 5b).

The dynamical model can be brought in agreement with the observational data using the same μ values as for NGC 6503. A mass ratio of $\mu \simeq 3$ (JF5) would imply that the disk possesses some margin of stability. We obtain a steady-state system whose velocity dispersion is most consistent with the observations when $\mu = 2.1$. This model, which we consider to be optimum, implies a stellar disk mass of about $1.75 \times 10^{10} M_\odot$, a factor of four higher that in NGC 6503.

NGC 891

This galaxy is viewed at an angle of $i = 89°$ [30], and its rotation curve differs from those considered above. The inner maximum of $V(r)$ is close to the center, implying a very small rigidly rotating portion ($r < 0.08L$). NGC 891 has the highest rotational velocity and largest photometric scale among the four galaxies considered.

Bottema [11] reports velocity dispersions for NGC 891 only within $r < 1.2L$, where $L = 4.6$ kpc. Through the appropriate fitting of parameters, the observed rotation curve can be explained over a wide interval of μ. In accordance with (9), we adopt $c_{\rm sim} = c_\varphi$ for this galaxy. However, since the galaxy is viewed nearly edge-on, formula (9) can lead to appreciable errors if it is used to compare the model and observed velocity dispersions, since the line of sight crosses regions located at different r. We therefore averaged the model velocity dispersion along the line of sight prior to comparing it with the observational estimates:

$$\hat{c}_{\rm sim} = \int \varrho\, c_{\rm sim}\, dl \Big/ \int \varrho\, dl, \qquad (8)$$

where ϱ is the volume density of the simulation particles and dl is a distance element along the line of sight.

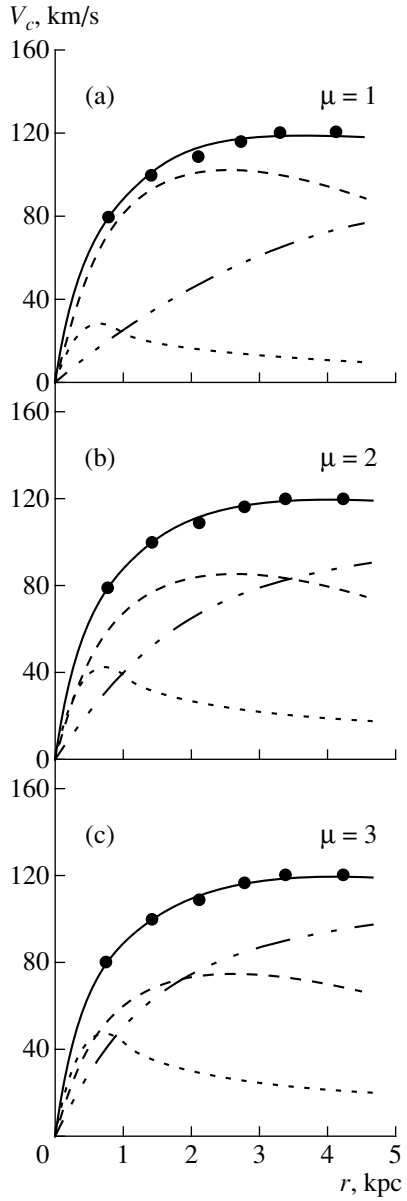


Fig. 4. Radial profiles of the model rotation velocity (solid curve) and observed rotation curve (dots) for the case of NGC 6503 for various μ. Also shown are the contributions of individual subsystems to the model rotation curve: bulge (dotted line), disk (dashed line), and halo (dashed and dotted line).

In the maximum-disk model (WF0 with $\mu = 0.16$), the system is strongly heated, so that the observed velocity dispersions are below the threshold required for gravitational stability throughout the galaxy (Fig. 7a), enabling us to reject models with low μ.

μ must be increased to ensure gravitational stability of the disk of the galaxy. Models with $1 \lesssim \mu \lesssim 3$ yield velocity dispersions that are consistent with the observational data. The model WF4 with $\mu = 1.77$ is in best agreement with the observations (Fig. 7a). The disk mass in this model is $M_d \simeq 7 \times 10^{10} M_\odot$, whereas the nucleus

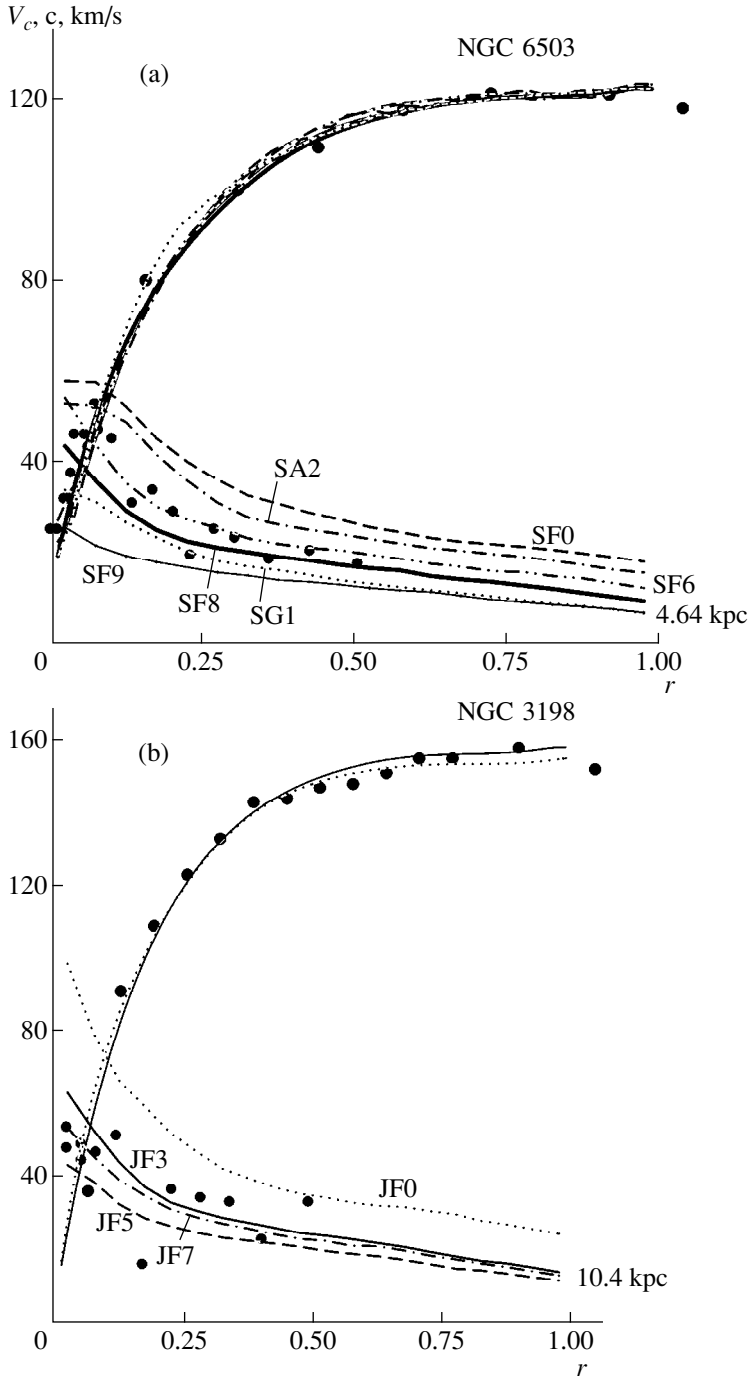


Fig. 5. Radial profiles of circular rotation velocity and line-of-sight velocity dispersion for various models for (a) NGC 6503 and (b) NGC 3198. The circles show the observational data [11].

and the bulge account for about one-fifth of the mass of the spherical subsystem.

NGC 1566

The rotation curve of this galaxy differs from those considered above: like NGC 891, its $V(r)$ curve shows a conspicuous maximum (Fig. 7a), however, its rigid-rotation portion extends to 2 kpc, which is close to the photometric scale $L = 2.25$ kpc. In contrast to the galaxies considered above, NGC 1566 is viewed almost face-on, at $i = 30°$. The slit used to measure the velocity dispersions was not aligned along the major axis of the galaxy image ($\alpha = 22°$) [32], so that the observed velocity dispersion includes contributions from all three components c_r, c_φ, and c_z.

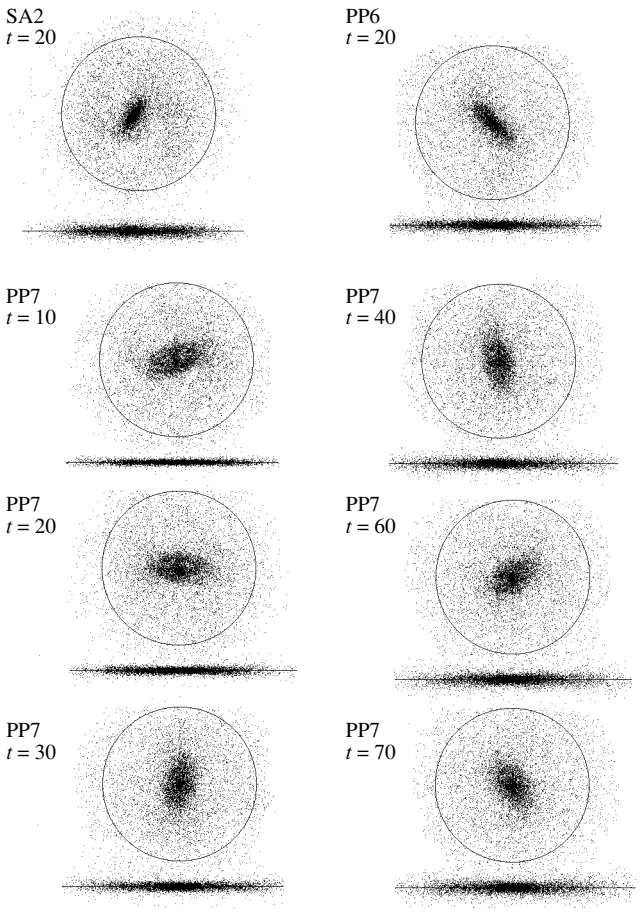


Fig. 6. Particle distributions in the plane of the disk for various dynamical models: SA2 (NGC 6503), PP6 (NGC 1566), and PP7 (NGC 1566). The numbers indicate time in units of the period of revolution at the outer disk boundary.

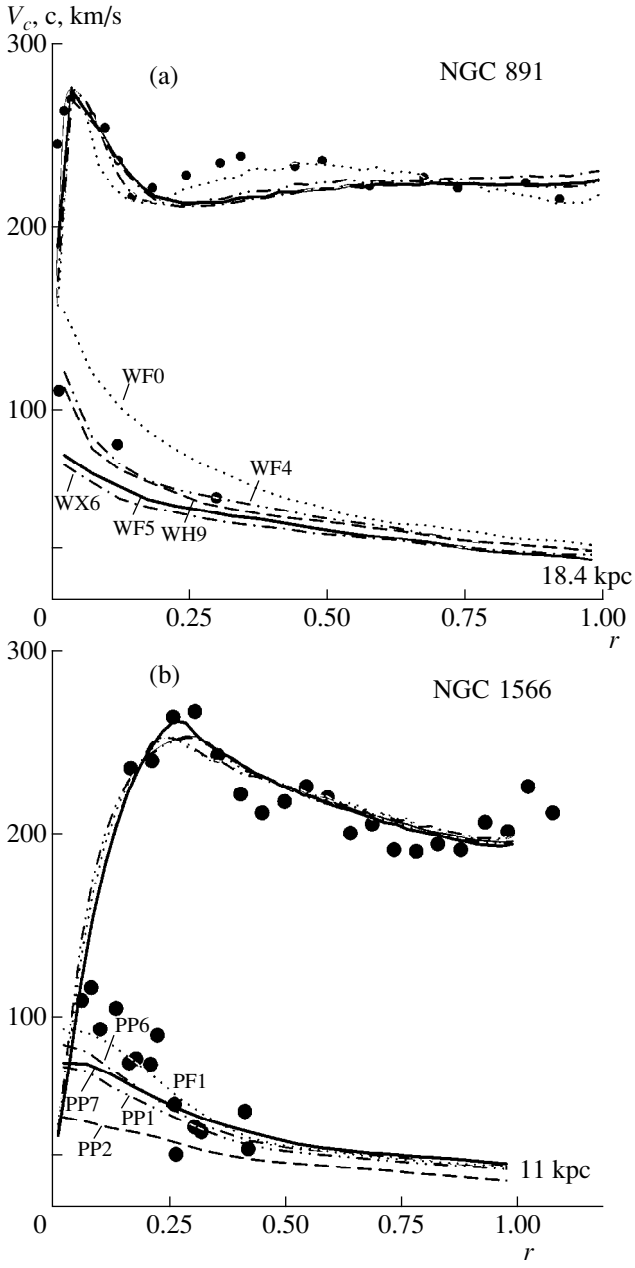


Fig. 7. Radial profiles of rotation velocity and line-of-sight velocity dispersion for models for (a) NGC 891 and (b) NGC 1566. The observed velocity dispersion in the central region of NGC 1566 is higher than the corresponding model velocity dispersions. The circles show the observational data [11].

Figure 7b shows the radial profiles of the rotational velocity and velocity dispersion for the observations and simulations. The galaxy possesses a prominent spiral structure, which is apparently responsible for the complex shape of the observed rotation curve, which has several local maxima. The galaxy displays substantial non-circular gas motions [39] with amplitudes reaching 20 km/s, implying that the observed $V(r)$ curve reflects the circular velocity only approximately.

For models PF1 and PF0, we chose the mass of the spherical subsystem to be no greater than that of the disk. Although the two models (PF1, PF0) satisfy the condition $c_{\rm sim} \gtrsim c_{\rm obs}$, neither they nor PP6, with $\mu = 1.33$, fit the data for NGC 1566. In all three cases, the dynamical modeling leads to a pronounced long-lived bar, which appears to be absent from the real galaxy. Intense heating and the development of a bar are accompanied by a substantial mass redistribution, which results in a very non-exponential surface-brightness profile, which is not observed. We must therefore consider models with more massive spherical subsystems (PP7 and PP1), which have different evolution for the velocity dispersion c_r. In the initial stage, which lasts three to four rotation periods, the system undergoes rapid heating similar to that shown in Fig. 1. However, instead of arriving at a steady state, the velocity dispersion continues to slowly increase in the inner part of the galaxy over at least 25 disk rotation periods (more than 5.5×10^9 years on the scale of the galaxy). This is clearly associated with the development of a long-lived asymmetric bar-like feature at the very center of the disk (Fig. 6),

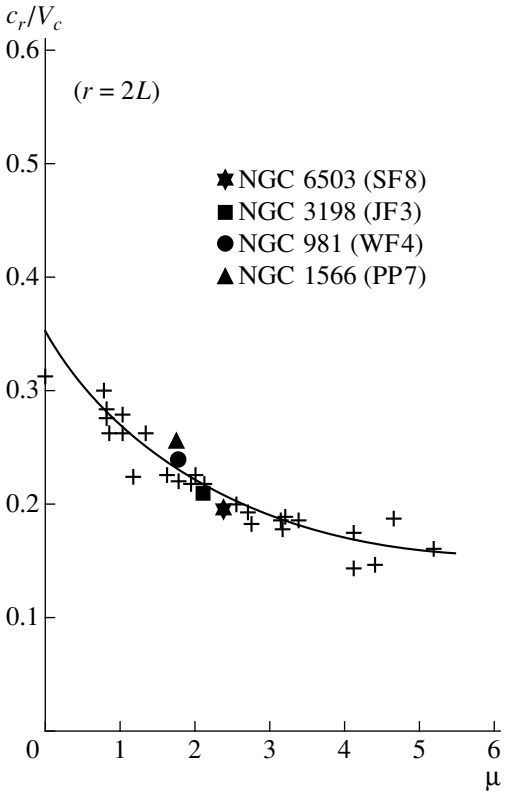


Fig. 8. Velocity dispersion ratio c_r/V_c as a function of the relative mass of the spherical subsystem μ at $r = 2L$ for models constructed in this paper.

which could be responsible for slow heating of the disk—a mechanism that, naturally, is more efficient at the center of the galaxy. After 30 revolutions, models PP1 and PP7 are in satisfactory agreement with the observed distribution of the velocity dispersion for $r = 0.2–0.4$.

No central non-axisymmetric feature develops in model PP2, which has a more massive spherical component ($\mu \gtrsim 2.5$); however, this solution underestimates the velocity dispersion in the inner part of the disk by a factor of two. The absence of secular heating in such models, which become axisymmetric after five revolutions, clearly demonstrates that the long-term heating is due to a small, bar-like feature. Although photometric observations have not detected the bar in this galaxy, Pence *et al.* [39] point out that such a feature could be responsible for the shape of the rotation curve in the central region. Model PP7, with $\mu = 1.74$, was in the best agreement with the observed velocity dispersions. This model implies a bulge-to-halo mass ratio of nearly three within $r < 4L$. The halo is "loose" ($a/L = 3$), so that the rotation curve is determined by the bulge throughout most of the disk. The mass of the disk in this model exceeds those of both the bulge and halo, although it remains less than the total mass of the spherical components.

4. DISCUSSION AND CONCLUSIONS

Our simulation results have yielded dynamical models for the galactic disks that agree simultaneously with the photometric disk scales, rotation curves, and estimated stellar-velocity dispersions of the galaxies studied. Similar patterns of the disk dynamical evolution computed with different numbers of particles (with fixed total disk mass), development of short-lived (transient) spirals during dynamic heating, and subsequent disk heating to a quasi-steady-state suggest that the increase of the velocity dispersion in axisymmetric disk models in the initial stage of their evolution is due to collective processes rather than particle scattering. In the latter case, comparing the numerical models with data for real galaxies would be problematic.

The numerical simulations enable us to estimate a lower limit to the mass of the spherical components that allows the galactic disks to remain in a stable state. If a disk is marginally stable, these estimates should correspond to the real component masses. Table 1 summarizes the μ values for the optimum models for the four galaxies considered. In all four galaxies, the minimum total halo and bulge mass in a sphere of radius $R = 4L$ is about twice the disk mass, although the disk-to-bulge mass ratios differ widely from one galaxy to another.

A comparison of the disk rotational velocities at $r = 2L$, where the disk contribution to the rotation curve is most significant, with model estimates of the disk component velocities shows that the maximum-disk solution overestimates the disk rotational velocity by 20–30% (see the last two columns in the table). Therefore, this model overestimates the disk mass by a factor of about 1.5 and underestimates the mass of the spherical components by about the same factor, providing further evidence for massive spherical components in these galaxies.

In the computed models, the ratio of the velocity dispersion to the rotational velocity c_r/V_c at a fixed r decreases as the relative mass of the spherical subsystem increases.[2] This pattern is clearly illustrated in Fig. 8, which compares model estimates of c_r/V_c at $r = 2L$ to the relative mass of the spherical components.

Figure 8 is based on the results of 35 simulations performed to search for optimum models for the galaxies considered (indicated by the large symbols). The strong differences of the halo and bulge concentrations of the models gives rise to the scatter of c_r/V_c values at a fixed μ. Since the systems analyzed had different maximum rotational velocities and rotation-curve shapes, this relation should be universal enough to allow coarse estimation of the masses of the spherical subsystems from the observed c_r/V_c ratios, without directly constructing dynamical models for objects studied. The form of this relation shows that the resulting μ values are most certain for galaxies with low relative masses for their spherical subsystems. Note, however, that the accuracies of local

[2] This relation was first pointed put by Morozov [37].

stellar velocity dispersions measured in old disks at fixed r are fairly low, so that the available data are consistent with a wide range of μ within the quoted errors. In this respect, our interpretations of the functional relations $c(r)$ yield more trustworthy results.

For three of the four galaxies, we succeeded in constructing dynamical models that are in good agreement with the assumption that their stellar disks are close to the gravitational stability limit. The discrepancy between the observed and modeled $c(r)$ and $V_c(r)$ relations are largest for NGC 1566, and cannot be eliminated by adjusting the parameters of the spherical subsystem. In any case, the central part of the stellar disk ($r < 0.75L$) in NGC 1566 is "overheated," and possesses some stability margin. This may be partially due to the prominent spiral pattern of this galaxy, which provides an additional source of heating and distorts the observed stellar rotation curve and velocity dispersions. The results of dynamical modeling do not rule out the presence of a central bar-like feature in this galaxy, which could be responsible for additional heating of the stellar disk.

Our simulations indicate that galaxy models that neglect the stellar velocity dispersion can yield appreciably underestimated masses for the spherical components for maximum-disk solutions. All dynamical models based on the maximum-disk solutions develop bars, and yield velocity dispersions that are in poor agreement with the observational data.

The model bulge masses are consistent with the photometric data: the bulges in NGC 6503 and NGC 3198 have both lower relative masses and lower relative luminosities compared to NGC 1563 and NGC 891.

The fact that our simulations yield rather thin models with $h/L \lesssim 0.2$ provides further evidence for a massive spherical subsystem in NGC 891. As first shown by Zasov *et al.* [4], the stellar-disk scale height h should decrease with increasing relative mass of the halo. The observational data yield $h/L \simeq 0.11$ for NGC 891 [33], which, according to the results of Zasov *et al.* [4], gives a similar estimate: $\mu \gtrsim 2$. The dynamical models of all four galaxies have velocity dispersion ratios c_z/c_r ranging from 0.4 to 0.7. This agrees well with available estimates for other spirals, and is somewhat higher than the value of 0.37 obtained from an analysis of the development of warping perturbations in the framework of a simple gravitating disk model [34].

ACKNOWLEDGMENTS

We modeled the rotation curves using the *GR* code written by A.N. Burlak. We are grateful to N.Ya. Sotnikova for valuable comments. The work was partially supported by the Russian Foundation for Basic Research (project code 98-02-17102) and the State Science and Technology Program in Astronomy.

REFERENCES

1. T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi, Astrophys. J. **295**, 305 (1985).
2. S. M. Kent, Astron. J. **91**, 1301 (1986).
3. P. Salucci and M. Persic, Astron. Astrophys. **351**, 442 (1999).
4. A. V. Zasov, D. I. Makarov, and E. A. Mikhaĭlova, Pis'ma Astron. Zh. **17**, 884 (1991).
5. R. Bottema and J. P. E. Gerritsen, Mon. Not. R. Astron. Soc. **290**, 585 (1997).
6. S. Courteau and H. Rex, Astrophys. J. **513**, 561 (1999).
7. B. Fuchs, in *Galaxy Dynamics*, Eds. by D. R. Merrit *et al.*; Astron. Soc. Pac. Conf. Ser. 182 (1999).
8. A. G. Morozov and A. V. Zasov, Astron. Zh. **62**, 475 (1985) [Sov. Astron. **29**, 277 (1985)].
9. A. V. Zasov, Pis'ma Astron. Zh. **11**, 730 (1985) [Sov. Astron. Lett. **11**, 307 (1985)].
10. E. A. Mikhaĭlova and A. V. Khoperskov, Astron. Zh. **69**, 1112 (1992) [Sov. Astron. **36**, 573 (1992)].
11. R. Bottema, Astron. Astrophys. **275**, 16 (1993).
12. E. Athanassoula, A. Bosma, and S. Papaioannou, Astron. Astrophys. **179**, 23 (1987).
13. G. Efstathiou, G. Lake, and J. Negroponte, Mon. Not. R. Astron. Soc. **199**, 1069 (1982).
14. J. C. Mihos, S. S. McGaugh, and W. J. G. de Block, Astrophys. J. Lett. **477**, L79 (1997).
15. A. Toomre, Astrophys. J. **139**, 1217 (1964).
16. J. A. Sellwood and R. G. Carlberg, Astrophys. J. **282**, 61 (1984).
17. R. G. Carlberg and J. A. Sellwood, Astrophys. J. **292**, 79 (1985).
18. J. M. Bardeen, in *IAU Symposium 69: Dynamics of Stellar Systems*, Ed. by A. Hayli (D. Reidel, Dordrecht, 1975), p. 297.
19. V. L. Polyachenko, E. V. Polyachenko, and A. V. Strel'nikov, Pis'ma Astron. Zh. **23** (8), 598 (1997) [Astron. Lett. **23**, 525 (1997)].
20. A. G. Morozov, Pis'ma Astron. Zh. **7**, 197 (1981) [Sov. Astron. Lett. **7**, 109 (1981)].
21. N. N. Gor'kavyĭ and A. M. Fridman, *Physics of Planetary Rings* [in Russian] (Nauka, Moscow, 1994).
22. A. G. Morozov, Astron. Zh. **58**, 734 (1981) [Sov. Astron. **25**, 421 (1981)].
23. R. Hockney and J. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1984; Mir, Moscow, 1987).
24. J. P. Ostriker and P. J. E. Peebles, Astrophys. J. **186**, 467 (1973).
25. A. G. Morozov, Astron. Zh. **58**, 34 (1981) [Sov. Astron. **25**, 19 (1981)].
26. K. G. Begeman, A. H. Broels, and R. H. Sanders, Mon. Not. R. Astron. Soc. **249**, 523 (1991).
27. J. V. Villumsen, Astrophys. J. **290**, 75 (1985).

28. P. C. van der Kruit and R. de Grijs, Astron. Astrophys. **352**, 129 (1999).
29. Y. Sofue and N. Nakai, Publ. Astron. Soc. Jpn. **45**, 139 (1993).
30. R. Bottema, P. C. van der Kruit, and E. A. Valentijn, Astron. Astrophys. **247**, 357 (1991).
31. R. C. Kennicutt, Astron. J. **86**, 1847 (1981).
32. R. Bottema, Astron. Astrophys. **257**, 69 (1992).
33. Y. I. Byun, N. D. Kylafis, E. V. Paleologou, *et al.*, Astron. Astrophys. **344**, 868 (1999).
34. V. L. Polyachenko and I. G. Shukhman, Pis'ma Astron. Zh. **3**, 254 (1977) [Sov. Astron. Lett. **3**, 134 (1977)].
35. P. C. van der Kruit and L. Searle, Astron. Astrophys. **95**, 105 (1981).
36. V. L. Polyachenko, Pis'ma Astron. Zh. **3**, 99 (1977) [Sov. Astron. Lett. **3**, 51 (1977)].
37. A. G. Morozov, Pis'ma Astron. Zh. **9**, 716 (1983) [Sov. Astron. Lett. **9**, 370 (1983)].
38. G. de Vaucouleurs, A. de Vaucouleurs, H. Corwin, *et al.*, *Third Reference Catalogue of Galaxies* (Springer-Verlag, New York, 1991).
39. W. D. Pence and P. Taylor, Astrophys. J. **357**, 415 (1990).

Translated by A. Dambis

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 195–202. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 229–237.

Turbulent Interstellar Plasma and Ultrahigh Angular Resolution in Radio Astronomy

V. I. Shishov

Pushchino Radio Astronomy Observatory, Astro Space Center of the Lebedev Institute of Physics, Pushchino, Moscow Region, 142292 Russia

Received March 10, 2000

Abstract—The effects of the modulation of radio waves during their passage through turbulent interstellar plasma on measurements of the structure of radio sources made with ultrahigh angular resolution using space radio interferometers are considered. Typical scattering angles $\theta_{scat,\,pl}$ for an extragalactic radio source at various wavelengths are estimated from scattering observations for pulsars and extragalactic sources. The $(\varphi_0\lambda)$ plane, where φ_0 is the source size and λ is the wavelength of the radio emission, can be divided into five regions, in which different regimes of radio-wave modulation and image reconstruction are realized. Possibilities for image reconstruction in each of these regions are investigated. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In 1968, shortly after the discovery of pulsars, Scheuer [1] proposed that the fine structure in the fluctuations of the pulsar emission on timescales of a few minutes and with a characteristic frequency scale of about 1 MHz is due to strong scintillations of the radio emission on inhomogeneities in the interstellar plasma. In 1969, Rickett [2] showed that the characteristic frequency scale for pulsar flux fluctuations decreases with increasing dispersion measure, demonstrating the interstellar nature of these fluctuations. According to Sieber [3], the parameters of pulsar flux fluctuations on timescales of days and months also correlate with the dispersion measure. In 1983, Rickett *et al.* [4] showed that short-period and long-period scintillations of pulsars correspond to refractive-index inhomogeneities with a single Kolmogorov power-law turbulence spectrum.

Over the 30 years of studying the interstellar scintillations of radio sources, extensive series of observational data have been accumulated, and we now know a fair bit about the properties of the turbulent interstellar plasma responsible for modulating the radio signals. In particular, it has been established that the turbulent plasma is concentrated in individual clouds, and the distribution of these clouds in the Galaxy has been studied [5–7]. The most important result of these studies from the point of view of very high angular resolution is that the radio signals of sources in all directions show modulation by turbulent plasma, so that the interstellar medium could impose restrictions on observations with high angular resolution. At the same time, these data are not sufficient to construct a complete and reliable model of the turbulence structure and distribution of turbulent plasma clouds in the Galaxy, and further measurements and analysis are required.

The most important parameter describing the modulation properties of a medium is the scattering angle θ_0. It is usually assumed that scattering on inhomogeneities of the interstellar medium simply increases the source angular size, and the angle θ_0 limits the attainable angular resolution. However, in some cases, the scattered emission partially maintains coherent properties associated with modulation by the turbulent medium, and we can separate the intrinsic angular size of a source from the component introduced by the turbulent medium [8, 9]. Below, we discuss in more detail the fluctuation properties of radio signals that have passed through a turbulent interstellar medium and probable restrictions imposed by the interstellar medium on the limiting angular resolution attainable in radio astronomy.

2. PARAMETERS OF CLOUDS OF TURBULENT INTERSTELLAR PLASMA

The modulation of radio waves that occurs during their passage through the interstellar medium is due to inhomogeneities of the refractive index. For a plasma, the refractive index at frequencies much higher than the plasma frequency ω_p is [10, 11]

$$\mu \approx 1 - (\omega/\omega_p)^2/2 = r_e\lambda^2 N_e/4\pi, \qquad (1)$$

where ω is the cyclic frequency, λ the wavelength, r_e the classical electron radius, and N_e the electron number density. We can see that the refractive index depends only on the electron density. The electron-density inhomogeneities are described by the square-law Fourier spectrum of the fluctuations of N_e:

$$\mathbf{\Phi}_{N_e}(\mathbf{q}) = \langle [\int d^3\mathbf{q} \exp(-i\mathbf{qr}) N_e(\mathbf{r})]^2 \rangle. \qquad (2)$$

1063-7729/01/4503-0195$21.00 © 2001 MAIK "Nauka/Interperiodica"

Analysis of extensive observational material on refractive and diffractive scintillations has shown that the square-law spatial spectrum of the electron-density fluctuations is close to a power law over a wide range of scales:

$$\mathbf{\Phi}_{N_e}(\mathbf{q}) = C_{N_e} q^n, \quad 1/L < q < 1/l, \tag{3}$$

where L and l are the outer and inner scales of the turbulence, respectively. Observational data indicate that $L > 10^{16}$ cm. Based on a combined analysis of diffractive and refractive scintillations of a large sample of pulsars, we showed in [12] that two kinds of spectra are realized in the interstellar plasma: type I—purely power-law Kolmogorov spectra—and type II—Kolmogorov spectra with inner scale $l \approx 10^{11}$ cm. Type I spectra are realized in dense, compact regions, while type II spectra occur in more uniformly distributed turbulent clouds. For both types of spectra, the exponent n is close to that of a Kolmogorov spectrum:

$$n \approx 3.67. \tag{4}$$

Large series of data have been accumulated on the scattering of radio signals on inhomogeneities of the interstellar plasma and interstellar scintillation of radio sources in various directions and at various distances. A statistical analysis of these data has established that the turbulent interstellar plasma is concentrated in clouds, which occupy approximately 20% of the total volume of the interstellar medium. To zeroth approximation, the distribution of turbulent plasma in the Galaxy has a three-component structure [5–7].

Component A is the name given to the turbulent interstellar plasma outside the spiral arms of the Galaxy. It has a statistically uniform distribution and consists of diffuse clouds with diameter <26 pc, electron density 0.1–0.3 cm^{-3}, cloud number density >10^{-5} pc^{-5}, and scale height of the cloud distribution above the Galactic plane ≅0.9 kpc [6]. Component BI consists of diffuse turbulent clouds within the Galactic spiral arms. These have electron density 0.2–1 cm^{-3}, size <50 pc, cloud number density >10^{-6} pc^{-3}, and a scale height above the Galactic plane of about 200 pc [7]. Component BII is made up of dense, compactly arranged clouds associated with H II regions around O stars. These have sizes of about 10–50 pc, electron density 1–10 cm^{-3}, cloud number density of about 10^{-8} pc^{-3}, and a scale height above the Galactic plane of about 100 pc [7].

The interstellar medium near the Galactic plane modulates the radio signals of distant sources much more strongly than the medium at high Galactic latitudes, and, accordingly, the highest angular resolution can be achieved only at sufficiently high Galactic latitudes. We will based our further estimates on the parameters of component A of the turbulent interstellar plasma, and assume that the observations are carried out high above the Galactic plane.

3. REGIMES FOR MODULATION OF RADIO SIGNALS BY THE TURBULENT INTERSTELLAR PLASMA

Depending on the observing technique used, we can measure the following effects of the modulation of radio waves by a turbulent medium [13]:

(1) Effects of scattering:

(a) increase of source angular size,

(b) pulse broadening.

(2) Effects of fluctuations:

(a) scintillations, flux fluctuations,

(b) phase fluctuations, wandering of the source position.

The simplest characteristic of the field of a wave modulated by a turbulent medium is its second moment, which is measured by the mean response of an interferometer. The second moment corresponds to observations of scattering effects alone. For an infinitely distant source, the second moment of the wave field can be written [13]

$$\begin{aligned} B_E(\rho) &= \langle \overline{E(\rho_1)E(\rho_1+\rho)} \rangle \\ &= B_{E,0}(\rho)\exp[D(\rho)/2], \end{aligned} \tag{5}$$

where $B_{E,0}$ is the intrinsic coherence function of the field (visibility function) unperturbed by the turbulent medium, and ρ is a two-dimensional vector in the plane perpendicular to the line of sight. Angular brackets denote averaging over the statistics of the turbulent medium. An overline designates averaging over the statistics of the source. The factor $\exp[-D_S(\rho)/2]$ describes scattering effects. The function $D_S(\rho)$ corresponds to the structure function of the phase fluctuations calculated in a linear approximation; it is determined by the power spectrum of the electron-density fluctuations (2) [14]:

$$\begin{aligned} D_S(\rho) &= 4\pi\lambda^2 r_e^2 \int dz \int d^2 q_\parallel \\ &\times [1 - \cos(\mathbf{q}\rho)]\Phi_{N_e}(\mathbf{q}_\perp, q_\parallel), \end{aligned} \tag{6}$$

where $q_\parallel$ is the line-of-sight component of the spatial frequency vector $\mathbf{q}$, $\mathbf{q}_\perp$ is the component normal to the line of sight, and the z axis is aligned with the line of sight. For a power-law power spectrum (3), the function $D_S(\rho)$ can be written [13]

$$D_S(\rho) = |k\theta_0\rho|^{n-2}, \tag{7}$$

where θ_0 is the characteristic scattering angle. In a statistically uniform medium [12],

$$\begin{aligned} \theta_0 &= (1/k)[A(n)\lambda^2 r_e^2 C_{N_e} R]^{1/(n-2)}, \\ A(n) &= 2^{4-n}\pi^3/\Gamma(n/2)\sin(\pi n/2), \end{aligned} \tag{8}$$

where R is the total depth of the medium. For a spectrum with inner scale l and $\rho < l$, the $n - 2$ in (7) should be replaced with 2; the relationship for the factor A is also somewhat different.

Measuring only the mean value of B_E, we cannot separate the brightness distribution intrinsic to the source from the contribution due to the propagation of the radio signals through the turbulent medium. However, the situation changes if we can measure the fluctuations of the received signal. The measured interferometer response can be represented

$$I(\rho, t) = (1/\tau_0)\int dt_1 E(\rho_1, t_1)E^*(\rho_1 + \rho, t_1), \quad (9)$$

where τ_0 is the integration time. Below, we will show that measurements of the fluctuations of I can, in some cases, separate out the components corresponding to the source and the turbulent medium. The properties of the fluctuations of the interferometer response depend on the scintillation index m, defined as [13]

$$m^2 = \langle [I(0, t) - \langle I(0, t)\rangle]^2\rangle / \langle I(0, t)\rangle^2, \quad (10)$$

where $I(0, t)$ is the received radiation flux. Scintillation theory distinguishes two scintillation regimes. In the weak-scintillation regime, the scintillation index $m < 1$, and we have for a spectrum of the form (3) [13]

$$m^2 \approx m_0^2 = K(n)D_S(\sqrt{R/k}) = K(n)(kR\theta_0^2)^{(n-2)/2}, \quad (11)$$

where $K(n)$ is a numerical factor of order unity and 1, $k = 2\pi/\lambda$ is the wave number. The characteristic spatial scale of the flux fluctuations is

$$b = (1/2)\sqrt{(R/k)}. \quad (12)$$

With increasing m_0, the scintillation index m grows only to $m_0 = 1$. When

$$m_0 > 1, \quad (13)$$

the scintillation index saturates at a level near unity [13]. The corresponding regime is that of strong, or saturated, scintillations. In the saturated-scintillation regime, the diffraction pattern consists of two components: small-scale (or diffractive) and large-scale (or refractive) [13].

The squared scintillation index for the diffractive component is

$$m_{\rm dif}^2 \approx 1, \quad (14)$$

and the spatial and temporal scales are [13]

$$b_{\rm dif} \approx 1/k\theta_0, \qquad T_{\rm dif} \approx b_{\rm dif}/V \approx 1/k\theta_0 V. \quad (15)$$

Here, V is the velocity of the medium with respect to the line of sight. One distinctive feature of diffractive scintillations is the presence of fine structure in the intensity fluctuations as the carrier frequency of the radio signals change. The characteristic frequency scale is

$$\Delta f_{\rm dif} = Ac/\pi R\theta_0^2, \quad (16)$$

where c is the velocity of light and A is a factor that depends on the model for the medium and the form of the turbulence spectrum. For a statistically uniform medium and a Kolmogorov spectrum, $A \cong 3.67$ [12].

The squared scintillation index of the refractive component is [12]

$$m_{\rm ref}^2 \approx [D_S(\sqrt{R/k})]^{-2(4-n)/(n-2)} \approx m_0^{-4(4-n)/(n-2)} \approx (1/kR\theta_0^2)^{4-n}, \quad (17)$$

and the spatial and temporal scales are [12]

$$b_{\rm ref} \approx R\theta_0, \qquad T_{\rm ref} \approx b_{\rm ref}/V \approx R\theta_0/V. \quad (18)$$

4. EFFECT OF THE TURBULENT INTERSTELLAR PLASMA ON AN INTERFEROMETER RESPONSE

Below, we consider in more detail the processes of averaging over the statistics of the source and the turbulent medium. In the absence of a turbulent medium and with

$$\tau_0 \Delta f \gg 1, \quad (19)$$

where Δf is the frequency band, the interferometer response I determined by (9) is close to the theoretical value of the field coherence function:

$$I_0(\rho, \mathbf{t}) \cong B_{E,0}(\rho, \mathbf{t}) = \overline{E_0(\rho_1, \mathbf{t})E_0(\rho_1 + \rho, t)}. \quad (20)$$

Here, the overline denotes averaging over the source statistics.

In the presence of a turbulent medium with

$$\tau_0 < t_{\rm dif}, \quad (21)$$

the quantity I fluctuates with characteristic timescales $T_{\rm dif}$ and $T_{\rm ref}$. The actual values of these timescales will be discussed later. In this case, we will call I the quasi-instantaneous interferometer response.

The process of averaging over the statistics of the turbulent medium corresponds to temporal integration with a long integration time T. We introduce two types of averaging over the statistics of the turbulent medium. If

$$T_{\rm dif} < T < T_{\rm ref}, \quad (22)$$

then the averaging is partial. We will designate such partial averaging by angular brackets with the subscript "dif." For example,

$$\langle I(\rho, t)\rangle_{\rm dif} = (1/T)\int_t^{t+T} dt_1 I(\rho, t_1). \quad (23)$$

Full averaging will be denoted by angular brackets with no subscript [see Eq. (5)]. Full averaging corresponds to a very long integration time

$$T \gg T_{\rm ref}. \quad (24)$$

4.1. Fluctuations of a Partially Averaged Interferometer Response

The mean value of $\langle I\rangle_{\rm dif}$ is

$$\langle\langle I\rho\rangle_{\rm dif}\rangle = \langle I(\rho)\rangle = B_E(\rho). \quad (25)$$

The coherence function B_E was considered above, and is defined by (5).

For strong or saturated scintillations, we can obtain for the second moment of $\langle I\rangle_{\rm dif}$ [15]

$$\langle|\langle I(\rho)\rangle_{\rm dif}|^2\rangle = \langle I(\rho)\rangle^2 + m_{\rm ref}^2 K(\rho)\langle I(0)\rangle^2, \quad (26)$$

where $m_{\rm ref}$ is the refractive scintillation index. Function $K(\rho)$ is determined by the intrinsic brightness distribution of the source and scattering effects; it has the characteristic scale ρ_K

$$\begin{aligned}\rho_K &\approx 1/k\theta_0, \quad \varphi_0 \ll \theta_0,\\ \rho_K &\approx 1/k\varphi_0, \quad \varphi_0 \gg \theta_0.\end{aligned} \quad (27)$$

Here, φ_0 is the intrinsic angular size of the source. For $\rho = 0$,

$$K(\rho) = 1. \quad (28)$$

When $\varphi_0 \ll \theta_0$, the refractive scintillation index $m_{\rm ref}$ is determined by (17), and when $\varphi_0 \gg \theta_0$, we should replace θ_0 in this relationship with φ_0.

We can see that the fluctuations of the partially averaged interferometer response $\langle I\rangle_{\rm dif}$ can be used to distinguish between the intrinsic angular source size φ_0 and the scattering angle θ_0. This separation of the intrinsic and scattering sizes works well if these quantities are fairly close to each other. Note also that function $K(\rho)$ is not equal to $I(\rho)$; hence, the structure of the partially averaged interferometer response $\langle I(\rho)\rangle_{\rm dif}$ also fluctuates.

4.2. Fluctuations of the Quasi-instantaneous Interferometer Response

The fluctuation parameters of $I(\rho)$ depend on the scintillation regime [13]. In the case of weak scintillations, the fluctuations of the amplitude of an interferometer response are small:

$$\begin{aligned}\langle|\Delta I(\rho)|^2\rangle &= \langle|I(\rho) - \langle I(\rho)\rangle|^2\rangle\\ &\ll \langle|I(\rho)|^2\rangle \cong |B_{E,0}(\rho)|^2.\end{aligned} \quad (29)$$

However, the phase fluctuations can be large:

$$\langle I(\rho)^2\rangle \cong B_{E,0}^2(\rho)\exp[-D_S(\rho)]. \quad (30)$$

This means that the turbulent medium makes the source position wander, without changing its brightness distribution. We can neglect the phase fluctuations as well if

$$\rho < \rho_0 = 1/k\theta_0. \quad (31)$$

In the case of strong or saturated scintillations, we can obtain for the second moment of $I(\rho)$ [16, 17]

$$\begin{aligned}\langle|I(\rho)|^2\rangle &= \langle|I(\rho)|\rangle^2\\ &+ \int d^2\theta \exp[ik(\rho\theta)]J_1(\theta)\exp[-\int drD(r\theta)],\end{aligned} \quad (32)$$

where

$$D(\rho) = (dD_s(\rho))/dr, \quad (33)$$

$$J_1(\theta) = \int d^2\theta_1 J_0(\theta_1)J_0(\theta_1 + \theta), \quad (34)$$

$$J_0(\theta) = (1/2\pi)^2\int d^2\rho \exp[-ik(\rho\theta)]B_{E,0}(\rho). \quad (35)$$

For a source with small angular size

$$\varphi < \theta_{\rm dif} = 2/kR\theta_0 \quad (36)$$

we approximately have [8, 17]

$$\langle|I(\rho)|^2\rangle \cong \langle|I(\rho)|\rangle^2 + |B_{E,0}(\rho)|^2. \quad (37)$$

Hence, in this case, we can reconstruct the intrinsic brightness distribution of the source. It is obvious that we can also do this for sources with large angular sizes

$$\varphi_0 > 2\theta_0 = \theta_{\rm scat,\,pl}. \quad (38)$$

Here, $\theta_{\rm scat,\,pl}$ is the apparent half-power angular size of an infinitely distant pointlike source. When the source has an intermediate angular size

$$\theta_{\rm dif} < \varphi_0 < 2\theta_0, \quad (39)$$

we lose information and cannot reconstruct the intrinsic brightness distribution.

Integration over angle θ in (32) corresponds to preaveraging the turbulent field fluctuations over the intrinsic brightness distribution of the source. This results in a loss of the possibility of separating the intrinsic brightness distribution from the modulation perturbation contributed by the turbulent medium.

The fluctuations of $I(\rho)$ have fine structure in the radio frequency, with the characteristic scale $\Delta f_{\rm dif}$ determined by (16). If the receiver frequency band Δf is broad, the interferometer response is preaveraged over fluctuations with a turbulent nature. In this case, we likewise cannot distinguish the intrinsic brightness distribution of the source from the modulation perturbation due to the turbulent medium. In order to preserve the diffractive fluctuations, the inequality

$$\Delta f < \Delta f_{\rm dif} \quad (40)$$

must be satisfied.

5. OBSERVATIONAL ESTIMATES OF SCATTERING AND SCINTILLATION PARAMETERS

In our recalculation and comparison of the observational data, we will use a model with a statistically uniform medium with a purely power-law Kolmogorov

turbulence spectrum, determined by (3) and (4). Our reasons for using this model are the following. It was shown in [6] that the spatial distribution of the turbulent interstellar plasma outside the spiral arms (component A) is close to statistically uniform. In Section 2, we cited two types of spectra for the interstellar plasma turbulence [12]. In our case of a medium with a statistically uniform distribution, we must use a type I spectrum.

The scattering angle is an important parameter describing the modulation properties of the turbulent interstellar plasma. This parameter has been estimated using several independent methods. In the first, the apparent angular sizes of a number of pulsars were measured [18]. Since the intrinsic angular sizes of pulsars are negligible, their apparent angular sizes correspond to the scattering angles in the directions toward these pulsars. The measurements were carried out at $f \cong 300$ MHz, and the observational data can be represented by the relationship

$$\theta_{\text{scat, sph}} \cong 8(\text{DM}/\text{DM}_0)\ \text{mas}, \quad \text{DM}_0 = 25\ \text{pc/cm}^3, \tag{41}$$

where $\theta_{\text{scat, sph}}$ is the scattering angle of a spherical wave and the apparent angular size corresponds to the half-power diameter of the source. Recall that θ_0 is the scattering angle of a plane wave corresponding to the half-power radius of the brightness distribution. Of primary importance for us are the conditions for the observation of extragalactic sources; therefore, we use the quantity $\theta_{\text{scat, pl}}$ defined above, which corresponds to the apparent half-power angular size of an infinitely distant pointlike source.

In the above model for the medium, the scattering angles of plane and spherical waves are related by the expression

$$\theta_{\text{scat, pl}} = (n-1)^{1/2}\theta_{\text{scat, sph}} = 1.6\theta_{\text{scat, spl}}. \tag{42}$$

Using the wavelength dependence of θ_0, according to (8), we obtain for the scattering angle of a plane wave

$$\theta_{\text{scat, pl}} = 2\theta_0 \cong 80(\lambda/10\ \text{cm})^{2.2}(\text{DM}/\text{DM}_0)^{1/2}\ \mu\text{as}. \tag{43}$$

In [6], we presented a plot of the measured dispersion measures DM of pulsars as a function of Galactic latitude *b*. At latitudes >10°, the maximum dispersion measures, which correspond to the most distant pulsars and probe the entire depth of the medium, can be represented by the approximate relationship

$$\text{DM}_{\max} \cong 25(\sin b)^{-1}\ \text{pc/cm}^3. \tag{44}$$

The DMs measured for pulsars in globular clusters also fit relationship (44) [19]. Substituting (44) into (43), we obtain

$$\theta_{\text{scat, pl}} \cong 80(\lambda/10\ \text{cm})^{2.2}(\sin b)^{-1/2}. \tag{45}$$

Note that this relationship reflects only the mean latitude dependence of the scattering angle, and there can be deviations toward either larger or smaller values against the background of this mean dependence.

The second method for estimation of the scattering angle uses measurements of pulsar pulse broadening τ and the frequency correlation scales of scintillations Δf_{dif}. In [6], we presented the dependence of τ on Galactic latitude *b*. The maximum values of τ for $b > 10°$ can be approximated by

$$\tau(300\ \text{MHz}) \approx 10(\sin|b|)^{-2}\ \mu\text{s}. \tag{46}$$

Using (16) for a statistically uniform medium [12], we obtain

$$\theta_{\text{scat, pl}} = 2\theta_0 \approx 12(\tau/10\ \mu\text{s})^{1/2}\ \text{mas}. \tag{47}$$

A similar relationship can be derived from a direct comparison of the τ and $\theta_{\text{scat, sph}}$ values measured for the same pulsars. Substituting (47) to (46), we obtain

$$\theta_{\text{scat, pl}} \cong 75(\lambda/10\ \text{cm})^{2.2}(\sin|b|)^{-1}\ \mu\text{as}. \tag{48}$$

Similar estimates of $\theta_{\text{scat, pl}}$ based on the model of Taylor and Cordes [21] are presented in [20].

The third method for estimation of the scattering angle is measurement of the angular sizes of extragalactic sources. It was shown in [22] that the minimum angular sizes of extragalactic sources depend on the emission measure ME, and increase with growing ME. However, such data are quite scarce at high Galactic latitudes. These data, derived from measurements at meter wavelengths and presented in [23, 24], can be approximated by

$$\theta_{\text{scat, pl}} \cong 20(\lambda/10\ \text{cm})^{2.2}(\sin b)^{-0.6}\ \mu\text{as}. \tag{49}$$

This estimate of the scattering angle is considerably lower than the estimates (45) and (47). This discrepancy is quite understandable. The pulsar estimates are based on a selection of maximum values at a given latitude, and accordingly refer to denser regions of the interstellar medium. The estimates based on extragalactic sources are based on minimum values, so that they reflect more rarefied regions. Below, we will use the estimate (49) of the scattering angle in our analysis. The numerical values of the scattering angle toward the Galactic pole at the frequencies for the *Radio Astron* project are given in the table.

The second important parameter is the critical wavelength λ_{cr} at which scintillations pass from the weak to the strong scintillation regime. The observed dependence of λ_{cr} on distance *R* is studied in [25]. This dependence can be approximated by

$$\lambda_{\text{cr, sph}} \cong 15(R/1\ \text{kpc})^{0.73}\ \text{cm}, \quad R \geq 1\ \text{kpc}. \tag{50}$$

Estimated scattering angle $\theta_{scat,\,pl}$ toward the Galactic pole at the *Radio Astron* frequencies

λ, cm	$\theta_{scat,\,pl}$, μas
1.35	0.24
6	6.4
10	20
18	72
92	2600

Within the observational errors, the exponent in (50) coincides with the theoretical exponent for a Kolmogorov spectrum:

$$\lambda_{cr} \propto R^{n/(n+2)} = R^{0.65}. \quad (51)$$

Relationship (50) has been derived for a spherical wave. For a plane wave, assuming that R = 0.9 kpc toward the Galactic pole [6], we obtain

$$\lambda_{cr,\,pl} \cong 10(\sin b)^{-0.65}\ \text{cm}. \quad (52)$$

A similar estimate of $\lambda_{cr,\,pl}$ is given in [20].

The next important modulation parameters are the scintillation timescales. Based on our analysis of the existing data carried out in [17], we can derive the following approximate relationships for the timescales for diffractive and refractive scintillations:

$$T_{dif,\,puls} \approx 0.75(10\ \text{cm}/\lambda)^{1.2}(\sin b)^{0.6}\ \text{h}, \quad (53)$$

$$T_{ref,\,puls} \approx 0.7(\lambda/10\ \text{cm})^{2.2}(\sin b)^{1.6}\ \text{h}. \quad (54)$$

Note that the temporal variations of pulsar scintillations are determined by the motions of the pulsars themselves; the typical velocity of a pulsar is $V_{puls} \cong 150$ km/s. For an extragalactic source, the velocity of the line of sight with respect to the interstellar medium will be determined by the Earth's orbital motion with velocity $V \cong 30$ km/s and by fluctuation motions in the interstellar medium. Taking into account this difference in the velocities and the fact that, for extragalactic sources, we must pass from a spherical wave to a plane wave approximation, we obtain

$$T_{dif} \cong 2.2(10\ \text{cm}/\lambda)^{1.2}(\sin b)^{0.6}\ \text{h}, \quad (55)$$

$$T_{ref} \cong 6.3(\lambda/10\ \text{cm})^{2.2}(\sin b)^{1.6}\ \text{h}. \quad (56)$$

All the observational data on the frequency correlations of scintillations can be approximated by the relationship [17]

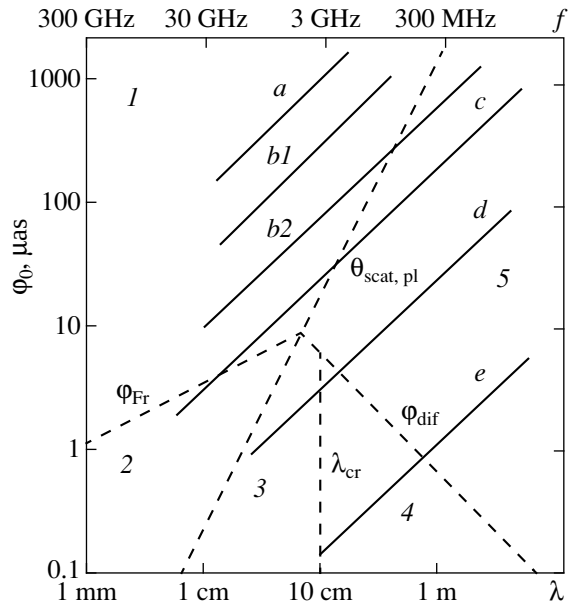
$$\Delta f_{dif} \cong 1/2\pi\tau \cong 9(f/1\ \text{GHz})^{4.4}(\sin b)^{2.2}\ \text{MHz}. \quad (57)$$

The parameter Δf_{dif} determines the limiting frequency band for which diffractive scintillations are not suppressed.

To conclude, we estimate a very important parameter: the diffraction angle. Assuming R = 0.9 kpc and using (48), we obtain

$$\varphi_{dif} = 2\theta_{dif} = 2b_{dif}/R = 4/k\theta_{scat,\,pl}R \approx 5(\lambda/10\ \text{cm})^{-2.2}(\sin b)^{1.6}\ \mu\text{as}. \quad (58)$$

The diffraction angle restricts the limiting intrinsic source size for which diffractive scintillations are not suppressed.



Domains with difference regimes for source-image reconstruction. The horizontal axis plots wavelength λ and the vertical axis the source angular size φ. The dashed lines correspond to the values of $\theta_{scat,\,pl}$, φ_{dif}, φ_{Fr}, and λ_{cr}. The solid lines show the angular resolutions of space interferometers with baselines (*a*) $B = 2 \times 10^4$ km (the VSOP project), (*b1*) $B = 7 \times 10^4$ km (previous working orbit for the *Radio Astron* project), (*b2*) $B = 3 \times 10^5$ km (proposed higher orbit for the *Radio Astron* project), (*c*) $B = 7 \times 10^5$ km, (*d*) $B = 7 \times 10^6$ km, (*e*) B = 1 AU.

6. REGIMES FOR THE RECONSTRUCTION OF THE SOURCE IMAGE

Our final results on the restrictions imposed by the turbulent interstellar plasma on source-image reconstruction using an interferometric array with maximum baseline D and angular resolution $\varphi_{res} = \lambda/D$ are shown in the figure. The horizontal axis plots wavelength λ, and the vertical axis plots the intrinsic source angular size φ_0. The dashed lines show the scattering angle $\theta_{scat,\,pl}$ and diffraction angle θ_{dif}, as well as the angular size of the first Fresnel zone

$$\varphi_{Fr} = (kR)^{-1/2}. \quad (59)$$

In addition, the vertical dashed line separates the weak and strong scintillation regimes. The dashed lines divide the plane of the figure into five regions (we will

call them domains) with different regimes for the reconstruction of the initial source image. The first domain is defined by the inequalities

$$\varphi_0 > \varphi_{Fr}, \quad \varphi_0 > \theta_{scat,\,pl}. \tag{60}$$

Here, the source angular size is larger than the scattering angle. In the left-hand part of this domain, at $\lambda < \lambda_{cr}$, fluctuations of the modulated emission intensity are small, and are additionally suppressed by the large angular size of the source. Here, we can neglect the effect of the turbulent interstellar plasma on the apparent source image. In the right-hand part of the domain, at $\lambda > \lambda_{cr}$, the strong-scintillation regime is realized, but the modulation fluctuations of the intensity are small, since they are suppressed by the large intrinsic size of the source φ_0. Since, $\varphi_0 > \theta_{scat,\,pl}$ in this zone, we can neglect the effect of the turbulent interstellar medium. One exception is the zone adjacent to the line $\varphi_0 = \theta_{scat,\,pl}$. Here, we will observe refractive scintillations, which are manifest as modulation of the total flux of the source without significant distortion of its brightness distribution.

The second domain is defined by the inequalities

$$\theta_{scat,\,pl} < \varphi_0 < \varphi_{Fr}. \tag{61}$$

In this domain, there are weak intensity fluctuations whose level is determined by the scintillation index m_0 [see Eq. (11)]. In addition, there are weak phase fluctuations whose level depends on the interferometer baseline and is determined by the magnitude of $D_S(\rho)$ [see Eq. (7)]. Here, the effect of turbulent interstellar plasma is manifested as additional weak sidelobes that vary randomly with the timescale of the weak scintillations.

The third domain is defined by the inequalities

$$\varphi_0 < \theta_{scat,\,pl}, \quad \lambda < \lambda_{cr}. \tag{62}$$

In this domain, there are no significant fluctuations of the flux or distortions of the brightness distribution. At the same time, there are fluctuations of the phase of the visibility function, which correspond to wandering of the source position. Here, we can separate out the intrinsic angular size of the source φ_0 and the angle $\theta_{scat,\,pl}$ describing the wandering of the position of the source barycenter. The observed response of the interferometer is given by (29) and (30).

The fourth domain is defined by the inequalities

$$\lambda > \lambda_{cr}, \quad \varphi_0 < \varphi_{dif}. \tag{63}$$

Here, the scintillations are strong, and the turbulent medium models the amplitude and phase of the visibility function. Since the intrinsic source size is small, there is no suppression of diffractive scintillations due to the intrinsic size of the source. The condition for preservation of the diffractive scintillations is also given by the inequalities

$$t_0 < t_{dif}, \quad \Delta f_0 < \Delta f_{dif}. \tag{64}$$

According to (32), we can reconstruct the amplitude of the initial visibility function.

The fifth domain is defined by the inequalities

$$\varphi_{dif} < \varphi_0 < \theta_{scat,\,pl}. \tag{65}$$

Under these conditions, diffractive scintillations are efficiently suppressed by the intrinsic source size, but refractive scintillations are preserved. As a rule, in this domain, we measure $\langle I \rangle_{dif}$. However, when

$$\varphi_0 \lll \theta_{scat,\,pl} \tag{66}$$

the influence of the intrinsic source size on the fluctuation properties of a partially averaged interferometer response is insignificant, and we lose information about the intrinsic brightness distribution. However, as the angular size of the source approaches the scattering angle,

$$\varphi_0 \longrightarrow \theta_{scat,\,pl}, \tag{67}$$

$\langle I \rangle_{dif}$ begins to appreciably depend on the intrinsic brightness distribution. This dependence is complicated and remains poorly studied [15]. In this boundary zone, we can partially reconstruct the intrinsic brightness distribution.

The solid lines in the figure show the limiting resolutions of instruments with baseline B. In particular, line *a* corresponds to the baseline of the VSOP space radio interferometer [26], line *b* to the baseline of the *Radio Astron* space interferometer for the previous working orbit [27], and line *c* to the baseline of *Radio Astron* for a proposed larger orbit. For baselines *a* and *b*, we can neglect the effect of the interstellar medium. In the case of baseline *c*, the effect of the interstellar medium will be appreciable only at long decimeter and meter wavelengths; we can recover the intrinsic brightness distribution by measuring the fluctuations of the interferometer response up to wavelengths as long as 1 m.

On the whole, our analysis shows that, at centimeter wavelengths, we can neglect the effect of the turbulent interstellar plasma for sources with angular sizes of the order of microarcseconds and larger. There are two regimes in the decimeter band. For sources with milliarcsecond angular sizes, we can neglect the influence of the interstellar medium; for sources with sizes of the order of microarcseconds and smaller, we must take the effect of the interstellar medium into account, though we are still able to reconstruct the intrinsic brightness distribution. For sources with such small angular sizes, measurements with baselines as long as $B = 1$ AU are plausible (line *e* in the figure).

Note that the most compact known extragalactic radio sources are those displaying flux variations on timescales shorter than one day (Intra Day Variation), whose sizes lie near the line $\varphi = \varphi_{Fr}$ [27]. The most compact known Galactic sources are pulsars, whose sizes lie within domains three and four.

7. CONCLUSIONS

(1) Turbulent interstellar plasma can seriously affect radio astronomical measurements with very high angular resolution. Its effect depends on the relationship between the intrinsic angular source size φ_0 and four parameters of the interstellar medium: the scattering angle θ_{scat}, diffraction angle φ_{dif}, angular size of the first Fresnel zone φ_{Fr}, and critical wavelength λ_{cr} separating the weak- and saturated-scintillation regimes.

(2) The (φ_0, λ) plane can be divided into five domains, in which different regimes for the modulation of radio signals and for source-image reconstruction are realized. The interstellar medium leads to an irrecoverable loss of information about the intrinsic brightness distribution only in the fifth domain, which is described by the inequalities $\varphi_{dif} < \varphi_0 < \theta_{scat,\,pl}$.

(3) At high Galactic latitudes, at centimeter and short-decimeter wavelengths, the angular sizes of sources resolved by VSOP and *Radio Astron* fall in the first domain, which is described by the inequalities $\varphi_0 > \varphi_{Fr}, \theta_{scat,\,pl}$. In this domain, the effect of the interstellar medium on the apparent image of a source can be neglected.

REFERENCES

1. P. A. G. Sheuer, Nature **218**, 920 (1968).
2. B. J. Rickett, Nature **221**, 158 (1969).
3. W. Sieber, Astron. Astrophys. **113**, 311 (1982).
4. B. J. Rickett, W. A. Coles, and G. Bourgois, Astron. Astrophys. **134**, 390 (1984).
5. J. M. Cordes, J. M. Weisberg, and V. Boriakoff, Astrophys. J. **288**, 221 (1985).
6. A. V. Pynzar' and V. I. Shishov, Astron. Zh. **74**, 663 (1997) [Astron. Rep. **41**, 586 (1997)].
7. A. V. Pynzar' and V. I. Shishov, Astron. Zh. **76**, 504 (1999) [Astron. Rep. **43**, 436 (1999)].
8. V. I. Shishov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **16**, 1796 (1973).
9. V. U. Zavorotny, in *Proceedings of a Workshop on Propagation Effects in Space VLBI*, Ed. by L. I. Gurvits (NAIC, Arecibo, 1993), p. 93.
10. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas* (Fizmatgiz, Moscow, 1967; Pergamon, Oxford, 1970).
11. E. N. Salpeter, Astrophys. J. **147**, 433 (1967).
12. T. V. Smirnova, V. I. Shishov, and D. R. Stinebring, Astron. Zh. **75**, 866 (1998) [Astron. Rep. **42**, 766 (1998)].
13. K. S. Gochelashvili and V. I. Shishov, in *Waves in Random Medium* [in Russian] (Vses. Inst. Nauchn. Tekh. Inf., Moscow, 1981), Itogi Nauki Tekh., Radiofiz., Fiz. Osnovy Élektron., Akust., Vol. 1.
14. V. I. Shishov and M. Tokumaru, J. Geomagn. Geoelectr. **48**, 1461 (1996).
15. V. I. Shishov, Waves Random Media **5**, 497 (1995).
16. V. I. Shishov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **33**, 443 (1990).
17. V. I. Shishov, V. M. Malofeev, A. V. Pynzar', and T. V. Smirnova, Astron. Zh. **72**, 485 (1995) [Astron. Rep. **39**, 428 (1995)].
18. C. R. Gwinn, N. Bartel, and J. M. Cordes, Astrophys. J. **410**, 673 (1993).
19. J. H. Taylor, R. N. Manchester, and A. G. Lyne, Astrophys. J., Suppl. Ser. **88**, 529 (1993).
20. M. A. Walker, Mon. Not. R. Astron. Soc. **294**, 307 (1998).
21. J. H. Taylor and J. M. Cordes, Astrophys. J. **411**, 674 (1993).
22. A. V. Pynzar', Astron. Zh. **70**, 480 (1993) [Astron. Rep. **37**, 245 (1993)].
23. V. S. Artyukh and T. V. Smirnova, Pis'ma Astron. Zh. **15**, 797 (1989) [Sov. Astron. Lett. **15**, 344 (1989)].
24. J. M. Cordes, in *IAU Symposium 110: Interstellar Scattering, VLBI and Compact Radio Sources*, Ed. by R. Fanti *et al.* (D. Reidel, Dordrecht, 1984), p. 303.
25. V. M. Malofeev, V. I. Shishov, W. Sieber, *et al.*, Astron. Astrophys. **308**, 180 (1994).
26. H. Hirabayashi, in *IAU Colloquium 164: Radio Emission from Galactic and Extragalactic Compact Sources*, Ed. by J. A. Zenus, G. B. Taylor, and J. M. Wrobel (Astronomical Society of the Pacific, San Francisco, 1997), Astron. Soc. Pac. Conf. Ser., Vol. 144, p. 11.
27. N. S. Kardashev, Exp. Astron. **7**, 329 (1997).
28. B. J. Rickett, A. Querrenbach, R. Wegner, *et al.*, Astron. Astrophys. **293**, 479 (1995).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 203–211. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 238–247.

Broadening of Electron Cyclotron Maser Emission Lines in a Nonuniform Magnetic Field

K. Yu. Platonov and G. D. Fleishman

St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

Ioffe Physicotechnical Institute, Russian Academy of Sciences, ul. Politekhnicheskaya 26, St. Petersburg, 194021 Russia

Received July 26, 1999

Abstract—The formation of cyclotron maser emission lines in a non-uniform (regular or random) magnetic field is studied. In the presence of sufficiently small inhomogeneity, the line shape can be described by a broadened Gaussian profile. In the case of stronger inhomogeneity, the initial Gaussian profile splits into two Gaussian components, which could be observationally perceived as "harmonics." A relation between the distribution of local magnetic trap sizes and the distribution of the spectral widths of solar radio spikes is derived. Possible applications of the results to the interpretation of solar radio spikes and related problems are discussed. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is currently accepted that electron cyclotron maser emission (CME) is responsible for many types of intense radio emission generated in the magnetospheres of planets and in the atmospheres of the Sun and stars [1–3]. As shown in review [4], the millisecond solar radio spikes that sometimes appear during solar flares are due to CME. Radio spikes represent one type of hyperfine temporal structure in the solar radio emission; they have the shortest durations and narrowest relative spectral widths, as well as the highest brightness temperatures and circular polarizations, among all known kinds of solar electromagnetic radiation (see [4]).

According to the model for radio-spike sources proposed in [4], a cluster of radio spikes (event) is generated by energetic electrons with an anisotropic distribution via the cyclotron maser mechanism acting in a large-scale magnetic trap (arch, loop). Each individual spike forms in its own local trap.

These small-scale traps are due to fluctuations of the magnetic field in a large-scale trap, associated with the primary energy release of a solar flare. Obviously, the properties of the CME generated in such a non-uniform medium differ from those for the case of a uniform medium. Intuitively, it is clear that inhomogeneity will first and foremost affect the spectral properties of the pulses of emission.

The spectral halfwidths of radio spikes have been considered in a number of studies [5–7]. Currently, it is known that the ratio $\Delta f/f$ is about a few per cent, and has a wide scatter for each event. Within a single event, $\Delta f/f$ is not correlated with frequency, and the number of spikes is often distributed in a power law in halfwidth (over a limited range of $\Delta f/f$). Sometimes, a tendency for the spike halfwidth to decrease with growth of spike intensity is observed. The absence of well-defined correlations of the halfwidth with other measurable parameters of radio spikes was interpreted in [6] as an effect of the non-uniform medium in which the spikes are generated.

In this paper, we study the role of both regular and stochastic non-uniformity of the magnetic field in an individual spike source, and derive a relationship between the distributions of radio-spike halfwidths and of microtrap sizes. Our results are consistent with the observations, and the predicted correlations can be verified via appropriate statistical analysis of already available observational data.

2. RADIATION TRANSFER IN A NON-UNIFORM MEDIUM (BASIC EQUATIONS)

Let us consider the propagation of transverse modes [8] in a plasma with either regular or random inhomogeneities of the magnetic field and/or density. In a linear approximation for the field intensity, the equation for the electric field has the standard form

$$\Delta E_i - \nabla_i E_j + \frac{(2\pi f)^2}{c^2}\int d\mathbf{r}'\epsilon_{ij}(\mathbf{r}-\mathbf{r}', \mathbf{r})E_j(\mathbf{r}') = 0, \quad (1)$$

where the dielectric constant $\epsilon_{ij}(\mathbf{r} - \mathbf{r}', \mathbf{r})$ takes into account spatial dispersion (argument $\mathbf{r} - \mathbf{r}'$) and non-uniformity of the medium (argument $\mathbf{r}$). In a randomly non-uniform medium, ϵ is also a random function. Under natural conditions, the inhomogeneity scale usually far exceeds the emission wavelength. This enables us to use an eikonal approximation for (1):

$$E_i(\mathbf{r}) = E_{0i}\exp(i\chi(\mathbf{r})). \quad (2)$$

1063-7729/01/4503-0203$21.00

The eikonal equation corresponding to (1) is [9]

$$\left|(\nabla\chi)^2\delta_{ij} - \nabla_i\chi\nabla_j\chi - \frac{(2\pi f)^2}{c^2}\tilde{\epsilon}_{ij}(\nabla\chi, \mathbf{r})\right| = 0, \quad (3)$$

where $\tilde{\epsilon}$ is the Fourier transform of ϵ_{ij} with respect to $\mathbf{r} - \mathbf{r}'$. We neglect the effects of spatial dispersion, since the thermal velocity of the background plasma electrons is much less than the velocity of light. We choose the direction of propagation of the electromagnetic wave to be along the z axis. Using the solution of (3), the field (2) can be written

$$E_i(\mathbf{r}) = E_{0i}\exp\left[i\int dz\frac{2\pi f}{c}n^\sigma(\mathbf{r})\right], \quad (4)$$

where n^σ is the refractive index of the mode with superscript σ. In the case of maser amplification, n^σ includes an imaginary part, which leads to (in a linear approximation) exponential growth of the field amplitude. Denoting the growth increment as $\kappa(\mathbf{r})$, we obtain the intensity of the emission from a magnetic trap along the z axis

$$I = I_0\int d\mathbf{r}_\perp\exp\left[2\int^z dz\kappa(\mathbf{r})\right], \quad (5)$$

where I_0 is the initial flux density of the intensity. In the exponential growth regime, the highest intensity will be emitted at those frequencies f where $\kappa(f, \mathbf{r})$ has a local maximum. Expanding $\kappa(f, \mathbf{r})$ in a series about the peak frequency f_0 (f_0 is obviously a function of $\mathbf{r}$) and retaining up to second-order terms, we obtain the profile of a radio-spike emission line in the form of a collection (integral of) local Gaussians:

$$I(\mathbf{r}) = I_0\exp\left[2\int^z dz\left(\kappa(f_0, \mathbf{r}) - \frac{\partial^2\kappa(\mathbf{r})}{\partial f_0^2}(f - f_0(\mathbf{r}))^2\right)\right]. \quad (6)$$

For the random functions $\kappa(\mathbf{r})$ and $f_0(\mathbf{r})$, (6) must also be averaged over the inhomogeneities. This procedure will be carried out in Section 5.

Thus, taking into account the inhomogeneities of the density and magnetic field can be reduced to determining the coordinate dependence of quantities in (6), subsequent averaging, and integrating (6) over the magnetic trap volume.

3. APPROXIMATION OF THE WAVE AMPLIFICATION INCREMENTS

The increments (both temporal and spatial) of CME-generated electromagnetic waves have been calculated in many previous studies [10–20]. The corresponding frequency dependences are fairly smooth functions [21], which admit the parabolic approximation

$$\kappa = A(f_B, Y, n_b/n_0)(1 - \alpha(s - s_0)^2), \quad (7)$$

near their maxima, where $s = f/f_B$, $s_0 = f_0/f_B$, and f_B is the gyrofrequency. The factor A does not depend on frequency; however, it is proportional to the gyrofrequency f_B, has an approximately power-law dependence on the ratio of the plasma frequency to the gyrofrequency $Y = \omega_p/\omega_B = f_p/f_B$ (the index of this power law is different for different eigenmodes and cyclotron harmonics), and is proportional to the relative number density of fast electrons n_b/n_0 [18, 19]. The parabolic-approximation parameter α (which determines the intrinsic half-width of the emission pulse in a uniform medium) depends mainly on the form and parameters of the fast-electron distribution function (degree of anisotropy, energy spectrum index, characteristic electron energies, etc.). Note that, in the case of a narrow emission line,

$$\alpha \gg 1. \quad (8)$$

The dependence of the spatial increment κ on the magnetic-field intensity includes the explicit dependences $A(f_B)$ and $s(f_B) = f/f_B$, and can also include implicit dependences if Y changes as the magnetic field in the source changes. We assume variations of the magnetic field B in the source to be small (specific criteria for this smallness will be obtained below), and approximate the dependences $Y(B)$ and $s_0(B)$ as power-law functions with different indices:

$$\kappa = \kappa_0 f_B(1 + \delta B/B)^\lambda \times\left\{1 - \alpha\left[\frac{f}{f_B(1 + \delta B/B)} - s_0(1 + \delta B/B)^\varepsilon\right]^2\right\}, \quad (9)$$

where κ_0, λ, and ε are constants and, if Y = const, then $\lambda = 1$ and $\varepsilon = 0$. B is the mean magnetic field in the source, f_B is the gyrofrequency, $s_0 = f_0/f_B$ is the frequency of the emission maximum in a uniform field B, and $\delta B(\mathbf{r})$ is the variation of the magnetic field in the source.

If the relative magnitude of the magnetic-field fluctuations is small,

$$\delta B/B \ll 1, \quad (10)$$

we can restrict our treatment to an expansion of κ in powers of $\delta B/B$, retaining up to second-order terms:

$$\kappa \approx \kappa_0 f_B\left\{[1 - \alpha(s - s_0)^2] + [\lambda + 2\alpha s_0(1 + \varepsilon)(s - s_0) + 2\alpha(2 - \lambda)(s - s_0)^2]\frac{\delta B}{B} + [\lambda(\lambda - 1)/2 - \alpha s_0^2(1 + \varepsilon)^2\right. \quad (11)$$

$$+\alpha s_0(1+\varepsilon)(1+\varepsilon)(\varepsilon-2(2-\lambda))(s-s_0)$$

$$-\alpha(\lambda-2)(\lambda-3)(s-s_0)^2/2]\left(\frac{\delta B}{B}\right)^2\Bigg\}.$$

Taking into account (6) and (8), $s - s_0 \ll 1$, so that

$$\alpha(s-s_0)^2 \le 1. \tag{12}$$

We can then neglect terms that do not contain α in the coefficient of $(\delta B/B)^2$:

$$\kappa \approx \kappa_0 f_B\{[1-\alpha(s-s_0)^2]$$

$$+[\lambda+2\alpha s_0(1+\varepsilon)(s-s_0)+2\alpha(2-\lambda)(s-s_0)^2]\frac{\delta B}{B} \tag{13}$$

$$-[\alpha s_0^2(1+\varepsilon)^2$$

$$-[\alpha s_0(1+\varepsilon)(\varepsilon-2(2-\lambda))(s-s_0)]\left(\frac{\delta B}{B}\right)^2\Bigg\}.$$

4. LINE BROADENING IN THE PRESENCE OF SMALL, REGULAR INHOMOGENEITY

We first consider the case where the inhomogeneity of the magnetic field can be described by a single sine wave

$$\delta B/B = \delta\sin(k_x x + k_z z), \tag{14}$$

where z is the coordinate along the line of sight, x is one of the coordinates transverse to the line of sight, and the dimensionless quantity δ is small enough that we can restrict our treatment to linear terms in δ in the expansion of κ (7).

According to (5), we have for the emission intensity

$$I = I_0\int d\mathbf{r}_\perp \exp\left[2\int_{-L/2}^{L/2}\kappa(z)dz\right], \tag{15}$$

where L is the source size along the line of sight. Substituting (13) and (14) into (15), we obtain

$$I = I_0\int d\mathbf{r}_\perp \exp\Bigg\{2\kappa_0 f_B\int_{-L/2}^{L/2}dz[1-\alpha(s-s_0)^2$$

$$+\delta\sin(k_x x+k_z z)(\lambda+2\alpha s_0(1+\varepsilon)(s-s_0) \tag{16}$$

$$+2\alpha(2-\lambda)(s-s_0)^2)]\Bigg\}$$

and, after integrating over dz in the exponent, we find

$$I = I_0\int d\mathbf{r}_\perp \exp\Bigg\{2\kappa_0 f_B L\Bigg[1-\alpha(s-s_0)^2$$

$$-\frac{2\alpha(\lambda-2)}{k_z L}\delta\sin(k_z L/2)\sin(k_x x)(s-s_0)^2$$

$$+\frac{4\alpha\delta s_0(\varepsilon+1)}{k_z L}\sin(k_z L/2)\sin(k_x x)(s-s_0) \tag{17}$$

$$+\frac{2\lambda\delta}{k_z L}\sin(k_z L/2)\sin(k_x x)\Bigg]\Bigg\}.$$

After integrating (17) over $d\mathbf{r}_\perp$, the exponent, which determines the emission-line profile, can be found using the saddle-point method [22]. In the exponent, terms that depend on x are proportional to $\sin(k_x x)$. The roots of the exponent derivative satisfy the equation $\sin(k_x x) = \pm 1$. It can readily be shown that the desired saddle point is provided by the positive root of this equation, which determines the asymptotic representation of integral (17) if $\sin(k_x x)$ is equal to unity in the integration region. In this case, it is even easier to find the solution when $k_x L \ll 1$. Then, we can assume that $\sin(k_x x) \approx k_x x$ in (17). Further, the integration is trivial, and yields

$$I = I_0\int d\mathbf{r}_\perp \exp\Bigg\{2\kappa_0 f_B L\Bigg[1-\alpha(s-s_0)^2$$

$$+\frac{k_x L}{2}\left(\frac{2\alpha(2-\lambda)}{k_z L}\delta\sin(k_z L/2)(s-s_0)^2\right. \tag{18}$$

$$\left.+\frac{4\alpha\delta s_0}{k_z L}\sin(k_z L/2)(s-s_0)+\frac{2\delta\lambda}{k_z L}\sin(k_z L/2)\right)\Bigg]\Bigg\}.$$

We assume here that the source sizes along and transverse to the line of sight are the same, and equal to L. If this is not the case, only the absolute value of the emission intensity will change, not the line profile in which we are interested here.

Isolating the perfect square in (18) and explicitly separating the contributions that do and do not depend on frequency, we find

$$I = I_m \exp[-(1-\Delta)(s-s_0-\Delta s_0)^2/2\Gamma_0^2], \tag{19}$$

where

$$I_m = \tilde{I}_0\exp\left[\tau_0\left(1+2\delta\lambda\frac{k_x L}{2}\frac{\sin(k_z L/2)}{k_z L}\right)\right], \tag{20}$$

$$\tau_0 = 2\kappa_0 f_B L \tag{21}$$

is the optical depth of the emission source when $\delta = 0$,

$$\Gamma_0 = \frac{1}{\sqrt{2\alpha\tau_0}} = \frac{1}{2\sqrt{\kappa_0 f_B L\alpha}} \tag{22}$$

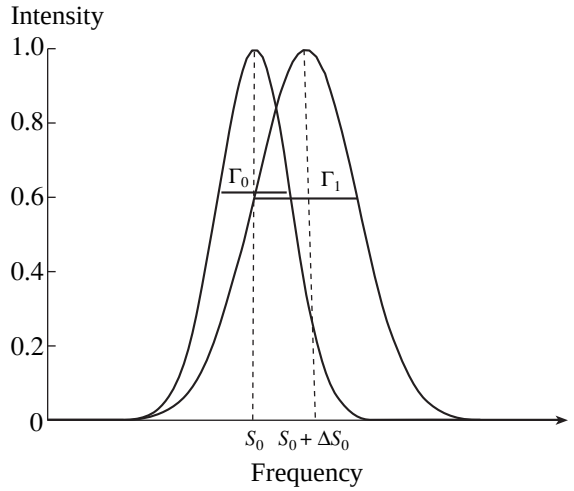


Fig. 1. CME lines in uniform and non-uniform magnetic fields. The linewidths in regularly non-uniform and random fields are determined by formulas (25) and (41), respectively. The intrinsic linewidth is determined by (22).

is the unperturbed halfwidth of the emission pulse, and

$$\Delta s_0 = 2\delta s_0(1+\varepsilon)\frac{k_x L}{2}\frac{\sin(k_z L/2)}{k_z L} \approx \delta s_0(1+\varepsilon)(k_x L) \ll 1 \quad (23)$$

is the shift of the central line frequency due to the inhomogeneity of the magnetic field (the second of equalities (23) is valid if, together with $k_x L \ll 1$, $k_z L \ll 1$). The quantity

$$\Delta = \frac{\delta(2-\lambda)}{k_z L}k_x L\sin(k_z L/2) \approx (2-\lambda)\delta\frac{k_x L}{2} \ll 1 \quad (24)$$

describes CME line broadening due to inhomogeneity of the magnetic field. When $k_x L > 1$, we must replace $k_x L$ with 1 in (23) and (24). If we introduce the Gaussian halfwidth Γ of the broadened line, it can be expressed in terms of Γ_0 as

$$\Gamma = \Gamma_0(1+\Delta/2). \quad (25)$$

Note the following results of the calculations (Fig. 1).

(1) In the case of small field inhomogeneity, the line retains a Gaussian profile.

(2) The shift of the central pulse frequency is proportional to the magnitude of the field inhomogeneity.

(3) The change in the line halfwidth is small, and is described by (25).

Note that, for estimates of Γ in a non-uniform medium in the literature, it is usually assumed that $\Gamma \sim \delta$, which would be valid in the absence of the maser effect. In the presence of the maser effect, the quantity that is of the order of δ is not the pulse halfwidth Γ, but its variation $\Gamma - \Gamma_0 \sim \delta$.

5. LINE BROADENING IN A RANDOM FIELD

Now let the quantity $\delta B/B$ be a random function with zero mean. In this case, the emission intensity is represented as the ensemble mean of (15):

$$I = I_0\left\langle \int d\mathbf{r}_\perp \exp\left(2\int_{-L/2}^{L/2}\kappa(z)dz\right)\right\rangle. \quad (26)$$

As in Section 4, we restrict our treatment to linear terms in $\delta B/B$ in the expansion of $\kappa(\mathbf{r})$ (13):

$$I = I_0\int d\mathbf{r}_\perp\left\langle \exp\left\{2\kappa_0 f_B L[1-\alpha(s-s_0)]^2 + 2\kappa_0 f_B[\lambda+\alpha(2-\lambda)(s-s_0)^2 + 2\alpha(s-s_0)s_0(1+\varepsilon)]\int_{-L/2}^{L/2}dz\frac{\delta B(\mathbf{r})}{B}\right\}\right\rangle. \quad (27)$$

Assuming the random process is Gaussian, we use the relation

$$\langle \exp(AX)\rangle = \exp(A^2\langle X^2\rangle/2). \quad (28)$$

Then, the emission intensity is

$$I = I_0\int d\mathbf{r}_\perp \exp\{2\kappa_0 f_B L[1-\alpha(s-s_0)^2]\} \times \exp\left\{2\kappa_0^2 f_B^2[\lambda+\alpha(2-\lambda)(s-s_0)^2 + 2\alpha(s-s_0)s_0(1+\varepsilon)]^2 \times \int_{-L/2}^{L/2}dz\int_{-L/2}^{L/2}dz'\frac{\langle\delta B(\mathbf{r})\delta B(\mathbf{r}')\rangle}{B^2}\right\}. \quad (29)$$

Thus, the CME intensity is expressed as a double integral over the coordinates of the pair correlator of the amplitudes of the random magnetic field $\langle\delta B(\mathbf{r})\delta B(\mathbf{r}')\rangle$. To calculate these integrals, we assume that the magnetic turbulence is statistically uniform and isotropic (i.e., the value of the correlator depends only on the distance between points):

$$\langle\delta B(\mathbf{r})\delta B(\mathbf{r}')\rangle = \langle\Delta B^2\rangle T(|\mathbf{r}-\mathbf{r}'|), \quad (30)$$

where $\langle\delta B^2\rangle$ is the rms value of the random magnetic field and $T(x)$ is the dimensionless pair correlation

function of the magnetic field. The integral of the correlation function is

$$\int_0^\infty T(x)dx = L_c, \tag{31}$$

where L_c is, by definition, the correlation length of the random field. If we make a change of variable from z and z' to their difference and half-sum when integrating (29), the quantity

$$\int_0^L T(x)dx \approx L_c \tag{32}$$

will enter the exponent if the correlation length L_c is less than the source size L. Calculation of the second integral yields simply the factor L, so that

$$I = I_1 \exp\{2\kappa_0 f_B L[1-\alpha(s-s_0)^2]\}$$
$$\times \exp\left\{2\kappa_0^2 f_B^2[\lambda+\alpha(2-\lambda)(s-s_0)^2 \right. \tag{33}$$
$$\left. + 2\alpha(s-s_0)s_0(1+\varepsilon)]^2 L L_c \frac{\langle \Delta B^2\rangle}{B^2}\right\},$$

where

$$I_1 = I_0 S, \tag{34}$$

and S is the source cross section. Let us expand the square in the last exponent, retaining terms up to $(s-s_0)^2$. Higher-order terms describe asymmetry of the line wings, and can readily be derived from (33) if required for the analysis of observed line profiles. Introducing the optical depth of a uniform source τ_0 (21) and neglecting small terms (keeping in mind that $\alpha \gg 1$), we find

$$I = I_m \exp\left\{-\tau_0\alpha\left[\left(1-2(1+\varepsilon)^2 s_0^2 \alpha\tau_0 \frac{\langle \Delta B^2\rangle}{B^2}\frac{L_c}{L}\right)\right.\right. \tag{35}$$
$$\left.\left.\times (s-s_0)^2 - 2\lambda(1+\varepsilon)s_0\tau_0\frac{\langle \Delta B^2\rangle}{B^2}\frac{L_c}{L}(s-s_0)\right]\right\},$$

$$I_m = I_1 \exp\left[\tau_0\left(1+\lambda^2\tau_0\frac{\langle \Delta B^2\rangle}{B^2}\frac{L_c}{L}\right)\right]. \tag{36}$$

As in the previous section, isolating the perfect square in the exponent, we finally obtain

$$I = I_m \exp[-(1-\Delta_{st})(s-s_0-\Delta s_0)^2/2\Gamma_0^2], \tag{37}$$

where, as before, Γ_0 is the intrinsic linewidth (22),

$$\Delta s_0 = \frac{\lambda(1+\varepsilon)s_0\tau_0\frac{\langle \Delta B^2\rangle}{B^2}\frac{L_c}{L}}{1-\Delta_{st}}, \tag{38}$$

$$\Delta_{st} = 2(1+\varepsilon)^2 s_0^2 \alpha\tau_0 \frac{\langle \Delta B^2\rangle}{B^2}\frac{L_c}{L}. \tag{39}$$

We have derived the expression for I (37) under the assumption that $\delta B/B \ll 1$. However, the value of Δ_{st} may not be small compared to unity, since $\alpha \gg 1$ and $\tau_0 \gg 1$. Nevertheless, if

$$(1-\Delta_{st}) > 0, \quad \Delta s_0 \ll s_0, \tag{40}$$

the amplified line still has a Gaussian profile (Fig. 1), with halfwidth

$$\Gamma = \frac{\Gamma_0}{\sqrt{1-\Delta_{st}}} \approx \Gamma_0(1+\Delta_{st}/2). \tag{41}$$

If Δ_{st} is sufficiently close to unity, so that $|\Delta s_0| \geq s_0$, the line profile will be very different from the initial Gaussian (for large deviations of s from s_0, the parabolic approximation for the increment κ (13) may not be valid). In particular, when $(1-\Delta_{st}) = 0$, quadratic terms of the form $(s-s_0)^2$ are completely absent from the exponent. Finally, if

$$(1-\Delta_{st}) < 0, \quad \Delta s_0 \ll s_0, \tag{42}$$

this analysis is valid near s_0, but the coefficient of $(s-s_0)^2$ in the exponent is now positive, so that we have a local minimum in the emission at $s \approx s_0$, not a maximum. This means that, in a medium with sufficiently strong inhomogeneities, the initial Gaussian profile is divided into two (or more) profiles; i.e., the CME line splits. To describe this effect quantitatively, we need more information about the statistical properties of the random field than is contained in the pair correlator (30). Nevertheless, we can obtain a quantitative description of line splitting in a regular field if we include second-order terms in $\delta B/B$ in (13).

6. CME LINE SPLITTING IN A REGULARLY NON-UNIFORM MAGNETIC FIELD

In the previous section, we showed that the change in the CME line profile in the presence of sufficiently strong (though small compared to unity) fluctuations of the magnetic field can be quite considerable. In particular, a single Gaussian line can be split into several lines. Let us consider this effect quantitatively in the case of regular inhomogeneity of the magnetic field, described by (14). We will take into account quadratic (in δ) corrections to the expansion of κ (13). The CME

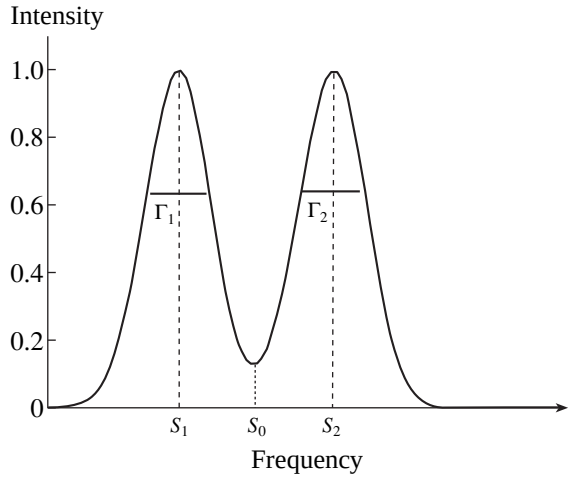


Fig. 2. Splitting of a CME line in a sufficiently strong, regularly non-uniform magnetic field.

intensity after integration over dz in the exponent is [analogous to (17)]

$$I = I_0 \exp\left\{2\kappa_0 f_B L\left[1-\alpha(s-s_0)^2 - \frac{\delta^2}{2}\left(1-\frac{\sin(k_z L)}{k_z L}\right)(\alpha s_0^2(1+\varepsilon)^2 + 2\alpha(2-\lambda)s_0(1+\varepsilon)(s-s_0))\right]\right\}$$

$$\times \int d\mathbf{r}_\perp \exp\left\{2\kappa_0 f_B L\left[-\frac{\delta^2}{2}\frac{\sin(k_z L)}{k_z L}\sin^2(k_x x) \times (\alpha s_0^2(1+\varepsilon)^2 + 2\alpha(2-\lambda)s_0(1+\varepsilon)(s-s_0)) + \delta\frac{\sin(k_z L/2)}{k_z L/2}\sin(k_x x) \times (\lambda + 2\alpha s_0(1+\varepsilon)(s-s_0) + \alpha(2-\lambda)(s-s_0)^2)\right]\right\}. \quad (43)$$

It is easy to see that, when

$$\alpha\delta^2 \geq 1 \quad (44)$$

the term proportional to δ^2 is not small compared to the term that is linear in δ. Note that, since $\alpha \gg 1$, condition (36) is consistent with the condition that δ be small.

To find the resulting line profile, we must correctly calculate the exponent after integration over $d\mathbf{r}_\perp$. This can be done using the saddle-point method [22]. The x-dependent terms in the exponent are

$$A\sin^2(k_x x) + B\sin(k_x x). \quad (45)$$

To find the saddle points, we must set the derivative of this quantity equal to zero. The roots of (45) are determined by the equations

$$\sin(k_x x) = -B/2A, \quad (46)$$

$$\sin(k_x x) = \pm 1. \quad (47)$$

It can readily be shown that (46) corresponds to a minimum of the exponent curve and (47) to a maximum. Therefore, the result of integrating over dr is the sum of two contributions corresponding to these roots:

$$I_{1,2} = I_{m1,2}\exp\{-2\kappa_0 f_B L\alpha[1\pm\delta(2-\lambda)] \times [s-s_0\pm(1+\varepsilon)\delta s_0]^2\}. \quad (48)$$

When deriving (48), we have omitted some small terms, set $\sin(k_z L)/k_z L = 1$ [this does not affect the functional form of (48)], and included terms that do not depend on s in $I_{m1,2}$. Introducing the central frequencies of the doublet components

$$s_{1,2} = s_0[1\mp(1+\varepsilon)\delta] \quad (49)$$

and the corresponding halfwidths

$$\Gamma_{1,2} = \frac{1}{2\sqrt{\kappa_0 f_B L\alpha(1\pm\delta(2-\lambda))}} \approx \Gamma_0(1\mp\delta(2-\lambda)/2), \quad (50)$$

where Γ_0 is the unperturbed halfwidth of the spectral line (22), we obtain for the doublet components

$$I_{1,2} = I_{m1,2}\exp[-(s-s_{1,2})^2/2\Gamma_{1,2}^2]. \quad (51)$$

In the model considered, the absolute value of the intensity I_{m2} of the high-frequency component of the doublet is greater than I_{m1}, so that $I_{m2}/I_{m1} = \exp(2\tau_0\delta\lambda)$. However, we emphasize that this result is very model-dependent: even small changes in the functional relation $\mathbf{B}(\mathbf{r})$ can appreciably affect the intensity ratio, since they enter the exponent of (43). In spite of the fact that the total broadening of the CME line $\Delta s_+ = 2(1+\varepsilon)s_0\delta \gg \Gamma_0$, the resulting emission is not described by a single Gaussian with width Δs_+, but is instead divided into two profiles with widths $\Gamma_{1,2}$ that are close to Γ_0 and central frequencies separated by Δs_+ (Fig. 2). Note that, in this case, the low-frequency component ($s \approx s_1$) has a smaller spectral halfwidth ($\Gamma_1 < \Gamma_2$).

7. WIDTH DISTRIBUTION AND DEPENDENCE ON PHYSICAL PARAMETERS

According to the model for radio-spike generation proposed in [4], the radio spikes that make up a cluster in a dynamic spectrum form in local magnetic traps

arising in a large-scale (global) magnetic loop due to fluctuations of the magnetic field. In this model, there should be a connection between the spike width distribution and the size distribution for the local traps.

Let us first consider the case where line broadening due to field inhomogeneity is small, so that the spike halfwidth Γ is related to the source size L as

$$\Gamma = \frac{1}{2\sqrt{\kappa_0 f_B L \alpha}} = \frac{1}{\sqrt{2\alpha\tau}}. \quad (52)$$

Since there exists a simple linear relationship between the optical depth τ and source size L, we will analyze dependences on τ rather than on L.

Let the distribution of optical depths τ for the local sources be given by function φ, such that

$$\int_{\tau_0}^{\tau_0+\Delta\tau} \varphi(\tau)d\tau = N_s(\Delta\tau), \quad (53)$$

where $N_s(\Delta\tau)$ is the number of local sources with optical depths in the interval from τ_0 to $\tau_0 + \Delta\tau$ and, accordingly,

$$\int_0^\infty \varphi(\tau)d\tau = N_s, \quad (54)$$

where N_s is the total number of local sources.

Let the distribution of the widths Γ for the radio spikes be described by the distribution function $\psi(\Gamma)$:

$$\int_0^\infty \psi(\Gamma)d\Gamma = N_r, \quad (55)$$

where N_r is the total number of radio spikes in a cluster.

It is natural to assume that each local source emits, on average, the same number of spikes; i.e.,

$$N_r = aN_s, \quad a \geq 1. \quad (56)$$

For the sake of definiteness, we assume that $a = 1$ (this does not affect the functional relation); i.e., each local source emits precisely one spike ($N_r = N_s$). Then,

$$\varphi(\tau) = \psi(\Gamma(\tau))\left|\frac{d\Gamma}{d\tau}\right| \propto \tau^{-3/2}\psi(\Gamma(\tau)) = \tau^{-3/2}\psi\left(\frac{1}{\sqrt{2\alpha\tau}}\right). \quad (57)$$

Accordingly, we have for the distribution of source sizes L

$$\varphi(L) \propto L^{-3/2}\psi\left(\frac{1}{2\sqrt{\alpha\kappa_0 f_B L}}\right). \quad (58)$$

If the function $\psi(\Gamma)$ is known from observations, (58) can be used to determine the size distribution for the local sources. For example, if we have in some interval of Γ

$$\psi(\Gamma) \propto \Gamma^{-\nu}, \quad (59)$$

then

$$\varphi(L) \propto L^{\frac{\nu-3}{2}}. \quad (60)$$

We now take into account line broadening in a random magnetic field (41):

$$\Gamma = \frac{1}{\sqrt{2\alpha\tau(1-\beta\tau)}}, \quad (61)$$

where

$$\beta = 2(1+\varepsilon)^2 s_0 \alpha \frac{\langle \Delta B^2 \rangle}{B^2}\frac{L_c}{L}, \quad (62)$$

and we adopt β = const. Then,

$$\left|\frac{d\Gamma}{d\tau}\right| = \left|\frac{\alpha(1-2\beta\tau)}{[2\alpha\tau(1-\beta\tau)]^{3/2}}\right| \quad (63)$$

and

$$\varphi(\tau) \propto \frac{|(1-2\beta\tau)|}{[2\alpha\tau(1-\beta\tau)]^{3/2}}\psi(\Gamma(\tau)). \quad (64)$$

When $\beta\tau \ll 1$, we have our previous result, and, when $\tau > 1/2\beta$, the spectral width of the pulse increases with τ, rather than decreasing, as earlier [see (52)]. In this case, the spectral width Γ as a function of τ reaches the minimum value

$$\Gamma_{\min} = 1/\sqrt{\alpha\tau} = \sqrt{2\beta/\alpha} \quad (65)$$

when

$$\tau = 1/2\beta. \quad (66)$$

We cannot use (64) for such τ values, since the representation of the discrete quantities N_r and N_s by the continuous functions ψ and φ is no longer correct. Formula (64) is applicable when $\tau < 1/\beta$, since the line profile is strongly modified at large values of τ (see Section 5). Together with the spectral halfwidth of the pulse, the optical depth is also determined by the pulse intensity [in accordance with (34)]. However, in general, the $\Gamma(\tau)$ dependence is non-monotonic: Γ can either increase or decrease as the emission intensity grows, and weak correlation (or absence of a correlation) is also possible.

We have for small inhomogeneity of the magnetic field

$$I \propto e^\tau, \quad \Gamma \propto \tau^{-1/2}, \quad (67)$$

therefore,

$$\Gamma \propto (\ln I)^{-1/2}. \quad (68)$$

The values of Γ can show some scatter about this correlation dependence due to variations of α from one

local source to another. In general, when τ is related to Γ by (61), we have the more complicated dependence

$$\Gamma = \frac{1}{\sqrt{2\alpha\ln(I/I_0)(1-\beta\ln(I/I_0))}}. \quad (69)$$

Note that we should now expect a wider scatter of Γ values, since both α and β can vary from one local source to another. In spite of the fact that $\Gamma(I)$ has a minimum, we cannot expect a correlation with minima in the analysis of any individual event, since the range for the variation of τ in each event is small (the flux varies by a factor of a few). We now make a transformation from the dimensionless quantities s and Γ to the dimensional ones f and γ. Then, expression (69) for Γ becomes

$$\gamma \propto \sqrt{\frac{f_B}{\alpha(1-Af_B)}}; \quad (70)$$

i.e., in the presence of small field inhomogeneities, the dimensional pulse halfwidth is proportional to the square root of the magnetic-field intensity.

8. APPLICATION TO ANALYSIS OF SPECTRAL DATA ON SOLAR RADIO SPIKES

The spectral properties of radio spikes detected at decimeter and centimeter wavelengths have been analyzed in a number of studies [5–7, 23, 24]. The considerable scatter of the spike spectral widths and lack of any clear dependence of Γ on other measurable parameters of radio spikes were interpreted in [6] as a reflection of the non-uniformity of the source medium. In their review [4], Fleishman and Mel'nikov note that, in general, the width of an observed spike consists of three different contributions:

(1) The intrinsic linewidth.

(2) Broadening due to non-uniformity of the source parameters.

(3) Broadening due to line-of-sight scattering of the waves between the source and observer.

Note that, in the case of CME (in contrast to, e.g., atomic line emission), the intrinsic width is not "fundamental"; on the contrary, it depends on many parameters (the characteristic energies of the emitting electrons, their energy and angular distributions, the source size, etc.). Therefore, the intrinsic width of a CME line can vary rather strongly from event to event, as well as from spike to spike within a single event. In our analysis, the dependence on the fast-electron distribution is described by the parameter α. We have not considered the role of the third factor (line-of-sight scattering). The relationships obtained for the CME linewidths depend on the parameters of the fast-electron distribution (through α), the type and magnitude of the magnetic-field fluctuations, the gyrofrequency, and the source size (through the optical depth τ).

Let us consider the application of our theoretical results to the analysis of observational data on radio spikes. Csillaghy and Benz [6] found no significant correlation between the dimensional halfwidth γ and the central frequency f_0 of the spikes for any individual event. We can understand this negative result, since our theory predicts a weaker than linear dependence between these quantities:

$$\gamma \propto f_0^{1/2} \quad (f_0 \propto f_B). \quad (71)$$

Taking into consideration the limited variation of f_0 within an individual event (values of $f_{max}/f_{min} < 1.4$ are typical), it becomes clear that the considerable scatter of the "intrinsic" CME linewidths could easily mask the correlation (71). When analyzing observational data, we should seek a dependence similar to (71), instead of a linear correlation between γ and f_0.

Csillaghy and Benz [6] also investigated the correlation between the spike widths and radio fluxes. In four of the eight events studied, there is a significant correlation between Γ and I; Γ decreases with growing flux I in three cases and increases with growing I in one case. In the previous section, we showed that the $\Gamma(I)$ dependence is not monotonic. Therefore, our theory admits both decreases and increases of Γ with growing I. In the case of small broadening, the most intense CME lines are the narrowest, so that the strongest correlation should be observed between Γ and $(\ln I)^{-1/2}$; this can be readily verified using available observations of radio spikes.

Karlicky *et al.* [7] obtained dependences of the power spectra of spike radio emission on $(2\pi/\Delta f)$ for two events. These dependences could be fit with power laws, with the spectral index in each case close to 5/3. This was interpreted as evidence for a relation between radio-spike generation and MHD turbulence. The theory we have developed here establishes a connection between the distributions of radio-spike widths and of spike source (local magnetic-trap) sizes. This latter distribution cannot in general be represented in terms of a turbulence spectrum (pair correlator), and this requires additional statistical information on the properties of the random magnetic field. Therefore, the spectral index of 5/3 obtained in [7] should not be considered special from a theoretical point of view. On the contrary, our theory admits a variety of spectral indices, as well as a variety of forms for this distribution.

In Section 6, we described the effect of CME line splitting in a non-uniform field. In observations, the components of this doublet would be manifest as "harmonics" with a frequency ratio close to unity (in any case, smaller than 2):

$$1 < f_2/f_1 < 2. \quad (72)$$

It is clear that these "harmonics" will have similar intensities and degrees of circular polarization (with the same sign). The model developed in Section 6 predicts a rather simple relation between the central frequencies

and spectral widths of the doublet components. Indeed, using (49),

$$\frac{s_2 - s_1}{s_2 + s_1} = (1 + \varepsilon)\delta, \tag{73}$$

and we find from (50)

$$\frac{\Gamma_2 - \Gamma_1}{\Gamma_2 + \Gamma_1} = \delta(2 - \lambda)/2; \tag{74}$$

therefore,

$$\frac{s_2 - s_1}{s_2 + s_1} = \frac{2(1 + \varepsilon)}{2 - \lambda}\frac{\Gamma_2 - \Gamma_1}{\Gamma_2 + \Gamma_1}. \tag{75}$$

The harmonic ratios of radio spikes were analyzed in detail in [24], where it was shown that the observed harmonic ratios $f_2 : f_1$ are not described by ratios of small integers, and, on the contrary, fill the interval of values

$$1.06 < f_2 : f_1 < 1.5. \tag{76}$$

For example, in the event of April 6, 1980, $f_2 : f_1$ is 1.06. If this corresponds to the splitting of the CME line considered above, this implies that

$$\delta B/B \sim 3 \times 10^{-2}. \tag{77}$$

Widths Γ_1 and Γ_2 should be close to the intrinsic width Γ_0, and their difference should be of order a few per cent, in accordance with (74). Although it is quite tempting to interpret the observed distribution of non-integer harmonic ratios in clusters of spikes as an effect of splitting of the CME lines in non-uniform magnetic fields, it will be possible to draw firm conclusions only after a detailed analysis of the observational data for events with harmonic structure, since non-integer harmonic ratios could also arise from the generation of CME in a uniform source [21].

9. CONCLUSION

We have considered effects accompanying the generation of CME in a non-uniform medium. Typical CME lines can be described by Gaussian profiles. If inhomogeneities (both regular and random) in the magnetic field are not too strong, the Gaussian line shape is retained, but the linewidth Γ increases. As the field inhomogeneity grows further, the line is split into two Gaussian components that form a doublet. We have discussed the relevance of our theoretical results for solar radio spikes, as well as directions for future analysis of the observational data.

ACKNOWLEDGMENTS

We are grateful to A. Benz for discussion of the formulation of the problem. This work was partially supported by the Russian Foundation for Basic Research (project codes 97-02-16972 and 98-02-17711).

REFERENCES

1. C. S. Wu, Space Sci. Rev. **41**, 215 (1985).
2. B. P. Ryabov and N. N. Gerasimova, *Decameter Sporadic Radio Emission of Jupiter* [in Russian] (Naukova Dumka, Kiev, 1990).
3. T. S. Bastian, J. Bookbinder, G. A. Dulk, and M. Davis, Astrophys. J. **353**, 265 (1990).
4. G. D. Fleishman and V. F. Mel'nikov, Usp. Fiz. Nauk **168**, 1265 (1998) [Phys. Usp. **41**, 1157 (1998)].
5. M. Stahli and A. Magun, Sol. Phys. **104**, 117 (1986).
6. A. Csillaghy and A. O. Benz, Astron. Astrophys. **274**, 487 (1993).
7. M. Karlicky, M. Sobotka, and K. Jiricka, Sol. Phys. **168**, 375 (1996).
8. *Plasma Electrodynamics*, Ed. by A. I. Akhiezer *et al.* (Nauka, Moscow, 1974; Pergamon, Oxford, 1975).
9. A. F. Aleksandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Principles of Plasma Electrocdynamics* (Vysshaya Shkola, Moscow, 1978; Springer-Verlag, Berlin, 1984).
10. D. B. Melrose and G. A. Dulk, Astrophys. J. **259**, 844 (1982).
11. R. R. Sharma and L. Vlahos, Astrophys. J. **280**, 405 (1984).
12. R. R. Winglee and G. A. Dulk, Sol. Phys. **104**, 93 (1986).
13. H. W. Li, Sol. Phys. **104**, 131 (1986).
14. P. Louarn, D. Le Queau, and A. Roux, Sol. Phys. **111**, 201 (1987).
15. D. B. Melrose, R. G. Hewitt, and G. A. Dulk, J. Geophys. Res. **89**, 897 (1984).
16. M. J. Aschwanden and A. O. Benz, Astrophys. J. **332**, 447 (1988).
17. M. J. Aschwanden, Astron. Astrophys., Suppl. Ser. **85**, 1141 (1990).
18. G. D. Fleishman and S. G. Yastrebov, Astron. Zh. **71**, 531 (1994) [Astron. Rep. **38**, 468 (1994)].
19. G. D. Fleishman and S. G. Yastrebov, Sol. Phys. **153**, 389 (1994).
20. V. G. Ledenev, Sol. Phys. **179**, 405 (1998).
21. G. D. Fleishman and S. G. Yastrebov, Sol. Phys. **154**, 361 (1994).
22. M. I. Fedoryuk, *The Saddle-Point Method* [in Russian] (Nauka, Moscow, 1977).
23. M. Gudel, Astron. Astrophys. **239**, L1 (1990).
24. S. Krucker and A. O. Benz, Astron. Astrophys. **285**, 1038 (1994).

Translated by G. Rudnitskiĭ

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 212–215. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 248–252.

Studies of the Evolutionary Relationship between Detached and Contact Binaries

G. N. Dremova and M. A. Svechnikov

Ural State University, pr. Lenina 51, Yekaterinburg, 620083 Russia

Received October 23, 1999

Abstract—We estimate the extent to which there is an evolutionary relationship between detached main-sequence binaries and ~KW, KW, and KP contact binaries in the first mass-exchange phase. The current and initial distributions of close binaries of these types are calculated per unit volume of space in the vicinity of the Sun and used to demonstrate evolutionary transitions from low-mass, short-period, detached systems to contact binaries. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In the current paper, we estimate the degree to which there is an evolutionary relationship between ~KW, KW, and KP close binary systems on the main sequence and in the first phase of mass exchange. The following abbreviations are adopted according to the classification introduced by Svechnikov *et al.* [1]:

DMS—detached main-sequence systems—close binaries in which both components are on the main sequence.

KW—W UMa contact systems—close binaries in which both components are comparatively close to filling their inner critical surfaces, and the primary components have periods $\leq 0^{d}.5$ and spectral types Sp_1 later then ≈F0.

~KW—KW-like systems—low-mass close binaries of spectral types F–K that are not in contact, but are closer to each other than DMS systems with the same mass and have many parameters similar to those of KW systems.

KP—early spectral-type contact binaries—close binaries in which the sizes of both components are close to those of the corresponding inner critical surfaces, with periods exceeding $0^{d}.5$ and Sp_1 no later then ≈F0.

Our study is based on the "Catalog of Approximate Photometric and Absolute Elements of Eclipsing Variable Stars" of Svechnikov and Kuznetsova [2]. The catalog contains data for 437 DMS, 153 ~KW, 215 KW, and 392 KP systems whose spectra are known or estimated from the mass–spectrum relation. The spectra of the primary components are known for 273 DMS, 67 ~KW, 66 KW, and 121 KP systems.

In this paper, we study the current and initial distributions of these systems per unit volume of space in the vicinity of the Sun. This makes it possible to determine qualitatively how closely areas of these distributions overlap, and to describe the DMS ⇒ ~KW ⇒ KW and KP ⇒ KW evolutionary transitions and the conditions under which they occur.

2. CURRENT DISTRIBUTIONS OF DMS, ~KW, KW, AND KP CLOSE BINARIES

Figure 1 presents the current distributions of DMS, ~KW, KW, and KP systems—the spatial density of these systems per pc^3 in the vicinity of the Sun as a function of primary mass and the rotation period of the system. The spatial density is derived from the observed close-binary distribution corrected for the geometric and photometric probability of detecting these types of close binaries as eclipsing variables and for the volume occupied by these systems. For DMS and ~KW systems, we derived the total detection probability $W(M_1, A, q, i)$ as a function of the primary mass M_1, the semimajor axis of the system A, the mass ratio q, and the orbital inclination i; for KW and KP systems, A is unambiguously determined by M_1 and q, so that the probability of their detection as eclipsing variables was calculated as a function of the three parameters M_1, q, and i. We calculated the detection probability $W^*(M_1, A, q, i)$ for each system independently via sequential linear interpolation between known table probabilities from [3, 4] for fixed M_1, A, q, and i and specific values of the primary mass, semimajor axis, mass ratio, and orbital inclination.

Estimating the spatial density of the systems is related to calculating the individual volumes occupied by stars of a given type per pc^3 in the vicinity of the Sun. For this purpose, we calculated the distance r^i from the observer to each (ith) system using the relation $M_b^i = m_v^i + \Delta m_b^i + 5 - 5\log r^i - \bar{A} r^i$, where $M_b^i = M_{b1}^i - \Delta m'$ is the absolute bolometric magnitude of the system at maximum brightness, and $\Delta m' = 0^{m}.48$ is the

1063-7729/01/4503-0212$21.00 © 2001 MAIK "Nauka/Interperiodica"

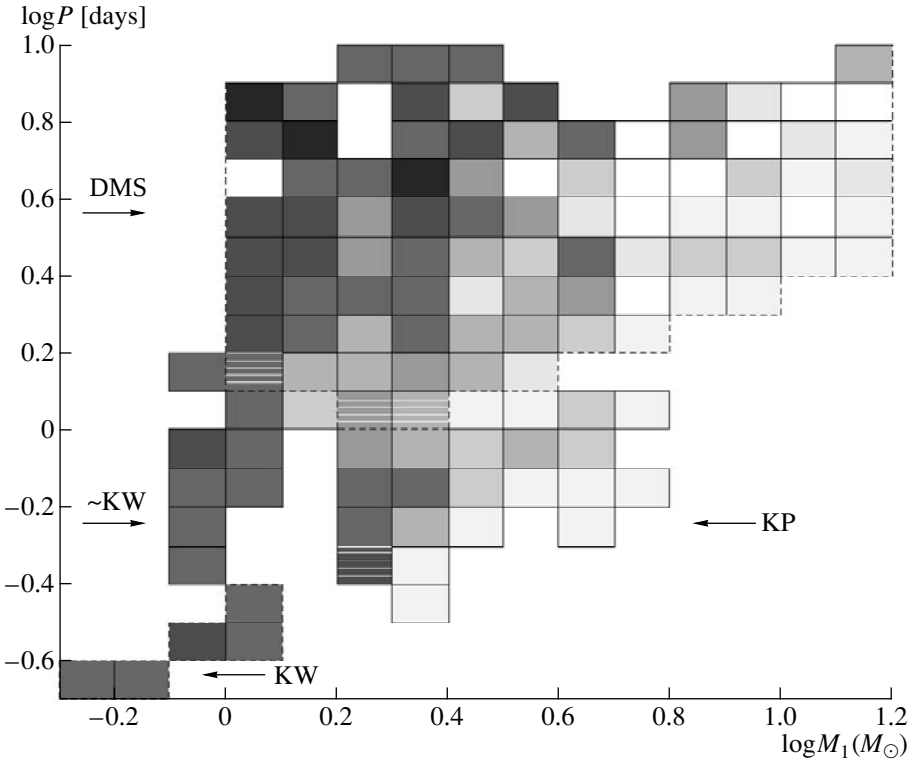


Fig. 1. Map of the current distributions of DMS, ~KW, KW and KP systems in the $\log M_1 - \log P$ diagram. Shaded areas correspond to regions of overlap for the distributions for different types of systems.

correction for the translation from the absolute bolometric magnitude of the primary M_{b1}^i to that of the system [5]. Due to the sharp peak in the q distribution for DMS systems and the small scatter in q [6], the corrections $\Delta m'$ for different DMS systems differ very little, and their mean value is quite suitable for statistical studies. $\bar{A} = 0.0019$ pc^{-1} is the mean interstellar absorption [7], m_v^i the apparent magnitude of the system at maximum brightness taken from the General Catalogue of Variable Stars (GCVS), and Δm_b^i the bolometric correction [8]. For each system, we estimated the volume V^i to be that of a spherical layer with outer radius r^i and thickness $h = 180$ pc [9].

The effects of observational selection for faint systems must be taken into account; in other words, a sample incompleteness factor f^i must be introduced. As m_v^i increases, the number of close binaries of a given type should increase in proportion to the volume of the Galactic layer containing the systems. However, in reality, due to observational selection effects, the number of detected systems increases more slowly than the layer volume. It was found for DMS systems that m_v^i $(f^i = 1) \leq 10\overset{\text{m}}{.}5$ [10]. This is also valid for ~KW systems, since they are detached binaries. We calculated individual sample incompleteness factors for KW and KP systems, which correspond to variations of the number of a given type of system per unit primary mass interval per unit volume of space as a function of the limiting apparent magnitude $m_{v\,\text{lim}}$. These graphs showed that m_v^i $(f^i = 1) \leq 11\overset{\text{m}}{.}5$ for KW and KP systems.

The density distributions of close binaries of the types considered can be graphically presented as "maps" of the current and initial distributions in a $\log M_1 - \log P$ diagram. These maps display non-monotonic behavior in the M_1 and P distributions for DMS systems, as well as individual peaks and valleys. Such non-monotonic behavior has been studied by numerous authors using various observational data [6, 11, 17–20], and we confirm its existence here.

As the period increases from $1\overset{\text{d}}{.}27$ to 10^{d}, there is a pronounced increase in the density distribution for DMS systems with primary masses in the intervals $1–1.6M_\odot$ and $2–2.5M_\odot$ (Fig. 1). We can clearly see maxima in the distribution at ($M_1 = 1.1\text{M}_\odot$, $P = 4\overset{\text{d}}{.}5$), ($M_1 = 1.1M_\odot$, $P = 7^{\text{d}}$), ($M_1 = 1.4M_\odot$, $P = 5\overset{\text{d}}{.}6$), ($M_1 = 1.4M_\odot$, $P = 10^{\text{d}}$), and ($M_1 = 2.2M_\odot$, $P = 4\overset{\text{d}}{.}5$). The diagram indi-

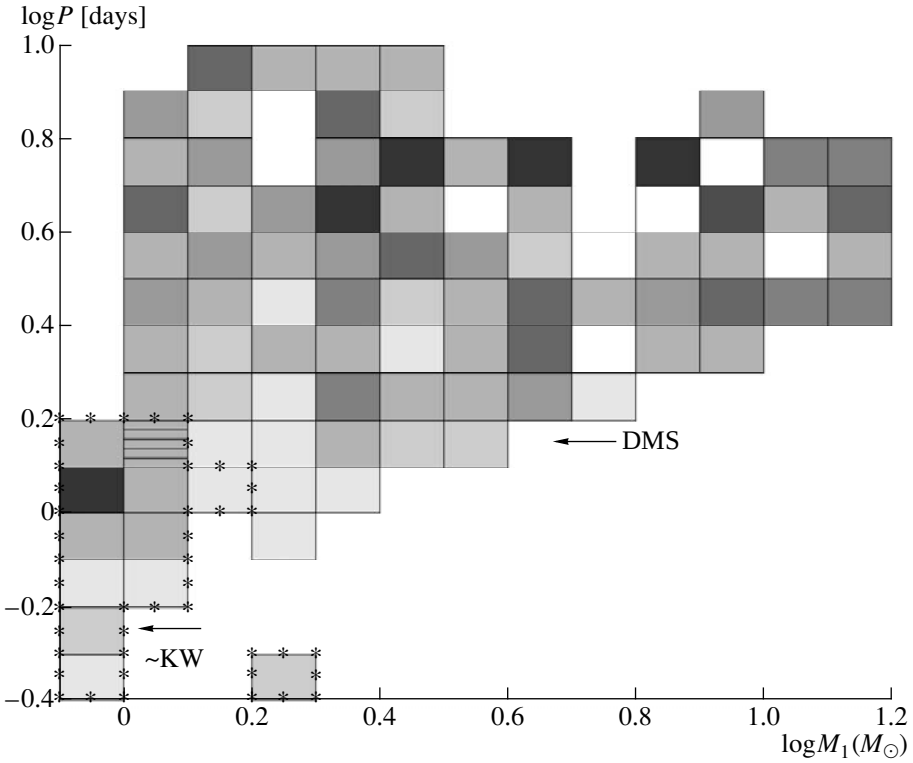


Fig. 2. Same as Fig. 1 for the initial distributions.

cates a decrease in the distribution for $M_1 \geq 5M_\odot$ and $P \geq 1^d.5$.

Dips in the current distribution for DMS systems are also present in the areas ($M_1 = 1.8M_\odot$, $P = 5^d$–8^d), ($M_1 = 3.5M_\odot, P = 4^d.5$ and $P = 10^d$), ($M_1 = 5.6M_\odot, P \geq 3^d$), ($M_1 \geq 3M_\odot$, $P = 10^d$). A dip at $M_1 \leq 1.6M_\odot$, $P \leq 1^d.27$ is quite characteristic and expected, corresponding to the maximum of the current distribution for ~KW systems with masses from 0.8 to $1M_\odot$ and periods from 1^d to $1^d.27$. The map for the current distribution of KW systems appears to continue the distribution for ~KW systems (Fig. 1); we can see a maximum for masses from 0.8 to $1M_\odot$ and periods of $0^d.27$, and a constant level for masses from 0.8 to $0.5M_\odot$ and periods of $0^d.25$–$0^d.2$.

The map for the current distribution of KP systems (Fig. 1) for masses from 1.5 to $6M_\odot$ and periods of $\leq 1^d.2$ clearly reflects the increase in the number of systems of this type with decreasing mass and period. For example, the maximum of this distribution lies in the mass interval 1.6–$2M_\odot$ and the period interval $0^d.4$–$0^d.8$, and the minima are concentrated in the intervals (2–$2.5M_\odot$, $0^d.3$–$0^d.4$), (2.5–$6M_\odot$, $0^d.5$–$0^d.8$), (5–$6.3M_\odot$, $0^d.63$–$1^d.26$).

3. INITIAL DISTRIBUTIONS OF DMS, ~KW, KW, AND KP CLOSE BINARIES

We can derive the initial distribution from the current spatial distribution of DMS and ~KW systems $F(M_1^j, P^k) = \sum_{i=1}^{n} \frac{1}{W^i V^i} f^i$ by separating it into the time τ^i a star with mass M_*^i and radius R_*^i spends on the main sequence. We determined τ^i individually for each DMS and ~KW star using isochrones constructed on the basis of evolutionary tracks obtained by Maeder and Meynet [12] with (**X**, **Y**, **Z**) = 0.70, 0.28, 0.02 for Population-I stars, taking into account convective penetration and mass loss by the components. The isochrone method is described in detail in [13].

The initial distribution of DMS systems (Fig. 2) is a sort of "negative image" of the corresponding current distribution, since the positions of local maxima and minima are interchanged. For example, in the initial distribution of DMS systems, maxima are concentrated in the area of large masses and periods: ($M_1 = 2.8M_\odot$, $P = 5^d.6$), ($M_1 = 3.5M_\odot$, $P = 7^d$), ($M_1 = 4.5M_\odot$, $P = 5^d.6$), ($M_1 = 7M_\odot, P = 5^d$–8^d), and ($M_1 = 15M_\odot, P = 10^d$). Minima in the initial distribution of DMS systems

are located in areas of small masses and periods: (M_1 = 1–4$M_\odot$, P = 1^d–$1^d.6$) and (M_1 = 1–2$M_\odot$, P= 1^d – 4^d). There remain dips in the initial distribution of DMS systems, due to the absence of these system in these intervals in the apparent distribution.

The initial distribution of DMS systems indicates that the formation of systems with masses $M_1 \geq 2.5M_\odot$ and periods $\geq 4^d$ is predominant; for masses larger than $8M_\odot$, periods can be around 2^d. Lucy and Ricco [14] concluded that most close binaries with initial periods $P \geq 25^d$ and total masses of the components from ≈ 0.5 to $10M_\odot$ form via a mechanism that creates components with similar masses—the fragmentation of a rotating protostar at the last stage of its dynamical collapse before its entry onto the main sequence. Depending on the thermal energy of the cloud, the fragmentation can occur either directly or via an intermediate stage of ring formation. The binary mode of fragmentation dominates [15, 16].

The initial distribution of ~KW systems (Fig. 2) displays levels close to zero, with the exception of the interval (M_1 = 0.8–1$M_\odot$, P = 1^d–$1^d.27$), where, as before, there is a maximum. This agrees with our current understanding of the evolutionary origin of ~KW systems. The maximum at (M_1 = 0.8–1.5$M_\odot$, P = 1^d–$1^d.5$), in the area where the current distributions of DMS and ~KW systems overlap, should probably be taken into account when interpreting the current distribution of DMS systems. In other words, ~KW systems form as a result of the evolution of DMS systems. Therefore, the initial distribution of ~KW systems can be derived from the current distribution of DMS systems with masses of $\leq 1.5M_\odot$, in which the components approach each other as a result of magnetic deceleration, virtually without filling their inner critical surfaces.

Thus, the location of ~KW systems in Fig. 1 can be interpreted as the initial distribution of ~KW systems, while the location of KW systems simultaneously reflects the current distribution of ~KW systems and the initial distribution of KW systems; i.e., the result of the transformation of ~KW systems as their components approaches due to a loss of mass and orbital angular momentum.

It follows that ~KW and KW systems, respectively, represent initial and current stages of evolution for the same class of contact binary system. Thus, based on the distributions analyzed, the evolutionary transition DMS $\Rightarrow$ ~KW $\Rightarrow$ KW is obvious.

4. CONCLUSIONS

Figure 1 presents the current distributions of all the types of systems we have considered. In overlapping areas, an evolutionary "migration" of systems due to loss of mass and orbital momentum, from the area of DMS stars to the area of KW systems via the stage of ~KW systems, can clearly be seen. The area of KP stars overlaps that of DMS systems, and seems to extend the boundaries of the DMS-system area to orbital periods of about 7.5^h ($\approx 0^d.32$).

No apparent intersection of areas for the KP and KW systems can be seen in Fig. 1. However, the proximity of these areas suggests a KP $\Rightarrow$ KW evolutionary transition via the same mechanism as for the DMS $\Rightarrow$ ~KW $\Rightarrow$ KW transition; i.e., loss of orbital angular momentum.

REFERENCES

1. M. A. Svechnikov, L. F. Istomin, and O. A. Grekhova, Perem. Zvezdy **21**, 399 (1980).
2. M. A. Svechnikov and É. F. Kuznetsova, *Catalogue of Approximate Photometric and Absolute Elements of Eclipsing Variables* [in Russian] (Yekaterinburg, Ural'sk. Gos. Univ., 1990).
3. M. A. Svechnikov, O. V. Eretnova, M. N. Ol'neva, and T. A. Taĭdakova, Nauchn. Inf. Astrosovet Akad. Nauk SSSR **67**, 1989.
4. M. A. Svechnikov and O. V. Eretnova, *Astronomical and Geodetic Studies. Variable Stars and Stellar Systems* [in Russian] (Yekaterinburg, 1995).
5. T. A. Taĭdakova, Masters thesis (Chelyab. Gos. Univ., Chelyabinsk, 1981).
6. M. A. Svechnikov, Doctoral Dissertation in Mathematical Physics (Chelyab. Gos. Univ., Chelyabinsk, 1985).
7. C. W. Allen, *Astrophysical Quantities* (Athlone Press, London, 1973; Mir, Moscow, 1977).
8. D. M. Popper, Ann. Rev. Astron. Astrophys. **18**, 115 (1980).
9. L. F. Istomin, *Star Clusters and Binary Systems* [in Russian] (Yekaterinburg, 1978).
10. M. A. Svechnikov and É. F. Kuznetsova, *Astronomical and Geodetic Studies. Dynamical and Physical Characteristics of Celestial Bodies* [in Russian] (Yekaterinburg, 1988).
11. M. A. Svechnikov, L. A. Bessonova, and N. A. Dubinina, Nauchn. Inf. Astrosovet Akad. Nauk SSSR, 1988.
12. A. Maeder and G. Meynet, Astron. Astrophys., Suppl. Ser. **76**, 411 (1988).
13. G. N. Dremova and M. A. Svechnikov, in *Catalogue of Orbital Elements, Masses, and Luminosities of DMS Eclipsing Variable Stars and some Statistical Results* [in Russian] (Ural. Gos. Univ., Yekaterinburg, 1999), p. 31.
14. L. B. Lucy and E. Ricco, Astron. J. **84**, 401 (1979).
15. P. Bodenheimer, J. E. Tohline, and D. C. Block, Astron. Astrophys., Suppl. Ser. **30** (4), 501 (1980).
16. R. B. Larson, Mon. Not. R. Astron. Soc. **214** (3), 379 (1985).
17. Z. T. Kraĭcheva, E. I. Popova, A. E. Tutukov, and L. R. Yungel'son, Astron. Zh. **55**, 1176 (1978) [Sov. Astron. **22**, 670 (1978)].
18. M. V. Popov, Perem. Zvezdy **17**, 412 (1970).
19. E. J. Devinney and J. H. Edward, Publ. Astron. Soc. Pac. **85** (505), 330 (1973).
20. Z. T. Kraĭcheva, E. I. Popova, A. E. Tutukov, and L. R. Yungel'son, Astron. Zh. **56**, 520 (1979) [Sov. Astron. **23**, 290 (1979)].

Translated by K. Maslennikov

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 216–229. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 253–268.

Formation of Complex Rotational Structures in Convective and Isothermal Magnetized Zones

Yu. V. Vandakurov

Ioffe Physicotechnical Institute, Russian Academy of Sciences, Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia

Received March 27, 2000

Abstract—We study the development of a complex rotation law in the magnetized convective and isothermal zones of stars and planetary atmospheres through the decomposition of vector quantities in terms of orthogonal vector spherical harmonics. In the case of a solar-type extended convective zone, it is assumed that (a) the transformation of thermal into magnetic energy is favorable from the viewpoint of energy balance, (b) the state that is supported with minimum energy loss is realized, and (c) the condition of minimum entropy production consistent with the two previous requirements is satisfied. To find the rotation law of a zone, weak interaction between variations in the rotation and magnetic-field distributions is assumed. Two possible zones of generation of the solar magnetic field are considered. The first is located in the lower half of the solar convective zone and possesses a latitude dependence of the rotational velocity similar to that observed. The second zone is located just below the surface, and has a rotational velocity that decreases sharply with height and depends only weakly on latitude. We also study simple equilibrium structures, in particular, those describing the super-rotation of the medium in a convective or isothermal zone. Realization of such super-rotation in an isothermal zone is associated with the outflow of matter and fields toward upper layers. © *2001 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

Observational data show that complex rotation patterns are often formed in the convective zones of stellar and planetary atmospheres. For example, the entire solar convective zone exhibits differential rotation (see, for example, [1]), along with cyclic magnetic-field generation processes. The differential rotation of the convective zone is associated with the excitation of strong turbulent viscosity, which is a strong source of entropy production (or energy dissipation). On the other hand, the entropy production for a system near equilibrium should be minimum [2], leading to the establishment of rigid-body rotation of the medium.

This paradox can be resolved if there is some additional constraint preventing the equalization of the angular velocities of different layers. In the case of the Sun, the huge size of the convective zone (whose thickness exceeds 100000 km) seems to be important, since it strongly increases the energy dissipation required to transport thermal flux via the excitation of convective motions. Recently, we suggested that the transformation of the energy of this thermal flux to magnetic-field energy is favorable from an energetic point of view [3]. In other words, this implies the formation of a state with minimum energy loss. In this case, the established rotation of the medium should correspond to the maximum conversion of thermal to magnetic energy. In addition, all modes preventing field generation must be suppressed. This leads to the excitation of strong turbulent viscosity, although probably somewhat less than that needed for direct support of the observed rotation (see the estimates in Section 3). As noted above, the direct support of differential rotation is in contradiction with the conclusions of irreversible thermodynamics.

Since any redistribution of the rotational velocity is quite difficult, we expect that the transformation of thermal to magnetic energy will be facilitated if the field variations are not accompanied by substantial changes in the rotational velocity. This is also confirmed by the small amplitude of variations in the solar rotation in the course of the solar activity cycle [4]. The condition of weak inter-dependence between variations in the magnetic field and the average rotational-velocity can be used to find the most probable rotation law for the medium. This approach was applied in [3] using an axially-symmetric model. It was established that, in the case of maximum extent of the zone of weak interdependence between the field and rotational variations, this layer will cover approximately the entire lower half of the solar convective zone, with the calculated latitude dependence of the rotational velocity similar to that observed for the Sun.

Nevertheless, it remains unclear how important the restriction on the character of the magnetic field assumed in [3] actually is, since the observed field does not possess axial symmetry. In addition, the possible existence of several regions of magnetic-field generation in the solar convective zone is of considerable interest. This problem is currently being widely discussed [5–7]. In particular, it has been proposed that, along with the main generation zone, located at a considerable depth and characterized by the 22-year period, there is also a subsurface zone characterized by quasi-two-year periods [6]. In standard dynamo theory,

1063-7729/01/4503-0216$21.00

the locations of the zones of generation of magnetic field are usually associated with zones having large angular-velocity gradients, assuming these gradients are formed by some other processes [6]. In our study, such gradients are determined in the process of finding the desired solution (see [3] and Section 3 below), so that the above-mentioned relation is not necessary. However, the existence of a zone with a sharp decrease in the rotational velocity in subsurface solar layers has been confirmed observationally (see, for example, [1]). Therefore, theoretical description of this phenomenon is very relevant. Some models describing rotation of a medium that strongly decelerates with height are studied in Section 4, although the latitude dependence for their rotational velocity differs from that observed for the Sun. We also discuss the possibility of realizing the structures obtained.

The supplementary condition mentioned above (preventing the establishment of rigid-body rotation of the convective zone) in not the only one possible. Effects facilitating convective heat transfer often play an important role. For example, in the case of the Venusian atmosphere, the planet rotates very slowly (a factor of 243 more slowly than the Earth), and convection is excited by large horizontal temperature gradients (see the reviews [8–10]). Under these conditions, the development of more rapid rotation of an atmosphere with meridional circulation of matter could strengthen horizontal heat transfer. Observations have demonstrated that the atmosphere is characterized by four-day circulation in both its lower and upper layers. The numerical model for the lower atmosphere of Young and Pollack [11] confirmed the possibility of spontaneous formation of such circulation. We will argue below that, in the presence of a magnetic field, structures whose rotational velocity increases rapidly with height can also form in isothermal upper atmospheric layers (see Section 5).

Mathematical difficulties associated with finding the equilibrium state of a rotating magnetized medium are due, first and foremost, to the presence of nonlinear forces. We present the vector fields of the hydrodynamical velocity **v** and magnetic induction **B** in the form of expansions in a complete set of orthogonal vector spherical harmonics (see Sections 2 and 3). This approach enables us to perform a separation of variables using exact relations, even in the presence of nonlinear forces. We also take into account the requirements for regular spatial behavior of the fields.

As illustrated by specific examples, the use of exact equations is quite important, since the most serious difficulties are often associated precisely with the nonexistence of a solution. Namely, the structure of the equations after the separation of variables is such that any transition from a given to a higher approximation is accompanied by an increase in the number of terms in the series representing the nonlinear force that is more rapid than the increase in the number of new coefficients of higher order (one such example will be described in Section 5). As a result, the number of coefficients may be insufficient to satisfy the equilibrium equations. This problem was analyzed in [12, 13] for the case of an axially-symmetric, non-rotating magnetic model. It was established that thermal equilibrium can be satisfied in the zone of radiative equilibrium only if the magnetic field is dipolar. This led to the hypothesis (put forward in [13]) that the formation of spots with different chemical compositions in magnetic stars is associated with deviations of their field geometry from dipolar.

The formation of complex rotational structures in stellar convective zones indicated by observations obviously points toward a truncation of the above-mentioned series due to suppression of higher-order rotational modes by convective motions. We will call modes that are not suppressed principal (or main) modes. The parameters of these modes can be determined by minimizing the total entropy production subject to the supplementary condition indicated above. In the case of a solar-type convective zone, we must find a configuration with maximum efficiency for transformation of thermal to magnetic energy and minimum total entropy production, with the first condition being more important. Our calculations testify that the conditions listed above are sufficient for unambiguous determination of both the location and parameters of the solar zones of magnetic-field generation.

Of particular interest are configurations with vector hydrodynamical-velocity and magnetic-induction fields described by finite series of vector spherical harmonics. The establishment of force balance is considerably simplified in this case, since no higher-order modes should be suppressed. We will show in Section 5 that truncation of this harmonics series is possible not only for rigid-body rotation of the medium. The general expressions involve both an azimuth-dependent component of the angular velocity and meridional circulation of the medium; the angular dependence of the magnetic field corresponds to a rotated dipole.

The equilibrium condition can also be satisfied in the case of an isothermal zone and an angular velocity that grows rapidly with height (see Section 5). Magnetic-field generation still takes place in such a model, so that the field and matter frozen into it must be carried to upper layers to satisfy the condition of a steady state. This process is facilitated if the increase in angular velocity with height is quite sharp. Note that observational data indicate the presence of super-rotating layers in the upper atmospheres of the Earth and Venus, as well as strong outflows of matter from the Sun. These problems are discussed in Section 5. A short summary is given in the last section.

2. BASIC EQUATIONS

The equilibrium equations in the case of a viscous magnetized medium can be written

$$\frac{1}{2}\nabla(\mathbf{v}\cdot\mathbf{v}) + \mathbf{Q} + \frac{1}{\rho}\nabla p + \nabla\Phi = \mathbf{F}, \qquad (1)$$

where

$$\mathbf{Q} = (\mathrm{curl}\,\mathbf{v})\times\mathbf{v} - (\mathrm{curl}\,\mathbf{B})\times\mathbf{B}/(4\pi\rho), \quad (2)$$

v is the hydrodynamical velocity, **B** the magnetic field, **F** the viscous force, p the pressure, ρ the density, and Φ the gravitational potential. In addition, a complete system of equations includes the relations

$$\mathrm{div}(\rho\mathbf{v}) = 0, \quad \mathrm{div}\,\mathbf{B} = 0 \quad (3)$$

and equations for the field and energy. The last equation in the problems considered below expresses the maintenance of either convective or isothermal structure of the zone. In addition, in an ideal magnetohydrodynamic approximation, the field is stationary if the vector curl $(\mathbf{v}\times\mathbf{B})$ is zero.

Note that the action of the rot operation on (1) yields

$$\mathbf{R} + (\nabla p\times\nabla\rho)/\rho^2 = \mathrm{curl}\,\mathbf{F}, \quad (4)$$

where

$$\mathbf{R} = \mathrm{curl}\,\mathbf{Q}. \quad (5)$$

We shall consider below an approximation in which the density in the formula for the magnetic force is assumed to be spherically symmetric. In addition, in studying equilibrium in the convective zone, we shall assume that the medium is close to a neutral convective state, so that the second term on the left-hand side of (4) can be neglected.

We perform a separation of variables in the above equations by writing any vector **f** (or scalar p) as an expansion in the orthogonal vector spherical harmonics $\mathbf{Y}_{JM}^{(\lambda)}$ (or the spherical functions Y_{JM}):

$$\mathbf{f} = \sum_{JM}\left\{\mathbf{i}_r f_{JM}^{(-1)} Y_{JM} + \mathbf{i}_\vartheta \frac{1}{[J(J+1)]^{1/2}}\right.$$
$$\times\left[f_{JM}^{(+1)}\frac{\partial Y_{JM}}{\partial\vartheta} - \frac{M}{\sin\vartheta} f_{JM}^{(0)} Y_{JM}\right] + \mathbf{i}_\varphi \frac{i}{[J(J+1)]^{1/2}} \quad (6)$$
$$\left.\times\left[\frac{M}{\sin\vartheta} f_{JM}^{(+1)} Y_{JM} - f_{JM}^{(0)}\frac{\partial Y_{JM}}{\partial\vartheta}\right]\right\},$$

$$p = \sum_{JM} p_{JM} Y_{JM}, \quad (7)$$

where $f_{JM}^{(\lambda)} = f_{JM}^{(\lambda)}(r)$, $Y_{JM} = Y_{JM}(\vartheta, \varphi)$, $\mathbf{i}_r$, $\mathbf{i}_\vartheta$, and $\mathbf{i}_\varphi$ are unit vectors of the spherical coordinate system (r, ϑ, φ); subscript J is an integer non-negative number; $M = -J$, $-J+1, \ldots, J$; and $\lambda = -1, 0,$ or $+1$. The coefficients $f_{JM}^{(\pm1)}$ and $f_{JM}^{(0)}$ describe the poloidal and toroidal components, respectively. In particular, when $\mathbf{f} = \mathbf{v}$, the coefficients $v_{J0}^{(0)}$ describe rotation of the medium. For example, if only the term with $J = 1$ is significant, then the rotation will not depend on latitude, and

$$v_{10}^{(0)} = -ir\Omega(8\pi/3)^{1/2}, \quad (8)$$

where Ω is the angular velocity. In the case of an axially-symmetric dipole magnetic field, $\mathbf{f} = \mathbf{B}$, and only the coefficient $B_{10}^{(\pm1)}$ will be non-zero.

The formulas for the representation of some nonlinear vector fields in terms of the expansions considered here were derived in [12], and additional results are given in the Appendix. When the vector **Q** in (2) is nonlinear, the general expressions for $Q_{JM}^{(\lambda)}$ are given by Eqs. (22)–(24) from [12]. These equations were studied in more detail in [3] under the assumption that the magnetic field and hydrodynamical velocity do not depend on the azimuth (i.e., $M = 0$).

In the present paper, we shall consider arbitrary toroidal vector fields. In this case, the vectors **v** and **B** have only the components $v_{JM}^{(0)}$ and $B_{JM}^{(0)}$. In particular, the expressions for the components of the vector $\mathbf{R} = \mathrm{curl}\,\mathbf{Q}$ have the form

$$R_{JM}^{(0)} = -\frac{i}{2r^2(2J+1)^{1/2}}\sum_{J_1M_1J_2} C_{J_1M_1J_2M_2}^{JM} K_{J_1J_2}^{J}$$
$$\times\left\{J(J+1)Z_{J_1J_2}^{J}\right.$$
$$\times\left[v_{J_2M_2}^{(0)}\frac{\partial}{\partial r} r v_{J_1M_1}^{(0)} - \frac{1}{4\pi\rho} B_{J_2M_2}^{(0)}\frac{\partial}{\partial r} r B_{J_1M_1}^{(0)}\right] \quad (9)$$
$$\left.+ J_1(J_1+1)Z_{JJ_2}^{J_1} r\frac{\partial}{\partial r}\left[v_{J_1}^{(0)} v_{J_2M_2}^{(0)} - \frac{1}{4\pi\rho} B_{J_1M_1}^{(0)} B_{J_2M_2}^{(0)}\right]\right\},$$

$$R_{JM}^{(-1)} = \frac{i}{r^2}\left[\frac{J(J+1)}{2J+1}\right]^{1/2}\sum_{J_1M_2J_2} C_{J_1M_1J_2M_2}^{JM}$$
$$\times\Theta_{J_1J_2}^{J} J_1(J_1+1)\left[v_{J_1M_1}^{(0)} v_{J_2M_2}^{(0)} - \frac{1}{4\pi\rho} B_{J_1M_1}^{(0)} B_{J_2M_2}^{(0)}\right], \quad (10)$$

$$R_{JM}^{+1} = -\frac{i}{r[J(J+1)]^{1/2}}\frac{\partial}{\partial r} r^2 R_{JM}^{(-1)}, \quad (11)$$

where

$$Z_{ab}^{c} = a(a+1) + b(b+1) - c(c+1), \quad (12)$$

$M_2 = M - M_1$, $C_{J_1M_1J_2M_2}^{JM}$ is a Clebsch–Gordan coefficient, and K_{J_1,J_2}^{J} and $\Theta_{J_1J_2}^{I}$ are expressed in terms of the analogous Clebsch–Gordan coefficients using formulas (10) and (11) from [12] (see also (59) from the

Appendix). We use these equations to study various numerical models in the next section.

3. THE LOWER ZONE OF SOLAR MAGNETIC-FIELD GENERATION

The generation of the solar magnetic field obviously involves transport of field to upper layers. If we wish to use the condition of weak inter-dependence between variations in the rotational velocity and field formulated above to determine the most probable distribution for the zone's rotation, we must compare the rotation laws for equilibrium models with various fields. In [3], we studied pairs of axially-symmetric models, one containing a toroidal field (the superscript of the field coefficients $\lambda = 0$) and the other containing a poloidal field ($\lambda = \pm 1$). We assumed that all modes with dimensions of velocity have the same radial dependence. More precisely, we took these dependences and the radial dependence of the equilibrium density to have the form

$$\rho = \text{const}\, r^{-s}, \quad v_{JM}^{(\lambda)} = \text{const}\, r^{\alpha}, \qquad B_{JM}^{(\lambda)}/(4\pi\rho) = \text{const}\, r^{\alpha}, \tag{13}$$

where s and α are constant coefficients, and the subscript M was assumed to be 0 for the models from [3]. Only rotation laws that were symmetric with respect to the equatorial plane were considered, and the parameter s was specified.

The coefficients $v_{J0}^{(\lambda)}$ and $B_{J0}^{(\lambda)}$ for the principal modes (i.e., those not suppressed by convection) were determined in [3] by the equilibrium condition in the absence of viscosity. The coefficient $v_{10}^{(0)}$ was taken to be unity, and the parameter α was varied instead of this coefficient. Note that the coefficients responsible for meridional motions turned out to be negligible. The main problem was to find those conditions, listed in (a)–(d) below, under which there were two models with approximately the same latitude distributions for the rotational velocity. To characterize the similarity of the rotation laws, we used the parameter σ—the rms difference between the relative angular velocities $\omega_a = (\Omega/\Omega_e)_a$ at latitudes $a = 5°, 10°, \ldots, 90°$, where Ω_e is the equatorial angular velocity of rotation. When σ is sufficiently small, the two models are characterized by approximately the same total angular momentum.

It was found that the extent of the zone of weak inter-dependence between variations in the field and rotation distributions is maximum when:

(a) the magnetic field is antisymmetric with respect to the equatorial plane,

(b) the lower half the solar convective zone (in radius) is considered,

(c) the number N of main modes with various values of the subscript J in (6) is four for both the rotational velocity and the field, and

(d) the parameter α in (13) is close to or slightly less than 1.3.

The calculated latitude dependence of the rotational velocity proved to be close to that observed for the Sun (see, for example, the thick solid and sparsely dotted curves in Fig. 3 from [3], which describe the latitude dependence of the angular velocities in models with toroidal and poloidal fields in the layer $s = 10$, which is located approximately in the middle of the solar convective zone). Note that σ begins to increase sharply toward the region $s > 10$ (where the relative radius $r/R_\odot > 0.85$), and the rotation laws for the two models of a pair become appreciably different.

However, it remains unknown if the parameters of the zone under investigation vary significantly when we consider the non-axially-symmetric magnetic field, which plays an important role in the solar cycle. We will study this problem below for the particular case of a toroidal magnetic field whose azimuth dependence is characterized by the first main harmonic. Namely, we shall compare parameters of a pair of models, one with an axially-symmetric toroidal magnetic field and the other with a field defined by a series of the form (6), containing the coefficients $v_{J\pm1}^{(0)}$ and $B_{J\pm1}^{(0)}$. We assume that modes with subscripts $J > 2N$ and $|M| > 1$ are suppressed by turbulent viscous forces. On the contrary, the viscosity should be negligible for principal modes, so that we obtain from (4) for a medium that is close to a neutral convective state

$$R_{JM}^{(0)} \approx 0, \quad R_{JM}^{(-1)} \approx 0, \tag{14}$$

where $J \le 2N$ and $|M| \le 1$.

Taking into account $R_{J0}^{(-1)} = 0$, we obtain for the coefficients in (14):

$$R_{J0}^{(0)} = -\frac{i}{2r^2(2J+1)^{1/2}}\sum_{J_1J_2} K_{J_1J_2}^{J}$$

$$\times\left\{C_{J_10J_20}^{J0}\left[J(J+1)Z_{J_1J_2}^{J}\left[v_{J_20}^{(0)}\frac{\partial}{\partial r}r v_{J_10}^{(0)}\right.\right.\right.$$

$$-\frac{1}{4\pi\rho}B_{J_20}^{(0)}\frac{\partial}{\partial r}rB_{J_10}^{(0)} + J_1(J_1+1)Z_{JJ_2}^{J_1}r\frac{\partial}{\partial r}$$

$$\left.\left.\left.\times\left(v_{J_10}^{(0)}v_{J_20}^{(0)} - \frac{1}{4\pi\rho}B_{J_10}^{(0)}B_{J_20}^{(0)}\right)\right]\right. \tag{15}$$

$$+ C_{J_11J_2-1}^{J0}\left[J(J+1)Z_{J_1J_2}^{J}\right.$$

$$\left.\times\left(\frac{1}{r}\frac{\partial}{\partial r}r^2 v_{J_11}^{(0)}v_{J_2-1}^{(0)} - \frac{1}{4\pi\rho r}\frac{\partial}{\partial r}r^2 B_{J_11}^{(0)}B_{J_2-1}^{(0)}\right)\right.$$

$$+ X^J_{J_1J_2} r\frac{\partial}{\partial r}\left(v^{(0)}_{J_11} v^{(0)}_{J_2-1} - \frac{1}{4\pi\rho} B^{(0)}_{J_11} B^{(0)}_{J_2-1}\right)\Bigg]\Bigg\},$$

$$R^{(0)}_{J1} = R^{(0)}_{J-1} = -\frac{i}{2r^2(2J+1)^{1/2}} \sum_{J_1J_2} K^J_{J_1J_2} C^{J1}_{J_11J_20}$$

$$\times\Bigg\{ J(J+1) Z^J_{J_1J_2}$$

$$\times\left[\frac{1}{r}\frac{\partial}{\partial r} r^2 v^{(0)}_{J_11} v^{(0)}_{J_20} - \frac{1}{4\pi\rho r}\frac{\partial}{\partial r} r^2 B^{(0)}_{J_11} B^{(0)}_{J_20}\right] \quad (16)$$

$$+ X^J_{J_1J_2} r\frac{\partial}{\partial r}\left(v^{(0)}_{J_11} v^{(0)}_{J_20} - \frac{1}{4\pi\rho} B^{(0)}_{J_11} B^{(0)}_{J_20}\right)\Bigg\},$$

$$R^{(-1)}_{J1} = -R^{(-1)}_{J-1} = \frac{i}{r^2}\left[\frac{J(J+1)}{2J+1}\right]^{1/2}$$

$$\times \sum_{J_1J_2} \Theta^J_{J_1J_2} C^{J1}_{J_11J_20} [J_1(J_1+1) - J_2(J_2+1)] \quad (17)$$

$$\times\left(v^{(0)}_{J_11} v^{(0)}_{J_20} - \frac{1}{4\pi\rho} B^{(0)}_{J_11} B^{(0)}_{J_20}\right),$$

where

$$X^J_{J_1J_2} = J_1(J_1+1) Z^{J_1}_{JJ_2} + J_2(J_2+1) Z^{J_2}_{JJ_1}. \quad (18)$$

Substituting these relations into (14), we obtain a system of equations for coefficients $v^{(0)}_{JM}$ and $B^{(0)}_{JM}$, where $J < (2N+1)$ and M equals 0 or ±1. We assume here, as above, that the radial dependences of the equilibrium density and all modes under consideration are specified by expressions (13). However, in the present case of nonlinear equations, we encounter several very difficult problems, such as non-existence of solutions, multivalued solutions, achievement of the required accuracy, and so on. Therefore, it is more expedient to formulate the problem of finding the solution of (14) in terms of the minimization of the following dimensionless functional, which is a sum of positive terms:

$$\epsilon^2 = \sum_J \epsilon^2_J = [r/v^{(0)}_{10}]^4 \sum_J \left[\sum_{\lambda M} R^{(\lambda)}_{JM} R^{(\lambda)*}_{JM}\right], \quad (19)$$

where an asterisk denotes complex conjugation and the summation limits are defined in agreement with the restrictions introduced above. Note that the term with subscript $J = 0$ in (19) vanishes.

The minimum of the functional (19) also corresponds to the minimum of the entropy production due to the principal modes in Eqs. (15)–(17). Investigation of models with large values of s and α showed that it is often impossible to choose $v^{(0)}_{JM}$ and $B^{(0)}_{JM}$ so that the minimum value of ϵ is negligible (our calculations were carried out to nine significant digits). We can assume that the comparatively small residuals can be compensated by excitation of the corresponding turbulent viscous forces (as mentioned above, such forces are necessary to suppress high-order modes). However, the problem of the validity of solutions for which ϵ is not negligible must still be justified in more detail.

Note that the vector field **f** under consideration (where **f** is either **v** or **B**) can be written in the form

$$\mathbf{f} = \sum_J \frac{1}{[J(J+1)]^{1/2}}\left[-\mathbf{i}_\varphi \frac{\partial Y_{J0}}{\partial\vartheta} i f^{(0)}_{J0}\right.$$

$$\left. + 2 f^{(0)}_{J1}\left(\mathbf{i}_\varphi \frac{\partial W_{J1}}{\partial\vartheta}\sin\varphi - \mathbf{i}_\vartheta \frac{\cos\varphi}{\sin\vartheta} W_{J1}\right)\right]. \quad (20)$$

Here, $W_{J1} = Y_{J1} e^{-i\varphi}$ is a function only of ϑ, and we have taken into account the relation $f^{(0)}_{J-1} = f^{(0)}_{J1}$.

Let us restrict our consideration to toroidal magnetic fields that are antisymmetric with respect to the equatorial plane and layers of the solar convective zone that are distant from the surface. We shall assume that the rotational velocity is symmetric with respect to the equatorial plane. It turns out that minimization of the functional (19) under the conditions formulated above leads to negligible values of the non-axially-symmetric velocity component; i.e., $v_{J1} \approx 0$. For the other components, it is expedient to introduce the real dimensionless coefficients u_J, b_J, and q_J via the relations

$$\frac{v^{(0)}_{J0}}{v^{(0)}_{10}} = u_J = u_{2k-1}, \quad \frac{B^{(0)}_{J0}}{v^{(0)}_{10}[4\pi\rho]^{1/2}} = b_J = b_{2k},$$

$$\frac{B^{(0)}_{J1}}{i v^{(0)}_{10}[4\pi\rho]^{1/2}} = q_J = q_{2k-1}, \quad (21)$$

where $k = 1, 2, \ldots, N$; J is expressed through k; and $u_1 = 1$. Because of this normalization, the average equatorial angular velocity is $(i v^{(0)}_{10}/r)[3/(8\pi)]^{1/2} d$, where d = const.

We construct a model with a non-axially-symmetric (*ns*) magnetic field by adjusting the coefficients $(u_J)_{J>1}$, q_J, and α so that the right-hand side of (19) is negligible ($\epsilon \sim 10^{-9}$) when the minimum is attained. The only difference when finding a model with an axially-symmetric (*si*) field is that the coefficients q_J are replaced by b_J. We consider models where $N = 4$ and the value of α is close to 1.3. In the case of the main models, s is in the interval $5 \le s \le 10$, which corresponds to the lower half of the solar convective zone. Next, we find pairs of models whose magnetic fields possess different azimuthal symmetry, taking either α or d to be approximately equal to each other in the two models. In other

Table 1. Characteristics of the axially-symmetric (*si*) and non-axially-symmetric (*ns*) models

Number	Type	s	α	d	A_{max}	ω_{30}	ω_{50}	ω_{70}	ω_{90}
1	*ns*	5	1.29781	1.04259	0.124	0.957	0.872	0.717	0.624
2	*si*	5	1.29782	1.04136	0.072	0.958	0.874	0.738	0.661
3	*si*	5	1.30630	1.04250	0.073	0.957	0.871	0.729	0.647
4	*ns*	8	1.29755	1.04269	0.121	0.956	0.871	0.731	0.651
5	*si*	8	1.29736	1.04184	0.063	0.958	0.873	0.729	0.645
6	*si*	8	1.30515	1.04278	0.061	0.957	0.870	0.721	0.634
7	*ns*	10	1.29591	1.04253	0.121	0.956	0.871	0.739	0.665
8	*si*	10	1.29591	1.04077	0.074	0.959	0.877	0.731	0.644
9	*si*	10	1.30262	1.04254	0.054	0.957	0.870	0.722	0.635
10	*ns*	20	1.28664	1.04153	0.119	0.955	0.872	0.776	0.732
11	*si*	20	1.29098	1.04218	0.043	0.958	0.872	0.726	0.640

words, this condition implies that either the radial gradients of all modes for the two models of the pair are similar, or that they have the same average equatorial velocities. Using more than one method for comparison could be useful, since the process for the generation of solar magnetic field remains to a large extent unclear.

The results of our study are summarized in Tables 1–3, which present the basic parameters of the models, including A_{max}, which is the largest Alfven velocity divided by the average equatorial rotational velocity Ω_e. The tables also present the relative angular velocities $\omega_a = (\Omega/\Omega_e)_a$ at several latitudes a (Table 1), and the rms differences σ between the relative angular velocities for the two models from each pair (Table 2). These differences are calculated at latitudes 5°, 10°, …, 90°. For the sake of comparison, Table 1 also presents models 10 and 11, for which $s = 20$; i.e., $r/R_\odot \approx 0.92$. Finally, Table 3 gives the coefficients u_J and q_J for models 1, 4, and 7 from Table 1, which have a non-axially-symmetric magnetic field.

Figure 1 shows distributions of the relative angular velocities for the three models noted above and for model 10 from the same table (lower to upper curves, respectively). The crosses denote the results of Doppler observations [4] and the bars, helioseismology data [14]. Analogous angular-velocity distributions for all the other models from Table 1 (except model 10) are located in Fig. 1 inside the region bounded by the solid curves; i.e., these curves represent envelopes for all variations in angular velocity for the models under consideration. We can see that these variations are small for $5 \le s \le 10$, and are concentrated at high latitudes. This indicates that the conclusions formulated above for models with axially-symmetric fields remain valid for zones for which the rotational variations are small when the equilibrium axially-symmetric toroidal magnetic fields are exchanged with non-axially-symmetric fields or vice versa. In the models discussed here, whose rotation is fairly close to rigid-body, it is not important which of the two criteria for comparison formulated above we use for the calculations.

Note also that, as follows from (20), the relative Alfven velocity in the case without axial symmetry can be written

$$\begin{aligned}\mathbf{A} &= [\mathbf{B}/(4\pi\rho)^{1/2}]/(r\Omega_e)\\ &= A_1\mathbf{i}_\varphi \sin\varphi + A_2\mathbf{i}_\vartheta \cos\varphi.\end{aligned} \tag{22}$$

Figure 2 shows the distribution of the A_j components for this relative Alfven velocity in the case of models 1 and 7 from Table 1 (solid/dashed curves and densely/sparsely dotted curves, respectively). The absolute values of the fields averaged over the azimuth are quite close to those considered in [3].

We also estimated ϵ_J in (19). Recall that the four first ϵ_J (with the lowest values of J) correspond to the principal modes, and all higher modes are assumed to be suppressed by turbulent viscous forces. Modes with the largest scales are most important among these. In the case of the model 7 from Table 1, which has a non-axially-symmetric magnetic field, we obtained for the first five ϵ_J the values 4.7×10^{-10}, 5.4×10^{-10}, 1.5×10^{-11}, 1.1×10^{-9}, and 3.8×10^{-2}. We can see that the principal modes are approximately non-viscous, and the last of the ϵ_J values can be used to estimate the turbulent viscosity of the medium. When we consider models with axially-symmetric magnetic fields and approximately the same basic parameters, the turbulent viscosity usually becomes appreciably lower.

Table 2. Rms differences between the relative angular velocities for models from Table 1

Pair	1–2	1–3	4–5	4–6	7–8	7–9	10–11
σ	0.0172	0.0103	0.0027	0.0079	0.0093	0.0142	0.044

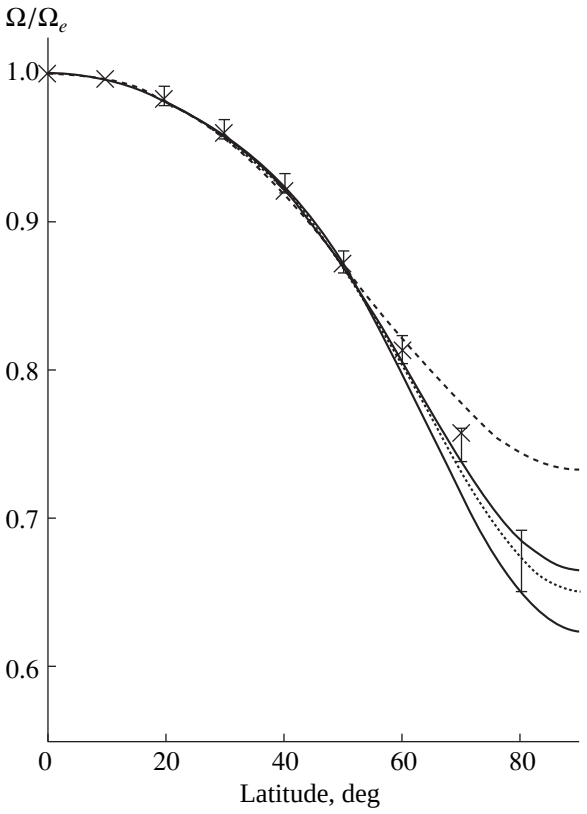


Fig. 1. Latitude dependence of the angular velocity Ω normalized to the average angular velocity at the equator Ω_e for models 1, 4, 7, and 10 from Table 1 (lower solid, densely dotted, upper solid, and sparsely dotted curves, respectively). Doppler data [4] are marked by crosses and helioseismology data [14] by vertical bars, whose length characterizes the data scatter.

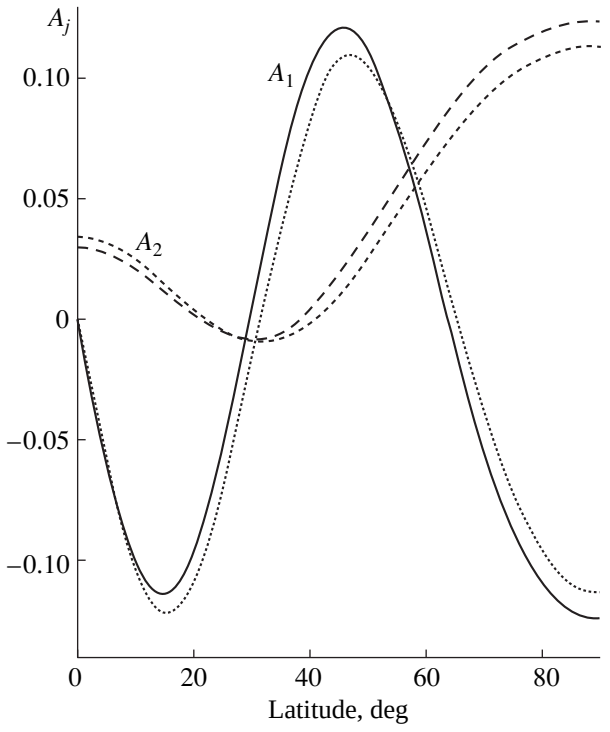


Fig. 2. Latitude distributions of the relative Alfven velocities A_1 and A_2 for model 1 (solid and dashed lines, respectively) and model 7 (densely and sparsely dotted curves, respectively) from Table 1.

4. THE UPPER ZONE OF FIELD GENERATION

In the convective-zone layers that are close to the surface, the parameter s (the ratio of the radius to the characteristic radial density scale) becomes large. In connection with the sharp decrease in the angular velocity of the solar rotation observed in this zone (see, for example, [1]), it is of interest to consider equilibrium configurations with large and negative α in equations (13). As noted in Section 1, the possibility of realization of a magnetic-field generation zone under such conditions is an important question. We shall restrict our consideration to the case of toroidal magnetic fields, when equations (13) and (15)–(21) are satisfied. As in the previous section, we shall find the model parameters by minimizing the functional (19) for a fixed value of s imposing the normalization $u_1 = 1$. It became clear that this minimum is reached at small values of N; therefore, we shall take $N = 2$ below.

Table 3. Parameters of models 1, 4, and 7 from Table 1

Number	1	4	7
s	5	8	10
α	1.297809	1.297546	1.295991
d	1.042592	1.042688	1.042529
u_3	−0.054386	−0.053666	−0.053039
u_5	−0.013068	−0.011237	−0.010330
u_7	−0.004146	−0.003142	−0.002641
q_1	0.010670	0.011812	0.010724
q_3	0.012449	0.005680	0.003552
q_5	0.053827	0.053847	0.053875
q_7	−0.018673	−0.017655	−0.016087

The parameters of the resulting models are given in Table 4, which does not contain the non-axially-symmetric velocity components, since they turned out to be negligible. As above, we identify two models as a pair if they have sufficiently similar rotational-velocity distributions, with one of them having a non-axially-symmetric (*ns*) and the other an axially-symmetric (*si*) field. We consider toroidal magnetic fields that are antisymmetric with respect to the equatorial plane. The ϵ_J defined by (19) are also presented in Table 4, where the parameter k is equal to $(J + 1)/2$ if J is odd or $J/2$ if J is even. The values of ϵ_J presented in the last line characterize the amplitude of the turbulent viscosity that is necessary to suppress higher-order modes. We can see that the modes under consideration are not completely free of viscosity, but the turbulent viscosity required to

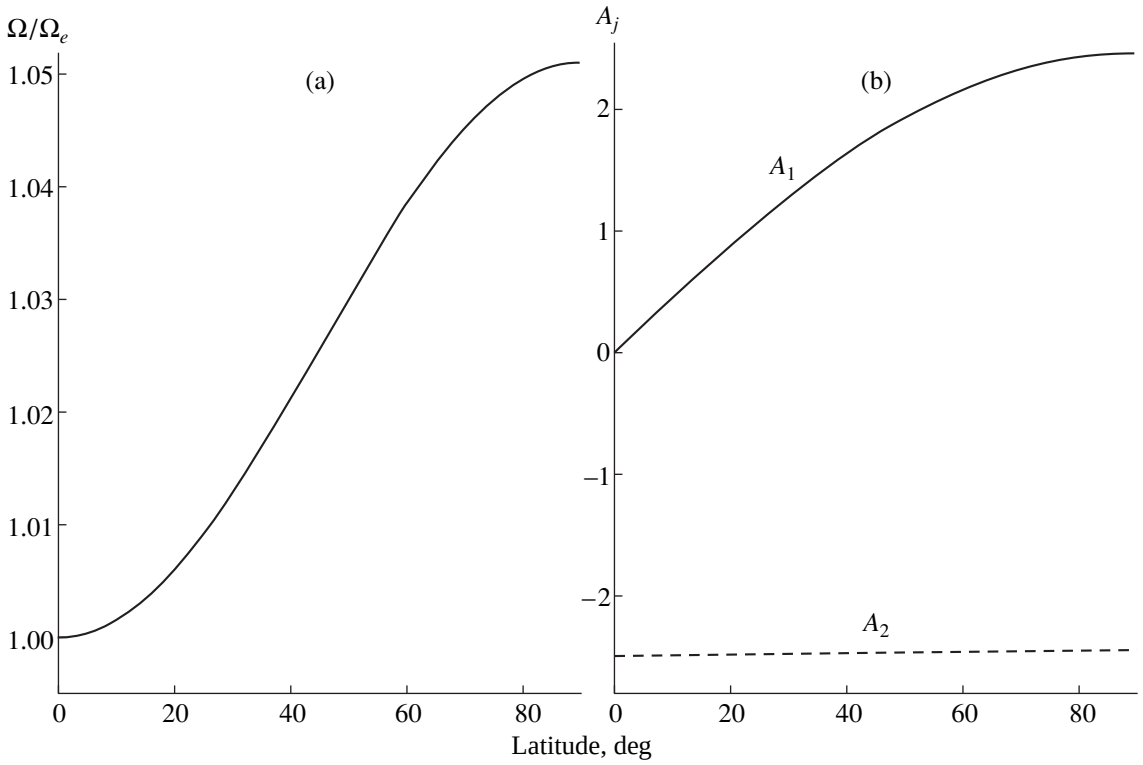


Fig. 3. Latitude dependences of (a) the relative angular velocity Ω normalized to the average angular velocity at the equator Ω_e and (b) the relative Alfven velocities A_1 and A_2 for model 1 from Table 4.

compensate these small residuals is over three orders of magnitude lower than the total viscosity. All other notation in Table 4 is the same as that used previously. Note also that the chosen value of $s = 50$ approximately corresponds to the relative radius $r/R_\odot \approx 0.96$. Realization of the models under consideration seems unlikely if s is appreciably less than 50.

Figure 3 presents distributions of the relative angular velocity and components of the relative Alfven velocity for model 1 from Table 4. The rotation obtained is quite close to rigid-body, and deviations from rigid-body rotation are even smaller for the other models in Table 4. Therefore, the latitude dependence of the rotational velocities for the modes under consideration does not coincide with the behavior observed for the Sun. Observations of the rotation of solar zones in which the magnetic field is carried to the surface reveal a more complex pattern for the solar rotation [15]. Namely, the magnetic field rotates slightly faster than the ambient medium at latitudes below approximately 45°, and the corresponding difference in rotational velocity becomes larger at higher latitudes, so that the angular velocity increases with latitude.

Since we are considering a rather thin subsurface layer, we cannot exclude the possibility that it is present inside a differentially-rotating medium. We can also consider a generation zone located only at latitudes above 45°. We can see in Fig. 3a that the latitude dependence of the angular velocity in this region is in qualitative agreement with that obtained by Stenflo (see Fig. 4 in [15]). Note also that realization of this theoretical model for the generation zone could result in a sharp deceleration of the rotation of matter in the subsurface layers of the solar convective. The characteristic amplitude of the magnetic field corresponding to the curves in Fig. 3b is several tens of kG.

Table 4. Models for the upper field-generation zone

Number	1	2	3	4
Type	*ns*	*si*	*ns*	*si*
N	2	2	2	2
s	50	50	50	50
α	−34.0043	−33.9826	−50.0000	−50.4109
d	0.98992	1.00014	1.00259	1.00001
$A_{\max}$	2.4875	0.67400	1.7243	0.72023
u_3	0.01078	−0.00014	−0.00277	−0.00001
q_1	−1.73609	0.58078	1.20893	0.64626
q_3	0.01328	0.04636	−0.03532	−0.00670
$(\epsilon_J)_{k=1}$	2×10^{-6}	2×10^{-8}	2×10^{-7}	9×10^{-9}
$(\epsilon_J)_{k=2}$	1×10^{-8}	9×10^{-11}	6×10^{-9}	1×10^{-10}
$(\epsilon_J)_{k=3}$	6×10^{-3}	7×10^{-1}	2×10^{-2}	7×10^{-2}

5. EQUILIBRIUM MODELS WITHOUT VISCOSITY

When we consider the general nonlinear system of differential equations for the equilibrium state of a rotating, magnetized, non-viscous medium under the action of gravity, serious problems arise due to the fact that, in many cases, there is no solution of this system. This can be illustrated using the example of a non-magnetized, rotating, isothermal medium possessing axial symmetry with respect to the equatorial plane. In this case, we obtain for the coefficients $v_{JM}^{(0)}$ with odd J a system of Eqs. (14)–(17) where only the coefficient $R_{J0}^{(0)}$ is not identically equal to zero, and the second term in braces in Eq. (15) is absent. Note that, in these equations, J_1 and J_2 are odd, but J is even; in addition, $R_{J0}^{(0)} = 0$ when $J = 0$.

We can see that the number of equations corresponding to the number of non-zero coefficients $C_{J_1 0 J_2 0}^{J0}$ is determined by the well-known triangle rule, $|J_1 - J_2| \le J \le J_1 + J_2$. When only coefficients with subscripts $J_1 = J_2 = 1$ are not equal to zero in the equation for $v_{JM}^{(0)}$—i.e., the number of modes $N = 1$—there is only one equation (15) when $J = 2$ for the quantity α. We assume here that equations (13) are valid, and that the normalization with fixed $v_{10}^{(0)}$ is used. On the other hand, if the coefficients J_1 and J_2 are equal to either 1 or 3—i.e., $N = 2$—then (15) reduces to three equations (for $J = 2$, 4, and 6) for the two unknowns α and $v_{30}^{(0)}$. Therefore, in this case, there is a discrepancy between the number of equations and the number of unknowns. Moreover, as can easily be shown, this discrepancy increases with N. Thus, the system of equations under consideration becomes unsolvable if $N > 1$. The same is true for more general conditions. For example, in the case of a rotating, magnetized configuration in the radiative zone, the second term on the left-hand side of (4), which depends on the temperature distribution, can play a substantial role. The variables appearing in the additional term under discussion are governed by other equations, so that the conclusion formulated above remains valid.

Thus, the single-mode model (where $\alpha = 1$ and the rotation is rigid-body) is distinguished from the other models by the fact that the series for $v_{J0}^{(0)}$ is truncated right at the beginning. It is interesting to study more general models of this kind, since the construction of such models is not associated with the suppression of high-order harmonics by some additional forces. In this section, we shall consider a particular solution of the equilibrium equations for a magnetized, rotating medium ($v_{10}^{(0)} \neq 0$) assuming that the first subscript $J = 1$ for all coefficients $v_{JM}^{(\lambda)}$ and $B_{JM}^{(\lambda)}$ (if this subscript equals two, the series will not be truncated).

In general, the coefficients of the nonlinear force $Q_{JM}^{(\lambda)}$ in (2) under these conditions can have $J = 0$, 1, and 2, with $\lambda = 0$ or ± 1 and $|M| \le J$. Explicit expressions for these coefficients are given by formulas (22)–(24) from [12]. It turns out that, when the conditions

$$\begin{aligned} &v_{10}^{(-1)} = 0, \quad v_{1\pm1}^{(-1)} = 0, \quad v_{10}^{(+1)} = 0, \\ &v_{1\pm1}^{(+1)} = 0, \quad B_{10}^{(0)} = 0, \quad B_{1\pm1}^{(0)} = 0, \end{aligned} \tag{23}$$

$$B_{1-1}^{(-1)} \frac{\partial^2}{\partial r^2} r^2 B_{11}^{(-1)} - B_{11}^{(-1)} \frac{\partial^2}{\partial r^2} r^2 B_{1-1}^{(-1)} = 0, \tag{24}$$

$$B_{10}^{(-1)} \frac{\partial^2}{\partial r^2} r^2 B_{1\pm1}^{(-1)} - B_{1\pm1}^{(-1)} \frac{\partial^2}{\partial r^2} r^2 B_{10}^{(-1)} = 0 \tag{25}$$

are satisfied, all the $Q_{JM}^{(\lambda)}$ will be identically equal to zero apart from $Q_{00}^{(-1)}$ and $Q_{2M}^{(\pm1)}$, where $\lambda = 0, \pm 1$ and $M \le J$ [16]. In this case, the problem of finding the equilibrium condition reduces to the standard form, in which the vector fields $\mathbf{v}$ and $\mathbf{B}$ have the form

$$\begin{aligned} \mathbf{v} = \frac{1}{4}\left(\frac{3}{\pi}\right)^{1/2} &\{\mathbf{i}_\vartheta (v_{11}^{(0)} e^{i\varphi} + v_{1-1}^{(0)} e^{-i\varphi}) \\ &+ i\mathbf{i}_\varphi [2^{1/2} v_{10}^{(0)} \sin\vartheta + (v_{11}^{(0)} e^{i\varphi} - v_{1-1}^{(0)} e^{-i\varphi}) \cos\vartheta]\}, \end{aligned} \tag{26}$$

$$\begin{aligned} \mathbf{B} = \frac{1}{4}\left(\frac{3}{\pi}\right)^{1/2} &\{\mathbf{i}_r [2B_{10}^{(-1)} \cos\vartheta \\ &- 2^{1/2} (B_{11}^{(-1)} e^{i\varphi} - B_{1-1}^{(-1)} e^{-i\varphi}) \sin\vartheta] \\ &- \mathbf{i}_\vartheta [2^{1/2} B_{10}^{(+1)} \sin\vartheta + (B_{11}^{(+1)} e^{i\varphi} - B_{1-1}^{(+1)} e^{-i\varphi}) \cos\vartheta] \\ &- i\mathbf{i}_\varphi (B_{11}^{(+1)} e^{i\varphi} + B_{1-1}^{(+1)} e^{-i\varphi})\}. \end{aligned} \tag{27}$$

Here, the right-hand sides are real if $v_{1-1}^{(0)} = v_{11}^{(0)*}$ and $B_{1-1}^{(\pm1)} = -B_{11}^{(\pm1)*}$; $i v_{10}^{(0)}$ and $B_{10}^{(-1)}$ are real; and an asterisk denotes complex conjugation. In addition, we find from the relation $\mathrm{div}\,\mathbf{B} = 0$ that $B_{JM}^{(+1)} = 1/\{r[J(J+1)]^{1/2}\}(\partial/\partial r)\,[r^2 B_{JM}^{(-1)}]$. In the case of field $\mathbf{v}$, we have $\mathrm{div}\,\rho\mathbf{v} = 0$.

Equations (24) and (25) are satisfied if the radial dependences of all magnetic modes in (27) are the same; i.e., if we consider a single field with dipolar geometry, whose axis can be inclined at some angle to the rotational axis. In this case, the field is represented in a rotated (primed) coordinate system as the sum over M' of the quantities $B_{1M'}^{(\lambda)} \mathbf{Y}_{1M'}^{(\lambda)}(\vartheta', \varphi')$, where the coefficients $B_{1M'}^{(\lambda)}$ are given and $\lambda = \pm 1$. If we denote the Euler rotation angles α_e, β_e, and γ_e, then, using the formulas presented by Varshalovich *et al.* [17], we find that the

sum over M' of the $a_{MM'}B_{1M'}^{(\lambda)}$ can be substituted into Eqs. (24), (25), and (27) in place of $B_{1M}^{(\lambda)}$, where

$$a_{MM'} = \frac{1}{2}e^{-i(M\alpha_e + M'\gamma_e)}[(1 + MM'\cos\beta_e)\delta_{|M|\,1}\delta_{|M'|\,1} + 2^{1/2}(\sin\beta_e)(M'\delta_{|M|\,0}\delta_{|M'|\,1} - M\delta_{|M|\,1}\delta_{|M'|\,0}) + 2(\cos\beta_e)\delta_{M\,0}\delta_{M'\,0}], \quad (28)$$

$\delta_{ab} = 1$ if $a = b$, and $\delta_{ab} = 0$ if $a \neq b$.

It follows from (26) that velocity can possess a component with a harmonic dependence on the azimuth angle. In this case, the meridional velocity, which is antisymmetric with respect to the equatorial plane, does not depend on latitude ϑ. The second term in the expression for the zonal velocity (in contrast to the first term) is also antisymmetric with respect to the equatorial plane. In the case of a magnetic field described by (27), the components that are dependent on and independent of the azimuth angle also have different symmetries with respect to the equatorial plane, with the axially-symmetric field being antisymmetric with respect to this plane. Note that (27) coincides with (26) from [13] in the case of a non-rotating medium and axially-symmetric dipolar magnetic field.

It can be shown that, in the absence of a magnetic field, the coefficients $v_{1M}^{(0)}$ in (26) are proportional to r. If the rotational velocity is negligible, then the $B_{1M}^{(-1)}$ are either constant or proportional to r^{-3} [16]. In general, the radial dependence is determined by the equilibrium conditions. We present an as example the solution for some isothermal, non-viscous layer in which

$$\mathbf{R} = \mathrm{curl}\mathbf{Q} = 0, \quad (29)$$

and we assumed that radial dependences of the equilibrium density and any modes with dimensions of velocity are approximately the same as in (13). Then, the coefficients $v_{1M}^{(0)}$ and $B_{1M}^{(-1)}$ in Eqs. (26) and (27) are determined by the two relations [16]

$$(\alpha - 1)v_{1M}^{(0)}v_{1M}^{(0)*} = \left(\alpha - \frac{s}{2}\right)\left(\alpha - \frac{s}{2} + 3\right) \times \left(\alpha + \frac{s}{2} - 2\right)\frac{1}{8\pi\rho}B_{1M}^{(-1)}B_{1M}^{(-1)*}, \quad (30)$$

where $M = 0$ or 1. We can see that the equations for the components with and without axial symmetry are separated. In the case $M = 0$, in accordance with (8), we obtain $v_{10}^{(0)}v_{10}^{(0)*} = (8\pi/3)(r\Omega)^2$. In addition,

$$B_{1M}^{(-1)}B_{1M}^{(-1)*} = 2B_{1M}^{(+1)}B_{1M}^{(+1)*}/(\alpha - s/2 + 2)^2. \quad (31)$$

Of course, relations (30) and (31) must be considered together with the field equation. However, it is obvious that there is no steady-state solution of this system of equations in a medium whose rotation differs from rigid-body. For example, if the fields (26) and (27) do not depend on longitude, the vector $\mathrm{curl}(\mathbf{v} \times \mathbf{B})$ in the field equation is equal to $\mathbf{i}_\varphi[3(2^{1/2})/(8\pi)]iB_{10}^{(-1)}r(\partial/\partial r)(v_{10}^{(0)}/r)\sin\vartheta\cos\vartheta$; i.e., it differs from zero if the ratio $v_{10}^{(0)}/r$ is not constant. Therefore, in the case of non-rigid-body rotation, we expect the generation of an azimuthal magnetic field that is proportional to $\sin 2\vartheta$.

From this point of view, it is interesting to consider structures whose angular velocity of rotation sharply increases with height, since the probability for the outflow of the generated magnetic field together with frozen-in charged particles toward the upper layers is increased in this case. We have in mind here structures formed in the upper isothermal layers of stellar and planetary atmospheres. These layers are usually partially or even completely ionized. The outflow of field and matter from the atmosphere corresponds to a decrease in the order (or an increase in the disorder) of the system; i.e., the total entropy increases. Consequently, the self-organization of this outflow of the field and frozen-in matter is in agreement with the laws of irreversible thermodynamics. It is not ruled out that the field outflow will be accompanied by the excitation of some irregular wave processes.

Taking into account the above, we shall consider models for which

$$s \gg 1, \quad \alpha \gg 1, \quad (32)$$

i.e., the ratio of the characteristic radial scale of the equilibrium density to the radius of a star or planet must be small and the amplitudes of all modes with dimensions of velocity must sharply increase with height. If, for example, the difference $\alpha - s/2$ is not very large, then the magnetic-field intensity will be close to uniform. In addition, we can see that equations (30) lead to the condition $(\alpha - s/2)(\alpha - s/2 + 3) > 0$ if the inequalities (32) are satisfied.

One characteristic feature of the isothermal models under consideration is super-rotation of the upper layers. In the case of the Earth's atmosphere, the presence of super-rotation at heights from 150 km to 400 km has been confirmed by observations of variations in the inclinations of satellite orbits [18]. On average, the angular velocity of rotation can reach a factor of 1.3 higher than its normal value. If the height of the lower boundary of the super-rotation zone is taken to be about 200 km, then the corresponding density will be of the order of $10^{-12.5}$ g/cm^3 [19]. In this case, in accordance with (30) and (31), the magnetic field will be about 10^{-1} G. The isothermal condition is apparently approximately satisfied in this zone. The resulting field amplitude is quite close to that expected for such heights.

Relations (30) and (32) can be satisfied if the extent of the zone of super-rotation is sufficiently small. For example, for a 30-km zone with $s \sim 100$, $\alpha \sim s/2$ (but $\alpha - s/2$ and $\alpha - s/2 + 3$ not close to zero), the angular

velocity of the upper layers of the zone should be approximately a factor of 1.4 higher than the velocity in lower layers. The variation in density of the medium within the zone is less than a factor of two. Therefore, the equations derived can be used to describe the comparatively moderate super-rotation observed in the Earth's atmosphere, which is characterized by a small angle of inclination of the dipolar magnetic field.

A strong zone of super-rotation forms in the Venusian atmosphere at heights above approximately 50 km (see the reviews [8–10]). In this case, the planet itself and its upper atmosphere rotate factors of 243 and four more slowly than the Earth, respectively, and at least the lower atmosphere is convective. A solution describing rapid rotation with circulation in the main convective zone (at heights below 64 km) was obtained by Young and Pollack [11], but the mechanism for global zonal (four-day) circulation in the upper atmosphere remains unclear [10]. Note in this connection that the symmetry of the circulation in our model differs from that chosen by Young and Pollack in [11].

In the Venusian atmosphere, there is apparently an isothermal zone at a height of about 100 km or slightly higher [9]. If the density in this zone is of the order of 10^{-9} g/cm^3, then, in accordance with (30) and (31), we find $B \sim 1$ G. This field will be lower if we consider either higher layers or a configuration in which, in accordance with (32), the ratio α/s is small. Moreover, if the super-rotation zone is sufficiently wide, the increase in the angular velocity will be substantial. Therefore, our suggestion that our solution can describe the strong super-rotation observed in the upper atmosphere of Venus seems quite reasonable. Meridional circulation of matter can also be excited in this case. The observational data indicate that the characteristic velocity of this circulation is only an order of magnitude lower than the main zonal velocity. This circulation facilitates heat transfer between the equatorial and polar regions.

It is known that strong outflows of matter are observed on the Sun. If a super-rotation regime is formed in some very high layers of the solar atmosphere, the associated increase in the outflow of matter could be appreciable. This problem requires more careful study.

6. DISCUSSION

Our theoretical studies indicate an important role for nonlinear interactions between various modes in rotating, magnetized stellar and atmospheric zones. It is also important that using expansions of vector fields in orthogonal vector spherical harmonics makes it possible to separate the variables in the framework of exact equations. The above-mentioned harmonics are characterized by very complex behavior near the poles, so that the accuracy required when solving these nonlinear equations using approximate methods can be high.

Our results show that the conditions for the formation of complex rotational patterns are substantially different in convective and non-convective zones. We do not consider here structures whose asymmetry is produced by chemical irregularities in the form of spots and so on. Our studies indicate that the formation of a differentially rotating configuration (for example, of the type observed for the Sun) is impossible in the case of a non-convective zone with a non-dipolar magnetic-field geometry, if the series describing the distributions of the rotational velocity and field are not truncated. We considered this problem assuming that the series are truncated right at the beginning. Of course, we have not obtained a rigorous proof that this type of truncation is unique. Nevertheless, it seems most probable that other types of truncation of the series are not possible.

In the non-magnetic case, the solution obtained is a sum of axially-symmetric and non-axially-symmetric components of the rotational velocity [see (26)]. Since, in this equation, $iv_{10}^{(0)}(r)$ is real and the coefficient $v_{11}^{(0)}(r)$ can also be taken to be real, we obtain for the hydrodynamical velocity

$$\mathbf{v} = \left(\frac{3}{4\pi}\right)^{1/2}\left[\frac{1}{2^{1/2}}\mathbf{i}_\varphi(iv_{10}^{(0)})\sin\vartheta + v_{11}^{(0)}(\mathbf{i}_\vartheta\cos\varphi - \mathbf{i}_\varphi\cos\vartheta\sin\varphi)\right]. \quad (33)$$

Note also that the vortex component of the Coriolis force equals zero if the radial dependence of the velocity **v** is linear.

If a magnetic field is present, its geometry is dipolar, and an azimuth-dependent component can exist. This solution is described by Eq. (27). It is important that the angle between the rotation and field axes can be arbitrary. The existence of magnetic stars whose field axes are rotated with respect to their rotational axes has been confirmed by observations (see the reviews [20, 21]). The observed strong field variations can be explained by the presence of a large longitude-dependent field component [see Eq. (27)].

In the case of an isothermal zone with uniform chemical composition, the equilibrium equations for the vortex and potential forces do not depend on each other if the density in the expression for the magnetic force is approximately spherically symmetric. Considering the equation for the vortex forces, we see that an equilibrium can be achieved, but the condition that the magnetic field be stationary will be violated. Nevertheless, we believe that, in the case of upper atmospheric layers and structures whose rotational velocity sharply increases with height, motions carrying the additional field together with frozen-in matter toward layers located at greater heights can be excited. The formation of such quasi-equilibrium structures, which results in a decrease in their order (or an increase in their disorder; i.e., an increase of the total entropy), is in agreement with the laws of irreversible thermodynamics. A steady-state

magnetic field in the structures under consideration possesses an inclined dipolar geometry, so that azimuth-dependent components of the field and angular velocity can be present.

Observations show that the quasi-equilibrium structures discussed here are often realized in stellar and planetary atmospheres. For example, in Section 5, we presented data indicating the formation of such structures in the upper atmospheres of Earth, Venus, and the Sun. It is not ruled out that the formation of stellar winds is based on the same phenomenon.

The models described in this section seem to represent the entire (very limited) set of rotating, magnetized configurations whose formation is allowed by the equilibrium equations when the viscosity of the medium is negligible. The fact that considerably more complex rotational structures are often observed in convective zones suggests a substantial role for turbulent convective viscosity in the formation of these structures. The differential rotation of a medium cannot be directly supported by convection from the viewpoint of irreversible thermodynamics; therefore, we are forced to conclude that the role of convective viscosity is to suppress certain higher-order rotational modes, whose presence would prevent the realization of some additional conditions associated with facilitating heat transfer or decreasing energy losses that support the convective zone itself. This suggests that this last condition is satisfied in the case of the solar convective zone, where the state with maximum efficiency of transformation of the input energy to magnetic-field energy seems to be realized.

To determine the parameters of the solar magnetic-field generation zone, we use the condition that the angular velocity of rotation be only weakly sensitive to any variations in the magnetic field. In the case of axially-symmetric magnetic fields, the most extensive layer where this condition is satisfied corresponds to the entire lower half of the solar convective zone [3]. The results in Section 3 testify that the conclusions of [3] for this lower region remain valid if we exchange the axially-symmetric field with a toroidal field characterized by a $\sin\varphi$ azimuth dependence. The upward motion of this field could result in the formation of sunspots.

Of course, it would be desirable to study the dimensions of the generation zone under consideration by comparing models with various fields without axial symmetry. The upper boundary of this zone seems to be located in or above the middle of the solar convective zone, and the amplitude of the asymmetric magnetic field has the same order of magnitude as in [3]. In this case, the conclusion of [3, 22] that approximately 0.1% of the total energy entering the zone is transformed into magnetic-field energy is valid. The fact that this efficiency for the energy transformation is consistent with the observational estimate cited in [23] was already noted in [3, 22]. It is also important that the latitude distribution of the rotational velocity obtained in our study (without introducing any fitting parameters) is quite close to that actually observed for the Sun.

At the same time, many important details of the field generation in our model remain unclear. In particular, there are large uncertainties in connection with the radial gradient of the rotational velocity, as well as the possible latitude dependence of this gradient. These questions, along with the problem of the cyclic variations in the direction of the rising field, should be clarified in the course of subsequent studies.

It is also unclear if a second subsurface zone of solar magnetic-field generation can exist in the region of large, negative, radial gradients of the rotational velocity. Section 4 considered equilibrium models whose realization was made possible by the presence of a thin zone where the rotational velocity sharply decreases with height. This zone could be self-supported if there is a transformation of thermal to magnetic energy in it. Here, we assume that this is a realization of the state that requires minimum energy loss for its support, as discussed above. However, the models obtained are characterized by angular velocities that differ from those observed for the Sun. On the other hand, if only some high-latitude zone of field generation is considered, then the rising magnetic fields could have nearly rigid-body rotation, as detected by Stenflo [15]. However, the question of how to realize such a model remains rather controversial.

Attention should be also given to the problem of super-rotation of the lower non-magnetized convective atmosphere of Venus. Neglecting high-order rotational modes, Young and Pollack [11] obtained a solution that is consistent with the observations, but the discarding of some modes was not justified [24, 25]. Note in connection with this that the suppression of high-order rotational modes by convection follows directly from the analysis of non-rigid-body rotation of the convective zones we presented above. In general, this problem requires further more detailed study.

APPENDIX

A representation of the vector curl $\mathbf{v}_1 \times \mathbf{v}_2$ as an expansion in the vector spherical harmonics was considered in [12]. In the case of the product of two arbitrary vectors, we obtain

$$\mathbf{X} = \mathbf{v}_1 \times \mathbf{v}_2, \tag{34}$$

where

$$X_{JM}^{(-1)} = -i \sum_{J_1 M_1 J_2} C_{J_1 M_1 J_2 M_2}^{JM} \left[\frac{J(J+1)}{2J+1}\right]^{1/2}$$
$$\times \{\Theta_{J_1 J_2}^{J} [v_{J_1 M_1}^{(+1)} v_{J_2 M_2}^{(+1)} - v_{J_1 M_1}^{(0)} v_{J_2 M_2}^{(0)}] \tag{35}$$
$$+ \frac{1}{2} K_{J_1 J_2}^{J} Z_{J_1 J_2}^{J} [v_{J_1 M_1}^{(+1)} v_{J_2 M_2}^{(0)} - v_{J_1 M_1}^{(0)} v_{J_2 M_2}^{(+1)}]\},$$

$$X_{JM}^{(+1)} = -i\sum_{J_1M_1J_2} C_{J_1M_1J_2M_2}^{JM}\left[\frac{1}{2J+1}\right]^{1/2}$$
$$\times\{\Theta_{J_1J_2}^{J}([J_1(J_1+1)]^{1/2}v_{J_1M_1}^{(-1)}v_{J_2M_2}^{(+1)}$$
$$+[J_2(J_2+1)]^{1/2}v_{J_1M_1}^{(+1)}v_{J_2M_2}^{(-1)}) \quad (36)$$
$$+\frac{1}{2}K_{J_1J_2}^{J}(-[J_1(J_1+1)]^{1/2}Z_{JJ_2}^{J_1}v_{J_1M_1}^{(-1)}v_{J_2M_2}^{(0)}$$
$$+[J_2(J_2+1)]^{1/2}Z_{JJ_1}^{J_2}v_{J_1M_1}^{(0)}v_{J_2M_2}^{(0)})\},$$

$$X_{JM}^{(0)} = -i\sum_{J_1M_1J_2} C_{J_1M_1J_2M_2}^{JM}\left[\frac{1}{2J+1}\right]^{1/2}$$
$$\times\{\Theta_{J_1J_2}^{J}([J_1(J_1+1)]^{1/2}v_{J_1M_1}^{(-1)}v_{J_2M_2}^{(0)}$$
$$+[J_2(J_2+1)]^{1/2}v_{J_1M_1}^{(0)}v_{J_2M_2}^{(-1)}) \quad (37)$$
$$+\frac{1}{2}K_{J_1J_2}^{J}([J_2(J_2+1)]^{1/2}Z_{JJ_1}^{J_2}v_{J_1M_1}^{(+1)}v_{J_2M_2}^{(-1)}$$
$$-[J_1(J_1+1)]^{1/2}Z_{JJ_2}^{J_1}v_{J_1M_1}^{(-1)}v_{J_2M_2}^{(+1)})\}.$$

Here, the coefficients $K_{J_1J_2}^{J}$ and $\Theta_{J_1J_2}^{J}$ are determined by formulas (10) and (11) from [12], and the last coefficient can also be written

$$\Theta_{J_1J_2}^{J} = C_{J_10J_20}^{J-1\,0}\left[\frac{(n_*+1)(n_*-J+1)(2n_*-2J_1+1)(2n_*-2J_2+1)}{4\pi KK_1K_2(2J-1)}\right]^{1/2}, \quad (38)$$

where $K_l = l(l+1)/(2l+1)$, $l = J$, J_1, or J_2; $n_* = (J+J_1+J_2-1)/2$; and $\Theta_{J_1J_2}^{1} = (-1)^{J_1}\{3(2J_1+1)/[8\pi J_1(J_1+1)]\}^{1/2}\delta_{J_1J_2}$.

The following expression represents a generalization of formula (25) from [12]:

$$\rho\mathbf{v} = \sum_{J_1M_1J_2} C_{J_1M_1J_2M_2}^{JM}C_{J_10J_20}^{J0}\left[\frac{(2J_1+1)(2J_2+1)}{4\pi(2J+1)}\right]^{1/2}$$
$$\times\rho_{J_2M_2}\left\{v_{J_1M_1}^{(-1)}\mathbf{Y}_{JM}^{(-1)} + \frac{1}{2J(J+1)}Z_{JJ_1}^{J_2}\right. \quad (39)$$
$$\left.\times\left[v_{J_1M_1}^{(+1)}\mathbf{Y}_{JM}^{(+1)} + \left(\frac{J(J+1)}{J_1(J_1+1)}\right)^{1/2}v_{J_1M_1}^{(0)}\mathbf{Y}_{JM}^{(0)}\right]\right\}.$$

Finally, a few words about misprints. The term $\Theta_{J_1J_2}^{J}$ [$v_{J_2M_2}^{(+1)}$] should be inserted after the brace in Eq. (22) from [12], and all quantities in Eq. (1) of [12] should be scalar. The superscripts of ξ in the denominator of the second formula from the Appendix to [12] should be LJ. Expressions (46) from [12] and (30) from [13] are very rough, and should be replaced by the exact Eq. (29) of [13]. In addition, the superscript of the coefficient Z in the third and fourth lines of the Table 2 A4 from [12] should be J (instead of J_2). The numbers 0.6894 and 0.7485 in columns 9 and 10 in the seventh line of Table 5 from [3] should be interchanged. In addition, the numbers 0, 8, and 1 in the 17th line from the bottom of the right column on p. 874 should be replaced by 20, 8, and 1. Finally, Fig. 1 in [21] should be replaced by Fig. 3 from [3].

ACKNOWLEDGMENTS

This work was supported by the "Integratsiya" Foundation (project no. KO854) and the Russian Foundation for Basic Research (project code 00-02-16939).

REFERENCES

1. J. Schou, H. M. Antia, S. Basu, *et al.*, Astrophys. J. **505**, 390 (1998).
2. I. Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences* (Freeman, San Francisco, 1980; Nauka, Moscow, 1985).
3. Yu. V. Vandakurov, Pis'ma Astron. Zh. **25**, 868 (1999) [Astron. Lett. **25**, 758 (1999)].
4. R. Howard, Ann. Rev. Astron. Astrophys. **22**, 131 (1984).
5. E. E. Benevolenskaya and V. I. Makarov, Pis'ma Astron. Zh. **18**, 195 (1992) [Sov. Astron. Lett. **18**, 108 (1992)].
6. E. E. Benevolenskaya, Astrophys. J. Lett. **509**, L49 (1998).
7. V. N. Obridko and Yu. Z. Rivin, Astron. Zh. **73**, 812 (1996) [Astron. Rep. **40**, 740 (1996)].
8. G. Schubert, C. C. Counselman, III, J. Hansen, *et al.*, Space Sci. Rev. **20**, 357 (1977).
9. M. Ya. Marov, *Planets of the Solar System* [in Russian] (Nauka, Moscow, 1981).
10. K. Ya. Kondrat'ev, N. N. Krupenio, and A. S. Selivanov, *The Planet Venus* [in Russian] (Gidrometeoizdat, Leningrad, 1987).
11. R. E. Young and J. B. Pollack, J. Atmos. Sci. **34**, 1315 (1977).
12. Yu. V. Vandakurov, Astron. Zh. **76**, 29 (1999) [Astron. Rep. **43**, 24 (1999)].
13. Yu. V. Vandakurov, Pis'ma Astron. Zh. **25**, 143 (1999) [Astron. Lett. **25**, 111 (1999)].
14. A. C. Birch and A. G. Kosovichev, Astrophys. J. Lett. **503**, L187 (1998).

15. J. O. Stenflo, Astron. Astrophys. **233**, 220 (1990).

16. Yu. V. Vandakurov, Zh. Tekhn. Fiz. (2000) (in press).

17. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).

18. D. G. King-Hele and D. M. C. Walker, Planet. Space Sci. **25**, 313 (1977).

19. C. W. Allen, *Astrophysical Quantities* (Athlone Press, London, 1973; Mir, Moscow, 1977).

20. E. F. Borra, J. D. Landstreet, and L. Mestel, Annu. Rev. Astron. Astrophys. **20**, 191 (1982).

21. V. L. Khokhlova, Itogi Nauki Tekh., Ser. Astron. **24**, 233 (1983).

22. Yu. V. Vandakurov, in *Proc. of the Conference "Large-scale Structure of Solar Activity: Advances and Aspects,"* St. Petersburg, 1999, p. 25.

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

24. W. B. Rossow, S. B. Fels, and P. H. Stone, J. Atmos. Sci. **37**, 250 (1980).

25. R. E. Young and J. B. Pollack, J. Atmos. Sci. **37**, 253 (1980).

Translated by Yu. Dumin

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 230–235. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 269–275.

The Thermal Evolution of Thorne–Zytkow Objects

M. V. Barkov[1], G. S. Bisnovatyĭ-Kogan[2], and S. A. Lamzin[1]

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

[2]*Space Research Institute, ul. Profsoyuznaya 84/32, Moscow, 117810 Russia*

Received January 25, 2000

Abstract—The evolution of stars with mass $5M_\odot$ with an initially cool neutron core (Thorne–Zytkow objects) is computed numerically, taking into account the heating of the neutron star by flows of heat released during the accretion of a surrounding envelope. The temperature of the neutron core does not rise to values at which the system could become unstable to rapid increases in the neutrino luminosity. In other words, the heating of the neutron core does not lead to a rapid collapse of the initial configuration. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Calculations of the evolution of close binary systems [1] suggest the possibility of scenarios in which a neutron star could be absorbed by a more massive companion, ending up at its center [2, 3]. Since detailed computations describing the penetration of the neutron star into the bulk of the normal star are very complicated, it is of independent interest to elucidate how the final state—a star with a neutron core—will appear after the relaxation of the various transitional processes.

The first attempt to construct a model in which a neutron star is surrounded by an optically thick and relatively massive envelope was made by Zel'dovich *et al.* [4], who formulated the problem as follows. They assumed that a cool neutron star was surrounded by a cloud with an arbitrary, specified density distribution at time $t = 0$ and considered the evolution of the system in the framework of one-dimensional hydrodynamics. These computations showed that, immediately after the onset of accretion, a shock wave forms and moves from the surface of the neutron star to the outer boundary of the cloud. The material behind the shock front is decelerated and heated to $T \approx 10^{10}$ K, leading to the generation of powerful neutrino emission, greatly exceeding the photon luminosity. The further evolution of the gas envelope consists of quasi-static (Kelvin) compression in a high-neutrino-luminosity regime.

In 1975, the work of Thorne and Zytkow [5] appeared (for more detail, see [6]), in which they construct models for stars with cool ($T \le 10^9$ K) neutron cores surrounded by massive ($M_{\rm env} > M_{\rm NS} = 1M_\odot$) envelopes. Their model for a star with $M_{\rm tot} = 5\,M_\odot$ outwardly resembles a red supergiant with $L \approx 4 \times 10^4 L_\odot$, $T_{\rm eff} \approx 2600$ K, and $R_* \approx 1000R_\odot$. An important characteristic property of this model is its low neutron luminosity relative to the photon luminosity: $L_\nu/L_{\rm ph} \approx 10^{-5}$. In contrast to [4], where a hydrostatic-equilibrium model was obtained as the result of the evolution of an initial cloud, Thorne and Zytkow [5, 6], having made a number of assumptions, were able to construct a self-consistent hydrostatic model without considering the processes leading to its formation. Their equilibrium model consisted of the following regions.

(1) The envelope (10^4 km $= r_{\rm om} \le r \le R_* \approx 1000R_\odot$)—a region absent of energy sources, in which it is assumed that $L(r)$ = constant. This zone contains most of the mass of the star $M_{\rm env} = 4M_\odot$, and serves as a reservoir of material that gradually settles onto the surface of the neutron star.

(2) The neutron core ($0 \le r \le R_{\rm NS} \approx 11$ km)—a cool neutron star with $M = 1M_\odot$. The computations of [6] adopted the radius corresponding to a density $\rho = 2 \times 10^8$ g/cm^3 in the envelope of the neutron star as the inner boundary for the model. The boundary conditions $M(r) = 1M_\odot$, $\rho(r) = 2 \times 10^8$ g/cm^3, and $L(r) = 0$ when $r = R_{\rm NS}$ were imposed at this boundary.

(3) The transition region $r_{\rm NS} < r < r_{\rm om}$—a zone containing $10^{-8}M_\odot$ of matter and located at the neutron-star surface.

It was assumed that accretion occurs in a quasi-steady-state regime; i.e., the Euler time derivatives of any quantity except for the mass were required to be zero. In this case, from the identity

$$\left(\frac{\partial}{\partial t}\right)_r = \left(\frac{\partial}{\partial t}\right)_{M_r} + \left(\frac{\partial M_r}{\partial t}\right)_r \left(\frac{\partial}{\partial M_r}\right)_t$$

flows the relation

$$\left(\frac{\partial}{\partial t}\right)_{M_r} = -\dot{M}\left(\frac{\partial}{\partial M_r}\right)_t,$$

where

$$\dot{M} = \left(\frac{\partial M_r}{\partial t}\right)_r$$

and the relation

$$\dot{M}(r) = \text{const} = \dot{M},$$

follows from the continuity equation. The assumption of the quasi-stationarity of the evolution of the star enabled Thorne and Zytkow [5, 6] to exclude partial derivatives in time from the stellar-structure equations, and reduce the problem of constructing the desired model to the integration of a system of ordinary differential equations.

Using the example of a star with $M_{tot} = 5M_\odot$, Bisnovatyĭ-Kogan and Lamzin [7] investigated the dependence of the structure of Thorne–Zytkow (T–Z) objects on the indicated initial assumptions. The evolution of the star's envelope was first computed, with a gradual movement of the lower boundary of the region with L_r = const and $\dot{M} = 0$ higher and higher and simultaneously a rejection of the condition $\dot{M}$ = const in lower-lying layers, but, as before, assuming that $L_r = 0$ at the neutron-star surface. It turned out that the structure of the inner regions remains virtually unchanged in this case, and the envelope begins to expand, probably flowing outward in the form of a stellar wind with a mass-loss rate of about $2 \times 10^{-6} M_\odot$/yr.

It was then verified in [7] how sensitive the structure of the star was to the value of ρ_{in} corresponding to the boundary condition $L_r = 0$. Several models were computed, identical to the models of T–Z but with successively increasing values of ρ_{in}. It turned out that it was not possible to construct a model with zero luminosity at the base when $\rho_{in} > 7 \times 10^9$ g/cm^3, since the temperature of the gas at the neutron-star surface becomes higher than 10^9 K, and the intensity of the neutrino outflow sharply increases. This was interpreted as evidence that the initial model was thermally unstable; however, is this really the case? In order to answer this question, it is necessary to include the thermal evolution of the neutron star in the computations, moving the boundary condition $L_r = 0$ to the only physical base in the problem—the center of the neutron star. A qualitative discussion of the instability of T–Z models is carried out in [8].

In the current paper, we present the results of computations of the thermal evolution of T–Z models taking into account the heating of the neutron core of the star by flows of heat released during accretion of the external envelope.

2. INITIAL MODEL FOR THE EVOLUTIONARY COMPUTATIONS

As an initial model for our evolutionary computations, we adopted a T–Z model with mass $M_{tot} = 5M_\odot$ and chemical composition for the envelope $X = 0.7$, $Y = 0.28$. We computed this model by solving the same equations as in [6], allowing for general relativistic effects, adopting the three conditions listed above for the structure of the model.

Our method for computing the outer regions of the star, where $L(r)$ = const, is nearly the same as that described in detail in [9]. The main differences are as follows: (1) the ionization balance for H_2, H, and He was determined via solution of a full system of the corresponding equations using a Newton method, and (2) numerical integration of the differential equations was carried out using a 4th-order Runge–Kutta method with automatic step selection.

When $T < 10^8$ K, we took into account the contributions of electrons, atoms, ions, molecules, and blackbody radiation in the equation of state. At higher temperatures, we considered a mixture of blackbody radiation and fully ionized plasma consisting of non-degenerate, non-relativistic atomic nuclei and electron–positron gas. All components make an additive contribution to the equation of state, and only the lepton gas must be treated in a non-trivial way; when $\log\rho < 10$, its thermodynamic functions were computed using the code of Blinnikov *et al.* [10].

When $\log\rho > 11$, the equation of the state of cool matter ($T = 0$) can be described by the approximation formula of [11], but with allowance for thermal corrections for the non-degenerate nuclei and degenerate electrons. We assumed that the chemical composition of the matter corresponded to so-called non-equilibrium neutronization occurring during compression of the cool material [12]. We used the later results [13] for a specified chemical composition with non-equilibrium neutronization. We matched the equations of state in the regions with $\log\rho < 10$ and $\log\rho > 11$ using a spline interpolation method.

Opacity coefficients in the range $3 \leq \log T \leq 8$ and $-12 \leq \log\rho \leq 12$ were interpolated from the table from [14] for the adopted envelope chemical composition. Beyond the range covered by this table, for temperatures above 10^8 K in the non-degenerate gas, the opacity is determined by Thomson scattering with a Kline–Nishina correction factor [15].

When the lepton gas becomes degenerate ($\log\rho \geq 6$), the main role in the transfer of heat begins to be played by the electron thermal conductivity [16]. We allowed for the possibility of crystallization of the matter, in which case the coefficient of thermal conductivity was computed in a self-consistent way [17]. In the transition region between the crystallization phase and the gas, we interpolated the thermal conductivity. When $\log\rho > 12$, we took into account the contribution of the degenerate neutron gas to the thermal conductivity in accordance with the method of [18]. We computed the convective gradient ∇_{conv} in accordance with mixing-length theory, with the characteristic length equal to α on the pressure height scale. The corresponding equations are

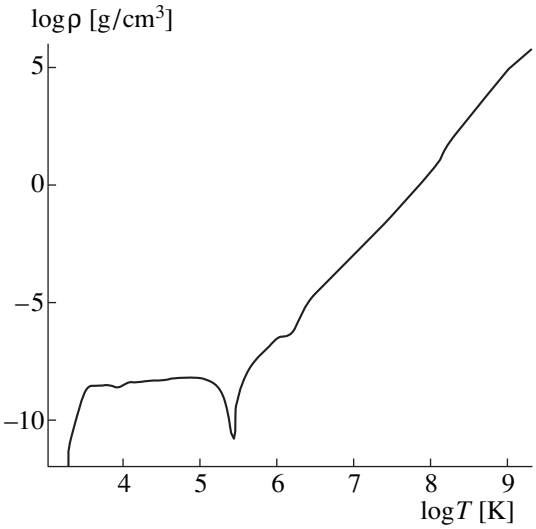


Fig. 1. Variation of the density as a function of temperature in the model with $\alpha = 1$. $L_{tot} = 36972.8L_\odot$, $R_{nr} = 1282.36R_\odot$, and $\dot{M} = 2.3172544 \times 10^{-8} M_\odot$/yr.

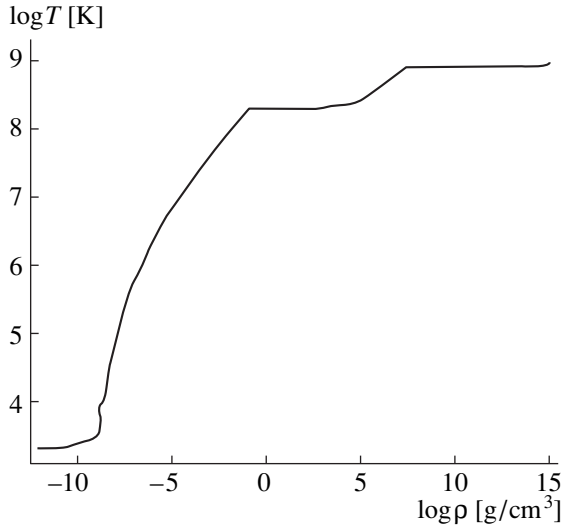


Fig. 2. Variation of the temperature as a function of density in the initial model for the evolutionary computations ($\alpha = 1.5$, $L_{tot} = 53842.64L_\odot$, $R_{nr} = 1229.6R_\odot$, and $\dot{M} = 2.5359 \times 10^{-8} M_\odot$/yr).

presented in [19]; the general relativistic effects reduce to correction factors in the expressions for the height scale H_P and the free-fall acceleration.

The computations took into account thermonuclear burning via 4H ⟶ He, 3He ⟶ C, and 2C ⟶ Mg [20]. Neutrino losses to pair annihilation and free–free, plasma, and Urca processes in free nucleons were computed using formulas from [21].

The only information about the neutron star needed to construct the initial model for the evolutionary computations is the radius r_{NS} for $\rho_{in} = 10^8$ g/cm^3 and a mass $M_{NS} = 1M_\odot$. We computed the value of r_{NS} by solving the Oppenheimer–Volkov equation. Note that, for a mass of $M_{NS} = 1M_\odot$, the values of r_{NS} for isothermal neutron stars with $T = 0$ K and $T = 10^9$ K differ by less than 3 m. In the thermal evolution computations, we assume that such variations of the radius will not influence the overall structure of the model, and can be neglected. This means that the radial temperature distribution of the neutron star can be chosen arbitrarily if $T_{max} < 10^9$ K.

We constructed the initial model as follows. We first chose trial values L_{tot} and T_{eff} and integrated the system of equations to a depth $r_{om} = 0.2R_\odot$ for the conditions $L(r) = L_{tot}$ and $\dot{M} = 0$. The evolutionary computations of [7] showed that, for this value of r_{om} $L(r)$ in the envelope and $\dot{M}$ in the lower-lying layers can indeed be considered virtually constant. We then chose a trial value of $\dot{M}$ and continued the integration to $r = r_{NS}$. In this way, construction of the model was reduced to the selection of values L_{tot}, T_{eff}, and $\dot{M}$ for which the boundary condition $L_r = 0$ was satisfied.

When using the new opacities of [14] with convection parameter $\alpha = 1$ (as did Thorne and Zytkow [6] and Bisnovatyĭ-Kogan and Lamzin [7]), the gas temperature exceeds 10^9 K at a radius of about 1000 km for comparatively small trial values L_{tot}. This leads to a high neutrino luminosity, making it impossible to satisfy the boundary conditions $L_r = 0$ at $\rho = \rho_{in}$ [7]. Increasing L_{tot} decreases the growth in temperature, but leads to a density inversion near $\log T = 5.5$ (Fig. 1). With further increase in L_{tot}, the amplitude of the inverse becomes anomalously large, which is clearly not physical. Thus, we were not able to construct a T–Z model using the new opacities and $\alpha = 1$: convection is not able to compensate the outflow of heat. However, if we adopt $\alpha = 1.5$ (i.e., we make the convection more efficient), it becomes possible to construct a model that satisfies the structure of the object described by the three conditions indicated above.

The integrated characteristics of the model with $M = 5M_\odot$ and $\alpha = 1.5$ are $L_{tot} = 53842.64L_\odot$, $R_{nr} = 1229.6R_\odot$, $\dot{M} = 2.5359 \times 10^{-8} M_\odot$/yr, $r_{NS} = 11.455$ km, and $T_{NS} = T(r_{NS}) = 9.3085 \times 10^8$ K. Figure 2 presents the dependence of $\log\rho$ on $\log T$, and Fig. 3 shows L as a function of $\log\rho$ along the star. Since a detailed qualitative analysis of the structure of the star is presented in [6], we do not present this here.

3. THERMAL EVOLUTION OF THE NEUTRON STAR

To describe the thermal evolution of the neutron star, we used the energy-generation equation re-written in the form

$$\left[\frac{\partial(\Omega^2 L_r)}{\partial M_r}\right]_t = -\Omega^2 \varepsilon_{ov}$$

$$-\left\{\left[\left(\frac{\partial E_T}{\partial \rho}\right)_{Mr} - \frac{P_T}{\rho^2}\right]\left(\frac{\partial \rho}{\partial t}\right)_{Mr} + \left(\frac{\partial E_T}{\partial T}\right)_{Mr}\left(\frac{\partial T}{\partial t}\right)_{Mr}\right\},$$

$$\Omega = \left(1 - \frac{2GM_{tr}}{rc^2}\right)^{1/2}.$$

The heat capacity of the neutron-star matter took into account the corresponding contributions from the degenerate relativistic electrons, degenerate neutrons, and non-degenerate atomic nuclei. In the computations, the time derivatives were replaced by finite differences. E_T and P_T are thermal corrections to the internal energy and pressure, respectively.

We computed the envelope using the method described above assuming $\dot{M}(r)$ = const, with a convection parameter $\alpha = 1.5$. When the luminosity L_{in} at the inner boundary of the envelope (i.e., where $\rho = 10^8$ g/cm^3) varies from $-100L_\odot$ to $100L_\odot$, the envelope mass varies by less than $10^{-8}M_\odot$, the radius by less than 3 m, and the temperature by no more than 60%. On the other hand, it became clear during the computations of the thermal evolution of the neutron core that the absolute value of the luminosity at the outer boundary of the neutron star does not exceed $100L_\odot$. This made it possible to avoid calculating all the envelope parameters at each time step, and to compute *a priori* a grid of envelope models and determine the values of L_{in} and T_{in} needed to match with the core via interpolation.

We computed the evolution of the neutron star using the method of Schwarzschild and a numerical scheme that is stable against the presence of various characteristic times [15]. The initial temperature distribution in the star was chosen from the condition that the heat flux inside the neutron star be zero: without allowing for general relativistic effects, this would imply an isotherm with $T(r) = 9.31 \times 10^8$ K. Taking general relativity into account raises somewhat the temperature at the center of the neutron star: $T_c = 1.1 \times 10^9$ K. Further, we chose the value of T_c so that the temperatures and luminosities at the matching point coincided within 1%. The time step was chosen so that T_c varied by no more than 10% over one step.

Figure 4a shows how the temperature distribution inside the neutron star varies with time. The cooling of the neutron star is associated with volume neutrino losses as a consequence of the modified Urca process. Due to the high thermal conductivity of the Coulomb crystals and degenerate neutrons, the temperature when $\log\rho > 11.5$ is nearly constant, although the heat flux reaches $3 \times 10^4 L_\odot$ near the center (Figs. 4b and 4c). The strong thermal isolation of the outer layers from the envelope prevents intense heating of the neutron star, so that the hot neutron star cools in the same way it would if it didn't have an envelope. Differences begin to be visible when the temperature at the center is decreased to $T \approx 4 \times 10^8$ K. Then, the heat flux from the envelope and the neutrino losses $L_\nu \approx 8 \times 10^{34}$ erg/s are comparable, and the neutron star makes a transition to a steady state (Fig. 4c).

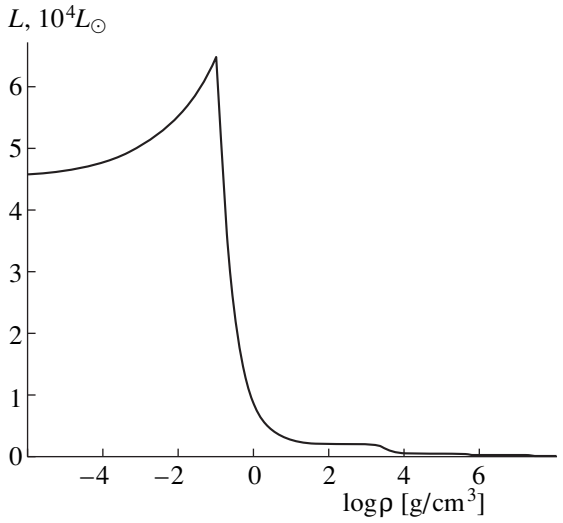


Fig. 3. Dependence of the luminosity on the density in the initial model for the evolutionary computations.

During the time when this steady state is being established, the temperature at the matching point varies within 20% (Fig. 5), and the luminosity at some times is reduced to $-70\ L_\odot$. At the same time, the value of $\dot{M}$ in the envelope differs from the M in the initial model by less than 0.1%. This suggests that the envelope essentially does not feel the variations in the thermal structure of the neutron core.

We can see in Fig. 4c that, in all cases, the luminosity becomes zero at densities lower than 10^8 g/cm^3; then, with the exception of the first time step, it remains everywhere negative, reaching zero only at the center. Bisnovatyĭ-Kogan and Lamzin [7], like Thorne and Zytkow [6], only sought solutions for which the luminosity reached zero when $\rho = \rho_{in}$, remaining everywhere positive. Thus, the fact that they were not able to construct a model with $L_r = 0$ at sufficiently high densities in [7] does not mean that stars with cool neutron cores cannot exist.

To elucidate how strongly the computation results depend on the initial temperature distribution, we carried out evolutionary computations for initial models with neutron stars with $T_c = 5.76 \times 10^8$, 4.5×10^8, and

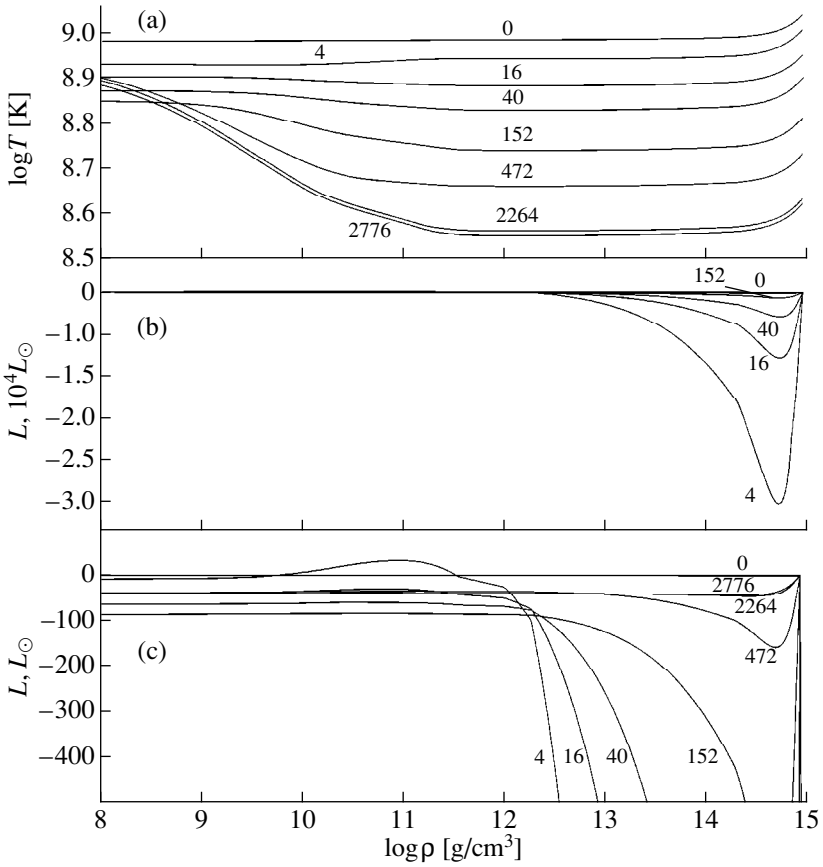


Fig. 4. Thermal evolution of the neutron star: (a) variation of the temperature profile; (b) variation of the luminosity distribution; (c) same as (b) on a coarser scale. The numbers near the curves show the time in years from the onset of the evolution.

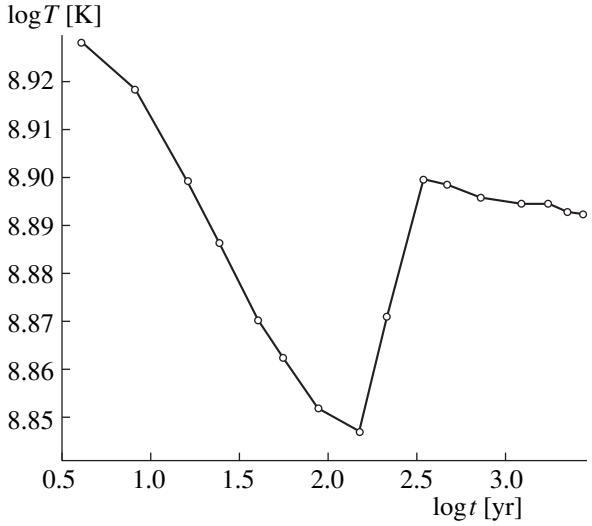


Fig. 5. Time dependence of the temperature at the matching point (logρ = 8).

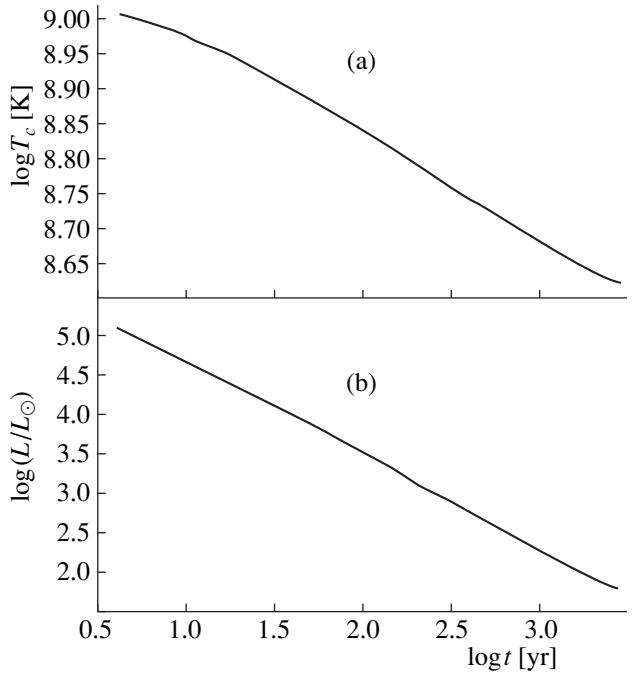


Fig. 6. Time dependence of (a) the central temperature and (b) the neutrino luminosity.

2.0×10^8 K. When T_c was higher than 4×10^8 K, the neutron star cooled, and it was heated when T_c was lower than 4×10^8 K; however, in both cases, it arrived at a steady state. The fact that the initial model had a temperature discontinuity at the neutron-star boundary in these cases did not affect the subsequent evolution, since, in the following time step, the temperature distribution in the neutron star took on the same form as that for the same T_c in the computations without a temperature jump in the initial model.

4. CONCLUSION

It follows from our analysis that, if the structure of the outer layers ($\log\rho < 8$) is treated under the assumptions of Thorne and Zytkow [5, 6] and moving the boundary condition $L_r = 0$ to the center of the star, the evolution of the star is such that the initial cool neutron core remains cool, although its neutrino luminosity during the cooling process may even exceed the photon luminosity of the T–Z object as a whole (Fig. 6). In particular, the low temperature of the neutron core means that the computations of the envelope evolution beyond the region of energy generation performed in [7] are correct. It follows that the maximum lifetime of a T–Z object is determined by the rate of mass loss via the stellar wind, and is $M_{\rm tot}/\dot{M} \sim 10^6$ yr.

However, the method used in [7], and here as well, is not entirely correct for evolutionary computations of the zone of energy generation. The point here is that this region has a mass $\sim 10^{-14} M_\odot$, and the accreted material passes through it in less than 1 s, while the minimum time step for the evolutionary computations was one year; clearly, this makes it impossible to study non-steady-state processes that could, in principle, occur in the zone of energy generation and lead to instability of the model. However, since the precision of the mantissa is only 15 significant digits, even in double-precision computations, it is not feasible to decrease the time step while using this type of algorithm. In other words, a special approach must be developed to investigate instabilities in the energy-generation zone, and we cannot currently consider the question of the instability of the models to rapid heating and sharp growth in the neutrino luminosity [4] to be definitively resolved.

ACKNOWLEDGMENTS

The authors thank S.I. Blinnikov, who presented us with a program to compute the equation of state of a lepton gas.

REFERENCES

1. B. Paczynski, Ann. Rev. Astron. Astrophys. **9**, 183 (1971).
2. G. S. Bisnovatyĭ-Kogan and R. A. Syunyaev, Astron. Zh. **48**, 881 (1971) [Sov. Astron. **15**, 697 (1972)].
3. J. P. Ostriker, in *IAU Symposium 73: The Structure and Evolution of Close Binary Systems*, Ed. by P. Eggleton, S. Mitton, and J. Whelan (Reidel, Dordrecht, 1973).
4. Ya. B. Zel'dovich, L. N. Ivanova, and D. K. Nadezhin, Astron. Zh. **49**, 253 (1972) [Sov. Astron. **16**, 209 (1972)].
5. K. S. Thorne and A. N. Zytkow, Astrophys. J. Lett. **199**, L19 (1975).
6. K. S. Thorne and A. N. Zytkow, Astrophys. J. **212**, 832 (1977).
7. G. S. Bisnovatyĭ-Kogan and S. A. Lamzin, Astron. Zh. **61**, 323 (1984) [Sov. Astron. **28**, 187 (1984)].
8. P. Podsidlowski, R. C. Cannon, and M. J. Rees, Mon. Not. R. Astron. Soc. **274**, 485 (1995).
9. B. Paczynski, Astrophys. J. **214**, 812 (1977).
10. S. I. Blinnikov, N. V. Dunina-Barkovskaya, and D. K. Nadyozhin, Astrophys. J., Suppl. Ser. **106**, 171 (1996).
11. N. V. Ardelyan, G. S. Bisnovatyĭ-Kogan, Yu. P. Popov, and S. V. Chernigovskiĭ, Astron. Zh. **64**, 761 (1987) [Sov. Astron. **31**, 398 (1987)].
12. G. S. Bisnovatyĭ-Kogan and V. M. Chechetkin, Astrophys. Space Sci. **26**, 25 (1974).
13. P. Haensel and J. L. Zdunik, Astron. Astrophys. **229**, 117 (1990).
14. M. J. Seaton, Y. Yan, D. Mihalas, and A. K. Pradhan, Mon. Not. R. Astron. Soc. **266**, 805 (S) (1994).
15. G. S. Bisnovatyĭ-Kogan, *Physical Problems of Theory of Stellar Evolution* [in Russian] (Nauka, Moscow, 1989).
16. V. A. Urpin and D. G. Yakovlev, Astron. Zh. **57**, 526 (1980) [Sov. Astron. **24**, 303 (1980)].
17. D. A. Baĭko and D. G. Yakovlev, Pis'ma Astron. Zh. **21**, 702 (1995) [Astron. Lett. **21**, 784 (1995)].
18. G. S. Bisnovatyĭ-Kogan and M. M. Romanova, Zh. Éksp. Teor. Fiz. **83**, 449 (1982) [Sov. Phys. JETP **56**, 243 (1982)].
19. K. S. Thorne, Astrophys. J. **212**, 825 (1977).
20. M. Harris, M. Fowler, G. Caughlan, and G. Zimmerman, Ann. Rev. Astron. Astrophys. **21**, 165 (1983).
21. P. Schinder *et al.*, Astrophys. J. **313**, 531 (1987).

Translated by D. Gabuzda

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 236–240. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 276–280.

Gamma-Ray Bursts as Standard-Energy Explosions

V. M. Lipunov[1, 2], K. A. Postnov[1, 2], and M. E. Prokhorov[1]

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

[2]*Moscow State University, Moscow, Russia*

Received September 14, 1999

Abstract—The distribution of observed energies for gamma-ray bursts with known redshifts can be explained as a consequence of events releasing a standard energy of $E_0 = 5 \times 10^{51}$ erg. Two situations are possible: the degree of collimation could vary from burst to burst, or there could be a universal radiation pattern for all bursts, with the observed differences being due to differences in the orientation of this pattern relative to the line of sight to the Earth.

1. INTRODUCTION

Events that release a standard amount of energy are a widespread phenomenon in astrophysics associated with sources whose masses are fixed (to within some accuracy). The best example is the collapse of the core of a massive star accompanied by a supernova explosion. When a neutron star forms as a result of such a collapse, its binding energy $\Delta E = 0.15 M_\odot c^2$ is radiated in the form of neutrinos. The four most accurately measured neutron-star masses lie within a narrow interval near $1.4 M_\odot$ [1], although the theoretically allowed range of masses is quite wide (from 0.1 to $3 M_\odot$). Another example is type Ia supernovae, which occur as a result of the thermonuclear explosion of a white dwarf with the Chandrasekhar mass. Similarly, a standard energy should be released during the merger of two neutron stars, in the form of gravitational waves (about 90% of the radiated energy) and neutrinos (the remaining 10%), as indicated by numerical computations [2].

Could cosmic gamma-ray bursts (GRBs) belong to such a class? Currently, their nature has not yet been firmly established (see the review [3]), although the most likely models involve the merger of two neutron stars (first proposed by Blinnikov *et al.* [4]) or the collapse of a very massive star (see, for example, Woosley [5] and Paczyński [6]). Note that considering GRBs to be standard candles enabled prediction of the redshift of GRB 970228 (the first GRB for which an optical afterglow was detected), $z = 0.7 \pm 0.1$, based on its position on the $\log N - \log S$ diagram; this prediction was made immediately after its discovery, before the first measurement of the redshift of the optical afterglow of the GRB registered by BeppoSax [7, 8]. This prediction proved to be in excellent agreement with the redshift of the GRB host galaxy obtained two years later, $z = 0.695$ [9]. This may not be a simple coincidence.

Intensive optical studies of the afterglows of GRBs have led to a rapid growth in the number of objects with measured redshifts. Currently, the redshifts of eight bursts are known (table), so that we can calculate the distances to these GRBs and the effective energy releases in gamma-rays E_γ of each of them (assuming their emission is isotropic). As we can see from the Table, these energies range over a wide interval from $\sim 5 \times 10^{51}$ to $\sim 2 \times 10^{54}$ erg. The real energy release E_0 could be appreciably lower than E_γ if the gamma-ray radiation is directed. If the half-opening angle of the cone into which the radiation is emitted is $\theta \ll 1$, the real energy released can be estimated as $E_0 \simeq E_\gamma \dfrac{\theta^2}{4}$.

In fact, high "apparent" energies of GRBs are not unexpected in models with merging neutron stars. The allowed theoretical merging rate, $R_{ns} \simeq 1\text{–}3 \times 10^{-4}$/yr per galaxy [7, 12–14], derived from first principles appreciably exceeds the observed rate of gamma-ray bursts, $R_{\mathrm{GRB}} \simeq 10^{-6}$/yr per galaxy [15]. We have inter-

Parameters of GRBs

GRB	z	$\Delta E_{\mathrm{obs}}^{(a)}$	E_γ/E_0	Directivity
970228	0.695[(b)]	5.2×10^{51}	~1	No
970508	0.835	5.3×10^{51}	~1	No
971214	3.418	2.5×10^{53}	~50	
980329	~5?	$\sim 2 \times 10^{54}$	~500	
980613	1.096	5×10^{51}	~1	
980703	0.966	9×10^{52}	~20	
990123	1.60	1.6×10^{54}	~300	Yes
990510	1.62[(c)]	1.4×10^{53}	~30	Yes

[(a)] Data taken from Briggs *et al.* [10].
[(b)] Djorgovski *et al.* [9].
[(c)] Vreesvijk *et al.* [11].

1063-7729/01/4503-0236$21.00

preted this large discrepancy as convincing evidence that the hard radiation of GRBs is collimated. The solid angle into which the energy is emitted is proportional to the ratio of the characteristic rates

$$\Omega \propto \frac{R_{\rm GRB}}{R_{ns}}.$$

It is thought that the maximum energy release in the form of gamma rays during the merging of two neutron stars can reach 10% of the rest energy (due to the annihilation of neutrinos and anti-neutrinos). Hence we obtain the following estimate for the maximum apparent energy release:

$$E_\gamma(\max) \simeq \gamma \frac{R_{ns}}{R_{\rm GRB}} \times 0.1 Mc^2 \lesssim \gamma \times 10^{56}\ \text{erg},$$

where $\gamma < 1$ is the coefficient for the conversion of the burst energy into gamma-rays.

In the general case, both the distribution of the real energy release $f(E_0)$ and the distribution of the directional pattern for the radiation $f(\theta)$ influence the distribution E_γ. Here, we will show that the existing observations of GRBs are consistent with the hypothesis that they have a standard intrinsic energy release E_0, and that the scatter in the effective burst energies can be explained purely by collimation E_0/E_γ. In this case, the observed distribution of GRBs in burst energy can be explained either by variation of the directional beam from object to object, or by the presence of a universal GRB directional beam made up of multiple components with appreciably different intensities.

2. THREE GROUPS OF OBSERVED GAMMA-RAY BURSTS

The Table presents the eight GRBs with measured redshifts. We have excluded GRB 980425, which was identified (probably mistakenly) with the supernova SN 1998bw in a nearby galaxy (at a distance of about 40 Mpc). On the other hand, the Table includes GRB 980329, which has a large redshift ~5 that was derived indirectly [16, 17]. Leaving this GRB out of our analysis does not appreciably affect our final conclusions.

Three of the eight events (GRB 970228, GRB 970508, and GRB 980613) have virtually identical effective energy releases, $(4.2–5.3) \times 10^{51}$ erg. They also form the group of weakest GRBs. Three other bursts (GRB 971214, GRB 980703, and GRB 990510) form an intermediate group with effective energy releases in the range 9×10^{52}–2.5×10^{53} erg; i.e., they are factors of 20–50 brighter than the bursts in the first group. The remaining two GRBs (GRB 980319 and GRB 990123) are the brightest of those presented in the Table, with effective energies of 2.4×10^{54} and 1.6×10^{54} erg, respectively (factors of 500 and 300 brighter than the bursts of the first group). Note that the scatter in the effective energy release for the GRBs of the first group is much smaller than that for the other two groups. The mean redshifts for the three groups show a tendency to grow with increase in E_γ: $\langle z_1 \rangle = 0.875$, $\langle z_2 \rangle = 2.0$, and $\langle z_3 \rangle = 3.3$. This correlation follows naturally from the hypothesis that the bursts in the different groups experience different collimation: narrower directional beams correspond to higher observed GRB energies, and such objects will be observed more rarely (in a fixed volume of space) but will be visible from larger distances, so that they will, on average, have larger redshifts z.

The observations of GRB 990123 show indications of substantial collimation of the radiation, $\theta \sim 0.1$ [18], as follows from the break in the light curve of the optical afterglow. A similar break should appear when the Lorentz factor Γ for the relativistically expanding shell is comparable to the inverse of the half-opening angle of the directional beam of the radiation, $\Gamma \sim 1/\theta$ [19]. The real energy release for GRB 990123 turns out to be $E_0 \simeq 4 \times 10^{51}$ erg, which is very close to the values for the first (weakest) group of GRBs. This could indicate that the gamma-ray bursts of the weakest group radiate spherically symmetrically, so that, for them, $E_0 = E_\gamma$, and an energy release of approximately 5×10^{51} erg is a fundamental value for *all* gamma-ray bursts.

A gradual variation in the slope of the light curve over a wide range of wavelength was detected in the optical afterglow of GRB 990510; this could also be explained by collimation. Consequently, the real energy release for this burst is lower than its effective energy release. On the other hand, there is currently no evidence for collimation in the afterglows of GRB 970508 and GRB 970228, which have been observed over the longest times; their energy releases are roughly the same, $E_\gamma \sim 5 \times 10^{51}$ erg.

3. STANDARD ENERGY RELEASE IN GAMMA-RAY BURSTS

Let us postulate that the real energy releases are the same for all GRBs, and are equal to $E_0 = 5 \times 10^{51}$ erg. The three groups of GRBs with differing ratios E_γ/E_0 are shown schematically in Fig. 1. This distribution is nearly flat, and can be interpreted in two different ways. It could reflect scatter in the collimation of the individual bursts, or it could be explained by a universal radiation pattern for all bursts (with a fairly complex form) together with a standard energy release.

3.1. Variable Collimation Angle

We will first consider the distribution of GRBs in collimation angle. Here, we suppose that the standard energy $E_0 = 5 \times 10^{51}$ erg is emitted into a cone whose half-opening angle θ is different for each of the three groups of GRBs. For the sake of simplicity, we will assume a Euclidean space. Then, the observed energy for each of the groups will be

$$E_{\gamma,i} = E_0/(\Omega_i/4\pi), \quad i = 1, 2, 3, \qquad (1)$$

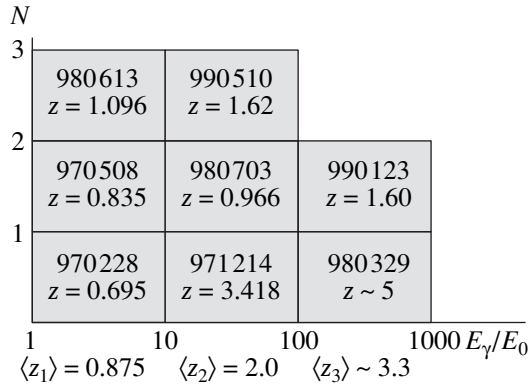


Fig. 1. Distribution of GRBs with known redshifts in the ratio E_γ/E_0. Individual and group-average redshifts are indicated.

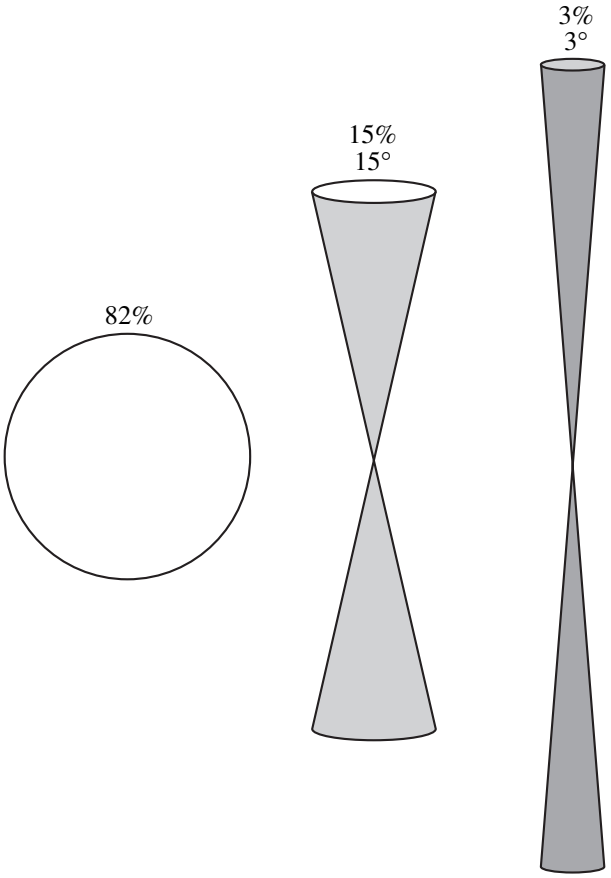


Fig. 2. Distribution of half-opening angles for the directional beams of GRBs θ assuming a standard energy $E_0 = 5 \times 10^{51}$ erg, derived from the observational statistics. The fractions of events with corresponding angles θ are indicated.

and consequently,

$$\frac{\Omega_1}{4\pi} = 1, \quad \frac{\Omega_2}{4\pi} = \frac{1}{30}, \quad \frac{\Omega_3}{4\pi} = \frac{1}{300} \tag{2}$$

(the second group of GRBs is a factor of 30 and the third group a factor of 300 brighter than the first group).

Let n_0 be the space density of GRB sources and n_i be the fraction of sources with collimation Ω_i, normalized to unity:

$$n_1 + n_2 + n_3 = 1. \tag{3}$$

The limiting distance at which a GRB of the *i*th group can be detected is $R_i = R_1(E_{\gamma,i}/E_0)^{1/2}$, since the number of observed events in the *i*th group is

$$N_i = n_0 n_i \left(\frac{\Omega_i}{4\pi}\right)\frac{4\pi}{3}R_1^3\left(\frac{E_{\gamma,i}}{E_0}\right)^{3/2}. \tag{4}$$

Solving Eqs. (1)–(4), we obtain $n_1 = 82\%$, $n_2 = 15\%$, and $n_3 = 3\%$, with the corresponding cone half-opening angles $\theta_2 = 15°$ and $\theta_3 = 3°$ (we used the relation $\theta_i \approx \sqrt{\Omega_i/2\pi}$, which is valid for $\Omega_i \ll 4\pi$). This distribution is schematically depicted in Fig. 2.

3.2. A Universal Directional Beam for the Radiation of Gamma-Ray Bursts

Here, we suppose that *all* GRBs have both a standard energy release E_0 and identical (complex) directional beams for their radiation. The differences in the observed energies and numbers of events occur because different bursts are viewed at different angles (relative to the beam symmetry axis; Fig. 3). In this case, Eq. (1) takes the form

$$E_{\gamma,i} = \epsilon_i E_0/(\Omega_i/4\pi), \quad i = 1, 2, 3, \tag{5}$$

where ϵ_i characterizes the fraction of the total burst energy radiated into a cone with solid angle Ω_i. Accordingly, formulas (2) can be written as

$$\frac{\Omega_1}{4\pi} = 1, \quad \frac{\Omega_2}{4\pi} = \frac{1}{30}\frac{\epsilon_2}{\epsilon_1}, \quad \frac{\Omega_3}{4\pi} = \frac{1}{300}\frac{\epsilon_3}{\epsilon_1}. \tag{6}$$

Taking into account the normalization

$$\epsilon_1 + \epsilon_2 + \epsilon_3 = 1 \tag{7}$$

(here, we have neglected overlapping of the radiation cones making up the overall directional beam, but this will contribute only a small error, since $\Omega_1 \gg \Omega_2 \gg \Omega_3$), Eq. (4) takes the form

$$N_i = n_0\left(\frac{\Omega_i}{4\pi}\right)\frac{4\pi}{3}R_1^3\left(\frac{\epsilon_i E_{\gamma,i}}{E_0}\right)^{3/2}. \tag{8}$$

Solving system (6)–(8), we find that the energy fractions $\epsilon_1 = 74\%$, $\epsilon_2 = 21\%$, and $\epsilon_3 = 5\%$ are emitted into cones with half-opening angles $\theta_1 = 90°$ (isotropic radiation), $\theta_2 = 20°$, and $\theta_3 = 3°$, respectively. Note that, in this case, the total energy release of the GRBs will be approximately 22% higher, $\sim 6 \times 10^{51}$ erg. Figure 3 provides an illustration of this scenario.

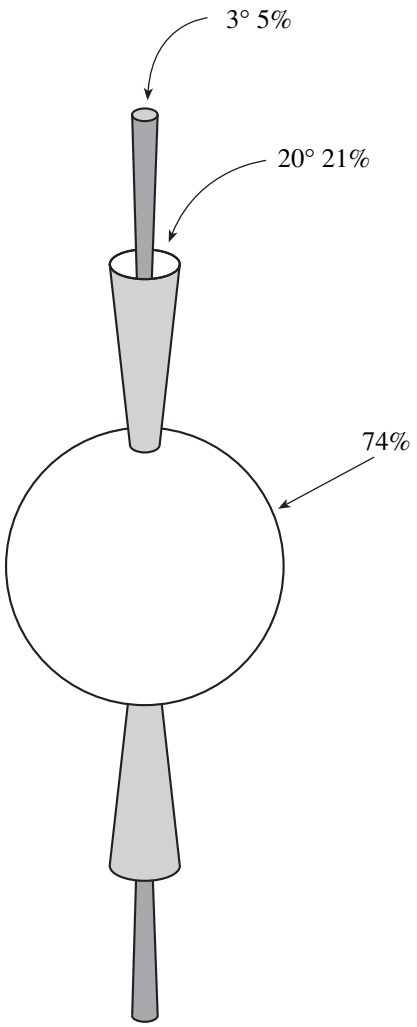


Fig. 3. The proposed universal radiation pattern for GRBs can be roughly divided into three parts: a (quasi-)spherically symmetrical part into which 74% of the total radiation is emitted and two narrower cones with half-opening angles of 20° and 3° into which 21 and 5% of the total energy is emitted, respectively. The total energy released is 6×10^{51} erg.

4. CONCLUSION

We have shown that the observed distribution of energies for GRBs can be explained by the hypothesis that GRBs have a standard energy release $E_0 = 5 \times 10^{51}$ erg. Two situations are possible here: either the collimation angle varies from burst to burst, or there is a universal multi-component radiation directional beam for all GRBs (and we view bursts with different strengths at different angles to the beam axis). At this stage, we have not discussed physical models for the beam pattern. It is possible that GRBs viewed at different angles (belonging, accordingly, to groups with different brightnesses) could have different physical properties (for example, spectral or temporal behavior).

There also exists the opposite point of view that the energy release of GRBs varies over orders of magnitude, and the radiation is not significantly collimated [21]. Currently, we cannot distinguish between these two possibilities (namely, whether the observed GRB energy release E_γ varies due to differences in collimation or due to differences in the real energy release E_0). The increase in the statistics of GRB observations expected in the near future with the launch of the specialized satellites HETE-2 and SWIFT may help determine which of these two points of view is correct.

However, the increase in observational statistics cannot help distinguish between the two hypotheses considered in this article, since both a distribution of GRBs in collimation angle and a universal directional beam could describe equally well any number of groups of GRBs with arbitrary numbers and brightnesses. We can also name several criteria for verification of the hypothesis of standard energy release in GRBs:

(a) there should be no GRBs with E_γ significantly lower than E_0;

(b) our theoretical predictions of the cone widths θ should coincide with those derived from observations;

(c) the mean redshifts of GRB groups should grow with growth in E_γ.

With the exception of GRB 980425, which was identified with SN 1998bw, there are no known bursts with observed energy releases below 10^{51} erg. Estimates of the opening angles for the GRB ejecta based on observations of the light curves of their optical afterglows are currently very inaccurate (for more detail, see [22]). The mean redshifts in the three groups of bursts we have identified increase with E_γ (Fig. 1).

ACKNOWLEDGMENTS

The authors thank S.I. Blinnikov, N.I. Shakura, and I.G. Mitrofanov for discussions. This work was partially supported by the Russian Foundation for Basic Research (project code 99-02-16205) and an INTAS grant (96-315).

REFERENCES

1. S. E. Thorsett and D. Chakrabarty, Astrophys. J. **512**, 288 (1999).
2. M. Ruffert and H.-Th. Janka, Astron. Astrophys. **338**, 535 (1998).
3. K. A. Postnov, Usp. Fiz. Nauk **169**, 545 (1999).
4. S. I. Blinnikov, I. D. Novikov, T. V. Perevodchikova, and A. G. Polnarev, Pis'ma Astron. Zh. **10**, 177 (1984) [Sov. Astron. Lett. **10**, 422 (1984)].
5. S. E. Woosley, Astrophys. J. **405**, 273 (1993).
6. B. Paczyński, Astrophys. J. Lett. **499**, L45 (1997).
7. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Mon. Not. R. Astron. Soc. **288**, 245 (1997).

8. V. M. Lipunov, in *Highlights of Astronomy*, Ed. by J. Andersen (Kluwer, Dordrecht, 1998), Vol. 11B, p. 779.
9. S. G. Djorgovski, S. R. Kulkarni, J. S. Bloom, and D. A. Frail, GCN 289 (1999).
10. M. S. Briggs, D. L. Band, R. M. Kippen, *et al.***,** Astrophys. J. **524**, 82 (1999).
11. P. M. Vreeswijk, T. J. Galama, E. Rol, *et al.***,** GCN 324 (1999).
12. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Astron. Astrophys. **176**, L1 (1987).
13. A. V. Tutukov and L. R. Yungelson, Mon. Not. R. Astron. Soc. **268**, 871 (1994).
14. E. P. J. van den Heuvel, in *Interacting Binaries*, Ed. by S. N. Shore, M. Livio, and E. P. J. van den Heuvel (Springer-Verlag, Berlin, 1994), p. 452.
15. V. M. Lipunov, K. A. Postnov, M. E. Prokhorov, *et al.*, Astrophys. J. **454**, 593 (1995).
16. E. Palazzi, E. Pian, N. Masetti, *et al.*, Astron. Astrophys. **336**, L95 (1998).
17. A. Fruchter, Astrophys. J. Lett. **512**, L1 (1999).
18. S. R. Kulkarni, S. G. Djorgovski, S. C. Odewahn, *et al.*, Nature **398**, 389 (1999).
19. J. E. Rhoads, Astrophys. J. **525**, 737 (1999).
20. F. A. Harrison, J. S. Bloom, D. A. Frail, *et al.*, Astrophys. J. Lett. **523**, L121 (1999).
21. T. Totani, Mon. Not. R. Astron. Soc. **307**, L41 (1999).
22. R. Sari, T. Piran, and J. Halpern, Astrophys. J. Lett. **519**, L17 (1999).

Translated by D. Gabuzda

Astronomy Reports, Vol. 45, No. 3, 2001, pp. 241–247. Translated from Astronomicheskiĭ Zhurnal, Vol. 78, No. 3, 2001, pp. 281–288.

Simulation of Neutrino Transport by Large-Scale Convective Instability in a Proto-Neutron Star

V. M. Suslin, S. D. Ustyugov, V. M. Chechetkin, and G. P. Churkina

Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Miusskaya pl. 4, Moscow, 125047 Russia

Received January 17, 2000

Abstract—Neutrino transfer via convective flow to the surface of a proto-neutron star is numerically simulated. The evolution of the neutrino distribution in a heated region rising from the center of the proto-neutron star to its surface is simulated using a kinetic equation with a Uehling–Uhlenbeck collision integral in a uniform, isotropic approximation. The composition of the matter in the region under consideration changes due to the "burning" of electrons and protons by beta processes. The simulation results enable the estimation of the characteristic time required for the rising medium to become optically thin to neutrinos and the characteristic spectrum of the neutrinos that are emitted. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of main problems in the theory of supernova explosions is the development of a scenario for the ejection of the envelope based on modern physical concepts. In particular, the source of the energy transfered to the ejected envelope must be identified and the mechanism for this energy transfer described. The gravitational energy of the collapsing core of the supernova progenitor when it reaches densities of the order of 10^{14} g/cm^3 could be the required energy source, but the mechanism for transferring this energy is unclear.

In the classical paper by Colgate and White [1], neutrino transport provides the needed energy-transfer mechanism, assuming that the matter is optically thin to neutrinos. However, as shown in [2], the propagation of neutrinos through matter with densities exceeding 10^{12} g/cm^3 is strongly affected by the opacity of the material, so that the characteristic time for neutrino diffusion considerably exceeds the lifetime of the envelope. Therefore, it is not surprising that numerous studies of mechanisms for type-II supernova explosions based on collapse of the supernova progenitor's core into a neutron star have not led to appreciable progress. The aim of the present work is to investigate the possibility that the mechanism for the supernova explosion is appreciably non-spherically symmetric.

Suslin *et al.* [3] showed that, for their chosen collapse scenario, non-equilibrium neutronization of the protostar leads to a substantial entropy increase, even on times of the order of 2×10^{-4} s, first and foremost in the central region of the star with dimensions of about 10^6 cm. The shortness of the characteristic time for entropy increase is due to the high rate of energy transport from ultra-relativistic, strongly-degenerate electrons to high-energy neutrinos at densities of the order of 10^{14} g/cm^3; the size of the primary region of entropy increase is determined by the extent of the central region with such densities. Consequently, hydrodynamic instability based on the buoyancy force can develop in this region. Our three-dimensional model calculations with a specified entropy excess in the central region of a star demonstrate the onset of large-scale instability in the form of bubbles rising toward the stellar surface on timescales considerably less than the characteristic neutrino-diffusion time. The dimensions of the bubbles produced are such that their optical depth to high-energy neutrinos exceeds unity [4]. Therefore, the neutrinos inside a bubble can be partially trapped, and rise to the surface until the bubble becomes fully optically thin.

We wish to determine the time required for the bubbles to become fully optically thin and the characteristic spectrum of the neutrinos at this time, and also to trace the evolution of some parameters of the bubble matter and the neutrinos that travel with it.

2. FORMULATION OF THE PROBLEM

Let us consider a bubble corresponding to a bounded region in the central part of a star about 6×10^5 cm in size, with a density of about 2×10^{14} g/cm^3; the density of the surrounding medium is approximately the same, while the bubble temperature substantially exceeds the temperature of the surrounding medium. A bubble rising due to the action of the buoyancy force expands, changing its size and density. We shall adopt the time dependence for the bubble density obtained in [4], based on numerical modeling of the onset of an initial entropy perturbation against the background of an equilibrium rotating configuration for a proto-neutron star. At the initial time, we specified a central region with perturbations using a Gaussian profile for the entropy and density variations. Next, the boundaries of the region occupied by the evolving perturbation were determined at each moment in time by comparing the initial background constant entropy with the entropy

1063-7729/01/4503-0241$21.00

inside the region. The mean bubble density was determined by averaging over all computational cells inside the bubble. Since we are primarily interested in relatively rough estimates, we shall consider a model that describes the matter in the perturbed region and relevant neutrino processes only approximately, but retains the features we consider to be most important.

To simplify the model, we assume that, at the initial time, the bubble is composed of iron nuclei ($A = 56$, $Z = 26$) and free ultrarelativistic electrons. At densities of 2×10^{14} g/cm^3, such a medium is characterized by intense beta processes, which are a source of neutrinos:

$$(A, Z) + e \Rightarrow (A, Z-1) + \nu.$$

Inverse processes that absorb neutrinos also occur:

$$(A, Z+1) + e \Leftarrow (A, Z) + \nu.$$

Both direct and inverse beta processes produce new elements, complicating the model for the medium. To avoid this, we introduce a neutron component and do not distinguish between different types of nuclei. Along with inelastic processes, neutrinos also participate in elastic interactions with electrons and nuclei:

$$(A, Z) + \nu \Rightarrow (A, Z)' + \nu',$$

$$e + \nu \Rightarrow e' + \nu'.$$

We emphasize that, due to the large difference between the masses of electrons and nuclei, neutrinos lose substantially greater energy in collisions with electrons. Since our treatment will be limited to a uniform and isotropic approximation for the neutrino distribution function (whereas scattering by nuclei contributes appreciably only to the anisotropic component of the distribution function), we can neglect scattering by nuclei in the collision integral. However, since scattering by nuclei appreciably affects the rate of escape of neutrinos through the bubble boundary, we will taken it into account in this process. We neglect all other processes involving neutrinos. In contrast to [5], we shall take into account the fact that there is also some distribution of neutrinos outside the bubble. Therefore, neutrinos can both leave and enter the region under consideration. The electron distribution function will be interpolated by a Fermi step function, which is obviously applicable only when $E_F \gg 1.5kT$ (i.e., when the Fermi energy of the electrons is considerably greater than their thermal energy).

3. THE MATHEMATICAL MODEL

In a uniform, isotropic approximation, the kinetic equation describing the evolution of the neutrino distribution in a bounded region whose characteristic size d, density, and partial concentrations of components vary with time can be written in the form

$$\frac{\partial f(p,t)}{\partial t} = 4\pi[(1-f(p,t))\int_0^\infty dp'(p')^2 \times f(p',t)K^{\text{in}}(p,p',t) - f(p,t)\int_0^\infty dp'(p')^2(1-f(p',t))K^{\text{out}}(p,p',t)] \quad (1)$$
$$+ \frac{d\rho(t)}{\rho(t)dt}f(p,t) + (1-f(p,t))S(p,t) - f(p,t)Y(p,t) - \frac{c}{d(t)[1+\gamma(p,t)]}(f(p,t) - f_g(p,p_g)).$$

Here, the functions $K^{\text{in}}(p, p' t)$ and $K^{\text{out}}(p, p', t)$ depend on details of the process of neutrino scattering by electrons, and $S(p, t)$ and $Y(p, t)$ are sources and sinks of neutrinos, determined by the direct and inverse beta processes. The term containing the logarithmic derivative of the density is responsible for variations in the neutrino distribution due to the changing dimensions of the region where the neutrinos are concentrated. The last term describes neutrino escape through the boundary. (The value $\gamma = 0$ corresponds to the case of free propagation.) We normalized the distribution function as follows:

$$n(t) = 4\pi(2\pi\hbar)^{-3}\int_0^\infty dp'(p')^2 f(p',t). \quad (2)$$

This equation for the neutrino distribution function must be supplemented by an equation describing evolution of the electron number density

$$\frac{dn_e(t)}{dt} = 4\pi\int_0^\infty dp\,p^2 \times [-(1-f(p,t))S(p,t) + f(p,t)Y(p,t)] + \frac{d\rho}{\rho dt}n_e(t), \quad (3)$$

and also by a relation between the densities of the electrons and neutrons, on the one hand, and the density of the medium, on the other:

$$m_n\left[n_n(t) + \frac{A}{Z}n_e(t)\right] = \rho(t). \quad (4)$$

In deriving (4), we have assumed that electrons make a negligible contribution to the density of the medium, and that the medium is electrically neutral.

To facilitate use of the above equations as a basis for numerical simulations, we introduce dimensionless variables, constructed using the following parameters: characteristic length 3×10^7 cm, characteristic time 10^{-3} s, characteristic density 2×10^{14} g/cm^3, characteristic

momentum $p_F(0) = s^{-1} E_F(0)$ (the Fermi-momentum of the electrons, where $E_F(0) = 219$ MeV), and characteristic number density of the particles $n_e(0) = 4.63 \times 10^{37}$ cm^{-3}. The dimensionless characteristic size $d(t)$ at the initial time is $d(0) = 2.13 \times 10^{-2}$.

We shall not introduce special notation for the dimensionless quantities (apart from $p \longrightarrow x$, $p' \longrightarrow y$, and $P_F \longrightarrow u$). The required dimensionless system of equations will then take the form

$$\frac{df(x,t)}{\partial t} = 4\pi[(1-f(x,t))\int_0^\infty dy y^2 f(y,t)K^{\text{in}}(x,y,t) - f(x,t)\int_0^\infty dy y^2(1-f(y,t))K^{\text{out}}(x,y,t)] + \frac{d\rho}{\rho dt}f(x,t) + (1-f(x,t))S(x,t) - f(x,t)Y(x,t) - \frac{1}{d(1+\gamma)}(f(x,t)-f_g(x,x_g)), \quad (5)$$

$$\frac{dn_e(t)}{dt} = \frac{3}{2}\int_0^\infty dx x^2 \times [-(1-f(x,t))S(x,t) + f(x,t)Y(x,t)] + \frac{d\rho}{\rho dt}n_e(t), \quad (6)$$

$$n_n(t) + \frac{A}{Z}n_e(t) = \frac{A}{Z}\rho(t). \quad (7)$$

The time dependence of the characteristic size d is given by the expression

$$d(t) = d(0)\sqrt[3]{\frac{\rho(0)}{\rho(t)}}. \quad (8)$$

The normalization of the neutrino distribution function becomes

$$n(t) = \frac{3}{2}\int_0^\infty dx x^2 f(x,t). \quad (9)$$

Formulas for $S(x, t)$ and $Y(x, t)$ are presented in [6]. We shall write here only the corresponding dimensionless relations:

$$S(x,t) = 1.874\times 10^3\left(\frac{Z}{A}\rho(0)\right)^{5/3} \times x^2\theta(u(t)-x)\frac{n_e(t)}{Z}, \quad (10)$$

$$Y(x,t) = 1.874\times 10^3\left(\frac{Z}{A}\rho(0)\right)^{5/3} x^2\theta(x-u(t)) \times\left[\frac{n_e(t)}{Z} + 100\frac{A}{Z}\left(\frac{\rho(t)}{\rho(0)} - n_e(t)\right)\right]. \quad (11)$$

(In addition, the relation $n_e(t) = u^3(t)$ is valid in the dimensionless notation.)

We used the following expressions for γ in the simulations:

$$\gamma(x,t) = 4\pi\int_0^\infty dy y^2(1-f(y,t)) \times[K^{\text{out}}(x,y,t) + K^n(x,y,t)], \quad (12)$$

and

$$\gamma(x,t) = 4\pi\int_0^\infty dy y^2 \times[(1-f(y,t))K^{\text{out}}(x,y,t) + K^n(x,y,t)]. \quad (13)$$

The function $K^n(x, y, t)$ describes scattering by nuclei:

$$K^n(x,y,t) = 1.69\times 10^6\left(\frac{A}{10}\right)^2 \times\left(\frac{Z}{A}\rho(0)\right)^{5/3}\frac{n_e(t)}{z}\delta(x-y), \quad K^{\text{in/out}}(x,y,t) = 2.326\times 10^6(xy)^{-2} \times\left(\frac{Z}{A}\rho(0)\right)^{5/3} r^{\text{in/out}}(x,y,t). \quad (14)$$

Let us write also the quantities $r(x, y, t) = r(x, y, u(t))$, taken from [6]:

(1) For $0 \le x \le u(t)$:

$$r^{\text{out}}(x,y,u) = \theta(x-y)y^3\left[\frac{8}{9}((u+x)^3-u^3) - y\left(2x^2+4ux+\frac{8}{3}u^2\right) + y^2\left(\frac{8}{5}x+\frac{4}{3}u\right) - \frac{22}{45}y^3\right], \quad (15)$$

$$r^{\text{in}}(x,y,u) = x^3\theta(u-y)\theta(y-x) \times\left[\frac{8}{9}y^3 + \frac{2}{3}(4u-3x)y^2 + \left(\frac{8}{5}x^2 - 4ux + \frac{8}{3}u^2\right)y + \left(-\frac{22}{45}x^3 + \frac{4}{3}ux^2 - \frac{8}{3}u^2x\right)\right] + x^3(u-x)\theta(y-u)$$

$$\times\left[\frac{8}{3}y^2+\left(\frac{8}{3}u-\frac{4}{3}x\right)y+\left(\frac{22}{45}x^2-\frac{10}{9}ux+\frac{8}{9}u^2\right)\right]$$
$$+\theta(y-u)\theta(u+x-y)\left[-\frac{4}{9}y^6+\frac{8}{5}(u+x)y^5\right.$$
$$-2(u+x)^2y^4+\frac{8}{9}(u+x)^3y^3+\frac{2}{3}x^4y^2+\frac{4}{15}x^5y$$
$$\left.-\frac{2}{45}((u+x)^6-x^6)\right]+\theta(u+x-y)$$
$$\times x^4\left(\frac{2}{3}y^2+\frac{4}{15}xy+\frac{2}{45}x^2\right). \quad (16)$$

(2) For $x > u(t)$:

$$r^{\rm in}(x,y,u) = \theta(y-x-u)u^4\left(\frac{2}{3}y^2+\frac{4}{15}uy+\frac{2}{45}u^2\right)$$
$$+\theta(y-x)\theta(u+x-y)\left[-\frac{4}{9}y^6+\frac{8}{5}(u+x)y^5\right.$$
$$-2(u+x)^2y^4+\frac{8}{9}(u+x)^3y^3+\frac{2}{3}u^4y^2$$
$$\left.+\frac{4}{15}u^5y+\frac{2}{45}(u^6-(u+x)^6)\right], \quad (17)$$
$$r^{\rm out}(x,y,u) = [\theta(2u-x)\theta(x-u-y)$$
$$+\theta(x-2u)\theta(u-y)]\left(\frac{2}{3}y^2+4xy+10x^2\right)\frac{y^4}{15}$$

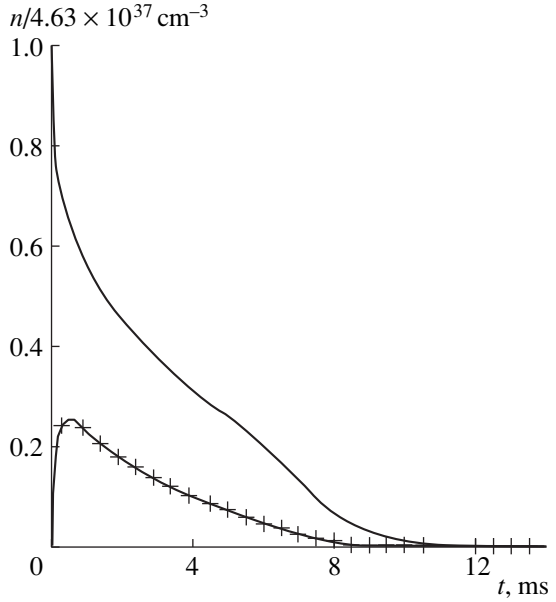


Fig. 1. Time dependence of the number density of neutrinos (marked curve) and electrons (unmarked curve).

$$+\theta(2u-x)\theta(u-y)\theta(y-x+u)\frac{(x-u)}{15}$$
$$\times\left[4y^5+10(u+x)y^4+\frac{40}{3}(x^2+ux+u^2)y^3\right.$$
$$+10(u-x)(u^2+2ux+3x^2)y^2$$
$$+4(u-x)^2(u^2+3ux+6x^2)y$$
$$\left.+\frac{2}{3}(u-x)^3(u^2+4ux+10x^2)\right]$$
$$+\theta(x-2u)\theta(y-u)\theta(x-u-y)$$
$$\times\frac{u^4}{15}\left(\frac{2}{3}u^2+4ux+10x^2\right) \quad (18)$$
$$+[\theta(2u-x)\theta(y-u)+\theta(x-2u)\theta(y-x+u)]$$
$$\times\theta(x-y)\frac{1}{15}\left[-\frac{2}{3}y^6-4uy^5-10u^2y^4-\frac{40}{3}(u^3-x^3)y^3\right.$$
$$-10(u-x)^2(u^2+2ux+3x^2)y^2$$
$$-4(u-x)^3(u^2+3ux+6x^2)y-\frac{2}{3}(u-x)^4$$
$$\left.\times(u^2+4ux+10x^2)+\frac{2}{3}u^4(u^2+6ux+15x^2)\right]$$
$$+\theta(u-y)y^3\left[-\frac{22}{45}y^3+\left(\frac{8}{5}u+\frac{4}{3}x\right)y^2\right.$$
$$\left.-\left(2u^2+4ux+\frac{8}{3}x^2\right)y+\frac{8}{9}((u+x)^3-x^3)\right].$$

We solved the system of equations (5)–(18) numerically on a uniform grid in x containing 101 points. The integrals were approximated by trapezoidal formulas. The time evolution was described by an implicit, second-order, two-layer scheme. We solved the implicit difference system using a successive-approximation method.

The time dependence of the bubble density computed in this way and used in the numerical simulations is described well by the formula

$$\rho(t) = \frac{167.14}{1+0.3\exp(t-5.5)}. \quad (19)$$

4. SIMULATION RESULTS

The results of our numerical solution of system (5)–(18) are presented in Figs. 1–7 for two cases: in the first (Figs. 1–4), we used formula (12) for the function $\gamma(x, t)$, and, in the second (Figs. 5–7), we used formula (13). The neutrino distribution function outside the specified region was taken to be zero in both cases. The number

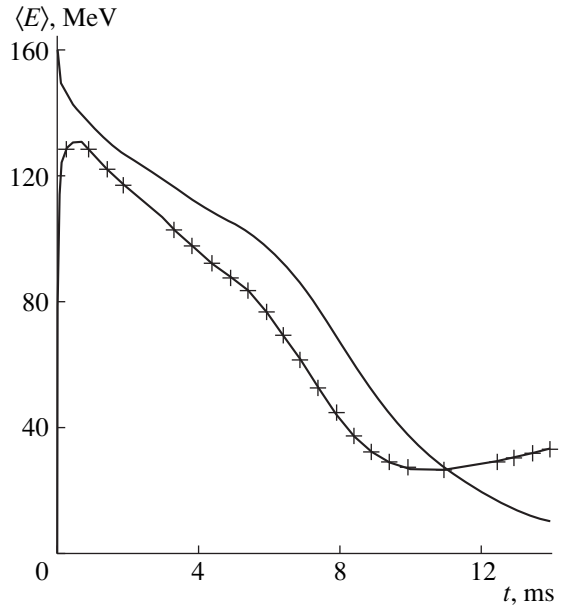


Fig. 2. Time dependence of the average energies of the neutrinos (marked curve) and electrons (unmarked curve).

of neutrinos emitted per unit volume per unit time in an interval dp is given in our model by the formula

$$dI(p,t) = \frac{4\pi c}{d(t)[1+\gamma(p,t)](2\pi\hbar)^3} \times [f(p,t) - f_g(p,p_g)]p^2 dp = n_e(0)I(x,t)dx. \quad (20)$$

The quantities $I(x, t)$ are presented in Figs. 4 and 7.

Although the difference between the two cases from the viewpoint of physical conditions seems negligible, the results of the numerical simulations are radically different. In the first case, the bubble is optically thin to the neutrino radiation from the very onset of the process. The emission maximum corresponds to time $t \approx 0.5$ ms, which is roughly equal to the characteristic time for the action of the source. The transparency of the bubble results from the substantial decrease in neutrino scattering due to the factor $(1 - f)$ in (12), when the neutrino distribution function is close to a Fermi step function.

However, under real conditions, the distribution function can be considerably different from a Fermi step distribution, for example, due to the heating of the electron component by neutrino scattering processes. Since the model we have used cannot be applied to the case of non-zero temperature for the medium, we simulated the influence of non-degeneracy of the neutrino distribution on the emission by neglecting the suppression factor $(1 - f)$ when describing scattering by nuclei. Since the neutrino energy remains virtually unchanged during scattering by nuclei (as compared to scattering by electrons), this should not affect the evolution of the neutrino distribution in momentum space, and should change only the optical depth of the boundary [see Eq. (13)].

Indeed, in this case, appreciable neutrino radiation appears only after $t \approx 10$ ms, and reaches its maximum at $t \approx 12$ ms. In accordance with (19), such times correspond to densities that are two orders of magnitude lower than the initial density. Therefore, in the second case, the bubble becomes optically thin to neutrinos only after it rises to the upper layers of the supernova core, which are characterized by considerably lower density. At earlier times, the rising bubble acts as an opaque neutrino trap. In this case (Fig. 6), the mean energies of the neutrino and electron components are equal from $t \approx 0.5$ ms (the characteristic time required for production of the neutrino component by beta pro-

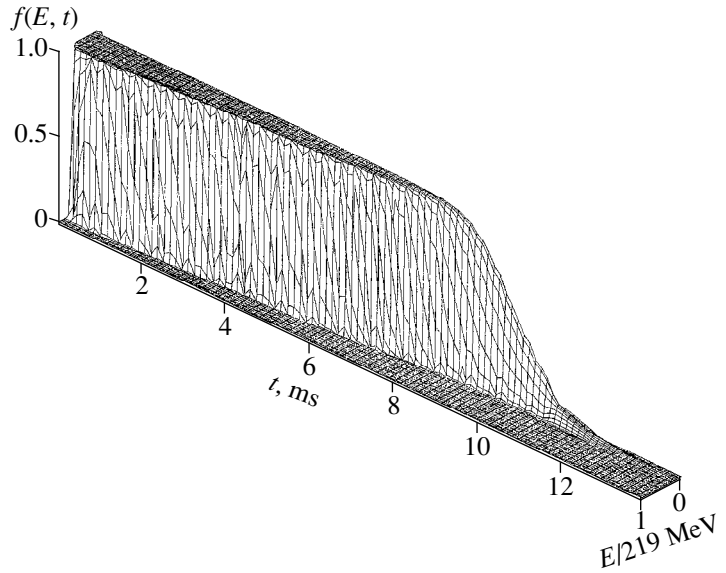


Fig. 3. Neutrino distribution function $f(E, t)$.

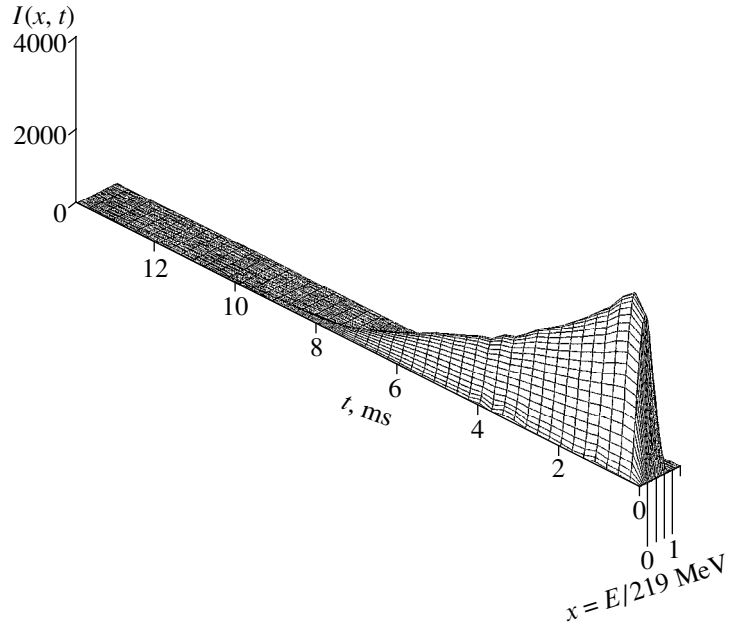


Fig. 4. The function $I(x, t)$.

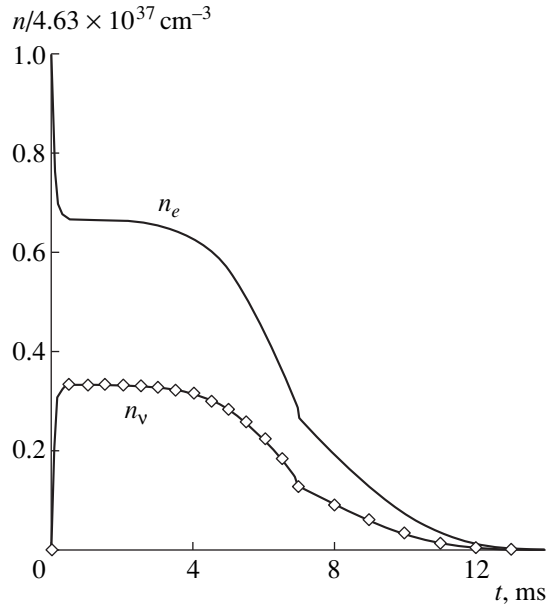


Fig. 5. Time dependence of the number density of neutrinos (marked curve) and electrons (unmarked curve).

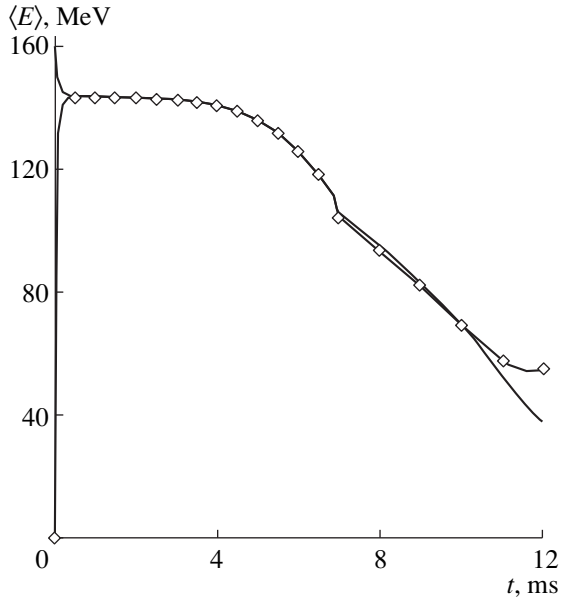


Fig. 6. Time dependence of the average energies of the neutrinos (marked curve) and electrons (unmarked curve).

cesses) to $t \approx 10$ ms (when the medium becomes optically thin). In the first case (Fig. 2), the average neutrino energy is less than the average electron energy, due to the emission of some of the neutrinos before the onset of the stage of "classical" transparency. Later, the mean neutrino energy exceeds the mean electron energy in both cases, since the degradation of neutrinos in *νe* processes is substantially decreased, while the Fermi energy of the electrons (and, consequently, their average energy) continues to decrease as the bubble expands.

In conclusion, let us estimate the spectrum of the neutrinos emitted from a supernova in the adopted scenario. In the second case, the spectra of the neutrino emission from the bubble and from the entire supernova should coincide to high accuracy, since the bubble radiation is emitted in the optically thin region of the

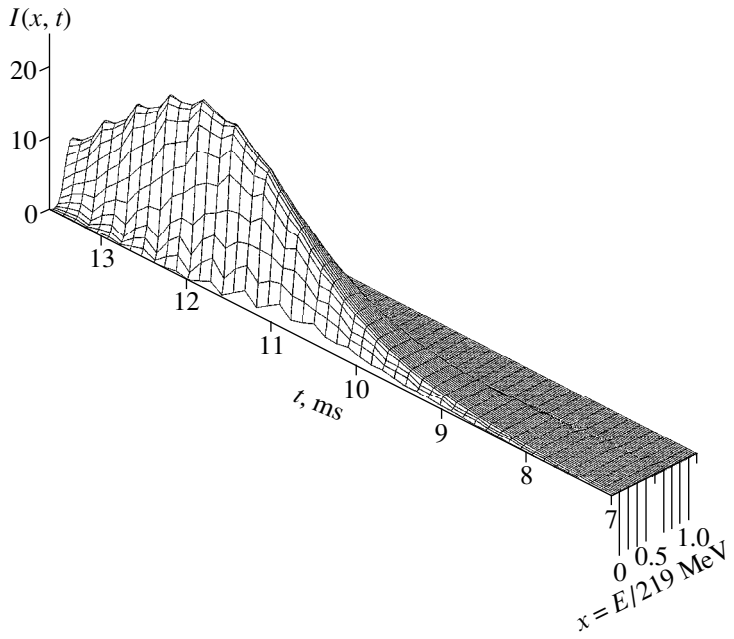


Fig. 7. Function $I(x, t)$.

star (Fig. 8c). It is impossible to accurately estimate the neutrino spectrum emitted by the supernova in the first case, since the neutrinos leave the bubble in the region of "classical" opacity, and we must simulate the neutrino processes over the entire stellar zone in order to obtain a correct description. We can only present here the hardest (Fig. 8a) and softest (Fig. 8b) possible spectra.

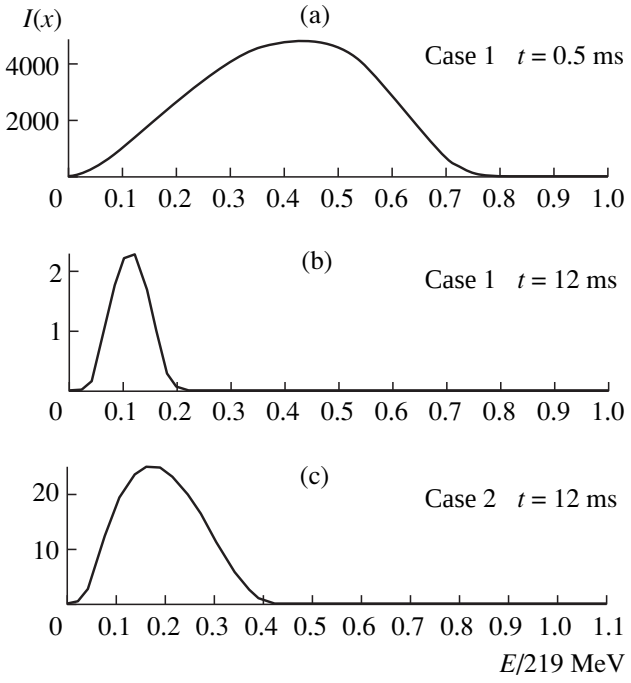


Fig. 8. Spectra of the emitted neutrinos $I(x)$.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project code 98-01-00956).

REFERENCES

1. S. A. Colgate and R. H. White, Astrophys. J. **143**, 626 (1966).
2. V. S. Imshennik and D. K. Nadezhin, Zh. Éksp. Teor. Fiz. **63**, 1548 (1972) [Sov. Phys. JETP **36**, 821 (1973)].
3. V. M. Suslin, M. Yu. Khlopov, V. M. Chechetkin, and V. A. Chuyanov, Astron. Zh. **73** (3), 399 (1996) [Astron. Rep. **40**, 358 (1996)].
4. M. V. Sazhin, S. D. Ustyugov, and V. M. Chechetkin, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 817 (1996) [JETP Lett. **64**, 871 (1996)].
5. V. M. Suslin, S. D. Ustyugov, and G. P. Churkina, Preprint No. 70, IPM im. M.V. Keldysha RAN (Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, 1998).
6. V. M. Suslin and M. Yu. Khlopov, Preprint No. 109, IPM im. M.V. Keldysha AN SSSR (Keldysh Institute of Applied Mathematics, Academy of Sciences of USSR, 1986).

Translated by Yu. Dumin