Astronomy Letters, Vol. 28, No. 8, 2002, pp. 495–508. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 563–578.

A Cold Front in the Galaxy Cluster A3667: Hydrodynamics, Heat Conduction, and Magnetic Field in the Intergalactic Medium

A. A. Vikhlinin[1*] and M. L. Markevitch[2]

[1] *Space Research Institute, Russian Academy of Sciences, Profsoyuznaya ul. 84/32, Moscow, 117810 Russia*

[2] *Harvard–Smithsonian Center for Astrophysics, USA*

Received March 14, 2002

Abstract—Chandra observations of the galaxy cluster A3667 are analyzed. The X-ray data strongly suggest that there is a giant cloud of cold plasma separated by a sharp boundary (cold front) from the hotter intergalactic medium. Analysis of the cold-front properties allows one to study a number of physical processes in the intergalactic medium: large-scale gas motions, transport processes, and magnetic-field structure and strength. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *galaxy clusters, intergalactic medium*

INTRODUCTION

The mergers of galaxy clusters are the most energetic events in the Universe since the Big Bang. A potential energy of 10^{63}–10^{64} erg is released during these events. Much of this energy goes into intracluster gas in the form of heating by shock waves, particle acceleration to relativistic energies, generation of turbulence and magnetic fields, etc. (Sarazin 2001). Because of their significant deviations from an equilibrium, the merging clusters are a unique laboratory for studying the physical processes in the hot intracluster gas through X-ray observations.

Since the merging clusters must fly into each other at a supersonic speed (Sarazin 2001), it has long been expected that the clearest observational manifestation of the merger must be the existence of powerful shock waves in the hot intracluster gas. Indeed, X-ray observations of several clusters revealed reliable evidence of shocks, with the clearest evidence being the surface-brightness discontinuities in the clusters A2142 and A3667. Markevitch *et al.* (1999) were the first to present physical applications of the shock observations: in the cluster Cygnus A, they estimated the velocity of the merging entities and the heating efficiency of the gas by a shock; in the cluster A1657, they managed to find independent evidence for the existence of a deep potential well that protected the cold gas at the cluster center from being disrupted by a shock; using A3667 as an example, they discussed the necessity of allowing for the difference between the shock structures in a normal gas and in plasma.

The launch of the Chandra X-ray (observatory, which has an angular resolution higher than 1″ and which can take an X-ray (0.5–10 keV) spectrum at each point with a typical resolution $\Delta E/E = 2$–3% at a sufficient source brightness, was expected to bring about a sharp improvement in our understanding of the shock structure in merging galaxy clusters. The two best (in this respect) objects, the clusters A2142 and A3667, were scheduled to be observed in the first months of the Chandra operation in the fall of 1999.

The Chandra observations yielded an unexpected result. The X-ray surface-brightness discontinuities in A2142, A3667, and several other clusters turned out to be not shock waves but phenomena of a different kind: contact-discontinuity surfaces between the clouds of cold dense plasma with a characteristic size of 500 kpc embedded in the hotter and more tenuous intergalactic medium.

The discovery of such a structure was first reported by Markevitch *et al.* (2000), who used the observations of A2142. The image obtained confirmed the existence of a sharp surface-brightness discontinuity with a regular cometary shape. The exposure time was too short to carry out detailed studies. However, the achieved accuracy of measuring the temperature allowed one to establish that the dense plasma within the discontinuity boundary is colder than the plasma outside. This rules out the possibility of the existence of a shock, because when the discontinuity boundary is crossed, the entropy changes in the direction opposite to the gas density. A similar structure was detected in the cluster A3667 (Vikhlinin *et al.* 2001a, 2001b) and was called a cold front; this observation

*E-mail: alexey@hea.iki.rssi.ru

1063-7737/02/2808-0495$22.00

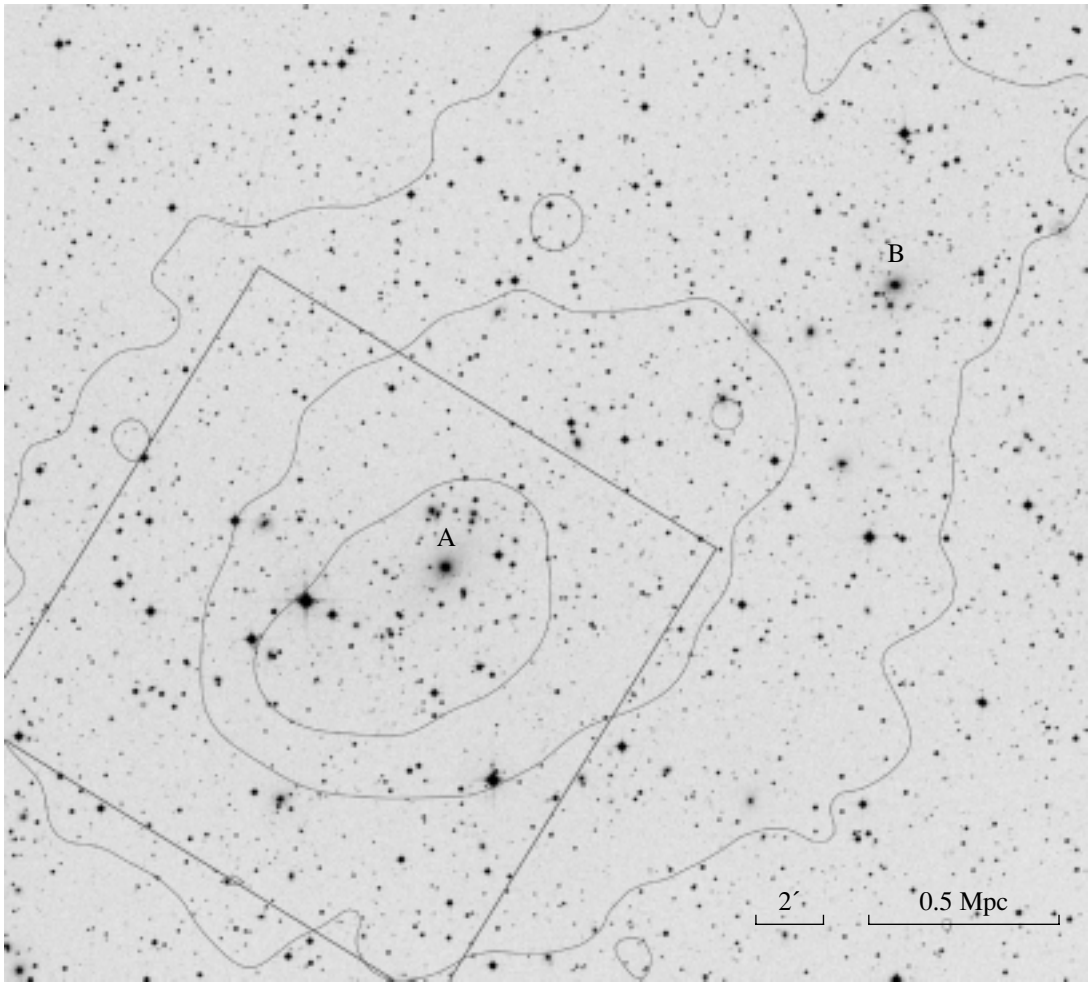


Fig. 1. An optical image of the cluster A3667 with superimposed X-ray contours.

is so far the only one that has been planned and performed with a long exposure time, which makes a detailed analysis possible. Therefore, below, we discuss the properties of cold fronts using A3667 as an example.

The cluster A3667 shows evidence of the current formation from two components in its optical, X-ray, and radio images (see Sodre *et al.* 1992; Knopp *et al.* 1996; Röttgering *et al.* 1997). Two concentrations of galaxies around the brightest galaxies, designated as A and B in Fig. 1, are clearly visible in the optical image. Weak-gravitational-lensing observations show that significant mass concentrations also correspond to them (Joffre *et al.* 2000). The X-ray image is elongated along the A–B axis, with the primary and weak secondary maxima corresponding to galaxies A and B, respectively. However, the most prominent structure in the X-ray image is a large brightness discontinuity southeast of galaxy A, which is oriented perpendicular to the A–B axis. The radio map contains giant diffuse sources, ~0.5–1 Mpc in size, which are also oriented perpendicular to the A–B axis. Their probable nature is synchrotron radiation from the relativistic electrons accelerated during the merger of the cluster components (Röttgering *et al.* 1997). Thus, all data indicate that the groups around galaxies A and B can be identified with the central parts of the merging cluster components. Although the components must move at a velocity of ≈2000–3000 km s^{-1} (Sarazin 2001), the observed light-of-sight velocity difference between galaxies A and B is only 120 km s^{-1} (Katgert *et al.* 1998). This suggests that the merger takes place along an axis that is almost perpendicular to the line of sight. Such a simple orientation considerably facilitates the interpretation of the observations.

CHANDRA OBSERVATIONS OF THE CLUSTER A3667

The central part of A3667 was observed by the Chandra observatory in December 1999. The field

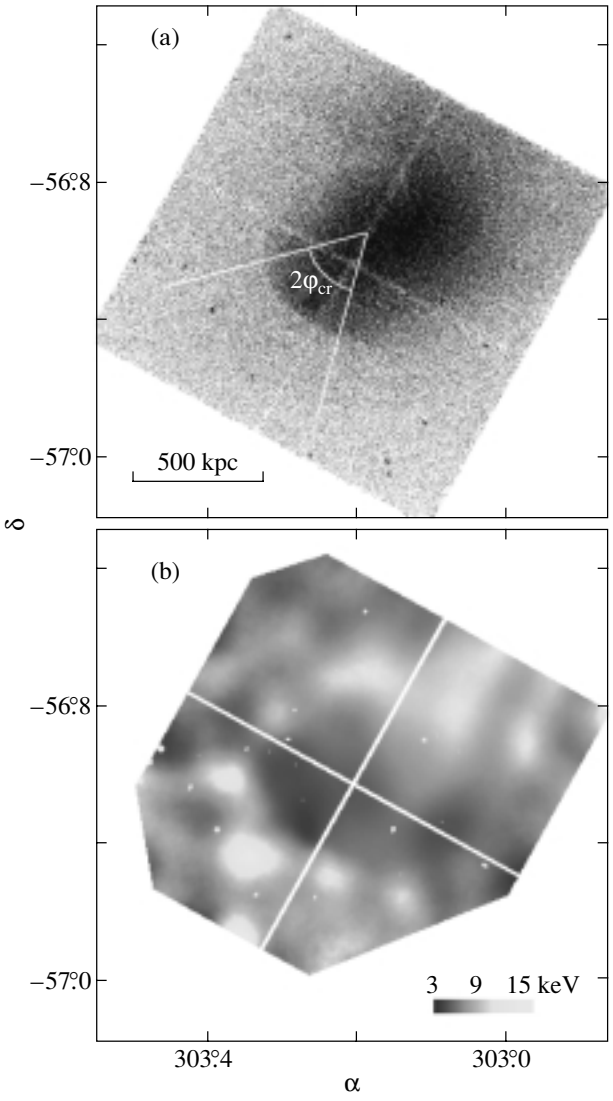


Fig. 2. (a) The Chandra X-ray image and (b) temperature map of the intergalactic plasma in A3667. Note that the surface-brightness discontinuity (cold front) is sharp within the sector $2\varphi_{cr} = 60^\circ$ and gradually dissolves outside.

of view (the square in Fig 1) is centered on the most interesting, from our viewpoint, object: the surface-brightness discontinuity about 8′ southeast of galaxy A. The angular size 1″, which is approximately equal to the Chandra angular resolution, corresponds to the physical size 1.46 kpc at the cluster redshift $z = 0.055$ (the Hubble constant is assumed to be $H_0 = 50$ km s^{-1} Mpc^{-1}). Technical aspects of our analysis of the Chandra data are described in detail in Vikhlinin *et al.* (2001a).

The Image and Temperature Map

We analyzed the X-ray images in the 0.7–4 keV energy band, in which the cluster-to-background

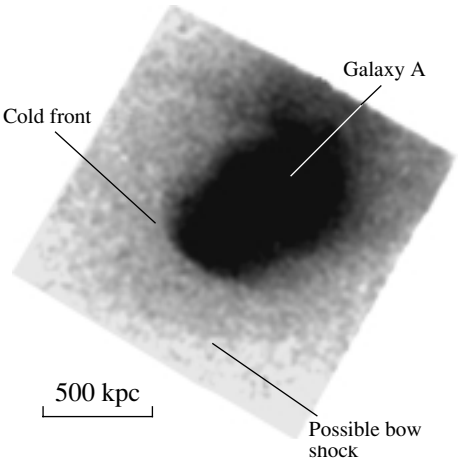


Fig. 3.

surface-brightness ratio is at a maximum. When processing the image, we took into account slight spatial nonuniformities of the detector background, as well as variations in the detector quantum efficiency and in the mirror effective area.

The image obtained is shown in Fig. 2a, and its slightly smoothed version with an indication of principal features is shown in Fig. 3. The most prominent structure is a sharp surface-brightness discontinuity whose shape is accurately described by a circumference of radius 410 ± 15 kpc. The brightness discontinuity is very sharp; it cannot be distinguished from an infinitely narrow one even with the Chandra unique angular resolution. The front thickness is considered in detail below.

The temperature map of the intracluster gas is shown in Fig. 2b. To suppress the noise, the map was obtained with an effective smoothing $\sigma = 48''$, except for a narrow region near the front, where the resolution is $\sigma = 24''$. All of the clear temperature variations to the right and from above the front location, in the region with a high surface brightness, are statistically significant. To the left and from below the front, the measurement error of the temperature is large because of the decrease in cluster surface brightness; thus, the spots with high temperatures (>12 keV) are statistically insignificant. However, it may be asserted with certainty that the gas temperature inside the discontinuity surface is much lower than that outside. Consequently, we deal with a cold front rather than with a shock.

The second, weak brightness discontinuity is clearly seen in the smoothed image south of the cold front. Below, we give arguments that allow this discontinuity to be identified with the bow shock ahead of the cold gas cloud whose boundary is the front, which moves at a slightly supersonic speed.

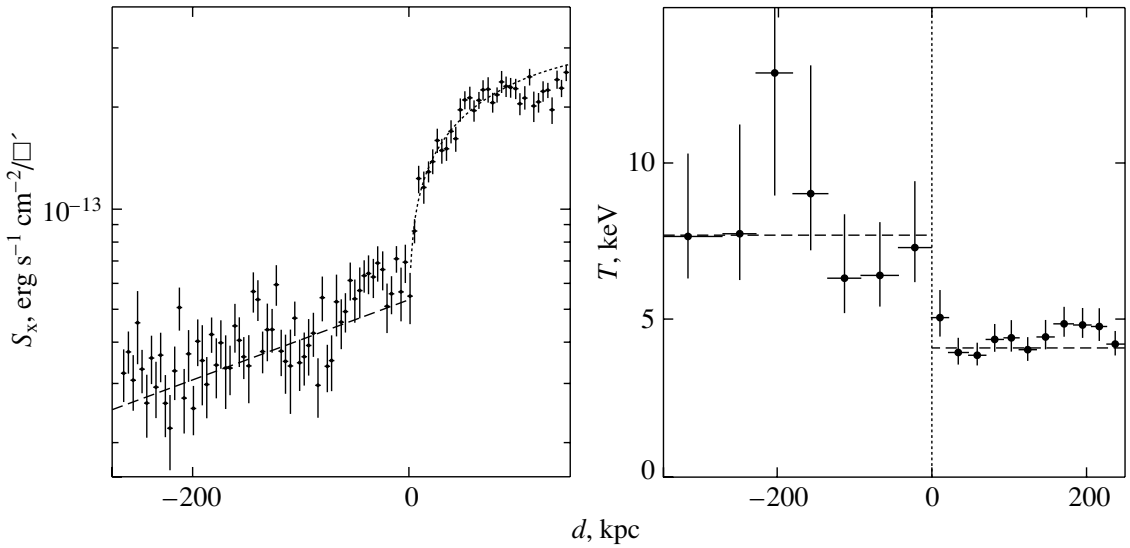


Fig. 4. X-ray brightness and gas-temperature profiles within the sector ±15° of the direction of cold-front motion. The dashed line in the brightness profile to the left of the front location corresponds to the model constructed from ROSAT data (without any renormalization)

In conclusion, we note a general similarity of our temperature map to the schematic map constructed from ASCA data (Markevitch *et al.* 1999).

Analysis of the Brightness and Temperature Profiles

The properties of the gas near the cold front can be studied in more detail by using the brightness and temperature profiles constructed within a narrow sector of ±15° relative to the front symmetry axis (Fig. 4).

The brightness profile contains a sharp discontinuity. When the discontinuity boundary is crossed, the brightness doubles within 7–10 kpc and continues to gradually rise by a factor of 2 within 40–80 kpc. Since the discontinuity is nearly circular in projection and since the merger of the cluster components appears to take place in the plane of the sky, it would be reasonable to assume that the three-dimensional shape of the gas cloud is an ellipsoid of revolution with the symmetry axis perpendicular to the line of sight. This assumption allows the gas density inside the front to be easily determined by deprojecting the brightness profile. Omitting the description of the deprojection procedure (for details, see Vikhlinin *et al.* 2001a), we present the final results. The plasma density inside the front is $n_e = (3.2 \pm 0.48) \times 10^{-3}$ cm^{-3}, where the uncertainty conservatively includes the measurement errors and uncertainty in the geometry. The plasma density outside the front is $n_e = 0.82 \times 10^{-3}$ cm^{-3}.

Plasma parameters near the cold front in the cluster A3667

Region	T, keV	n_e, 10^{-3} cm^{-3}	$p = Tn_e$, 10^{-2} keV cm^{-3}
Outside the front	7.7 ± 0.8	0.82 ± 0.12	0.63 ± 0.11
Inside the front	4.1 ± 0.2	3.2 ± 0.5	1.32 ± 0.21

Although this has no physical meaning, we note that the ratio of the gas densities inside and outside the front is 3.9, which is close to the maximum compression in a strong shock propagating in a monoatomic gas. However, the temperature behavior definitely rules out the interpretation of the front as a shock. When the boundary of the X-ray brightness discontinuity is crossed, the temperature also changes sharply from about 8 to 4–5 keV (Fig. 4). The measured temperatures on both sides of the discontinuity are compatible with constant values within some distance from the front. Fitting the spectra collected within 275 kpc outside the front and 125 kpc inside by the optically thin plasma emission model yields $T_{\rm out} = 7.7 \pm 0.8$ keV and $T_{\rm in} = 4.1 \pm 0.2$ keV.

Thus, the gas density and temperature both undergo an abrupt change at the surface-brightness discontinuity. In this case, the pressure ratio is $p_{\rm in}/p_{\rm out} = 2.1 \pm 0.5$ (table); it is clear from the preceding discussion that the error of this value was estimated conservatively. If the cold front were at rest, then the equality $p_{\rm in} = p_{\rm out}$ would necessarily hold. An enhanced internal pressure implies that the front moves and the gas inside the front feels the drag force.

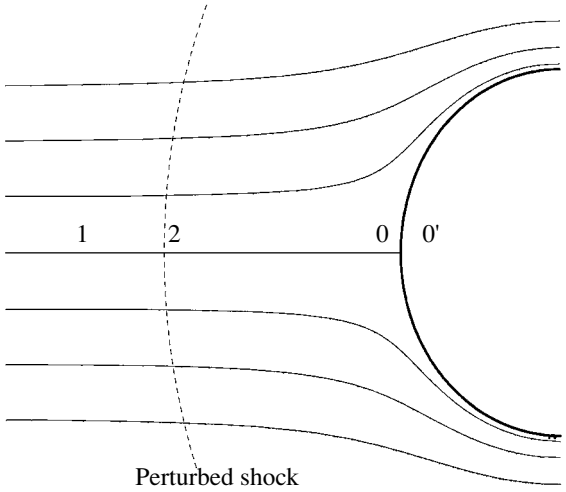


Fig. 5.

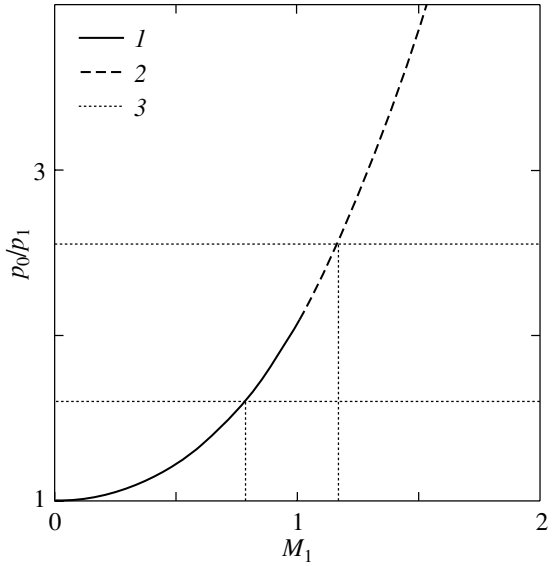


Fig. 6. Pressure ratio inside the cold front and in the outer gas versus velocity: *1* for $M < 1$, *2* for $M > 1$; *3* uncertainty in the measurement of p_0/p_1.

HYDRODYNAMIC EFFECTS NEAR THE COLD FRONT

Most of our subsequent analysis is performed assuming that the gas flow outside the cold front is adiabatic, i.e., that the plasma heat conduction may be disregarded. This assumption can be justified as follows. We are mainly concerned with scales that are much larger than the Coulomb mean free path; the heat conduction on such scales is slower than the gas motion. A more important reason is that the plasma heat conduction is most likely strongly suppressed by a magnetic field.

The Front Velocity

Determining the velocity of a blunt body of revolution from the drag force is a simple problem whose solution can be found in the book by Landau and Lifshitz. Let us consider a streamline far from the leading edge of the body (Fig. 5). The gas flows freely with Mach number M_1 far from the body in zone 1 and stops in front of its nose in zone 0. The pressure ratio p_0/p_1 unambiguously determines the flow velocity, $M_1 < 1$ (Fig. 6).

X-ray observations allow the gas parameters in zone 1 to be deduced. The pressure in zone 0 cannot be determined because of its small size. However, it is clear that the pressure in zone 0 must be equal to the pressure in zone 0′ (inside the cold front), which can be reliably measured. Thus, the $p_{0'}/p_1$ ratio can be used instead of p_0/p_1 to determine the front velocity. We see from Fig. 6 that the experimental uncertainty in the pressure ratio $p_0/p_1 = 2.1 \pm 0.5$ gives a narrow interval, $M_1 = 1 \pm 0.2$, which corresponds to the absolute front velocity $v = 1400 \pm 300$ km s^{-1} at the outer gas temperature $T = 7.7$ keV. Presently, this is the only more or less direct measurement of the velocity of large gas volumes in merging clusters.

Thus, by measuring the ratio of the pressures outside and inside the cold front, we can establish that it moves at a transonic speed relative to the hot outer gas. Consequently, we should expect a manifestation of gas-compressibility effects: (i) an enhanced density ahead of the leading edge; and (ii) a bow shock if the velocity even slightly exceeds the speed of sound.

The bow shock. A weak surface-brightness discontinuity is clearly seen in the X-ray image south of the cold front (Fig. 3), for which the identification with a bow shock suggests itself. At a given flow velocity, the bow shock must be characterized by a specific amplitude, location, and asymptotic angle with respect to the direction of motion. Let us consider these characteristics individually.

Amplitude. The surface-brightness profile within the sector 230°–280° contains a characteristic jump that can be described by the model of a spherical density discontinuity (Fig. 7). Fitting gives the density discontinuity $\rho_1/\rho_2 = 1.15 \pm 0.05$, whose substitution in the equation for a shock adiabat yields the shock velocity $M_1 = 1.10 \pm 0.03$.

Location. The faster the body moves, the closer to it the bow shock must be. The shock location for any velocity can be predicted either by the semi-analytical method of Moeckel (1949) or by numerically solving the hydrodynamic equations (see below). The distance between the front and the hypothetical bow shock, $d \approx 350$ kpc, for a given shape of the cold cloud (the diameter is 500 kpc and the radius of the

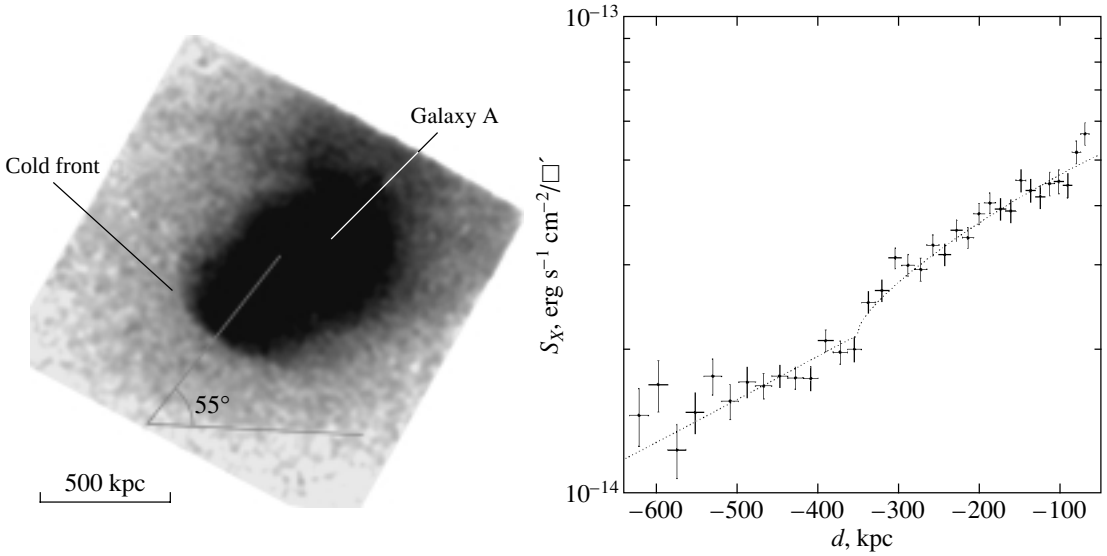


Fig. 7. Measuring the velocity from the bow shock (see the text).

spherical head is 410 kpc) corresponds to the velocity $M = 1.2$.

Asymptotic angle. At large distances, the bow shock front must be inclined with respect to the direction of motion at the angle $\alpha = \arcsin M^{-1}$. The asymptotic angle cannot be reliably determined from the image, but it is close to $\alpha = 55°$ (see Fig. 7). Hence, the travel velocity is $M_1 = 1/\sin\alpha = 1.2$.

Thus, the amplitude, location, and orientation of the weak density discontinuity are in reasonable agreement with the assumption that it is a bow shock for the cold front moving at a slightly supersonic speed, $M = 1.1$–1.2, which is within the limits of the error in determining the velocity from the amplitude of the pressure discontinuity at the very front. A major problem in identifying the density discontinuity with a bow shock is its absence (or at least it is indistinct) in the upper half of the image (Fig. 3). This could be explained by several factors: (1) the outer gas in the upper half is slightly hotter, which reduces the Mach number; since the bow shock rapidly recedes from the body and decreases its amplitude at small Mach numbers, it is unseen in the upper half; (2) the cold front does not move along a straight line, with the bow shock orientation reflecting the direction of motion in the past.

Compression of the outer gas near the leading edge. As the outer gas approaches the leading edge, its velocity significantly decreases, while its density and pressure increase. As a result, since the plasma emissivity is $\varepsilon \propto \rho^2$, excess X-ray emission must emerge in a zone near the leading edge. It is easy to calculate that the emissivity at the flow stagnation point increases by a factor of 2.4 at $M_1 = 1$. Hence, one might expect that the effect must generally be observable.

For the rise in X-ray brightness to be calculated, we must know the exact gas density distribution near the front. Since the flow velocity is transonic, no analytical method can be used to solve the hydrodynamic equations. Therefore, we used a numerical calculation to find the solution. Integration of the derived plasma emissivity distribution along the line of sight shows that the zone of enhanced X-ray brightness for M close to unity has a characteristic thickness of ≈90 kpc and that the amplitude of the brightness rise in the energy band under consideration is 1.6×10^{-14} cm^{-2}/□′. Amazingly, precisely such a rise in surface brightness above the profile extrapolation from large radii is observed (Fig. 4).

Thus, the following conclusions about the motion of the cold front can be reached:

(1) The velocity is close to the speed of sound, $M = 1 \pm 0.2$, as follows from the ratio of the gas pressures inside and outside the front.

(2) This velocity is confirmed by the observation of additional structures expected during flow with $M \sim 1$:

(i) the bow shock;

(ii) the outer-gas compression near the cloud nose.

(3) The existence of a bow shock allows the uncertainty range for the velocity to be narrowed: $1.0 < M < 1.2$.

Stationary Shape of the Cold Front

In the previous discussion, we tacitly assumed that the cold gas cloud moves as a rigid body with a constant shape. Is this assumption even roughly

valid? If hydrodynamic[1] instabilities are allowed to grow freely, then they significantly deform the moving cloud to the point of its complete destruction in a time that corresponds to the passage to a distance of the order of several cloud diameters (Jones *et al.* 1996). Thus, the regular boundary of the cold cloud suggests that the instabilities are suppressed and, hence, the front can move with an almost constant shape.

The constancy of the front shape can be judged more definitely from the dependence of the pressure on the angle of attack. The point is that all streamlines in the immediate vicinity of the front emerge from a small region near the leading edge. If the front shape is constant, then the pressure distribution along a streamline follows the Bernoulli equation—the pressure decreases with increasing flow velocity as one recedes from the leading edge; the gas pressure inside the front must closely follow the variations in external pressure; otherwise, the front shape will change with time. Thus, at a given front velocity, which can be determined by examining the central streamline and, thereby, does not depend on the assumption of constant shape, a clear-cut prediction of the angular gas-pressure distribution inside the front follows.

For the angular gas-pressure distribution to be measured, we must know the gas-density and temperature distributions. The density distribution can be easily measured by deprojecting the X-ray brightness profiles within narrow sectors at various angles from the direction of motion. The temperature distribution can be measured with a much lower accuracy: the measured value is 3.8 ± 0.40 keV within the sector $-10^\circ < \theta < 10^\circ$ of the direction of motion, 3.8 ± 0.55 keV within the sector $-30^\circ < \theta < -10^\circ$, and 4.7 ± 0.70 keV for $10^\circ < \theta < 30^\circ$.[2] Within the limits of the measurement errors, the temperatures are equal to the mean value $T = 4.1 \pm 0.2$ keV on the inside of the front (table). Therefore, we assume below that the gas temperature at the inner front boundary is constant.

At a constant temperature, the pressure is proportional to the gas density inside the front. The latter can be easily determined by analyzing the brightness profiles within various sectors of the direction of motion. The dependence of the gas pressure at the inner front boundary on the angle of attack obtained in this way is shown in Fig. 8. We see that the pressure decreases appreciably, by almost a factor of 1.5 at $\theta = 30^\circ$, where the front begins to smear. Note that an exact hydrodynamic calculation of the flow around a rigid body for Mach numbers slightly larger than unity yields precisely this pressure distribution (the solid line in Fig. 8). Thus, the angular dependence of the cold-gas pressure indirectly but reliably suggests that the shape of the cold front is stationary.

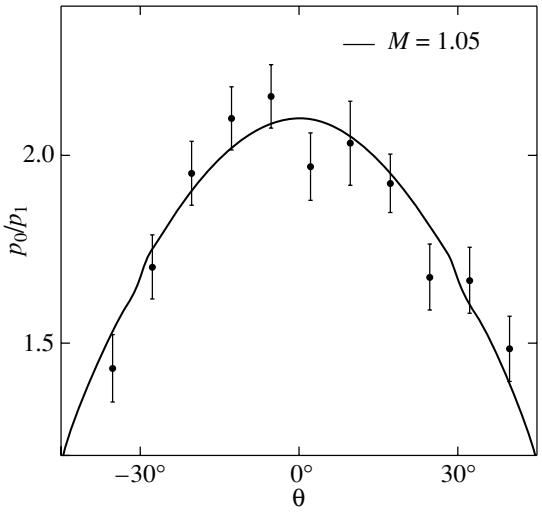


Fig. 8. The angular pressure distribution inside the front.

It should also be noted that if we assume *a priori* the shape of the cold front to be stationary and the gas temperature at its inner boundary to be constant, then the dependence $p(\theta)$ is a convenient experimental diagnostic for the velocity. The point is that under our assumptions, the velocity can be unambiguously determined from the dependence $\rho(\theta)/\rho(0)$. The latter, in turn, can be easily reconstructed from the observed X-ray brightness distribution. At the same time, $p(\theta)$ is sensitive to the velocity near $M = 1$; thus, the data shown in Fig. 8, can be described by the model of a steady flow only in the range $0.7 < M < 1.5$.

The Dark Matter Flying with the Gas Cloud

Since the cold-front shape is stationary, it would be reasonable to assume that the cold gas inside the front is in hydrostatic equilibrium in the gravitational potential produced by the matter that moves with the gas. In this case, the equation for gravitational equilibrium

$$\nabla\varphi = -\frac{\nabla p}{\rho} = -\frac{kT}{\mu m_p}\frac{\nabla\rho}{\rho}, \qquad (1)$$

where $\mu = 0.6$ (the mean molecular weight of the ionized cosmic mixture of hydrogen and helium) yields the gravitational-potential distribution and, hence, information on the distribution of the matter that flies with the front. Of course, the data quality is too low to directly deduce the gas density from the surface-brightness distribution and to determine the matter

[1] In our case, Rayleigh–Taylor instability is of primary importance.

[2] The spectra were measured in a 70-kpc-wide stripe on the inside of the front; the angles are counted off from north to south through east.

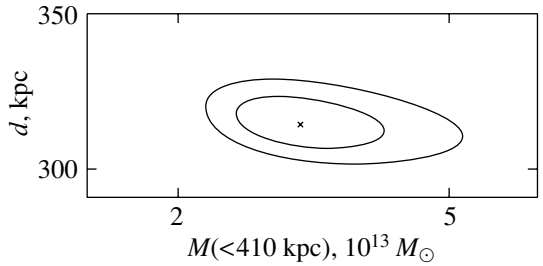


Fig. 9. 68 and 95% confidence regions for the parameters of the King profile of the matter: the center-of-mass position d, counted off from the center of curvature of the cold front, and the mass M within 410 kpc. The cross denotes the best-fit parameters.

density by differentiating the latter. Instead, we can specify a model density distribution for the gravitating (dark) matter, predict the surface-brightness distribution for it, and then fit the distribution parameters until the best agreement with the data is achieved.

Using this method, we assume that the density distribution for the gravitating matter is described by the King profile: $\rho_m = \rho_{m0}(1 + r^2/r_c^2)^{-3/2}$. The plasma-density distribution can be calculated by substituting the gravitational potential of the King profile $\varphi = 4\pi G\rho_0\,(1 - \mathrm{asinh}x/x)$, where $x = r/r_c$, in Eq. (1) (or in the Boltzmann distribution). Integrating the plasma emissivity, which is equal to the square of the density, along the line of sight, we find the model surface-brightness distribution.

By maximizing the likelihood function calculated from the data at the cloud nose (approximately within 200 kpc of the front), we can determine the King-profile parameters: the center-of-mass position (d), the core radius (r_c), and the central density, which we parametrize through the total mass within 410 kpc, the radius of the cold front. The use of only the cloud nose is explained by several factors: (1) from the viewpoint of the cold-front dynamics, it is desirable to find the gravitational-potential distribution precisely in this region; (2) the assumption that the gas distribution is cylindrically symmetric holds here; and (3) this region is so small that the King law has enough freedom to describe virtually any reasonable distribution of the gravitational potential.

The fit obtained by the maximum-likelihood method is shown in Fig. 9. The center of mass of the dark matter is located at $d = 315 \pm 7$ kpc from the center of curvature of the cold front (i.e., at 95 kpc from the cloud nose), the mass within 410 kpc is $(3.2 \pm 0.8) \times 10^{13}\ M_\odot$, and the core radius lies in the range $50 < r_c < 200$ kpc. Note that the derived mass and core radius are typical of the central part of a moderately rich galaxy cluster with a temperature of 3–4 keV. The total mass within 100 kpc (the distance from the center of mass to the front) is $4.6 \times 10^{12}\ M_\odot$. This value is a factor of 15 larger than the gas mass within the same radius and, hence, the gas contribution to the gravitational potential may be disregarded.

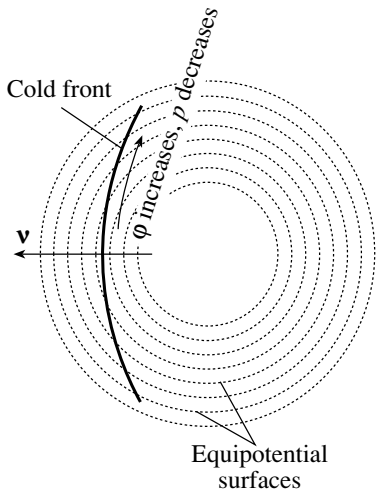


Fig. 10.

The distance from the center of mass to the front is only 25% of its radius of curvature. Thus, the dark matter may be said to drag the cold gas. Note that such a mutual arrangement of the gravitating matter and the gas is required for the cold front to be globally stable. The situation is illustrated in Fig. 10. In the absence of a gravitational field, the gas pressure inside the front would be the same everywhere, but at the same time, the pressure outside the front decreases from the center to the periphery according to the Bernoulli equation. Consequently, the front must have gradually straightened out. The presence of a gravitational field not only allows the pressure inside the front to be changed simultaneously with the external pressure but also provides stability against global changes in the shape. Indeed, we deform the front in Fig. 10 in such a way that its curvature increases. In this case, the pressure outside on the front periphery decreases because the flow velocity increases and the pressure inside increases, because the front passes to a lower potential level; the emerged pressure difference restores the original front shape. Thus, a substantial concentration of the dark matter that flies with the cold gas is a necessary condition for the cold-front stability. Figuratively speaking, when flying into the cluster, the cold gas cloud must eject its outer layers until the cloud boundary approaches an equilibrium position.

A massive concentration of dark matter is required not only to provide global front stability but also to suppress the small-scale Rayleigh–Taylor instability. The point is that because of the drag by the outer

gas, the flying cloud must gradually decelerate, which gives rise to an effective gravity directed from the inside of the cold front outward. Thus, we achieve a situation where a layer of dense gas overlies a layer of less dense gas and this configuration is unstable. Let us make numerical estimates. The drag force acting on the cloud is $F_{\rm drag} = C \times \rho_{\rm out} v^2 A/2$, where $\rho_{\rm out}$ is the density of the outer gas, A is the cloud cross-section area, and $C \approx 0.4$ is the drag coefficient for a cylinder with a rounded head at transonic speed. The acceleration of the gas cloud is

$$g_{\rm drag} = \frac{F_{\rm drag}}{M} \approx \frac{0.2\rho_{\rm out} v^2 \pi r^2}{4/3\pi r^3 \rho_{\rm in}} = 0.15 \frac{\rho_{\rm out}}{\rho_{\rm in}} \frac{v^2}{r}. \quad (2)$$

Substituting numerical values yields $g_{\rm drag} \approx 8 \times 10^{-10}$ cm s^{-2}. At the same time, the acceleration of gravity near the front surface is

$$g = \frac{GM(<100\ {\rm kpc})}{(100\ {\rm kpc})^2} = 6.4 \times 10^{-9}\ {\rm cm\ s^{-2}} \quad (3)$$

and, thus, it dominates over the acceleration by drag. Note that in the absence of dark matter, the acceleration of gravity produced by the gas would be a factor of 15 lower, which is not enough to compensate for the acceleration by drag. Thus, a purely gas cloud would be unstable according to Rayleigh–Taylor, which is observed in numerical experiments (Jones *et al.* 1996).

SUPPRESSION OF THE TRANSPORT PROCESSES

Using the Chandra unique angular resolution, we can establish that the front width is smaller than the Coulomb mean free path in the plasma. Figure 11 shows the brightness profile through the front constructed in narrower (2.2-kpc-wide) rings. We see that the brightness abruptly rises by a factor of 1.7 within only 2 kpc. Let us compare this distance with the mean free path in plasma on both sides of the front. Four different characteristic mean free path lengths can be identified near the front: for thermal particles on each side of the front, $\lambda_{\rm in}$ and $\lambda_{\rm out}$, as well as the lengths corresponding to the scattering of the particles that fly from one side of the front to the other, $\lambda_{\rm in\to out}$ and $\lambda_{\rm out\to in}$. Using relations from Spitzer (1962), we have for $\lambda_{\rm in}$ and $\lambda_{\rm out}$

$$\lambda = 15\ {\rm kpc} \left(\frac{T}{7\ {\rm keV}}\right)^2 \left(\frac{n_e}{10^{-3}\ {\rm cm^{-3}}}\right)^{-1} \quad (4)$$

and for $\lambda_{\rm in\to out}$ and $\lambda_{\rm out\to in}$

$$\lambda_{\rm in\to out} = \lambda_{\rm out} \frac{T_{\rm in}}{T_{\rm out}} \frac{G(1)}{G(\sqrt{T_{\rm in}/T_{\rm out}})}, \quad (5)$$

$$\lambda_{\rm out\to \mathit{in}} = \lambda_{\rm in} \frac{T_{\rm out}}{T_{\rm in}} \frac{G(1)}{G(\sqrt{T_{\rm out}/T_{\rm in}})},$$

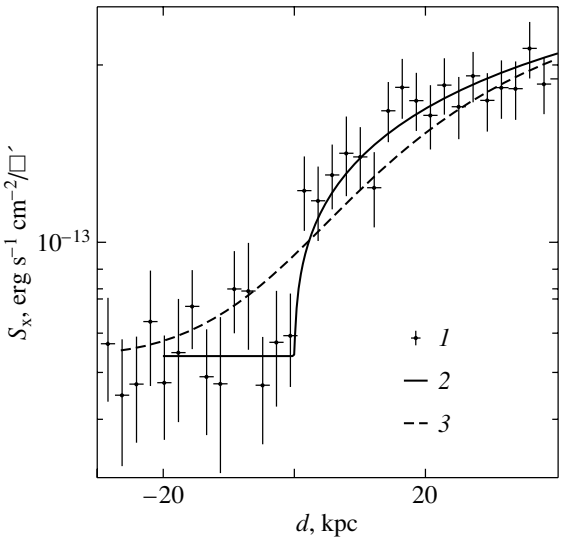


Fig. 11. The brightness profile near the cold front (*1*); (*2*) the fit by the model of an infinitely sharp spherical density discontinuity and (*3*) by the model of the same discontinuity smoothed with a $\sigma = 20$ kpc Gaussian, which roughly corresponds to one Coulomb mean free path length.

where $G(x) = [\Phi(x) - x\Phi'(x)]/2x^2$ and $\Phi(x)$ is the error function. For the measured gas parameters (table), the numerical values are $\lambda_{\rm out} = 22$ kpc, $\lambda_{\rm in} = 1.6$ kpc, $\lambda_{\rm in\to out} = 13$ kpc, $\lambda_{\rm out\to in} = 3.5$ kpc.

The outer gas near the leading edge of the front moves slowly. Consequently, the diffusion processes must take place freely and they must smear the front to a width of at least several mean free path lengths in a short time. In our case, diffusion must arise mainly from the inside of the front outward, because the particle flux per unit area is proportional to $nT^{1/2}$. Therefore, one might expect the front to smear to a width of at least several $\lambda_{\rm in\to out}$. However, this is clearly in conflict with the observed front sharpness. A quantitative characteristic of the sharpness can be derived by fitting the observed profile by a model that corresponds to the projection of a spherical density discontinuity smoothed in three-dimensional space with a Gaussian $\exp(-\Delta r^2/2\sigma^2)$. The best fit (the solid line in Fig. 11) is achieved for $\sigma = 0$, with the formal upper limit on σ at a 95% confidence level being only 5 kpc; the model of a discontinuity smoothed with the Coulomb mean free path length does not describe the data (the dashed line in Fig. 11). Thus, the observed front sharpness requires a substantial suppression of the transport processes.

The suppression of transport processes in the intracluster medium is generally attributed to the presence of a magnetic field. The magnetic field is efficient in two cases. The first is its great twisting; in this case, an upper limit on the effective size of magnetic

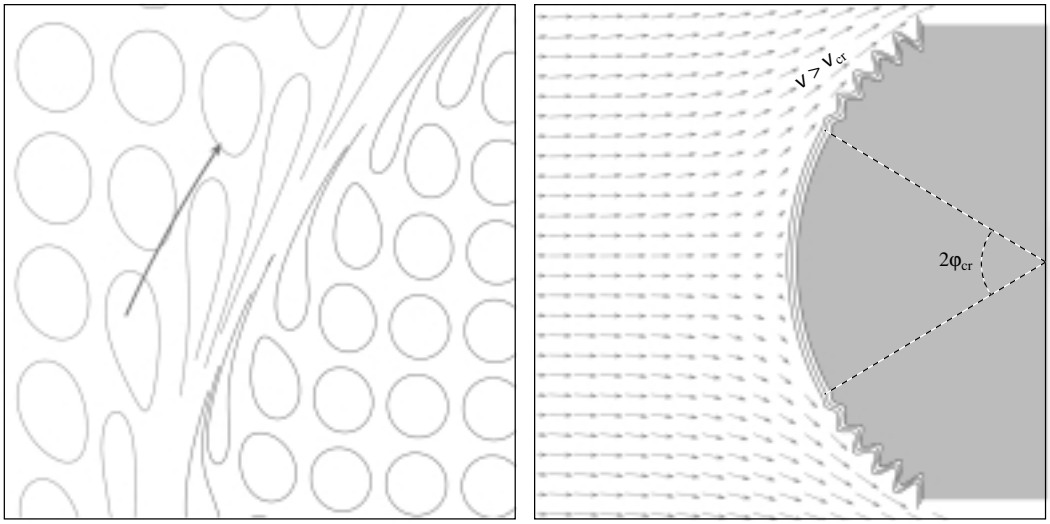


Fig. 12. A scenario for the formation of a magnetic layer near the cold front. The initially tangled magnetic field lines in the hot outer gas are stretched along the front because of the tangential velocity. The magnetic field lines inside the front are also stretched in the same direction, because the cold gas must be washed away through viscosity without complete magnetic isolation. In the long run, a narrow layer in which the magnetic field is parallel to the interface between the gases and to the flow direction is formed. The interface is subject to the hydrodynamic instability of a tangential discontinuity. As long as the flow velocity is lower than a certain critical value v_{cr}, the surface tension produced by the layer magnetic field suppresses the front instability.

loops is the observed front sharpness, $\sigma < 5$ kpc. The other possibility is the existence of an ordered magnetic field parallel to the cold front, at least in a narrow layer, which virtually isolates the cold gas and does not allow the strong density and temperature gradients to dissolve.

THE MAGNETIC-FIELD STRUCTURE AND STRENGTH

Even moderately weak magnetic fields can significantly change the properties of the hot gas in clusters. Examples of invoking magnetic fields to explain the mismatch between theory and experiment are the suppression of plasma heat conduction, which is required to account for the presence of a radiatively cooled gas at the cluster centers (for a review, see Fabian 1994), and the discrepancy between the cluster masses measured from X-ray data and from strong gravitational lensing (Loeb and Mao 1994; Miralda-Escudé and Babul 1995). Observationally, the intergalactic magnetic field is ill-famed because of the difficulties involved in its measurement. Up until now, the two principal measurement methods were the use of the Faraday effect in radio sources observed through the intergalactic medium (Kim *et al.* 1991) and the interpretation of the diffuse radio and hard X-ray cluster emissions resulting from the synchrotron and inverse Compton mechanisms from the same population of relativistic electrons (Fusco-Femiano *et al.* 1999). These two methods yield a magnetic-field strength at a μG level, but the uncertainty in the measurements and in their interpretation is high. It turns out that the properties of the cold front in A3667 make it possible to determine the magnetic-field strength by a radically new method, from its effect on the intergalactic plasma dynamics.

The idea is as follows. The cold front is very sharp and smooth in shape and, hence, it must be mechanically stable. At the same time, it may be shown that at the observed velocity of the outer-gas flow, the front must be rapidly perturbed through the instability of the tangential discontinuity. An examination of the possible mechanisms for the suppression of this instability points to the tension produced by a magnetic field as the most probable candidate.

The magnetic configuration required for this can be produced according to the scenario schematically shown in Fig. 12. We assume that a randomly oriented magnetic field is frozen in the intracluster gas. Because of the tangential plasma motions during the flow around the front, the magnetic loops are stretched along the surface. Sooner or later, the oncoming magnetic fields are reconnected to produce a layer in which the magnetic field is parallel to the interface and to the flow direction. This magnetic configuration not only stops the transport processes but is also required to suppress the instability of the tangential discontinuity.

As one recedes from the leading edge, the tangential flow velocity increases. If the magnetic-field

strength in the layer is moderately high, then sooner or later the field will be incapable of stabilizing the instability, and starting from this point, the front begins to smear; this position is denoted by φ_{cr} in Fig. 12. It is difficult not to perceive a parallel with the X-ray image of A3667, from which we see that the front at $\varphi \approx 30^\circ$ ceases to be sharp and gradually smears completely (Fig. 2). The identification of the point $\varphi = 30^\circ$ with the point at which the magnetic field is close to its limiting value that provides mechanical stability allows its strength to be determined.

Let us now consider in detail the entire line of reasoning. For simplicity, we assume that the front velocity is exactly equal to $M = 1$.

Hydrodynamic Instability of the Front

The interface between two fluids can generally be perturbed by two types of hydrodynamic instability: (i) Rayleigh–Taylor instability, which arises if the effective gravity is directed from the heavier fluid to the lighter one; and (ii) tangential-discontinuity (Kelvin–Helmholtz) instability, which arises when there is a tangential velocity of the relative fluid motion. As was discussed above, the dark matter flying with the front produces a gravitational field that suppresses the Rayleigh–Taylor instability. Therefore, we consider below only the tangential-discontinuity instability.

Numerical calculation gives the following dependence of the flow velocity near the front on the angle from the direction of motion:

$$M \simeq 1.1 \sin\varphi. \qquad (6)$$

At $\varphi = 30^\circ$, the Mach number reaches $M = 0.55$. Since the flow velocity is high, one might expect a rapid growth of the tangential-discontinuity instability followed by the formation of a turbulent layer that smears the front. Let us formulate several widely known properties of the tangential-discontinuity instability that are tacitly used below.

(1) The modes whose wave vector is parallel to the flow, at least at moderately high flow velocities, $M \lesssim 1.5$ ($M < \sqrt{3}$ for equal densities on both sides of the interface), grow most rapidly. We restrict our subsequent analysis to such parallel perturbations.

(2) The short-wavelength modes grow faster than the long-wavelength modes.

(3) Since the perturbations grow exponentially, a nonlinear stage is reached rapidly. When the nonlinear stage is reached, the amplitude is approximately equal to the wavelength of the initially linear perturbation; subsequently, the perturbations with a given wavelength merge into large vortices. Thus, if we will be able to show that the perturbations with wavelength λ are unstable, then it can be asserted with certainty that the formed turbulent layer must have a thickness of at least λ.

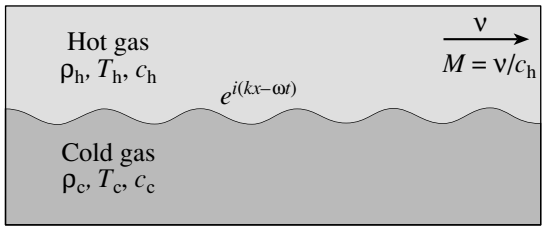


Fig. 13. Analysis of the tangential-discontinuity instability. The hot, low-density gas moves at velocity ν relative to the cold, denser gas. Perturbations in the form of plane waves, $\exp(i(kx - \omega t))$, are superimposed on the interface, with the perturbation wave vector being parallel to the flow. For a given wave vector, ω can be derived from the dispersion relation (7) in pure hydrodynamics and from (15) in MHD.

The dispersion equation for perturbations of the gas interface. It is convenient to analyze the interface stability by considering small perturbations in the form of plane waves and by solving the corresponding dispersion equation. The dispersion equation for incompressible fluids or compressible gases of equal densities can be found in the book by Landau and Lifshitz. After cumbersome transformations and after rejecting stable solutions, a similar analysis for gases of different densities yields the dispersion equation

$$-\frac{1}{\omega^2} - \frac{c_h^2/c_c^2}{(\omega - Mc_h k)^2} + \frac{1}{k^2 c_c^2} = 0, \qquad (7)$$

where M is the flow velocity of the hot gas, expressed in terms of the Mach number; and c_c and c_h are the speeds of sound in the cold and hot gases (Fig. 13), respectively. This equation was found in an equivalent form by Miles (1958).

Application to A3667. For reasonable M, Eq. (7) has two complex solutions for the frequency, one of which corresponds to the exponentially growing mode. The instability growth time is $\tau = (\mathrm{Im}\omega)^{-1}$. Although the perturbations with any wavelength are formally unstable, the growth time for some of them may to be short compared to the lifetime. The time scale with which τ should be compared is the time it takes for the cold cloud to cross the cluster[3]: $t_{cross} = L/M_\infty c_\infty$, where L is the characteristic cluster size. The quantity $\exp(t_{cross}/\tau)$ is the factor of perturbation growth in the time it takes for the cloud to traverse the distance L; if $t_{cross}/\tau > 1$–10, then the corresponding perturbation may be considered to be

[3] Since the velocity of the cold cloud exceeds the speed of sound in its gas, the cloud must fly into the cluster from the outside rather than accelerate somewhere near its current position.

effectively unstable, because it will have enough time for its growth to the nonlinear stage.

For the observed temperatures and $M_\infty = 1$, the solution to Eq. (7) is given by the approximate relation $t_{\rm cross}/\tau \simeq 11.8\ M\ L/\lambda$, where $\lambda = 2\pi/k$ is the perturbation wavelength. Taking into account the distribution of the local Mach number near the front [Eq. (6)], we have

$$\frac{t_{\rm cross}}{\tau} \simeq 3.3\,\frac{L}{\lambda}\,\sin\varphi. \qquad (8)$$

For the typical cluster size $L = 1$ Mpc, it follows from this relation that the perturbations with $\lambda = 10$ kpc are unstable ($t_{\rm cross}/\tau = 10$) at $\varphi = 1.7°$; the modes with $\lambda < 57$ kpc and $\lambda < 150$ kpc are unstable at $\varphi = 10°$ and $\varphi = 30°$, respectively.

Another effect that limits the perturbation lifetime should be considered. The point is that the linear perturbations have a nonzero group velocity $v_{\rm dr} = d{\rm Re}\omega/dk$ and, thus, drift parallel to the gas flow as they grow. In some cases, the perturbation can reach the front periphery before it grows to the nonlinear stage. Let us consider this question in detail. Analysis of the solutions to the dispersion equation (7) shows that the frequency is always directly proportional to the wave vector and, hence, the group velocity is equal to the phase velocity, $v_{\rm dr} = {\rm Re}(\omega)/k$. Moreover, over the entire range of flow velocities near the cold front in A3667, the imaginary part of the complex solution for the frequency is proportional, with good accuracy, to its real part:

$${\rm Re}\omega/{\rm Im}\omega \simeq 0.70. \qquad (9)$$

It thus follows that there is a simple relation between the growth factor of the perturbation and its displacement due to the drift:

$$\text{Growth factor} = \exp\left(\int {\rm Im}\omega dt\right) \qquad (10)$$
$$\simeq \exp\left(1.43\int {\rm Re}\omega dt\right) = \exp\left(1.43k\int \frac{{\rm Re}\omega}{k}dt\right)$$
$$= \exp\left(1.43k\int dl\right) = \exp\left(4.7\frac{R}{\lambda}\frac{\Delta\varphi}{30°}\right),$$

where R is the radius of curvature of the front and $\Delta\varphi$ is the displacement of the wave packet in angle of attack. Assuming, as previously, that a growth factor of e^{10} is required for the perturbation to pass to the nonlinear stage, we obtain the condition of effective stability for a front with the radius $R = 410$ kpc:

$$\lambda < 190\ \text{kpc}\ \frac{\Delta\varphi}{30°}. \qquad (11)$$

Perturbations with a longer wavelength reach the front periphery before reaching the nonlinear stage. It is clear, however, that for a perturbation with $\lambda =$ 10–30 kpc, the instability condition (11) is definitely satisfied.

Thus, it may well be asserted that the perturbations with wavelengths <20–30 kpc are unstable over the entire front surface and, hence, one might expect a turbulent layer with at least this thickness to be formed. Since the observed front thickness is <5 kpc, the tangential-discontinuity instability must be suppressed.

The universally known suppression mechanisms of the tangential-discontinuity instability are (i) a gravitational field and (ii) surface tension at the gas interface. Let us consider these mechanisms for the cold front in A3667.

The Role of Gravity in Suppressing the Instability

If a gravitational field acting in the direction from the lighter fluid to the heavier fluid is applied to the tangential discontinuity, then it is capable of suppressing the tangential-discontinuity instability. The condition for stability against sinusoidal perturbations with a wave vector k parallel to the flow is the inequality

$$\frac{g}{k} > v^2\frac{\rho_{\rm c}\rho_{\rm h}}{\rho_{\rm c}^2 - \rho_{\rm h}^2}. \qquad (12)$$

Thus, only the long-wavelength modes are suppressed, while the short-wavelength modes that do not satisfy this condition grow almost freely. Since $\rho_{\rm c}$ is much larger than $\rho_{\rm h}$, we transform condition (12) into

$$\frac{g}{k} > v^2\frac{\rho_{\rm c}\rho_{\rm h}}{\rho_{\rm c}^2 - \rho_{\rm h}^2} \approx v^2\frac{\rho_{\rm h}}{\rho_{\rm c}} = v^2\frac{T_{\rm c}}{T_{\rm h}} \qquad (13)$$
$$= v^2\frac{T_{\rm c}}{c_s^2\mu m_p/\gamma} = \frac{T_{\rm c}}{\mu m_p/\gamma}M^2.$$

The free-fall acceleration near the front was obtained above; its substitution in the derived inequality yields the stability condition

$$\lambda[\text{kpc}] > 3500\ M^2 = 4200\sin^2\varphi. \qquad (14)$$

Using this condition, we find that the perturbations with $\lambda = 10$ kpc are stable only within the sector $\varphi < 3°$; the stability is achieved only for $\lambda > 32$ kpc at $\varphi = 5°$ and for $\lambda > 1050$ kpc at $\varphi = 30°$.

Thus, gravity cannot prevent the formation of a turbulent layer with a thickness of 10–20 kpc on virtually the entire front surface. In this light, the most likely suppression mechanism of the front instability seems to be the surface tension produced by a magnetic field.

Suppression of the Instability by a Magnetic Field

A magnetic field suppresses the tangential-discontinuity instability in a perfectly conducting plasma if it is directed roughly along the flow and if its strength is high enough. Even if the magnetic field outside and inside the cold front was initially random, the tangential plasma motions that arise during the flow around the front are capable of producing the required magnetic configuration (see Fig. 12).

Analysis of the tangential-discontinuity stability in MHD is radically simplified if the plasma is assumed to be incompressible. In our case, this can be done, because the hot gas in the region under consideration, ($\varphi \lesssim 30^\circ$), moves with a moderately large Mach number ($M \lesssim 0.5$), while the cold gas feels only the phase velocity of the growing perturbations, which is low. The possibility of using the incompressibility approximation is confirmed by the fact that the growth time of hydrodynamic instability derived from Eq. (7) virtually matches the value obtained in the incompressible limit, i.e., at $Mc_\mathrm{h} \to v$, $c_\mathrm{c}/c_\mathrm{h}^2 \to \rho_\mathrm{h}/\rho_\mathrm{c}$, $c_\mathrm{c} \to \infty$.

The dispersion equation for small tangential-discontinuity perturbations in a perfectly conducting, incompressible plasma can be written as (Syrovatskiĭ 1953)

$$\rho_\mathrm{h}(\omega - kv)^2 + \rho_\mathrm{c}\omega^2 = k^2\left(\frac{B_\mathrm{h}^2}{4\pi} + \frac{B_\mathrm{c}^2}{4\pi}\right), \qquad (15)$$

where B_h and B_c are the magnetic-field strengths in the hot and cold gases (the field is assumed to be parallel to the flow), and ρ_h and ρ_c are the gas densities. The roots of this equation are real (i.e., the discontinuity is stable) if

$$B_\mathrm{h}^2 + B_\mathrm{c}^2 > 4\pi\,\frac{\rho_\mathrm{h}\,\rho_\mathrm{c}}{\rho_\mathrm{h} + \rho_\mathrm{c}}\,v^2. \qquad (16)$$

The Magnetic-Field Strength Near the Front in A3667

Before applying the stability condition (16) to the cold front in A3667, let us rewrite it in a more convenient form by assuming that the magnetic pressure is either low compared to the gas pressure or the p_mag/p ratio is the same on both sides of the discontinuity:

$$\frac{B_\mathrm{h}^2}{8\pi} + \frac{B_\mathrm{c}^2}{8\pi} > \frac{1}{2}\,\frac{\gamma M^2}{1 + T_\mathrm{c}/T_\mathrm{h}}p_\mathrm{gas}. \qquad (17)$$

Thus, at the observed temperatures, for the front within the sector $\varphi < 30^\circ$, where the flow velocity is $M \le 0.55$, to be stable, the condition $(B_\mathrm{h}^2 + B_\mathrm{c}^2)/8\pi > 0.17p$ must be satisfied. If the front smearing outside this sector is interpreted as the generation of instability, then the inferred lower limit is an estimate of the total magnetic pressure on both sides of the discontinuity. So,

$$p_\mathrm{mag,h} + p_\mathrm{mag,c} = 0.17p_\mathrm{gas}. \qquad (18)$$

In principle, we must take into account the fact that even if the magnetic field is too weak to provide the stability, it still decelerates the growth of perturbations,

$$\tau = (\mathrm{Im}\omega)^{-1} \qquad (19)$$
$$= k^{-1}\left[v^2\frac{\rho_\mathrm{c}\rho_\mathrm{h}}{(\rho_\mathrm{c} + \rho_\mathrm{h})^2} - \frac{B_\mathrm{h}^2 + B_\mathrm{c}^2}{4\pi(\rho_\mathrm{c} + \rho_\mathrm{h})}\right]^{-1/2},$$

and, hence, can make the discontinuity effectively stable [$t_\mathrm{cross}/\tau < 1$, cf. condition (8)]. However, on closer examination, it turns out that as the magnetic-field strength decreases below the critical value that provides stability, its effect on the growth of perturbations rapidly becomes negligible. Thus, for B equal to 0.7–0.8 of the critical value, the growth time increases by only a factor of 1.1–1.4 compared to $B = 0$.

The uncertainty in estimate (18) is mainly attributable to the uncertainty in determining the front stability boundary φ_cr and, consequently, the gas velocity at this point. For $\varphi_\mathrm{cr} = 30^\circ \pm 10^\circ$ (which is a conservative estimate; see Fig. 2), we find from Eqs. (16) and (6) that the magnetic-field strength must lie within the range

$$0.09p_\mathrm{gas} < p_\mathrm{mag,h} + p_\mathrm{mag,c} < 0.23p_\mathrm{gas}. \qquad (20)$$

The maximum magnetic-field strength, $B = \max(B_\mathrm{h}, B_\mathrm{c})$, lies within the range $(4\pi p_\mathrm{mag})^{1/2} < B < (8\pi p_\mathrm{mag})^{1/2}$, where $p_\mathrm{mag} = p_\mathrm{mag,h} + p_\mathrm{mag,c}$. Substituting the numerical values for the gas pressure from the table and Fig. 8 yields a numerical value of $6 < B < 14\ \mu$G.

Thus, the main conclusion that follows from our analysis of hydrodynamic instabilities of the front in A3667 is that the front sharpness near the axis and its gradual smearing at large angles from the direction of motion are most likely explained by the existence of a plasma layer in which the magnetic field is parallel to the front and has a strength of $\approx 10\ \mu$G. As was mentioned above, such a layer can be formed when the magnetic field lines are drawn through shear plasma motions. As a result, the magnetic-field strength in the layer is most likely higher than that in the remaining part of the cluster. Thus, we reach the following conclusion: *the magnetic pressure accounts for only a small fraction of the gas pressure:* $p_\mathrm{mag}/p_\mathrm{gas} \sim 0.1$–$0.2$ *in the magnetic layer and is even lower outside the layer.* This conclusion is reliable: if, for example, the magnetic pressure is assumed to be equal to the gas pressure,

then, according to Eq. (17), the front would be stable at velocities up to $M \sim 2(1 + T_c/T_h)^{1/2}\gamma^{-1/2} = 1.9$, i.e., over the entire surface of the cold cloud.

The above estimate of the magnetic-field strength is difficult to compare with the measurements by other methods, for example, by using Faraday rotation, which generally yields $B \sim 1$ μG (Kim *et al.* 1991). On the one hand, our estimate refers to the plasma near the front, where the field is enhanced, but, on the other hand, the Faraday rotation directly gives only the integral of the magnetic-field component parallel to the line of sight weighted with the plasma density. Therefore, the derived B strongly depend on the estimates of the poorly known degree of magnetic-field topology.

Note that complete magnetic-field ordering is not required to suppress the instability; it will suffice only to draw the magnetic field lines so as to achieve field uniformity on scales of several hundred kpc. The perturbations with wavelengths of ~ 10 kpc will then be suppressed, while the perturbations with $\lambda > 100$ kpc will simply have no time to grow.

The drawing of magnetic field lines through shear plasma motions near the tangential discontinuity is not the only mechanism that can give rise to a layer with a strong uniform magnetic field. Thus, for example, Frank *et al.* (1996) and Jones *et al.* (1997) showed that even a weak seed magnetic field, which is incapable of suppressing the tangential-discontinuity instability, is significantly enhanced in forming nonlinear vortices, and after active field-line reconnections, a thin layer is formed. In this layer, the magnetic field lines are aligned with the flow and the field strength increases to such an extent that the interface becomes unstable; as a result, the plasma flow becomes quasi-laminar. At any rate, the formation of a magnetic layer seems virtually inevitable; having formed, the layer begins to act as a kind of magnetic isolation, which stops diffusion and heat transfer, and also as a magnetic lubrication, which ensures a nonviscous, laminar pattern of flow around the cold front.

In conclusion, note that structures similar to the cold fronts in the clusters A3667 and A2142 have now also been detected in several other objects: RXJ 1720.1+2638 (Mazzotta *et al.* 2001), A1795 (Markevitch *et al.* 2001), 1E0657–056 (Markevitch *et al.* 2002), and in A496, A478, A2029, Zw3146. Thus, we are dealing with a common phenomenon.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 00-02-17124) and the Committee of the Russian Academy of Sciences on Work with Youth.

REFERENCES

1. A. C. Fabian, Annu. Rev. Astron. Astrophys. **32**, 277 (1994).
2. A. Frank, T. W. Jones, D. Ryu, and J. B. Gaalaas, Astrophys. J. **460**, 777 (1996).
3. R. Fusco-Femiano, D. dal Fiume, L. Feretti, *et al.*, Astrophys. J. Lett. **513**, L21 (1999).
4. M. Joffre, P. Fischer, J. Frieman, *et al.*, Astrophys. J. Lett. **534**, L131 (2000).
5. T. W. Jones, D. Ryu, and I. L. Tregillis, Astrophys. J. **473**, 365 (1996).
6. T. W. Jones, J. B. Gaalaas, D. Ryu, and A. Frank, Astrophys. J. **482**, 230 (1997).
7. P. Katgert, A. Mazure, R. den Hartog, *et al.*, Astron. Astrophys., Suppl. Ser. **129**, 399 (1998).
8. K. Kim, P. P. Kronberg, and P. C. Tribble, Astrophys. J. **379**, 80 (1991).
9. G. P. Knopp, J. P. Henry, and U. G. Briel, Astrophys. J. **472**, 125 (1996).
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics,* Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1988; Pergamon, New York, 1987).
11. A. Loeb and S. Mao, Astrophys. J. Lett. **435**, L109 (1994).
12. M. Markevitch, C. L. Sarazin, and A. Vikhlinin, Astrophys. J. **521**, 526 (1999).
13. M. Markevitch, T. J. Ponman, P. E. J. Nulsen, *et al.*, Astrophys. J. **541**, 542 (2000).
14. M. Markevitch, A. Vikhlinin, and P. Mazzotta, Astrophys. J. Lett. **562**, L153 (2001).
15. M. Markevitch, A. H. González, L. David, *et al.*, Astrophys. J. Lett. **567**, L27 (2002).
16. P. Mazzotta, M. Markevitch, A. Vikhlinin, *et al.*, Astrophys. J. **555**, 205 (2001).
17. J. W. Miles, J. Fluid Mech. **4**, 538 (1958).
18. J. Miralda-Escudé and A. Babul, Astrophys. J. **449**, 18 (1995).
19. W. E. Moeckel, NACA Technical Note No. 1921, http://naca.larc.nasa.gov/reports/1949/nacatn-1921 (1949).
20. H. J. A. Röttgering, M. H. Wieringa, R. W. Hunstead, and R. D. Ekers, Mon. Not. R. Astron. Soc. **290**, 577 (1997).
21. C. L. Sarazin, in *Merging Processes in Clusters of Galaxies,* Ed. by L. Feretti, I. M. Gioia, and G. Giovannini (Kluwer, Dordrecht, 2001); astro-ph/0105418.
22. L. J. Sodre, H. V. Capelato, J. E. Steiner, *et al.*, Mon. Not. R. Astron. Soc. **259**, 233 (1992).
23. L. Spitzer, *Physics of Fully Ionized Gases* (Interscience, New York, 1962).
24. S. I. Syrovatskiĭ, Zh. Éksp. Teor. Fiz. **24**, 622 (1953).
25. A. Vikhlinin, M. Markevitch, and S. S. Murray, Astrophys. J. **551**, 160 (2001a).
26. A. Vikhlinin, M. Markevitch, and S. S. Murray, Astrophys. J. Lett. **549**, L47 (2001b).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 509–521. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 579–592.

Analysis of TTE/BATSE Time Profiles for Short Gamma-Ray Bursts

I. V. Arkhangelskaja*

Moscow State Engineering Physics Institute, Kashirskoe sh. 31, Moscow, 115409 Russia

Received February 18, 2002

Abstract—The TTE/BATSE time profiles for short gamma-ray bursts (GRBs) are analyzed. A sample of 287 short GRBs and a sample of 143 background regions are studied. Bursts similar to BRBs with precursors and bursts with time profiles that are not encountered among the bursts whose time profiles were investigated by using a combination of DISCSC and PREB data. In addition, there are fewer events with single-peak time profiles among short GRBs than among long GRBs (many double and triple bursts). A fractal analysis of the TTE time profiles for short GRBs is performed. According to the TTE data, the range of fractal dimensions for short bursts is $0.80 \le D \le 2.25$. The derived fractal-dimension distribution exhibits two peaks that correspond to a similar distribution obtained previously by reducing the DISCSC data for short GRBs ($D = 1.44 \pm 0.02$ and $D = 1.90 \pm 0.03$) and a third peak ($D = 1.05 \pm 0.03$). The bursts with $\langle D \rangle = 1.90 \pm 0.03$ correspond to events in whose sources the processes with long-term variations and (or) quasi-periodicity take place, while the event time profiles with a fractal dimension $\langle D \rangle = 1.05 \pm 0.03$ correspond to events in whose sources many extremely short random processes take place. A more detailed analysis of a subgroup of bursts with $D = 1.44 \pm 0.02$ shows that its fractal dimension distribution is broader than that for a group with the same (within the limits of the measurement errors) D. At least two more GRB subgroups can be distinguished in this subgroup: (1) bursts with $\langle D \rangle = 1.51 \pm 0.04$; according to the TTE data, their fractal dimensions correspond to those of the background; i.e., these are events with smooth time profiles without any variability in the sources on the time scales on which the fractal dimension is analyzed; and (2) bursts with $\langle D \rangle = 1.31 \pm 0.05$, whose time profiles can correspond to those of the events obtained from the fireball model with internal shock waves. We present time profiles for the events obtained by using this model. The range of fractal dimensions for the modeled time profiles is $1.213 \le D \le 1.400$, with the fractal dimensions for such an event and for the real GRB 990208 being equal, within the error limits, for some model parameters. A study of the TTE and DISCSC fractal-dimension distributions for the background indicates that the fractal dimension distributions obtained by analyzing these two types of data can be processed simultaneously. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *BATSE, gamma-ray bursts, fractal analysis, time profiles*

INTRODUCTION

Gamma-ray bursts (GRBs) were first discovered by the VELA series of satellites, which was designed to detect gamma-ray radiation from nuclear explosions in the atmosphere and to localize gamma-ray sources. On July 2, 1967, the VELA4a satellite detected a short (with a duration of ~10 s) increase in the count rate in the energy range 0.1 to 1 MeV; the gamma-ray source in this case was clearly extraterrestrial in origin (Klebesadel *et al.* 1973). The existence of such gamma-ray sources was soon confirmed by the processing of data from other instruments (Mazets *et al.* 1974; Cline 1974). Subsequently, such events were registered quite often.

A total of approximately 4000 such events have been observed to date by detectors onboard more than twenty different spacecraft (Fenimore *et al.* 1993; Hurley *et al.* 1999; Tkachenko *et al.* 1998; Costa *et al.* 1999; Briggs *et al.* 1999; Catalog 2001). The most complete catalog of GRBs was compiled after the processing of data from the BATSE (Burst and Transient Source Experiment) instrument, which was specially designed to investigate GRBs.

THE BATSE INSTRUMENT AND ITS DATA

The BATSE instrument is one of the four instruments onboard the CGRO (Compton Gamma-Ray Observatory) (Bunner 1989), which was launched on April 5, 1991, and completed its operation on June 4, 2000.

BATSE consists of eight uncollimated detector modules arranged on the CGRO corners to provide the best unobstructed view of the sky. Each BATSE module consisted of two NaI(Tl) scintillator detectors (Fishman *et al.* 1989): a large-area detector (LAD),

*E-mail: irene.belousova@usa.net

Table 1. The boundaries of four energy channels for the eight LAD detectors (keV)

Detector number	Channel number			
	1	2	3	4
0	21.6–56.1	56.1–105	105–319	319–10^5
1	25.0–58.1	58.1–105	105–307	307–10^5
2	25.8–59.4	59.4–108	108–326	326–10^5
3	23.7–65.2	65.2–114	114–322	322–10^5
4	22.2–60.5	60.5–110	110–318	318–10^5
5	24.4–59.1	59.1–107	107–316	316–10^5
6	23.5–66.0	66.0–110	110–321	321–10^5
7	14.4–60.0	60.0–110	110–321	321–10^5

Note: The band boundaries are given according to the inflight calibration data for 1997 (Catalog 2001).

which was designed to obtain the time profiles of events with a high time resolution and which was a disk 20″ in diameter and 0.5″ in thickness mounted on a 0.75″-thick quartz plate that was viewed by three photomultipliers; and a spectroscopic detector (SD), which was designed to take the spectra of events with a high energy resolution over a wide energy range and which was a disk 5″ in diameter and 3″ in thickness with a 50-mm-thick beryllium window located at the center of the crystal casing to provide high detection efficiency of radiation with an energy ~15 keV. The LAD and SD axes were inclined to each other at an angle of 19°. A 0.25″-thick plastic scintillator detector located in front of the LAD was used as the anticoincidence detector.

The BATSE instrument operated in two modes: background and burst modes. To switch BATSE to the mode of detection and analysis of a burst event, it was necessary that the count rate in the second and third energy channels (see Table 1) exceeded the mean count rate in the previous 17.4 s by 5.5σ at least in two LAD detectors. In addition, the LAD shield, in which the count rate was at a maximum for a given event, should not record excesses of the charged-particle flux above some threshold level. This threshold level was determined through detector activation in the South Atlantic Anomaly by the presence of radioactive calibration sources onboard the CGRO, etc. Bursts were searched for in several time intervals (64, 256, and 1024 ms). Whenever a burst trigger occurs, BATSE provides the other CGRO instruments with a signal that can be used to initiate special burst data processing modes.

Below, we list the types of information on event time profiles from the LAD detectors in burst mode (Horack 1991), because here we discuss the processing of data precisely from these detectors:

The DISCSC data containing information on the count rate in four energy channels with a 64 ms time resolution from each LAD detector switched to burst mode;

The PREB data containing information on the count rate in four energy channels from each of the eight LAD detectors with a 64-ms time resolution accumulated within 2.048 s before burst-cell triggering;

The DISCLB data containing information similar to the PREB data but accumulated within 16.384 s before burst trigger with a 2.048-s time resolution and within 49.154 s after burst-cell triggering with a 1.024-s time resolution;

The TTE data containing the detection time of the first 32 768 photons from each of the triggered LAD detectors with a 2-μs time resolution in four energy bands;

The TTS data containing the accumulation time of a fixed number of counts in four energy bands from triggered LAD detectors with a 1-ms time resolution (usually 64 events);

The MER data containing information on the count rate in 16 energy channels from each LAD detector switched to burst mode with a 16-ms time resolution.

Unfortunately, the MER data are not available for all of the registered GRBs and the boundaries of the energy ranges for this type of data are such that four ranges similar to the energy ranges in which the DISCSC, TTE, and TTS data are accumulated cannot be separated out. The boundaries of the ranges in which the DISCSC, TTE, and TTS data are accumulated are identical (see Table 1), which makes it possible to simultaneously analyze these types of data.

BASIC GRB CHARACTERISTICS

All GRBs seem to be very different at first glance. Some GRBs have one peak in their time profiles, while other GRBs have several peaks. However, several common features can be distinguished for all GRBs (see, e.g., Luchkov *et al.* 1996; Piran 1999; Postnov 1999):

(1) the angular distribution is isotropic for the entire sample obtained after analyzing the BATSE data;

(2) the burst duration lies within the range 10^{-2}–10^3 s;

(3) the detection rate is ~1 burst per day;

(4) the observed GRB fluxes near the Earth are 10^{-8}–10^{-3} erg s^{-2};

(5) the spectra of most GRBs are described by Band's model (Briggs *et al.* 1999; Band 1993):

$$N_E(E), \frac{\text{counts}}{\text{cm}^2\text{s keV}} = \begin{cases} A \times E^{\alpha} \times e^{-E/E_0}, & E < E_0 \\ B \times E^{\beta}, & E > E_0 \end{cases} \quad (1)$$

where $-1.5 < \alpha < 0.5$; $-3.5 < \beta < -2$; $50 < E_0 < 800$ keV.

In addition, several different groups of GRBs can be separated by using burst duration t_{90}, burst hardness H_{32}, which is defined as the ratio of the fluences in the third and second BATSE energy channels, and the type of time profile as the criteria.

The first statistical analysis of the GRB duration distribution revealed two groups of bursts: short GRBs with $t_{90} < 2$ s and long GRBs with $t_{90} > 2$ s. Analysis of the third BATSE catalog also revealed two groups of GRBs (Meegan *et al.* 1996). A more detailed analysis of the fourth BATSE catalog and the current BATSE catalog, which contains almost five times more bursts than the third catalog, allows three groups of GRBs to be separated by taking into account the GRB hardness distribution (Meszaros *et al.* 2000; Belousova *et al.* 1999):

short $t_{90} < 2$ s, $\langle t_{90} \rangle \sim 0.7$ s, $H_{32} > 1.85$,

long $t_{90} > 5$ s, $\langle t_{90} \rangle \sim 20$ s, $H_{32} < 6.00$,

intermediate 0.8 c $< t_{90} < 25$ s, $\langle t_{90} \rangle \sim 3$ s.

Several types of GRBs can be qualitatively identified by the shape of the time profiles (Kouveliotou 1995). Two groups of bursts are generally separated: simple GRBs and those composed of several episodes. Next, smooth, symmetric, erratic, quasi-periodic, complex bursts and bursts with a fast rise in count rate and an exponential decay are separated in these groups. Besides, some of the long bursts have precursors; i.e., a short (with a duration up to 20 s) precursor burst occurs first, the count rate subsequently decreases to the background level, and another peak with a duration of several tens of seconds is observed after 50–100 s (Koshut *et al.* 1995).

THE FRACTAL DIMENSION AS A PARAMETER OF THE GRB TIME PROFILES

GRBs with time profiles of almost all the types listed above are present in the groups of short, long, and intermediate bursts. This suggests that there are several subgroups of GRBs that can be separated by simultaneously analyzing the GRB distribution in duration and the types of GRB time profiles. Unfortunately, the type of GRB time profile is only a qualitative and subjective characteristic, which severely complicates detailed analysis.

A decomposition of the function $f(x)$ into a complete orthogonal basis is commonly used to determine quantitative criteria for describing this function. For real functions (and the GRB time profile is a real function of time), there are an infinite number of families of functions that form such bases, for example, a basic of Dirac δ-functions specified at each point x with amplitude $f(x)$ or a Fourier basis. However, for nonperiodic functions (which the GRB time profiles mostly are), a Fourier analysis turns out to be useless in most cases, because it gives an infinite spectrum of harmonics. A wavelet analysis (Abarbane 1991) is commonly used as an alternative method for representing nonperiodic functions. In this case, a complete orthogonal basis of functions $\phi_{jk}(x)$ can be obtained from the principal wavelet function $\phi(x)$ by extending both the principal function and the set on which it is defined by a factor of $2^{j/2}$, where j is a nonnegative integer. Unfortunately, for a wavelet analysis to be used, the time scales of the GRB time profiles must be identical, which is difficult to achieve even within one group of GRBs. Nevertheless, an attempt to perform a wavelet analysis of the TTE/BATSE GRB time profiles was made by Walker and Schaefer (2000). They analyzed 20 GRBs (for which a similar scale of the time profiles can be chosen) and found rapid variations in the count rate on a time scale in the range $0.256 < \tau_{\min} < 33$ ms, where $\tau_{\min}$ is the minimum variability time, but found no statistically significant correlation between t_{90} and $\tau_{\min}$.

Here, we use a different method for studying the time profiles of events: fractal analysis, which is sensitive to the shape and irregularity of the event time profile. In addition, the fractal dimensions must be different for the time profiles of events caused by different physical processes (Feder 1988). A fractal analysis has several peculiarities that allow it to be used to study sets with characteristics varying over a wide range, for example, GRBs. One peculiarity is scaling; i.e., two events with similar time profiles but with different durations have a similar fractal dimension, which allows one to use the TTE data for bursts with a duration <3 s and the DISCSC data for the remaining bursts. Another peculiarity of fractal analysis is the possibility of using data from different detectors as a single sample if the background fractal dimension for these detectors is the same (*Fractals in Physics* 1988).

The fractal dimension of a set ξ is defined as such a number $D \geq 0$ that when this set is covered with a finite or infinite set of neighborhoods of radius $r_i < \varepsilon$ for any $a > 0$

$$\inf_{\xi} \sum r_i^a \xrightarrow[\epsilon \to 0]{} 0 \text{ for each } a > D.$$

$$\inf_{\xi} \sum r_i^a \xrightarrow[\epsilon \to 0]{} \infty \text{ for each } a < D.$$

An attempt to study the fractal dimension of GRB time profiles was first made by Shakura *et al.* (1994); they analyzed four events registered by different detectors and determined the range of fractal dimensions for them, $1.9 < (D+1) < 2.6$. In addition, these authors analyzed in detail the background fractal dimension for the gamma-ray detectors onboard the Venera 11 and Prognoz 7 spacecraft. This analysis showed that the background fractal dimension for these detectors was 1.5 and matched the fractal dimension of a set of 512 random numbers in the range 0.000 to 1.000.

The DISCSC data are commonly used to analyze GRB time profiles,[1] because the time profile in the data of this type is contained in explicit form as a time dependence of the count rate in a certain energy channel (see Table 1).

A fractal analysis of the time profiles based on the DISCSC data from the current BATSE catalog was performed by Arkhangel'skaya *et al.* (2001) and Kotov *et al.* (2001). These authors found that two peaks at $D = 1.45 \pm 0.02$ and $D = 1.80 \pm 0.02$ and three peaks at $D = 1.45 \pm 0.02$, $D = 0.90 \pm 0.02$, and $D = 1.17 \pm 0.02$ could be distinguished in the fractal dimension distributions for short and intermediate GRBs, respectively. They also found that the fractal-dimension distribution for long bursts was significantly wider than that for a homogeneous sample (by a homogeneous sample, we mean a group with the same D within the error limits, e.g., the detector background). The latter leads us to conclude that there may be several subgroups in this group of bursts. The first two subgroups of GRBs in the groups of short and intermediate bursts correspond to events with smooth time profiles; their fractal dimension is close to the background fractal dimension $D = 1.5 \pm 0.02$.

ANALYSIS OF THE TTE TIME PROFILES FOR GRBs

As was noted above, the TTE, DISCSC, and TTS data from BATSE can be analyzed simultaneously, because the boundaries of the energy bands for these types of data are the same. Since the GRB duration varies over the range 10^{-2}–10^3 s, it will suffice to use data of two types, DISCSC and TTE or TTS, to comprehensively analyze the GRB time profiles. Here, we analyze the time profiles of short GRBs for a larger sample than that in Arkhangel'skaya *et al.* (2001) and Kotov *et al.* (2001) by using the results of DISCSC-data processing, because the entire event time profile consists of about 10 points in the data of this type for most short bursts. In contrast to the DISCSC data, the TTE and TTS data do not contain an event time profile in the form of a time dependence of the count rate in some energy band; when using these types of data, the time profile must be reconstructed. When using the TTS data to reconstruct the time profile, the width of the time bin (accumulation time) depends on the registered GRB flux; i.e., it may be different for different bursts. The TTE data contain the number of the detector that registered a gamma-ray photon, the number of the energy channel in which the gamma-ray photon was registered, and its detection time. Thus, to reconstruct the time profile of an event from the TTE data, we may choose the same accumulation time Δt and the count rate in a given energy band is defined as the number of events registered by the detector in the accumulation time. Here, the accumulation time was taken to be $\Delta t = 8$ ms.

The current BATSE catalog contains 451 GRBs for which the TTE data are available. About 300 of these bursts are short, i.e., bursts with $t_{90} < 2$ s, $\langle t_{90} \rangle \sim 0.7$ s, $H_{32} > 1.85$ (which is more than a factor of 3 larger than the size of the sample used to analyze the fractal-dimension distribution based on the DISCSC data). We studied a sample of 287 short GRBs and a sample of 143 background regions; the starting time of a background region was assumed to be the time $3t_{90}$ from the end of the GRB for which the TTE file was taken. Since the TTE data generally contain information on the photon detection time within ~3 s after BATSE burst cell triggering, we failed to distinguish background regions for all bursts.

Our analysis of the time profiles for short bursts revealed that there are fewer events with single-peak time profiles among the short GRBs than among the long GRBs; there are many double and triple bursts. When analyzing the time profile, we separated the first peak with a dip to the background level with a deviation by no more than 1σ; the remaining portion of the time profile was then tested for correspondence of the amplitude to random values by using two criteria (Gaĭdaliev 2001; Dyuk 1997):

(1) if the amplitude of the time profile corresponds to random values, then the number of time intervals with count rates above and below the median must be equal to 1/2 of the number of time intervals analyzed;

[1] The event time profile proper is constructed as the sum of DISCSC, PREB, and DISCLB data, but only the first two types of data are used for a temporal analysis, because their time resolution is 64 ms. In the BATSE catalogs, these data are stored in the DISCSC_BFITS_N.FITS file, where N is the trigger number. Therefore, below, we speak about the DISCSC data, implying precisely the contents of these files.

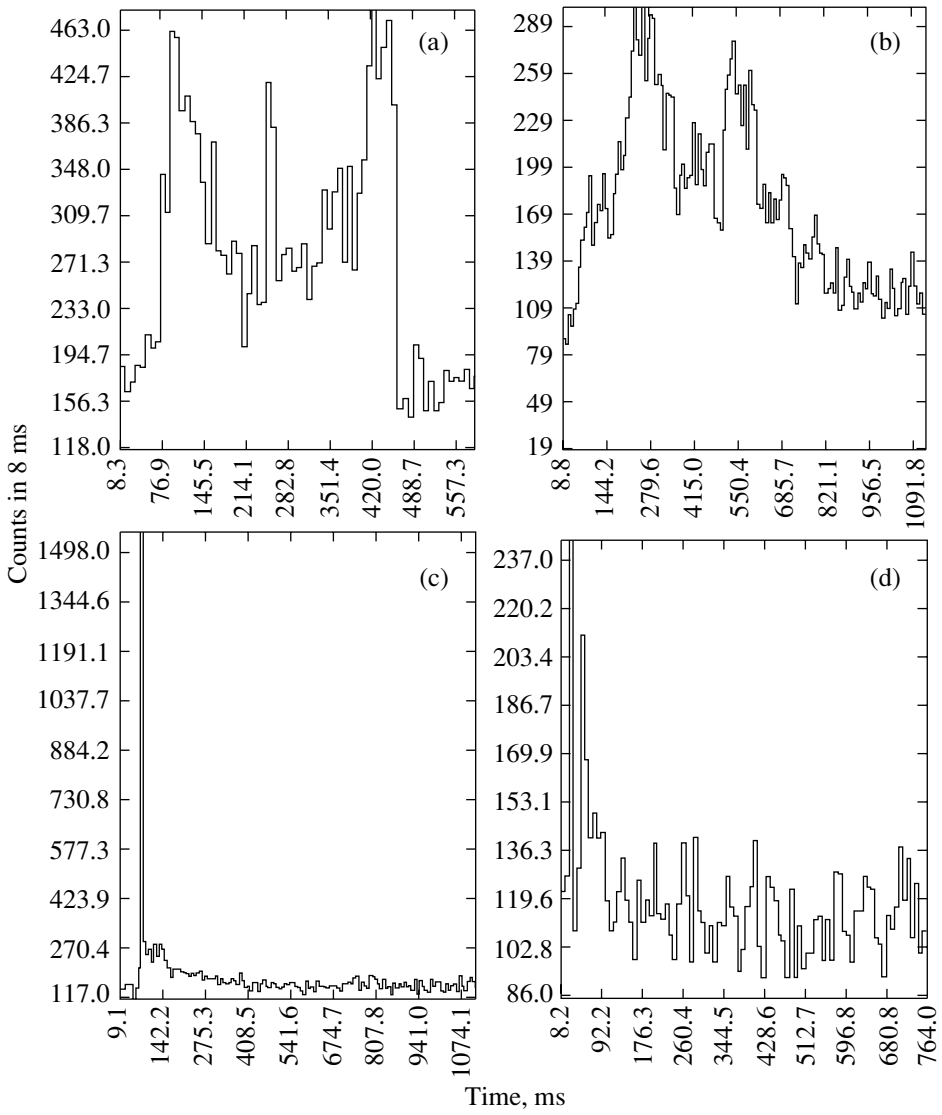


Fig. 1. The GRB time profiles corresponding to the types that were not encountered previously when processing the DISCSC data: (a) and (b) the M-shaped burst time profiles (BATSE trigger numbers 7456 and 7663, respectively); (c) the burst time profile with a fast rise, a fast decay in the count rate to a level that is a factor of 2 or 3 higher than the background, and a long flat dip in the count rate (BATSE trigger number 7427); and (d) the burst time profile with quasi-periodic (a period of $\sim$80 ms) variations in count rate. For this burst, the number of time intervals with count rates above and below the median is 33, the number of rises and declines is 45; the expected number of time intervals with count rates above and below the median must be 41, while the number of rises and declines must be 53 (we analyzed 83 time intervals) if the amplitude of the time profile corresponds to random values. The probability that the amplitude for a given time profile will not correspond to random values is 92% according to the first criterion and 97% according to the second criterion. A Fourier analysis gives a peak in the periodogram near 82 s.

(2) if the amplitude of the time profile corresponds to random values, then the number of rises and declines in this time profile must be equal to 2/3 of the number of time intervals analyzed.

In our case, the width of the time interval being analyzed for the TTE data was taken to be equal to the accumulation time $\Delta t = 8$ ms; for the DISCSC data, we took 64 ms. If the amplitude of the part of the time profile that remained after the peak separation for a given event corresponded to random values with a probability of more than 95%, then the event was considered to be a single-peak one. Sixty of the 287 short bursts analyzed are single-peak ones. For comparison, we analyzed two samples of bursts with $t_{90} >$ 25 s taken one after another from the current BATSE catalog ordered by trigger numbers, each containing 287 events. We found that the first and second samples include 104 and 120 single-peak events, respectively. In addition, we detected bursts with time

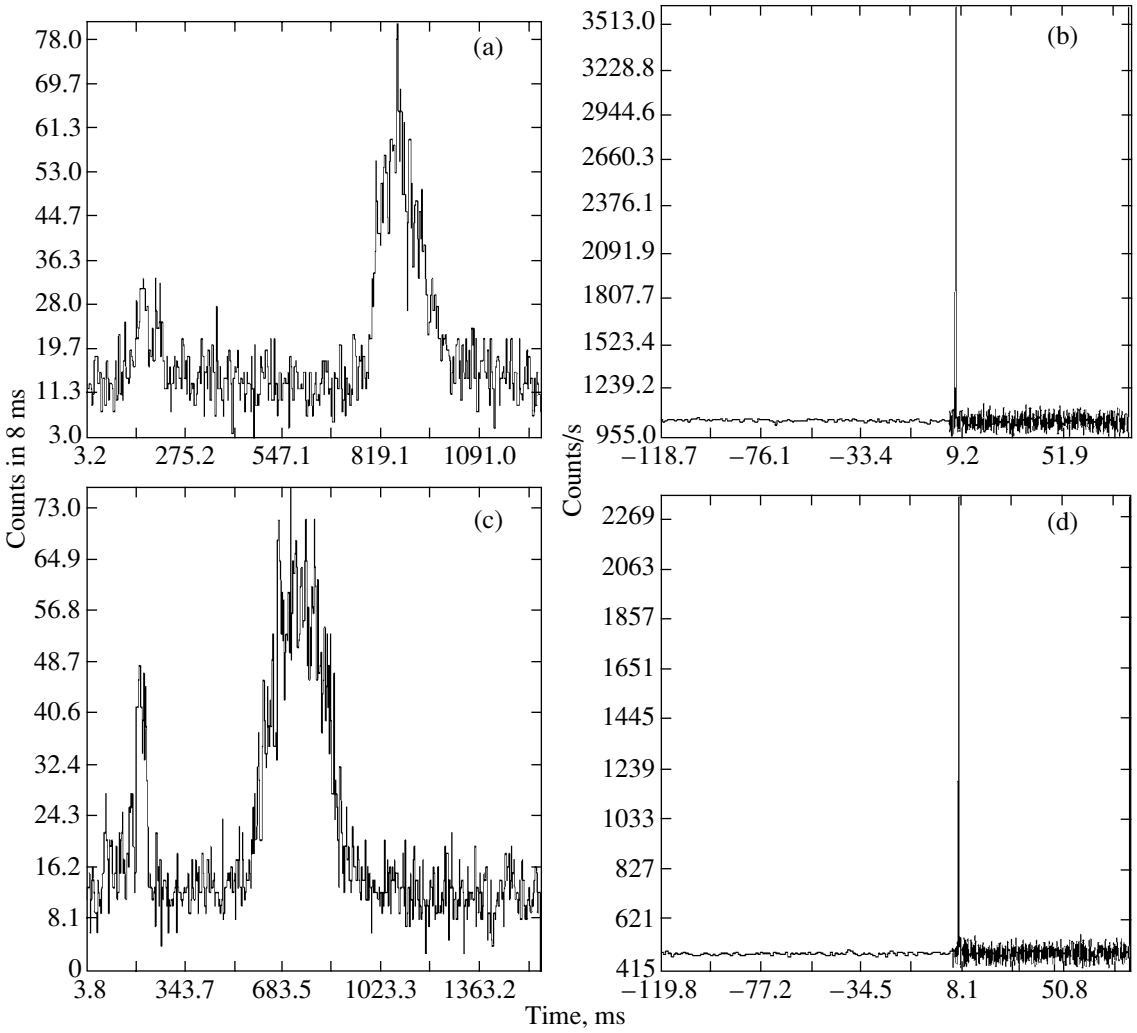


Fig. 2. The time profiles of GRBs similar to bursts with precursors but on a time scale of several tens of milliseconds: (a) for trigger number 2330 using the TTE data; (b) for trigger number 2330 using the DISCSC data; (c) for trigger number 2834 using the TTE data; and (d) for trigger number 2834 using the DISCSC data.

profiles that are not encountered among the bursts whose time profiles were investigated by using the DISCSC data (see Fig. 1):

bursts with a fast rise, a fast decay in the count rate to a level that is a factor of 2 or 3 higher than the background, and a long flat dip in the count rate;

bursts with M-shaped time profiles;

several bursts with quasi-periodic (a period of $\sim$80 ms) variations in the count rate;

and bursts similar to those with precursors (see Fig. 2).

The bursts with quasi-periodic variations in the count rate were chosen from the results of Fourier analysis for the time profiles of events among the bursts for which the amplitude of part of the time profile that remained after the separation of the first peak did not correspond to random values with a probability $> 95\%$.

The methods for determining the fractal dimension of an event time profile are generally based on a decomposition of the time profile into bins and on an analysis of the count-rate fluctuations in each bin. If the number of points in a bin is not enough for analysis, as with the TTE data, then a cell algorithm of determining the fractal dimension is used (Feder 1988): part of the plane in which the curve being analyzed lies is covered with cells with a side δ. If $N(\delta)$ is the number of cells that have at least one common point with a given curve, then a gauge can be assigned to this curve:

$$L = N(\delta)\delta^D \qquad (2)$$

For an ordinary (nonfractal) curve, $L \to 0$ for $\delta \to 0$. For a fractal curve, however, the corresponding gauge (2) is nonzero for some $D \neq 1$.

Table 2. Parameters for the peaks of the fit to the fractal dimension distribution for the background and short burst time profiles obtained using TTE data

Type of group	Peak position	FWHM (2σ)	Amplitude
With the same, within the limits of the measurement errors, values of D (background)	1.49 ± 0.02	0.082 ± 0.006	0.38 ± 0.03
Short bursts (a three-peak fit)	1.05 ± 0.03	0.08 ± 0.02	0.12 ± 0.01
	1.44 ± 0.02	0.45 ± 0.06	0.22 ± 0.02
	1.90 ± 0.03	0.09 ± 0.02	0.09 ± 0.01
Short bursts (a four-peak fit)	1.05 ± 0.03	0.08 ± 0.02	0.12 ± 0.01
	1.31 ± 0.05	0.17 ± 0.02	0.21 ± 0.01
	1.51 ± 0.04	0.19 ± 0.02	0.18 ± 0.02
	1.90 ± 0.03	0.09 ± 0.02	0.09 ± 0.01

However, among other things, the GRB time profile is also characterized by the fact that the cells have different dimensions in width (time) and in length (count rate). In this case, the corresponding gauge (2) is constructed as follows (see Fig. 3): a set of n δ_j is chosen, the curve being analyzed is broken down into κ δ_j-wide stripes, and

$$N(\delta_j) = \sum_{i=1}^{k} [(\sigma_{i_{\max}} - \sigma_{i_{\min}})/\delta_j + 2], \qquad (3)$$

where $\sigma_{i_{\max}}$ and $\sigma_{i_{\min}}$ are the maximum and minimum count rates in stripe i.

In this case, we use a plot of $N(\delta)$ for a sequence of different δ. If this dependence appears as

$$N(\delta) = \delta^{-D} \qquad (4)$$

for $a > 0$, then the fractal dimension of the curve being analyzed is D. It can be determined from the plot of $\log N(\log \delta)$ by the least-squares method and the errors in the fractal dimension are attributable to this method.

It should be noted that the derived D is not the real fractal dimension of the time profile, because the fractal is generally the limit of an infinite process, while the process of analyzing the time profile for an real event is limited below by the detector time resolution; i.e., we have a predfractal, and the range of fractal dimensions can be shifted from the theoretical values for $1 < D < 2$, for example, $0.80 < D < 2.42$ (Kotov *et al.* 2001).

Thus, before using the cell algorithm for determining the fractal dimension, it is first necessary to prove that a straight line is the best fit to the function

$$\log N = -D \log \delta. \qquad (5)$$

To determine the fractal dimensions of GRB time profiles from the TTE data, we used five values of δ, starting from 20 ms at steps of 12 ms, and analyzed the data in the energy range defined as the sum of the second and third BATSE energy channels, because the energy release for GRBs is at a maximum precisely in this range.

In our case, dependence (5) is such that the χ^2 test does not allow us to determine what the best fit to this dependence is: a straight line or a polynomial of a different degree (see Figs. 4b and 4d). It is well known that a sequence of n points can be fitted by a polynomial of degree $r = n - 1$. A polynomial of degree r is the best fit to dependence (5) if the following conditions are satisfied:

$$f_r = \frac{\omega_r}{R_r/(n-1-r)} > F_{1,n-r-1} \qquad (6)$$

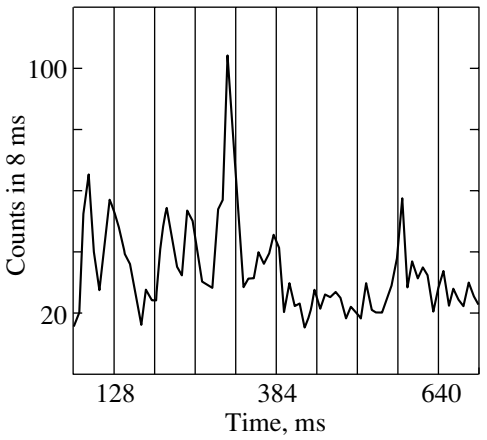


Fig. 3. A decomposition of the event time profile into $\delta = 64$-ms-wide stripes for the following construction of gauge (3).

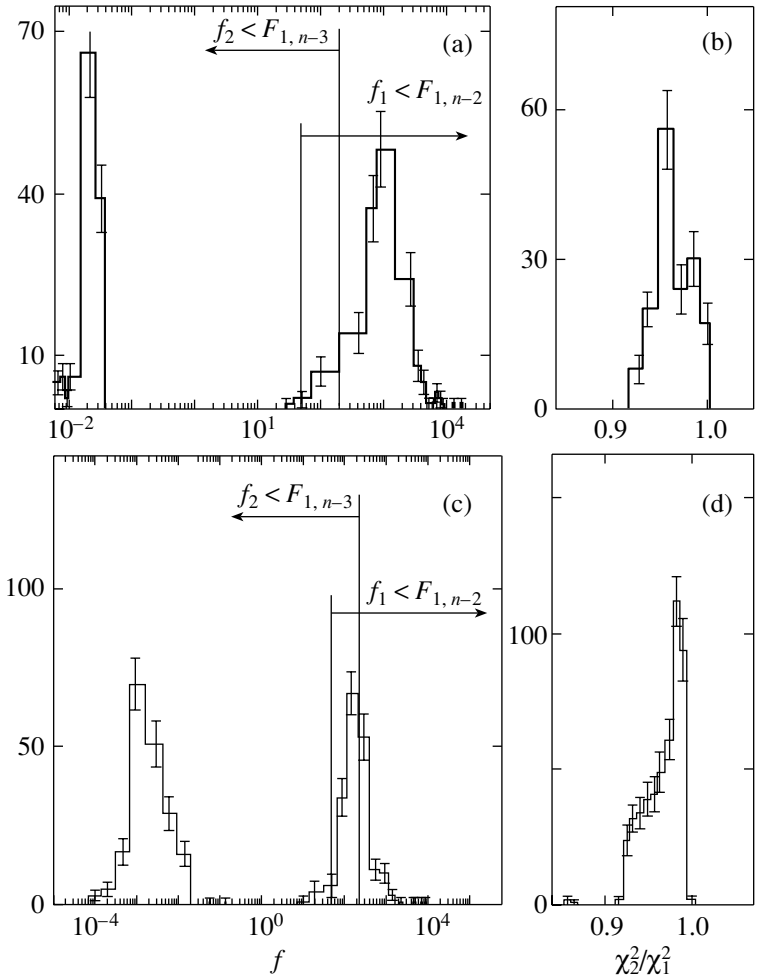


Fig. 4. The distributions of f_1, f_2, and the χ^2 ratio for a parabolic and linear fit to dependence (5): (a) f_1 and f_2 for a sample of 143 background regions; (b) the distribution of the ratio of χ^2_2 for a parabolic fit to χ^2_1 for a linear fit (χ^2_2/χ^2_1) for a sample of 143 background regions; (c) f_1 and f_2 for a sample of 287 short bursts; (d) the distribution of the ratio of χ^2_2 for a parabolic fit to χ^2_1 for a linear fit (χ^2_2/χ^2_1) for a sample of 287 short bursts.

and

$$f_{r+1} = \frac{\omega_{r+1}}{R_{r-1}/(n-1-(r+1))} < F_{1,n-(r+1)-1}, \tag{7}$$

where F is the Fisher function and ω_j is a set of weights for the orthonormalized polynomials Q_{ij} for which $N(\delta) = \sum_{j=0}^{r} \omega_j Q_j$. Thus, the cell algorithm can be used to determine the fractal dimension of GRBs and the background from the TTE data if conditions (6) and (7) are satisfied at $r = 1$ for dependence (5) for each GRB and for each background region, i.e.,

$$f_1 > F_{1,n-2} \tag{8}$$

and

$$f_2 < F_{1,n-3}, \tag{9}$$

where n is the number of different δ and $F_{1,n-2}$ and $F_{1,n-3}$ are the corresponding values of the Fisher function.

The f_1 and f_2 distributions for short GRBs and the background are shown in Figs. 4a and 4c. We see from the figure that the cell algorithm is applicable for determining the fractal dimensions of the GRB and background time profiles by using the TTE data from the current BATSE catalog.

As was pointed out above, when determining the fractal dimension, data of different types can be used as a single sample, provided that the background fractal dimensions for these data are equal. In our case, the time resolutions of the types of data used are different, but the fractal dimension distributions obtained by analyzing the DISCSC and TTE data can be reduced simultaneously if the background fractal dimensions for these data are equal. The distribu-

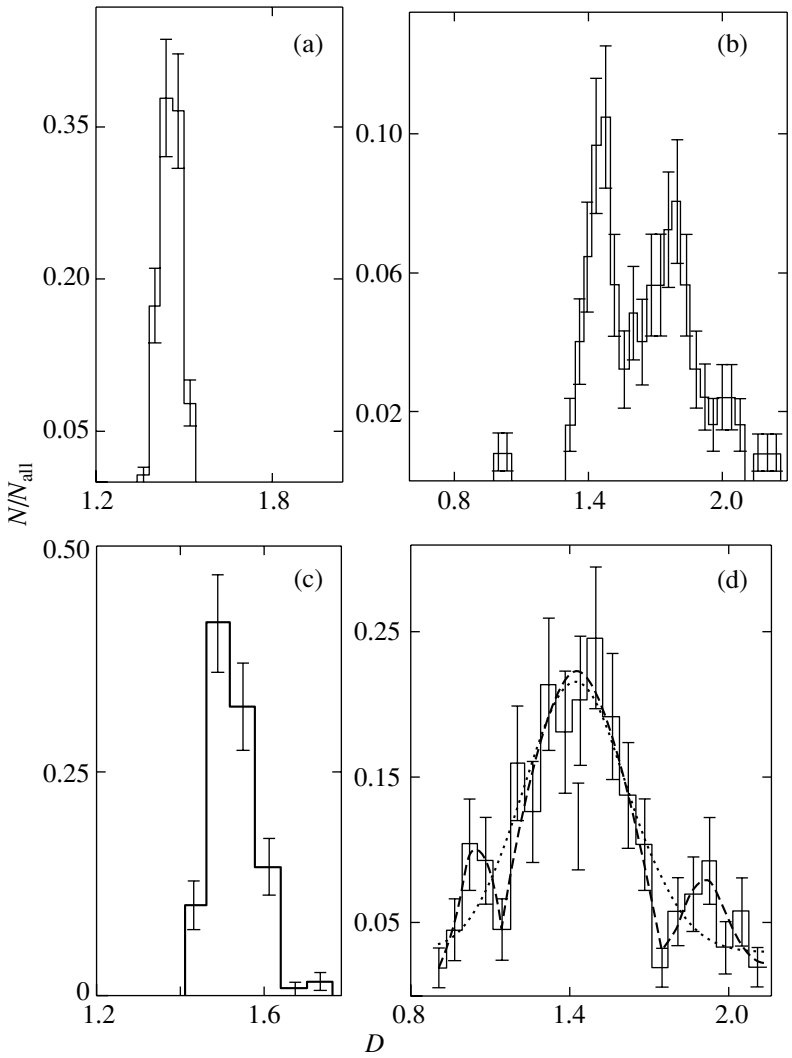


Fig. 5. The fractal dimension distributions for background regions and short GRBs: (a) for a sample of 120 background regions constructed from the DISCSC data; (b) for a sample of short GRBs time profiles constructed from the DISCSC data; (c) for a sample of 143 background regions constructed from the TTE data; and (d) for a sample of 287 short GRBs time profiles constructed from the TTE data (the dotted line represents a single-peak fit to this distribution; the dashed line represents the envelope of a three-peak fit to this distribution). All of the distributions shown here were normalized to the sample size.

tion of the background fractal dimension is shown in Fig. 5c for the TTE data (a sample of 143 background regions was analyzed) (for the distribution parameters, see Table 2) and in Fig. 5a for the DISCSC data (a sample of 120 background regions was analyzed). Analysis of these distributions indicates that the fractal dimension distributions obtained by analyzing these types of data can really be processed simultaneously, because the peaks in their distributions of the background fractal dimension coincide. In addition, the width of the fractal dimension distribution for the background of the group with the same, within the limits of the measurement errors, values of D can be determined from the analysis of this distribution; in our case, the FWHM is 0.082 ± 0.006.

Figure 5d shows the fractal dimension distribution for short GRBs constructed from the TTE data (a sample of 287 bursts was analyzed). According to the TTE data, the range of fractal dimensions for short bursts is $0.80 \leq D \leq 2.25$. We fitted this distribution by single-peak (Gaussian) and three-peak (the sum of three Gaussians) functions; $\chi^2 = 31$ for the single-peak function for 18 degrees of freedom and $\chi^2 = 19$ for the three-peak function for 12 degrees of freedom; i.e., this distribution has three peaks at $D = 1.44 \pm 0.02$, $D = 1.05 \pm 0.03$, and $D = 1.90 \pm 0.03$ (for the remaining parameters of the function, see Table 2). The fractal dimension distribution for short GRBs time profiles constructed from the DISCSC data (Kotov *et al.* 2001) is shown in Fig. 5b. Analysis of these distributions indicates that two peaks remained

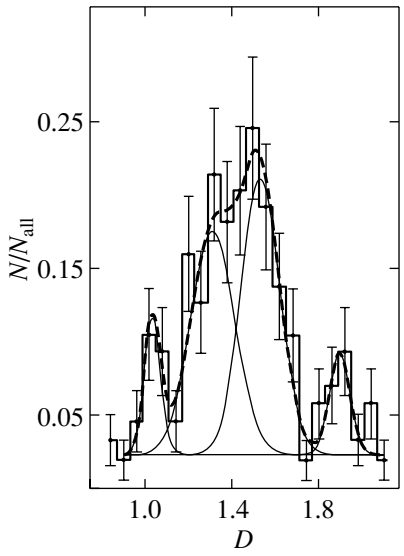


Fig. 6. The fractal dimension distribution for a sample of 287 short GRBs time profiles constructed from the TTE data (the thin line represents a four-peak fit to this distribution; the dashed line represents the envelope for this fit).

unshifted when using data of a different type; i.e., two groups of GRBs can really be separated. In addition, as the sample size increased, a third group of bursts with a small fractal dimension appeared.

However, the peak with a maximum at $D = 1.44 \pm 0.02$ is broader than the distribution for groups with the same, within the limits of the measurement errors, values of D. Therefore, we attempted to fit this peak by the sum of two Gaussians. A four-peak fit to the fractal dimension distribution for short bursts constructed from the TTE data is shown in Fig. 6 (for the parameters of the fit, see Table 2); $\chi^2 = 9$ for the four-peak function for 8 degrees of freedom. Thus, the peak with a maximum at $D = 1.44 \pm 0.02$ splits up and two more subgroups of GRBs can be separated: bursts with $\langle D \rangle = 1.51 \pm 0.04$ whose fractal dimension corresponds to the background fractal dimension, as inferred from the TTE data (i.e., these are events with smooth time profiles and without any variability in the sources on the time scales on which the fractal dimension is analyzed); and bursts with $\langle D \rangle = 1.31 \pm 0.05$.

The bursts with fractal dimensions smaller than the background fractal dimension correspond to events in the sources of which many short random processes take place (Crownover 1995). One of the models for the sources of such events is the fireball model with internal shock waves (Kobayashi *et al.* 1997; Sari *et al.* 1996; Piran 1999). The latter arise from variability of the internal source, which periodically ejects shells of matter with a mass m_i, thickness l_i, and Lorentz factor γ_i. If the succeeding $(i+1)$th shell has a higher velocity than the previously ejected ith shell, then they will collide and merge together; the Lorentz factor and internal energy of the shell produced after the merging will be (Sari *et al.* 1995)

$$\gamma \approx \sqrt{(m_{i+1}\gamma_{i+1} + m_i\gamma_i)/\left(\frac{m_{i+1}}{\gamma_{i+1}} + \frac{m_i}{\gamma_i}\right)} \quad (10)$$

and

$$E_{\rm int} = m_{i+1}c^2(\gamma_{i+1} - \gamma) + m_i c^2(\gamma - \gamma_i). \quad (11)$$

In addition, two shock waves are produced in this case, direct and reverse, and the observed burst emission results mainly from the interaction of the reverse shock with the matter of the shell that formed after the merging (Sari *et al.* 1995; Kobayashi *et al.* 1997). The duration of this emission for an observer toward which the shell moves is proportional to time, as long as the reverse shock crosses the fast shell (Kobayashi *et al.* 1997):

$$\delta t = \frac{l_{i+1}}{2\gamma^2 c(\beta_{i+1} - \beta_{\rm rs})}, \quad (12)$$

where the Lorentz factor of the reverse shock is $\gamma_{\rm rs} \approx \gamma\sqrt{(1+2\gamma/\gamma_{i+1})/(2+\gamma/\gamma_{i+1})}$, $\beta_{\rm rs} = \sqrt{1-1/\gamma_{\rm rs}^2}$ and the observed intensity depends on time as follows:

$$\begin{cases} 0, & t < 0 \\ K(1 - 1/(1+2\gamma^2 ct/R)^2), & 0 < t < \delta t \\ K[1/(1+2\gamma^2 c(t-\delta t)/R)^2 - 1/(1+2\gamma^2 ct/R)^2], & t > \delta t, \end{cases} \quad (13)$$

where $K = (1 - (1 + l_{i+1}/L_{i,i+1})^{-2}) \times E_{\rm int}/\delta t$, $R = \gamma^2 L_{i,i+1}$ is the distance at which the collision took place and $L_{i,i+1}$ is the distance between shells i and $i+1$.

Here, to generate the time profiles, we used a model with M shells colliding in pairs supplemented with Poissonian noise with the mean P = 30 (this mean background value in the LAD detectors was obtained by analyzing 143 background portions of the time profiles reconstructed from the TTE/BATSE data with the accumulation time $\Delta t = 8$ ms). The

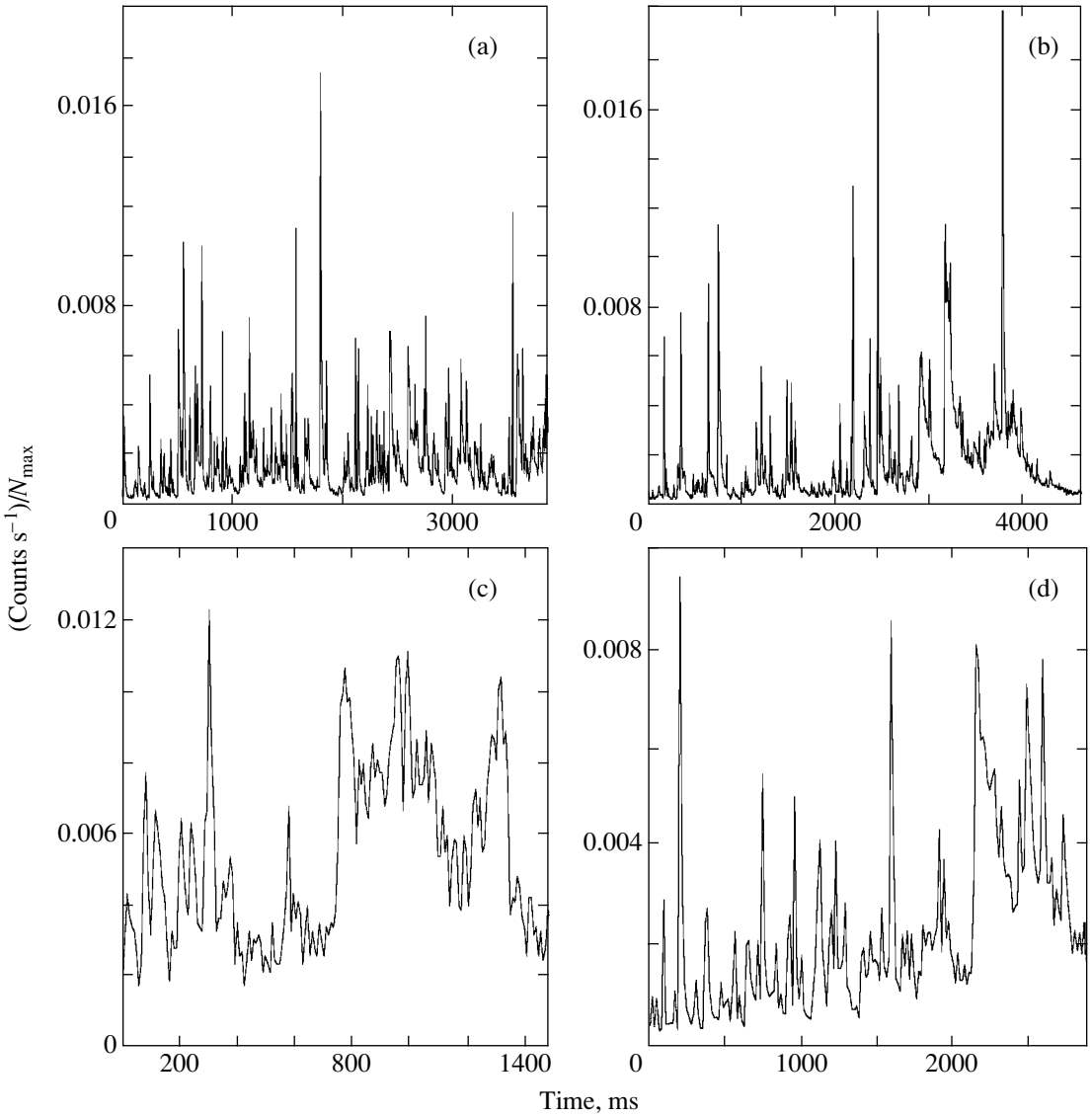


Fig. 7. Modeled burst time profiles and the time profile of an real burst: (a) the time profile obtained in a model with $M = 400$, $dt = 10^{-5}$, $\gamma_{min} = 10$, $\gamma_{max} = 50$, $L/l = 5.0$; (b) the time profile obtained in a model with $M = 200$, $dt = 10^{-5}$, $\gamma_{min} = 30$, $\gamma_{max} = 80$, $L/l = 1.0$; (c) the time profile of the real GRB990208 (BATSE trigger number 7378) constructed from the TTE data, $t_{90} = 1.25 \pm 0.07$ s, $D = 1.38 \pm 0.03$; (d) the time profile obtained in a model with $M = 300$, $dt = 10^{-5}$, $\gamma_{min} = 100$, $\gamma_{max} = 1000$, $L/l = 0.1$; the values obtained are $t_{90} = 2.75 \pm 0.05$ s, $D = 1.393 \pm 0.005$. All of the time profiles shown here were normalized to unity.

model parameters were the following: the period dt during which the source operated; the Lorentz factor γ_i of the shell being formed, which was generated as a random variable uniformly distributed over the range γ_{min} to γ_{max}; the L/l ratio, which was taken to be the same for all M shells; and the mass m_i of the shell being formed, which was generated as a random variable uniformly distributed over the range $0.01m$ to m, where m is a certain mass. The numerical value of m, as well as the numerical value of the distance to the burst source (if the redshift is disregarded) will be of minor importance after the normalization of the derived time profile to unity.

Typical modeled time profiles are shown in Figs. 7a and 7b. The fractal dimension of the derived time profiles varies between $D = 1.213 \pm 0.005$ and $D = 1.400 \pm 0.006$ as M varies between 200 and 400. Figure 7c shows the time profile for the real GRB990208 with the fractal dimension $D = 1.38 \pm 0.03$, and Fig. 7d shows the modeled time profile with $D = 1.393 \pm 0.005$ for $M = 300$. We see that the time profiles of these bursts are similar.

CONCLUSIONS

In this paper, we have analyzed GRB time profiles using TTE data from the BATSE instrument. Our data processing revealed that, first, there are fewer events with single-peak time profiles among the short GRBs than among the long GRBs (many double and triple bursts) and, second, there are bursts with time profiles that were not encountered previously in the DISCSC-data processing:

bursts with a fast rise, a fast decay in the count rate to a level that exceeds the background by a factor of 2 or 3, and a long flat dip in the count rate;

bursts with M-shaped time profiles;

several bursts with quasi-periodic (a period of $\sim$80 ms) variations in the count rate;

and bursts similar to GRBs with precursors but on a time scale of several tens of milliseconds.

We performed a fractal analysis of the time profiles for short GRBs using TTE data. The range of fractal dimensions for short burst, time profiles obtained using TTE data, is $0.80 \le D \le 2.25$. Our study of the fractal dimension distributions for the background based on the TTE and DISCSC data shows that the distributions obtained by analyzing these types of data can really be reduced simultaneously, because the peaks in their background fractal dimension distributions coincide. The fractal-dimension distribution for the time profiles of short GRBs time profiles constructed from the TTE data exhibit two peaks that correspond to the fractal dimension distribution for the time profiles of short GRBs time profiles constructed from the DISKCS data ($D = 1.44 \pm 0.02$ and $D = 1.90 \pm 0.03$); yet another peak is observed at $D = 1.05 \pm 0.03$. A more detailed analysis of the subgroup of bursts with $\langle D \rangle = 1.44 \pm 0.02$ indicates that the fractal dimension distribution for this subgroup is broader than that for the group with the same values of D, within the error limits. At least two more subgroups of GRBs can be separated in this subgroup: bursts with $\langle D \rangle = 1.51 \pm 0.04$ and with $\langle D \rangle = 1.31 \pm 0.05$.

It is commonly assumed that the time profiles of events with fractal dimensions above 1.5 correspond to events in whose sources the processes with long-term variations and/or quasi-periodicity take place, while the time profiles of events with fractal dimensions below 1.5 correspond to events in whose sources many short random processes take place (Crownover 1995). Hence, the peak at $D = 1.05 \pm 0.03$ for short bursts could really appear only when analyzing the time profiles with a sufficiently high time resolution.

Thus, several subgroups of bursts with different fractal dimensions can be separated in the group of short GRBs. The bursts with $\langle D \rangle = 1.90 \pm 0.03$ correspond to events in whose sources the processes with long-term variations or quasi-periodicity take place; the bursts with $\langle D \rangle = 1.51 \pm 0.04$ correspond to events with smooth time profiles, without any variability in the sources of the time scales on which the fractal dimension is analyzed (the fractal dimension of such bursts is completely determined by detector noise); the time profiles of events with fractal dimensions $\langle D \rangle = 1.05 \pm 0.03$ and $\langle D \rangle = 1.31 \pm 0.05$ correspond to events in whose sources many short random processes take place. The fractal dimension of an event time profile gives an idea of the type of processes that take place in its source and can be used to construct models for the sources of specific GRBs or to choose the type of model. In this case, if a model can predict the GRB time profile with a sufficient time resolution [this is possible, for example, in fireball models with internal shock waves (Kobayashi *et al.* 1997)], then only noise from the real detector must be added to the derived profile, and the time profiles obtained from this model can be compared with those for real GRBs. The fractal dimensions of the event time profiles obtained by using the fireball model with internal shock waves lie within the range $1.213 \le D \le 1.400$, which may correspond to the subgroup of short GRBs with $\langle D \rangle = 1.31 \pm 0.05$; for some model parameters, the fractal dimensions of such an event and the real short GRB990208 are equal, within the error limits.

REFERENCES

1. H. Abarbane, Waveletes, Technical Report JSR-90-150, JASON, McLean, 1991.
2. I. V. Arkhangel'skaya, I. V. Afonina, Yu. D. Kotov, and I. L. Rozental', Izv. Akad. Nauk, Ser. Fiz. **65**, 458 (2001).
3. D. Band, J. Matteson, L. Ford, *et al.*, Astrophys. J. **413**, 281 (1993).
4. I. V. Belousova, A. Mizaki, T. M. Roganova, *et al.*, Izv. Akad. Nauk, Ser. Fiz. **64**, 615 (1999).
5. M. Briggs, R. Preece, G. Pendlenton, *et al.*, astro-ph/9908119 (1999).
6. A. N. Bunner, in *Proceedings of the CGRO Science Workshop, GSFC, 1989,* p. 12.
7. Catalog http://cossc.gsfc.nasa.gov/cossc/batse/data/current (2001).
8. T. Cline, Astrophys. J. Lett. **185**, 77 (1974).
9. E. Costa, G. Guidorzi, F. Frontera, *et al.*, AIP Conf. Proc. **526**, 18 (2000).
10. R. M. Crownover, *Introduction to Fractals and Chaos* (Jones and Bartlett, Boston, 1995; Postmarket, Moscow, 2000).
11. V. A. Dyuk, *Data Processing on Computer on Examples* (Piter, St. Petersburg, 1997).
12. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1991).

13. E. Fenimore, R. Epstein, R. W. Klebesadel, *et al.*, Nature **366**, 40 (1993).
14. G. Fishman, C. Meegan, R. Wilson, *et al.*, in *Proceedings of the CGRO Science Workshop, GSFC, 1989,* p. 45.
15. *Fractals in Physics: Proceedings of the VI International Symposium on Fractals in Physics,* Ed. by Ya. B. Sinaĭ and I. M. Khalatnikov (Nauka, Moscow, 1988).
16. I. Gaĭdaliev, *Analysis and Data Processing* (Piter, St. Petersburg, 2001).
17. J. Horack, NASA Ref. Publ. **1268**, 142 (1991).
18. K. Hurley, M. Lund, R. A. Sunyaev, *et al.*, Astrophys. J., Suppl. Ser. **128**, 549 (2000).
19. R. W. Klebesadel, I. B. Strong, and R. A. Olson, Astrophys. J. Lett. **182**, 185 (1973).
20. S. Kobayashi, T. Piran, and R. Sari, Astrophys. J. **490**, 92 (1997).
21. T. Koshut, C. Kouveliotou, W. Pasiesas, *et al.*, Preprint Space Science Laboratory (October 10, 1995).
22. Yu. D. Kotov, I. V. Arkhangelskaja, I. B. Afonina, and E. A. Borodina, in *Proceedings of the International Cosmic Ray Conference, 2001,* p. 675.
23. C. Kouveliotou, in *Proceedings of the 17th Texas Symposium of Relativistic Astrophysics, Munich, Germany, 1995,* Ed. by H. Böhringen, G. E. Morfill, and J. E. Trumper; AIP Conf. Proc. **759**, 411 (1995).
24. B. I. Luchkov, I. G. Mitrofanov, and I. L. Rozental', Usp. Fiz. Nauk **166**, 743 (1996) [Phys. Usp. **39**, 695 (1996)].
25. E. P. Mazets, S. V. Golenetskiĭ, and V. N. Il'inskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **19**, 77 (1974) [JETP Lett. **19**, 4 (1974)].
26. C. A. Meegan, G. N. Pedlenton, M. S. Briggs, *et al.*, Astrophys. J., Suppl. Ser. **106**, 65 (1996).
27. A. Meszaros, Z. Bagoly, I. Horvath, *et al.*, Astrophys. J. **539**, 98 (2000).
28. T. Piran, Phys. Rep. **314**, 575 (1999).
29. K. A. Postnov, Phys. Usp. **42**, 469 (1999).
30. R. Sari and T. Piran, Astrophys. J. Lett. **455**, L143 (1995).
31. R. Sari, R. Narayan, and T. Piran, Astrophys. J. **473**, 204 (1996).
32. N. N. Shakura, M. E. Prokhorov, and N. I. Shakura, Pis'ma Astron. Zh. **20**, 168 (1994) [Astron. Lett. **20**, 137 (1994)].
33. A. Yu. Tkachenko, O. V. Terekhov, R. A. Sunyaev, *et al.*, Pis'ma Astron. Zh. **24**, 833 (1998) [Astron. Lett. **24**, 722 (1998)].
34. C. Walker and B. Schaefer, Astrophys. J. **537**, 264 (2000).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 522–526. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 593–598.

Neutrino Radiation from a Collapsing Galactic Nucleus

V. I. Dokuchaev*

Institute for Nuclear Research, Russian Academy of Sciences,
pr. 60-letiya Oktyabrya 7a, Moscow, 117312 Russia

Received February 13, 2002

Abstract—We discuss the possible observational manifestation of the formation of massive black holes in galactic nuclei in the form of an intense high-energy neutrino flux. A short-lived ($\leq$10 yr) hidden neutrino source results from the natural dynamicalal evolution of a central star cluster in the galactic nucleus before its gravitational collapse. The central star cluster at the final evolutionary stage consists of degenerate compact stars (neutron stars and stellar-mass black holes) and is embedded in a massive gaseous envelope produced by destructive collisions of normal stars. Multiple fireballs from frequent collisions of neutron stars give rise to a tenuous quasi-stationary cavity in the central part of the massive envelope. The cavity is filled with shock waves on which an effective cosmic-ray acceleration takes place. All the accelerated particles, except the secondary high-energy neutrinos, are absorbed in the dense envelope. The neutrino signal that carries information on the dynamicals of the collapsing galactic nucleus can be recorded by a neutrino detector with an effective area $S \sim 1$ km^2. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *active galactic nuclei, quasars, and radio galaxies; high-energy neutrinos; massive black holes*

INTRODUCTION

An accreting massive black hole (MBH) with a mass $M \gg M_\odot$ is the universally accepted standard energy source of active galactic nuclei (AGNs) and quasars (see, e.g., Zeldovich and Novikov 1967; Rees 1984). The existence of MBHs in AGNs at cosmologically large distances is confirmed by the entire set of astronomical data and theoretical models. In turn, recent detailed observations with the Hubble Space Telescope strongly suggest that there are MBHs with masses $M \gg M_\odot$ in almost all the galactic nuclei (Richstone *et al.* 1998; Kormendy and Gebhardt 2001) of nearby normal galaxies. The nucleus of our Galaxy apparently contains an MBH with a mass of $2.6 \times 10^6 M_\odot$ (Melia and Falcke 2001).

The formation mechanism of stellar-mass black holes is the gravitational collapse of evolved massive stars, as observationally evidenced by supernova explosions and remnants. For MBHs, in contrast to stellar-mass black holes, only the possible theoretical mechanisms of their formation have been formulated. At the same time, observational evidence for a particular MBH formation mechanism is currently lacking. In principle, MBHs could have been formed either at the pregalactic evolutionary stage of the Universe through the collapse of primordial density fluctuations (see, e.g., Ivanov *et al.* 1994; García-Bellido *et al.* 1996; Dokuchaev and Eroshenko 2002) or protogalactic halos of dark matter (Balberg and Shapiro 2001), or even after the formation of galaxies via the dynamical evolution and merging of their nuclei (see, e.g., Rees 1984; Silk 2001). The mechanisms of MBH formation from primordial density fluctuations are based on the assumptions that the inflaton field potential on which the fluctuation spectrum at the postinflation epoch depends has a peculiar form. However, these assumptions are purely hypothetical and, presently, there is no theoretical justification for them. The best developed scenario for the MBH formation is currently the dynamical evolution of central star clusters in galactic nuclei. According to this scenario, which requires no additional assumptions, MBHs must inevitably be formed in sufficiently dense galactic nuclei in a short time compared to the galaxy lifetime (Rees 1984; Dokuchaev 1991) via the collapse of a central star cluster or a supermassive star. Our subsequent discussion is based on the assumption that most MBHs in the Universe emerged after the formation of galaxies through the evolution of central star clusters in galactic nuclei. In this case, an observational manifestation of the MBH formation in a collapsing galactic nucleus can be a short (about several years) intense burst of high-energy neutrino radiation.

*E-mail: dokuchaev@inr.npd.ac.ru

1063-7737/02/2808-0522$22.00 © 2002 MAIK "Nauka/Interperiodica"

THE DYNAMICAL EVOLUTION OF GALACTIC NUCLEI

The most simple and natural formation mechanism of MBHs with masses $M \gg M_\odot$ is the collapse of a central star cluster in a galactic nucleus during its dynamical evolution (Zel'dovich and Podurets 1965; Quinlan and Shapiro 1987, 1990). The main evolutionary factor in the dynamical evolution of compact stellar systems are the relaxation pair interactions of stars, which lead to the escape (evaporation) of the fastest stars from the stellar system (see, e.g., Saslaw 1989; Spitzer 1990). The escape of the fast stars is accompanied by contraction of the stellar system and by an increase in the dispersion of stellar velocities v. Accordingly, the gravitational potential of the stellar system $\phi \sim GM/R \sim v^2$, where M and R are the mass and radius of the stellar system, respectively, increases evolitionally. Qualitatively, the evolutionary contraction of an evaporating stellar system follows from the virial relation for its total energy $E = -Nmv^2/2$, where N and m are the total number of stars in the system and the mass of an individual star, respectively. The threshold of star evaporation from the system is reached when its total energy increases to zero through interaction with other stars. Therefore, we may assume, to a first approximation, that the system energy does not change during star evaporation, i.e., $E \simeq$ const. Hence, it follows that the stellar-velocity dispersion in the remnant of the stellar system increases with decreasing N as $v \propto N^{-1/2}$. Accordingly, the system radius decreases as $R \propto N^2$. Any dissipative processes in the system only accelerate this contraction. As the stellar-velocity dispersion formally increases to the speed of light, $v \to c$, the size of the stellar-system remnant tends to the gravitational radius $R_g = 2GM/c^2$. Due to general-relativity effects, the system becomes unstable at $v \simeq (0.1{-}0.3)c$ and its dynamicalally rapid contraction [the avalanche of Zel'dovich and Podurets (1965)] and transformation into an MBH begin.

On the way to the formation of an MBH, compact galactic nuclei inevitably go through the stage of destructive collisions between normal stars due to the evolutionary increase in stellar-velocity dispersion (Spitzer and Saslaw 1966; Colgate 1967; Sanders 1970; Spitzer 1971; Dokuchaev 1991). A collision of two stars ends with their total destruction if the kinetic energy exceeds the gravitational binding energy of the stars. This condition is satisfied if the stellar-velocity dispersion is sufficiently high, more specifically, if $v \geq v_{\rm esc} = (2Gm_*/r_*)^{1/2}$, where $v_{\rm esc}$ is the escape velocity from the surface of a star with mass m_* and radius r_*. For a solar-type star, the escape velocity is $v_{\rm esc} = 620$ km s^{-1}. In a stellar system with $v > v_{\rm esc}$, normal stars are inevitably destroyed during mutual collisions or during collisions with compact stellar remnants—neutron stars (NS) or stellar-mass black holes.

The dynamical evolution time scale of a star cluster is determined by the time of relaxation through pair interactions of stars:

$$t_{\rm r} = \left(\frac{2}{3}\right)^{1/2} \frac{v^3}{4\pi G^2 m^2 n \Lambda} \simeq 4.6 \times 10^8 N_8^2 \left(\frac{v}{v_{\rm esc}}\right)^{-3} \text{ yr}, \quad (1)$$

where $N = 10^8 N_8$ is the number of stars in the cluster, $\Lambda = \ln(0.4N)$ is the gravitational Coulomb logarithm, v is the stellar-velocity dispersion, n is the number density of stars, and $M \simeq M_\odot$ is the typical mass of an individual star. The first equality in expression (1) is valid for the local parameters, while the second equality is valid only for the mean (virial) parameters of a self-gravitating cluster.

Only the compact stellar remnants, which below we call neutron stars for simplicity, survive in a stellar system with $v \geq v_{\rm esc}$ that passed through the evolutionary stage of destructive collisions between normal stars. At $v > v_{\rm esc}$, the collisional self-destruction time scale for normal stars is $t_{\rm coll} = (v_{\rm esc}/v)^4 \Lambda t_{\rm r}$ (see, e.g., Dokuchaev 1991). At $v \gg v_{\rm esc}$, the normal stars are completely destroyed in mutual collisions (for numerical simulations of catastrophic stellar collisions, see, e.g., Benz and Hills 1992; Lai *et al.* 1993). We use the quantity $v \simeq v_{\rm esc}$ as the characteristic threshold for the total destruction of two colliding stars and the formation of gravitationally unbound gas clouds.

The gas of the destroyed normal stars, which constituted the bulk of the initial star cluster, forms a gravitationally bound massive envelope in which the survived NSs are embedded. Due to dynamical friction, these NSs sink to the central part of the envelope and form their own short-lived cluster. The formation of a self-gravitating NS cluster takes place simultaneously with the formation of a massive gas envelope from the matter of destroyed normal stars. The formation time scale for the NS cluster (through dynamical friction) is of the same order of magnitude as the formation time scale for the massive envelope itself (through the collisional destruction of normal stars). The case where the formation of the NS cluster ends after the formation of the massive gas envelope, a supermassive star, will be considered elsewhere (Berezinskiĭ and Dokuchaev 2002). In the opposite case where the formation of a short-lived NS cluster ends before the formation of the massive envelope, the envelope parameters are fixed by the time of its formation at $v = v_{\rm esc}$ (Berezinsky and Dokuchaev 2001).

The virial radius of this envelope is

$$R_{\rm env} = \frac{GM_{\rm env}}{2v_p^2} = \frac{1}{4}\frac{M_{\rm env}}{m_*}r_* \qquad (2)$$
$$\simeq 0.56M_8\left(\frac{m_*}{\rm M_\odot}\right)^{-1}\left(\frac{r_*}{R_\odot}\right)\ \ {\rm pc},$$

where $M_{\rm env} = 10^8 M_8 M_\odot$ is the envelope mass. The gas of the destroyed normal stars makes up the bulk of the initial cluster in the galactic nucleus. Therefore, the natural range of masses for the massive envelope matches the corresponding range of masses for the central star clusters in galactic nuclei, $M_{\rm env} = (10^7{-}10^8)M_\odot$. Accordingly, the envelope radius is fixed by the radius of a central cluster at the time of evolution when $v = v_{\rm esc}$. The massive envelope will be maintained in a virial equilibrium by the kinetic motion of its constituent gas clouds from the destroyed stars. The conditions, at least in the outer parts of the envelope, may be similar to the conditions in the formation regions of broad emission lines in AGNs. For our model, only the mean (integrated) envelope parameters are of importance. The mean gas density in the envelope is

$$n_{\rm env} = \frac{3}{4\pi}\frac{1}{R_{\rm env}^3}\frac{M_{\rm env}}{m_p} \qquad (3)$$
$$\simeq 5.4\times 10^9 M_8^{-2}\left(\frac{m_*}{M_\odot}\right)^3\left(\frac{r_*}{R_\odot}\right)^{-3}\ \ {\rm cm}^{-3},$$

where m_p is the proton mass. The surface density (thickness) of the envelope is

$$X_{\rm env} = m_p n_{\rm env} R_{\rm env} \qquad (4)$$
$$\simeq 1.6\times 10^4 M_8^{-1}\left(\frac{m_*}{M_\odot}\right)^2\left(\frac{r_*}{R_\odot}\right)^{-2}\ \ {\rm g\,cm}^{-2},$$

and, hence, such an envelope completely absorbs electromagnetic radiation and high-energy particles, except for neutrinos and gravitational waves.

We assume that the total mass of the NS cluster inside the massive envelope of the gas from destroyed stars accounts for $\sim$1–10% of the total mass of the initial star cluster in the galactic nucleus (Spitzer and Saslaw 1966; Colgate 1967; Sanders 1970; Spitzer 1971; Rees 1984) and, hence, the total mass of the envelope, i.e., $M \sim (0.01{-}0.1)M_{\rm env}$. Let us consider a cluster of identical NSs with masses $m = 1.4M_\odot$ and with the total NS number $N = M/m = 10^6 N_6$. This cluster is located deeply inside the massive envelope that formed at the previous evolutionary stage of the galactic nucleus. The virial radius of such a cluster is

$$R = \frac{GNm}{2v^2} = \frac{1}{4}\left(\frac{c}{v}\right)^2 N r_g \qquad (5)$$
$$\simeq 1.6\times 11^{13} N_6 (v/0.1c)^{-2}\ \ {\rm cm},$$

where $r_g = 2Gm/c^2$ is the gravitational NS radius. Before the collapse into a MBH, the NS cluster for $N\sim 10^6$ and $v\sim 0.1c$ has a virial radius of $\sim$1 AU. The dynamical time $t_{\rm dyn} = R/v = (1/4)N(c/v)^3 r_g/c \simeq 0.95 N_6 (v/0.1c)^{-3}$ hrs and the relaxation (evolution) time $t_{\rm r} \simeq 0.1(N/\ln N)T_{\rm dyn} \simeq 19N_6^2(v/0.1c)^{-3}$ yr correspond to this state. This evolution time determines the duration of the most active phase of the hidden source under consideration, $t_{\rm s} \sim t_{\rm r} \sim 10$ yr.

THE CENTRAL ENERGY SOURCE

The principal feature of our model is an evolutionary increase in the rate of chance NS collisions in the cluster. The corresponding rate of NS collisions with losses through gravitational radiation (Quinlan and Shapiro 1987; Dokuchaev *et al.* 1998) is

$$\dot N_{\rm c} = 36\sqrt{2}\left(\frac{v}{c}\right)^{31/7}\frac{1}{N}\frac{c}{r_g} \qquad (6)$$
$$\simeq 4.4\times 10^3 N_6^{-1}(v/0.1c)^{31/7}\ \ {\rm yr}^{-1}.$$

The time between successive NS collisions is

$$t_{\rm c} = \dot N_{\rm c}^{-1} \simeq 7.3\times 10^3 N_6 (v/0.1c)^{-31/7}\ \ {\rm s}. \qquad (7)$$

We assume that each NS collision is accompanied by the release of energy $E = 10^{52}E_{52}$ erg with $E_{52}\sim 1$ in the form of a relativistic fireball, which after deceleration in the massive envelope transforms into a Newtonian shock. These numerous recurrent fireballs serve as the central energy source inside the massive envelope. Note that this central energy source is not restricted by the Eddington luminosity limit for the NS cluster. The total energy released by the central source in the form of fireballs over the time of the highest activity of the hidden source, $t_{\rm s}\sim 10$ yr, is

$$E_{\rm tot} \sim \dot N_{\rm c} E_0 t_{\rm s} \sim 8\times 10^{56}\ \ {\rm erg}. \qquad (8)$$

The physics of fireballs has been intensively developed in recent years for modeling cosmological gamma-ray bursts (Piran 2000). An important model parameter is the total baryon mass of the fireball

$$M_0 = E_0/\eta c^2 \simeq 5.6\times 10^{-6} E_{52}\eta_3^{-1} M_\odot, \qquad (9)$$

where the baryon mass parameter is $\eta = 10^3\eta_3$. The maximum Lorentz factor of the fireball is $\Gamma_f = \eta + 1$ at the expansion stage, when matter dominates (Cavallo and Rees 1978; Shemi and Piran 1990). The interaction of the fireball with ambient matter determines the spatial scale of its relativistic expansion. In our case, the fireball propagates inside a dense envelope with the mean density $n_{\rm env} = \rho_{\rm env}/m_p = n_9 10^9$ cm^{-3} according to (3). The fireball remains

relativistic with $\Gamma \gg 2$ up to the distance determined by the Sedov length

$$l_S = \left(\frac{3}{4\pi}\frac{E_0}{\rho_{env}c^2}\right)^{1/3} \simeq 1.2\times10^{15} n_9^{-1/3} E_{52}^{1/3} \text{ cm}. \quad (10)$$

At $r \gg l_S$, the fireball is transformed into a strong Newtonian shock wave, which expands in accordance with a self-similar Sedov–Taylor solution (Landau and Lifshits 1988) as $R_{cav}(t) = (E_0 t^2/\rho_{env})^{1/5}$ with decreasing velocity $u_{cav} = (2/5)[l_S/R(t)]^{3/2}c$. The distance between two successive fireballs, according to (7), is

$$R_c = ct_c \simeq 2.2\times10^{14} N_6 (v/0.1c)^{-31/7} \text{ cm}. \quad (11)$$

For sufficiently frequent NS collisions, each new fireball propagates in a tenuous medium produced by the previous fireball. As a result, a quasi-stationary tenuous cavity is formed at the center of the massive envelope; its radius is much larger than the radius of the NS cluster, and its outer boundary is maintained in a quasi-steady state by the constant impacts of recurrent fireballs (Berezinsky and Dokuchaev 2001). The direct and reverse shocks inside the cavity produce a strongly turbulent medium, which is favorable for the generation of magnetic fields and for Fermi cosmic-ray acceleration. The strength of the magnetic field generated by the dynamo mechanism can be estimated from the condition for equipartition of the magnetic and turbulent energies inside the cavity, $H_{eq}^2/(8\pi) \sim \rho u_t^2/2$, where ρ and u_t are the gas density and the velocity of turbulent motions in the gas, respectively. Accordingly, the maximum energy of the accelerated protons is $E_{max} \sim eH_{eq}R_{cav}$ and the acceleration time is $t_{acc} \sim (R_{cav}/c)(c/u_{sh})^2$, where R_{cav} is the cavity size and u_{sh} is the shock velocity. For a strongly turbulent medium with relativistic shocks inside the cavity, $u_t \sim u_{sh} \sim c$ and $\rho \sim \rho_{env}$, we obtain $H_{eq} = 4\times10^3$ G, $E_{max} = 2\times10^{21}$ eV, and $t_{acc} = 4\times10^4 E_{52}^{7/0} n_9^{-1/3}$ s (Berezinsky and Dokuchaev 2001). The typical time of the energy losses determined by pp collisions considerably exceeds t_{acc} and, hence, does not prevent the acceleration:

$$t_{pp} = \left(\frac{1}{E}\frac{dE}{dt}\right)^{-1} = \frac{1}{f_p\sigma_{pp}n_{env}c} \simeq 8\times10^6 n_9^{-1} \text{ s}, \quad (12)$$

where $f_p \approx 0.5$ is the fraction of the energy lost by a high-energy proton in a single collision, σ_{pp} is the cross section for the pp interaction, and n_{env} is the gas number density at the cavity boundary.

THE GENERATION AND DETECTION OF NEUTRINOS

The particles accelerated inside a tenuous cavity interact with the gas of the massive envelope and generate a high-energy neutrino flux. As a result, the galactic nucleus at the stage of gravitational collapse into an MBH is a hidden source of high-energy neutrinos (Berezinsky and Ginzburg 1981). Its central "engine" is hidden under a thick layer of matter and, apart from gravitational waves, only neutrinos can yield information on the MBH formation dynamics.

According to the standard models of particle acceleration by shock waves, we assume that half of the total power of the central source $L_{tot} = \dot{N}_c E_0$ is transformed into accelerated particles, $L_p \sim 7\times10^{47}$ erg s^{-1}. The optical depth of the gas of the surrounding massive envelope (4) is large enough for the emergent cosmic rays to be completely absorbed. The charged pions with Lorentz factors up to $\Gamma_c \sim 1/(\sigma_{\pi N} n_{env} c\tau_\pi) \sim 4\times10^{13} n_9^{-1}$ produced by pp collisions freely decay inside the massive envelope (here, $\sigma_\pi N \sim 3\times10^{-26}$ cm^2 is the cross section for the πN interaction, τ_π is the lifetime of a charged pion, and $n_{env} = 10^9 n_9$ cm^{-3} is the gas number density in the envelope). We assume a standard E^{-2} spectrum of the accelerated protons, $Q_p(E) = L_p/(\zeta E^2)$, where $\zeta = \ln(E_{max}/E_{min}) \sim 20 - -30$. The intensity of the neutrino radiation L_ν from the central source (given the generation kinematics of muon neutrinos) is $\sim (1/3)(3/4)L_p$, and, hence, the total generation power of muon neutrinos and antineutrinos is

$$Q_{\nu_\mu+\bar{\nu}_\mu}(>E) = \frac{L_p}{4\zeta E^2}. \quad (13)$$

Crossing the Earth, these neutrinos produce a steady muon flux (see, e.g., Berezinskiĭ *et al.* 1990):

$$F_\mu(>E) = \frac{\sigma_0 N_A}{b_\mu} Y_\mu(E_\mu)\frac{L_p}{4\xi E_\mu}\frac{1}{4\pi r^2}, \quad (14)$$

where the normalization cross section is $\sigma_0 = 1\times10^{-34}$ cm^2, $N_A = 6\times10^{23}$ is Avogadro's number, $b_\mu = 4\times10^{-6}$ cm^2 g^{-1} is the rate of muon energy losses, and $Y_\mu(E)$ is the total muon moment of the $\nu_\mu N$ interaction. The most effective energy range for muon detection is $E_\mu \geq 1$ TeV. The corresponding detection rate of muon events $\dot{N}(\nu_\mu) = F_\mu S$ in a neutrino detector with an effective area S at a distance r from the source is

$$\dot{N}(\nu_\mu) \simeq 70\left(\frac{L_p}{10^{48}\text{ erg s}^{-1}}\right) \times \left(\frac{S}{1\text{ km}^2}\right)\left(\frac{r}{10^3\text{ Mpc}}\right)^{-2} \text{ yr}^{-1}. \quad (15)$$

Consequently, one migh expect $\sim 10^2$ muon events per year from an individual source at a distance of 10^3 Mpc. The duration of the active stage of the neutrino source in a collapsing galactic nucleus, t_s, is determined by the NS relaxation time in the cluster: $t_s \sim t_r \sim 10$–20 yr. This stage can take place in the life of a galaxy only once, more specifically, before MBH formation. If we assume that after the MBH formation, the nucleus of a normal galaxy is transformed into an AGN, we can obtain an optimistic estimate for the number of hidden neutrino sources that simultaneously exist in the observable Universe, $N_{HS} \sim 10$–100 (Berezinsky and Dokuchaev 2001). In the infrared and optical ranges, the massive envelope around such an object must radiate at the Eddington luminosity level. Therefore, the hidden neutrino source in a collapsing galactic nucleus can manifest itself as a powerful quasar or AGN without rapid variability.

REFERENCES

1. S. Balberg and S. L. Shapiro, Phys. Rev. Lett. **88**, 101301 (2002); astro-ph/0111176 (2001).
2. W. Benz and J. G. Hills, Astrophys. J. **389**, 546 (1992).
3. V. S. Berezinskiĭ, S. V. Bulanov, V. A. Dogel', V. L. Ginzburg, and V. S. Ptuskin, *Astrophysics of Cosmic Rays* (Nauka, Moscow, 1990).
4. V. S. Berezinsky and V. L. Ginzburg, Mon. Not. R. Astron. Soc. **194**, 3 (1981).
5. V. S. Berezinsky and V. I. Dokuchaev, Astropart. Phys. **15**, 87 (2001).
6. V. S. Berezinskiĭ and V. I. Dokuchaev (in press) (2002).
7. G. Cavallo and M. J. Rees, Mon. Not. R. Astron. Soc. **183**, 359 (1978).
8. S. A. Colgate, Astrophys. J. **150**, 163 (1967).
9. V. I. Dokuchaev, Usp. Fiz. Nauk **161** (6), 1 (1991) [Sov. Phys. Usp. **34**, 447 (1991)].
10. V. I. Dokuchaev and Yu. N. Eroshenko, Zh. Éksp. Teor. Fiz. **121**, 5 (2002) [JETP **94**, 1 (2002)].
11. V. I. Dokuchaev, Yu. N. Eroshenko, and L. M. Ozernoy, Astrophys. J. **502**, 192 (1998).
12. J. García-Bellido, A. Linde, and D. Wands, Phys. Rev. D **54**, 6040 (1996).
13. P. Ivanov, P. Naselsky, and I. Novikov, Phys. Rev. D **50**, 7173 (1994).
14. J. Kormendy and K. Gebhardt, in *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics,* Ed. by H. Martel and J. C. Wheeler (American Inst. of Physics, New York, 2001); astro-ph/0105230 (2001).
15. D. Lai, F. A. Rasio, and S. L. Shapiro, Astrophys. J. **412**, 593 (1993).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics,* Vol. 6: *Fluid Mechanics* (Pergamon, New York, 1987; Nauka, Moscow, 1988), Chap. 10, Section 106.
17. F. Melia and H. Falcke, Annu. Rev. Astron. Astrophys. **39**, 309 (2001).
18. T. Piran, Phys. Rep. **333**, 529 (2000).
19. C. D. Quinlan and S. L. Shapiro, Astrophys. J. **321**, 199 (1987).
20. C. D. Quinlan and S. L. Shapiro, Astrophys. J. **356**, 483 (1990).
21. M. J. Rees, Annu. Rev. Astron. Astrophys. **22**, 471 (1984).
22. D. Richstone, E. A. Ajhar, R. Bender, *et al.*, Nature **395**, A14 (1998).
23. R. H. Sanders, Astrophys. J. **162**, 791 (1970).
24. W. C. Saslaw, *Gravitational Physics of Stellar and Galactic Systems* (Cambridge Univ. Press, Cambridge, 1985; Mir, Moscow, 1989).
25. A. Shemi and T. Piran, Astrophys. J. Lett. **365**, L55 (1990).
26. J. Silk, astro-ph/0109325 (2001).
27. L. Spitzer, in *Galactic Nuclei,* Ed. by D. Q'Connel (North-Holland, Amsterdam, 1971), p. 443.
28. L. Spitzer, Jr., *Dynamicalal Evolution of Globular Clusters* (Princeton Univ. Press, Princeton, 1987; Mir, Moscow, 1990).
29. L. Spitzer and W. C. Saslaw, Astrophys. J. **143**, 400 (1966).
30. Ya. B. Zeldovich and I. D. Novikov, *Relativistic Astrophysics* (Nauka, Moscow, 1967; University of Chicago Press, Chicago, 1971).
31. Ya. B. Zel'dovich and M. A. Podurets, Astron. Zh. **42**, 963 (1965) [Sov. Astron. **9**, 742 (1965)].

Translated by G. Rudnitskii

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 527–535. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 599–608.

Relationship between the Thickness of Stellar Disks and the Relative Mass of a Dark Galactic Halo

A. V. Zasov[1*], D. V. Bizyaev[1], D. I. Makarov[2], and N. V. Tyurina[1]

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

[2] *Special Astrophysical Observatory, Russian Academy of Sciences, pos. Nizhniĭ Arkhyz, Russia*

Received March 13, 2002

Abstract—We analyze the R- and K_s-band photometric profiles for two independent samples of edge-on galaxies. The thickness of old stellar disks is shown to be related to the relative masses of the spherical and disk components of galaxies. The radial-to-vertical scale length ratio for galactic disks increases (the disks become thinner) with increasing total mass-to-light ratio of the galaxies, which reflects the relative contribution of the dark halo to the total mass, and with decreasing central deprojected disk brightness (density). Our results are in good agreement with numerical models of collisionless disks that evolved to a marginally stable equilibrium state. This suggests that, in most galaxies, the vertical stellar-velocity dispersion, on which the equilibrium-disk thickness depends, is close to a minimum value that ensures disk stability. The thinnest edge-on disks appear to be low-brightness galaxies in which the dark-halo mass far exceeds the stellar-disk mass. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *galactic structure, galactic dynamics, edge-on galaxies*

INTRODUCTION

Galaxy disks are complex structural components that include the bulk of the stellar mass in most of the spiral galaxies. Their masses and internal structure are crucial factors that determine all large-scale active processes in galaxies, such as propagation of density waves, star formation, and the associated phenomena.

As a rule, the brightness (and, consequently, the surface density) of disks at large galactocentric distances R decreases with increasing R according to an exponential law with a radial scale length h of the order of several kpc. Another geometrical parameter of the stellar disk—its thickness—can be characterized by vertical scale length z_0. In an isothermal disk, the decrease of density with the distance from the galactic plane can be described by the following law:

$$\rho(z) = \rho_0 \mathrm{sech}^2(z/z_0), \qquad (1)$$

although some other alternative approximations are possible (de Grijs and van der Kruit 1996) such as the exponential or sech(z)-model of brightness decrease.

The thickness z_0, or the vertical disk scale height, is primarily determined by the local disk density and stellar velocity dispersion. However, as we can see in our own Galaxy, young and old stars have different velocity dispersions resulting in a rather complex vertical disk structure. Actually, since the bulk of the disk mass in spiral galaxies consists of stars that are several billion years old, hereafter we assign the photometrically determined thickness to the old stellar disk. Nonexistent or small color gradients in edge-on galaxies in the direction perpendicular to the disk plane beyond the narrow lane along the major axis where dust is concentrated (see de Grijs 1998 and references therein) are indicative of the rather homogeneous stellar content of old disks.

In contrast to radial scale length h, the disk thickness can be measured directly only in galaxies where disks are seen edge-on. The relative disk thickness can be characterized, to a first approximation, by the photometric isophotal axial ratios b/a of these galaxies, although inferring the vertical-to-horizontal scale length ratio z_0/h from photometric data requires modeling the 3D luminosity distribution of the disk (to make corrections for projection effect). The observed brightness distribution $\mu(r, z)$ (determined by neglecting absorption) in a finite-thickness disk seen edge-on is related to the parameters h and z_0 via modified Bessel's functions of the first kind, $K_1(r/h)$ (van der Kruit and Searle 1981a):

$$\mu(r, z) = \mu(0, 0)\mathrm{sech}^2\left(\frac{z}{z_0}\right)\frac{r}{h}K_1\left(\frac{r}{h}\right), \qquad (2)$$

*E-mail: zasov@sai.msu.ru

1063-7737/02/2808-0527$22.00

where r and z are the sky-plane coordinates.

Given relation (2), z_0 and h can be determined from the vertical and major-axis photometric cross sections, respectively. At the peripheral regions of galaxies, such a parameter as the so-called disk cutoff radius, R_c, beyond which the decrease of the disk brightness is described by a shorter exponential scale length than at smaller galactocentric distances, may be of great importance. According to different estimates, R_c is typically equal to (3–5) h (see de Grijs *et al.* 2001; de Grijs and van der Kruit 1996; and references therein).

Indirect estimates of z_0 may be based on measurements of the stellar velocity dispersion of the old disk population. However, such estimates require certain assumptions about the surface brightness or integrated mass of the exponential disk (Bottema 1993). The reverse is also true: given the disk thickness, velocity dispersion measurements make it possible to estimate the local surface brightness of the disk and, consequently, its total mass.

The observations of edge-on galaxies showed that the disk thickness varies over a wide range from one galaxy to another, and the apparent axial ratio can be as high as 10–20 for the thinnest disks (Kudrya *et al.* 1994; Karachentsev *et al.* 1997). What determines the relative disk thickness remains an open question. It appears to correlate with the morphological type, although the latter is determined rather uncertainly for edge-on galaxies—it is inferred not from the shape of spirals but only from the relative size and luminosity of the bulge. The disks of late-type galaxies (Sc–Sd) are, on the average, "thinner" than those of early-type objects (Karachentsev *et al.* 1997; de Grijs 1998; Ma *et al.* 1997, 1999). According to de Grijs (1998), the h/z_0 ratio in his sample of edge-on galaxies varies from 1.5–2 for early-type spirals to 3–8 for Sc–Sd galaxies, and in addition, the relative disk thickness shows no direct correlation with the rotation velocity or luminosity. To illustrate these conclusions, in Fig. 1 we compare the (B-band) a/b ratio according to the Flat Galaxy Catalog (RFGC) (Karachentsev *et al.* 1999) with the known HI line width (W_{50}), which is approximately equal to twice the maximum velocity V of disk rotation, and with absolute magnitude M_B (both parameters adopted from LEDA catalog).

It can be expected, however, from the most general considerations that the relative thickness of the equilibrium disk (at least its minimum possible value) must reflect its kinematic characteristics. The disk thickness is indeed determined by its surface density and by the dispersion C_z of stellar velocities in the direction perpendicular to the disk plane. On the other hand, C_z and radial dispersion C_r are interrelated quantities[1] with the minimum C_r determined by the condition of local gravitational stability of the disk. Zasov *et al.* (1991) showed that if the radial dispersion C_r of stellar velocities in an old stellar disk is close or proportional to the critical threshold for the gravitational (Jeans) instability of the rotating disk and if velocity dispersion C_z along the z-coordinate is proportional to C_r, then the relative disk thickness should increase with decreasing relative mass of the galactic halo.

To a first approximation (neglecting the z-component of the acceleration due to the spherical component of the galaxy), $z_0 \approx C_z^2/\pi G\sigma$ (here σ is the disk surface density). Let the radial velocity dispersion be equal to $C_r = Q \times 3.36G\sigma/\kappa$, where $\kappa \sim V/R$ is the epicyclic frequency and the Toomre parameter $Q = 1$ corresponds to a thin uniform disk that is stable (in Toomre's sense) with respect to radial perturbations. In general, Q is a function of radius. Beyond the central bulge-dominated region, it varies slowly with R and gradually increases toward the periphery (Bottema 1993). Numerical models of marginally stable disks also show that parameter Q remains almost constant over a wide R interval beyond the central region and its value ($Q \approx 1.2$–1.5) depends only slightly on the mass of the spherical and disk components of the galaxy (Khoperskov *et al.* 2002). Hence, for the region where one can assume $Q(R) \approx$ const, the vertical-to-radial disk scale length ratio can be easily expressed in terms of other parameter ratios:

$$\frac{z_0}{h} \sim \frac{C_z^2}{\sigma h} \sim \frac{\sigma}{h\kappa^2} \sim \frac{\sigma h^2}{V^2 h} \sim \frac{M_\mathrm{d}}{M_\mathrm{t}}. \qquad (3)$$

Here, $M_\mathrm{d} \sim \sigma h^2$ and $M_\mathrm{t} \sim V^2 h$ are the disk mass and the galaxy total mass, respectively, within the fixed radius (in the units of h). The thinnest galaxies can therefore be expected to be those with the highest mass fraction of the spherical halo. This conclusion agrees well with the results of the 3D N-body numerical simulations of collisionless disks (Zasov and Morozov 1985; Zasov *et al.* 1991; Mikhailova *et al.* 2001).

When applied to real galaxies, the situation is further complicated by a a number of factors that can potentially increase the thickness of quasi-equilibrium disks in the process of their long evolution and due to the slow increase of C_z (Gerssen *et al.* 2000;

[1] According to Gerssen *et al.* (2000), direct estimates obtained for several galaxies yield $C_z/C_r \approx 0.5$–0.8; within approximately the same interval (0.35–0.8) fall the ratios obtained by numerical simulation of the dynamical evolution of initially "cold" collisionless disks (Mikhailova *et al.* 2001). The condition of stability against bending perturbations for collisionless disk yields $C_z/C_r \approx 0.37$ (Polyachenko and Shukhman 1977)

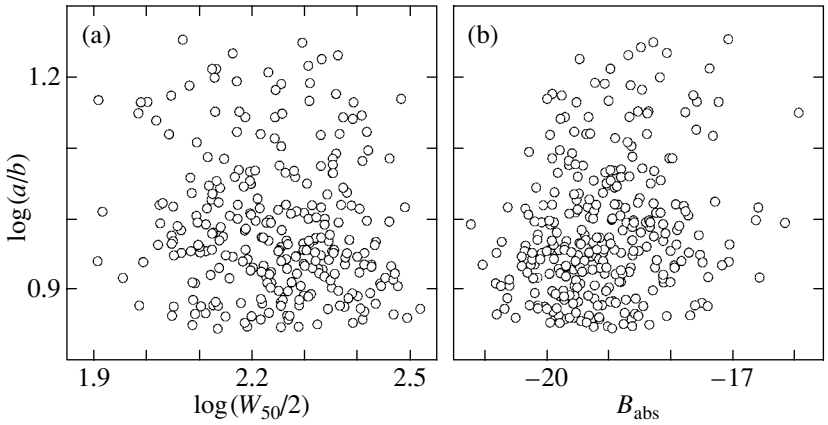


Fig. 1. Diagrams illustrating the absence of correlation between the observed axial ratio a/b and the HI line halfwidth $W_{50}/2$ (a) or absolute magnitude $B_{\rm abs}$ (b) for 340 galaxies from the RFGC catalog (Karachentsev *et al.* 1999).

Binney 2000). These factors include the scattering of disk stars during their interaction with giant molecular clouds or globular clusters; interaction of stars with density waves; merging of small satellites, which could repeatedly pass over the disk; star formation in the process of gas accretion onto the disk, which has not yet reached equilibrium; and gravitational perturbations due to neighboring galaxies. The latter effect shows up conspicuously in the fact that the relative thickness of disks in interacting systems is about twice greater than in galaxies without close neighbors (Reshetnikov and Combes 1997).

Note that the efficiency of all the processes mentioned above should be different at different galactocentric distances, whereas photometric measurements of edge-on galaxies imply that the disk thickness varies only slightly with radius (van der Kruit and Searle 1981a, 1981b; Barnaby and Thronson 1992). (However, some galaxies appear not to obey this rule—see de Grijs and Peletier 1997). The conclusion about the disk thickness remaining constant over a wide interval of galactocentric distances also follows from numerical N-body simulations of the dynamical evolution of initially cold (along the z-coordinate) collisionless disks (Mikhailova *et al.* 2001).

To clarify the processes that determine the vertical scale height of a stellar disk, it is worth verifying whether the relative thickness of disks seen edge-on correlates with the dark halo mass, and this is precisely the aim of this work.

GALAXY SAMPLES

We chose the galaxies satisfying the condition $a/b \geq 7$, which is the underlying criterion of the Flat Galaxy Catalog (RFGC, Karachentsev *et al.* 1999). The objects obeying this criterion are mostly Sc–Sd galaxies ($\sim$75%). These are disk-dominated galaxies with a small bulge contribution to the integrated luminosity (although the bulge presence may be seen in central regions), making it easier to determine their vertical and radial scale lengths and the total disk luminosities.

In this work, we use two samples of edge-on galaxies.

The first sample consists of 121 late-type UGC galaxies and includes objects with $a/b > 7$ (in the B band). For these galaxies, R-band surface CCD photometry was performed with the BTA telescope (Karachentsev *et al.* 1992). We excluded from the initial sample objects whose outer isophotes had an uncertain shape and those with isophotal asymmetry in the inner region, which might indicate that the the disk inclination differs appreciably from 90°. We also excluded nearby galaxies ($V < 750$ km s^{-1}), Virgo members, and galaxies with large internal extinction ($A_R > 0.5$). Our final analysis was based on a final sample of 51 galaxies.

Karachentsev *et al.* (1992) report their estimates of R-band axial ratios a/b, angular sizes of the semi-major axes of the 23 and $24^m/\mathrm{arcsec}^2$ isophotes, the corresponding isophotal magnitudes, and photometric profiles of the observed galaxies.

We estimated the radial scale length h by fitting the photometric major-axis profile to a function implied by relation (2) at $z_0 = 0$. Given h, the vertical scale height z_0 can be determined by measuring the semi-major (a) and semi-minor (b) axes of a certain isophote of the galaxy (sufficiently far from the center to minimize bulge effects) and using relation (2). The latter implies the points lying along the major $(0, a)$ and minor $(b, 0)$ axes:

$$\mathrm{sech}^2\left(\frac{b}{z_0}\right) = \frac{a}{h} K_1\left(\frac{a}{h}\right). \qquad (4)$$

Unfortunately, the available photometric data were insufficient for allowing a more refined approach that makes use of the entire pattern of the two-dimensional brightness distribution of a galaxy. We found the galaxies of our sample to have the ratio of the mean semimajor axis to the radial disk scale, a/h, equal to 2.9 and 3.7 for the 23^m and $24^m/\mathrm{arcsec}^2$, respectively.

As a second sample, we analyzed the largest 60 RFGC galaxies, whose vertical (z_0) and radial (h) disk scale lengths could be determined in a more rigorous way—by modeling photometric cross sections along and across the major axis of the galaxy. To this end, we used the 2MASS survey K_s-band near-infrared images available from the NASA Extragalactic Database (NED). A detailed description of the procedure we used to determine the photometric parameters is given by Bizyaev and Mitronova (in press). The above authors obtained vertical disk scale heights averaged over 20–30 vertical cross sections. We fitted each profile to a $I = I_0\mathrm{sech}^2(z/z_0)$ law with allowance for atmospheric blurring. We determined the radial disk scale lengths from the cross sections parallel to the major axis of the galaxy (but not closer than $\sim 2''$ to avoid the dustiest regions). To minimize bulge effects in the estimated photometric disk parameters z_0 and h, we also excluded the centermost regions in the cases where the isophotal ellipticity decreased centerward (due to the presence of the bulge). We treated z_0 and h in equation (2) and the central brightness of the exponential disk as free parameters of the photometric model.

For a comparative analysis of disk scales, we selected the 24 brightest ($K_\mathrm{s} < 10^m.5$) and relatively distant ($V > 750\ \mathrm{km\ s^{-1}}$) galaxies of the second sample, excluding the objects with the strongest galactic extinction $A_K > 0.25$ and probable Virgo cluster members.

The first sample of galaxies with R-band photometry and the initial sample of galaxies from the 2MASS catalog have 28 objects in common. When comparing the two samples, we excluded two objects with supposedly nonexponential profiles yielding highly discrepant scale length estimates obtained in two samples (UGC 542 = RFGC 206 and UGC 7774 = RFGC 2336). In Fig. 2, we compare the independently determined z_0 and h (in acrseconds). The median radial and vertical scale height ratios for the two samples are equal to $h(\mathrm{BTA})/h(\mathrm{2MASS}) = 1.21 \pm 0.08$ and $z_0(\mathrm{BTA})/z_0(\mathrm{2MASS}) = 1.66 \pm 0.07$, respectively. The relation between the radial scale lengths agrees well with the conclusion of de Grijs (1998) that near-infrared (K) photometric scale lengths are systematically smaller than those measured at shorter wavelengths (by a factor of about ≈ 1.2 and ≈ 1.6 compared to the I- and B-band data, respectively). The scale length ratio z_0/h also decreases as one passes to longer wavelengths (see Fig. 6 of de Grijs (1998)). According to our measurements, the relative thickness of the galaxies of the first sample (R-band) is also greater than that of the second sample (K_s-band) (Fig. 3): the means (h/z_0) are equal to 3.52 ± 0.1 and 4.93 ± 0.34 for BTA and 2MASS samples, respectively. This effect, however, can be partially due to systematically overestimated z_0 based on BTA data, because the method employed is sensitive to the eventual bulge effects in the computed minor axes of the isophotes used to determine the vertical scale height.

Our photometry showed that the galaxies of the first sample (BTA data) have a mean integrated color index of $B-R = 1.06 \pm 0.05$. The mean color indices of the edge-on galaxies common for the two samples are $B-K_\mathrm{s} = 3.34 \pm 0.17$ and $R-K_\mathrm{s} = 2.24 \pm 0.12$, respectively. These results agree well with the integrated colors of Sc–Sd galaxies of 86 galaxies observed almost face-on (de Jong 1996). The absence of strong reddening is no surprise: the dust lane extending along the major axis of an edge-on galaxy can strongly decrease the observed luminosity while having little effect on the color if the optical depth of the galaxy $\tau \gg 1$.

VERIFICATION OF THE DEPENDENCE OF THE RELATIVE DISK THICKNESS ON A M/L RATIO OF A GALAXY

If the above assumption about the decrease of the relative disk thickness with the mass fraction of the spherical component (dark halo) is true,[2] one should expect the z_0/h ratio to be the lowest in galaxies with a high ratio of the integrated mass to the integrated red (infrared) luminosity: the latter is only slightly sensitive to ongoing star formation and therefore better corresponds to the total mass of the stellar population of the disk.

Hereafter, we determine the masses of galaxies inside a fixed radius of $R_m = 4h$, within which the luminosity (actually, the disk luminosity) was determined from photometric data. Less than 10% of the mass of the exponential disk is located beyond $4h$—even in the absence of the usually observed steepening of the radial distribution at large R. We assume that the total mass M_t of the galaxy within R_m is approximately

[2] The samples considered consist mostly of late-type galaxies without massive bulges, and, therefore, the bulk of the mass of the spherical component belongs to the dark halo.

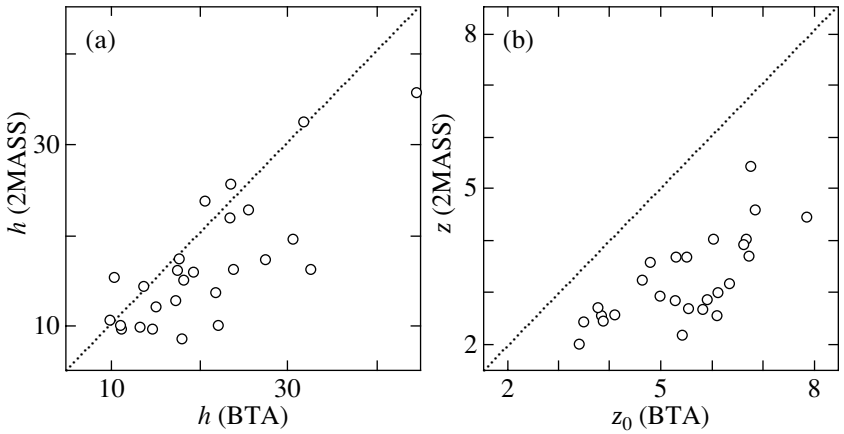


Fig. 2. Relations between the estimated radial (a) and vertical (b) disk scale lengths in different color bands (BTA—R band; 2MASS—K_s band) for the galaxies common for both samples.

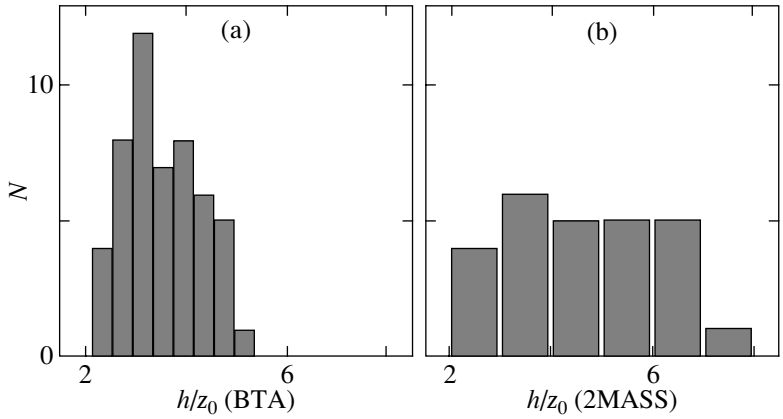


Fig. 3. Histogram of the radial-to-vertical disk scale length ratio, h/z_0, for the galaxy samples considered: (a) R-band (BTA data); (b) K_s-band (2MASS data).

equal to $W_H^2 R_m/4G$, where W_H is the HI line width. [3] This simple expression for mass is, strictly speaking, correct for spherically-symmetric systems, however, this assumption introduces a rather small error. Numerical simulations of galaxies with the measured velocity dispersion of the old disk stellar population imply halo masses significantly exceeding the disk masses within chosen R_m in most of the cases (Zasov *et al.* 2000; Khoperskov *et al.* 2001). However, even if the mass of the halo is equal to that of the thick disk within $R_m = 4h$, the above formula overestimates the mass M_t only by $\approx 25\%$.

In the first sample, we estimated the galaxy luminosities from the $24^m/\mathrm{arcsec}^2$ isophotal magnitudes extrapolating them if necessary out to R_m based on the radial brightness scale length.

If the relative component masses are unknown, the mass of the disk cannot be estimated from the rotation velocity and therefore we infer it from the disk luminosity assuming that $M_d = A(\lambda)L_d \times (M/L)_d$, where $A(\lambda) > 1$ is the factor that allows for internal extinction (which is important for the R band); L_d, the observed disk luminosity, and $(M/L)_d$, the integrated mass-to-luminosity ratio of the stellar population in the chosen photometric band. The total-to-disk mass ratio can therefore be written in the following form:

$$\frac{M_t}{M_d} \approx \frac{W_{50}^2 h}{A(\lambda) G L_d (M/L)_d}. \qquad (5)$$

The luminosity underestimation of an edge-on galaxy is difficult to take into account: it can be important even in the infrared. The extinction correction applied to reduce the R-band magnitudes of edge-on galaxies to those of face-on galaxies exceeds, on

[3] In this work, we use the W_{50} width at 50% of the maximum (adopted from LEDA database). However, the choice between W_{20} and W_{50} is of no fundamental importance, because both quantities are close to twice the maximum velocity of gas rotation.

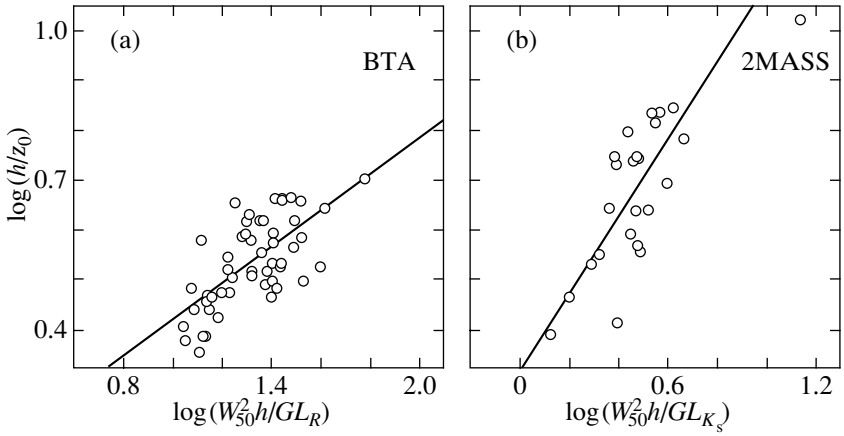


Fig. 4. Relation between the photometrically determined radial and vertical disk scale length ratio and the quantity $W_{50}^2 h/GL_{(R,K_s)}$, which determines the M_t/L_d ratio within $R = 4h$: (a) R band (BTA data); (b) K_s band (2MASS data).

the average, 1^m (Tully *et al.* 1998). The problem of estimating M_d from the observed luminosity is further complicated by the large scatter of coefficient $A(\lambda)$, which, in turn, can depend on the disk mass and thickness. For the second galaxy sample, the photometric estimates should be much less affected by dust. The reasons for this are twofold: (1) K_s-band extinction in galaxies resulting from their edge-on orientation does not, on the average, exceed $0^m.3$ (Tully *et al.* 1998), and (2) when estimating the scale lengths we excluded the regions close to the Galactic plane, which suffer from the strongest extinction. However, in spite of the simplifying assumptions adopted here both the first and the second galaxy samples exhibit conspicuous relations between h/z_0 and M_t/L_d (or, to be more precise, a quantity proportional to this ratio)—see Fig. 4—with the correlation coefficients equal to 0.68 and 0.73, respectively. This relation, which corroborates the conclusion that the disk thickness decreases with the relative mass of the spherical component is the main result of this work.

The scatter of data points on the diagrams shown here is due to the errors in the estimates of the parameters used the differences of A_λ and M/L_d of the stellar populations of individual galaxies, and the disregarded physical factors, which may increase the disk thickness (see the Introduction). The differences between the slopes based on two galaxy samples must be real despite the uncertainty of the inferred slope of the relation in Fig. 4a (photometric estimates based on 2MASS data are more reliable): the shallower behavior of the R-band relation agrees qualitatively with the fact that thinner galaxies (in the upper part of the diagram) are more extinction-affected and thus have their M/L overestimated.

DISCUSSION AND CONCLUSIONS

This study is probably the first to show that a relation exists between the relative disk thickness and the masses of the spherical and disk components of real galaxies. The conclusion that marginally stable collisionless disks become thinner with increasing mass fraction of their spherical components (in the absence of external gravitational perturbations) was earlier reached from N-body numerical simulations of three-dimensional disks in a fixed field of the spherical component starting from an unstable state with low vertical velocity dispersion (Zasov *et al.* 1991; Mikhailova *et al.* 2001). During the time equal to several rotation periods, at the outer disk edge the parameter C_z increases reaching a certain level (which decreases with R) as a result of the development of

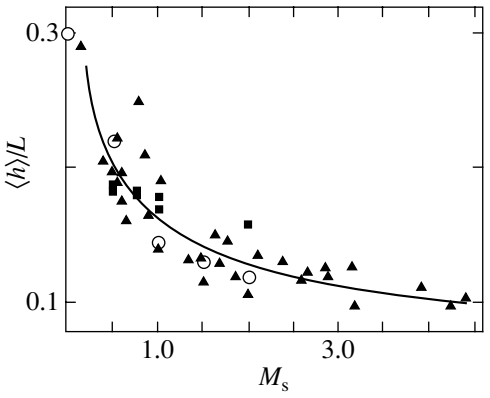


Fig. 5. Relation between the rms distance $\langle h \rangle$ of points from the disk plane expressed in the units of the radial disk scale length and the halo-to-disk mass ratio, M_s, obtained from N-body numerical simulations of galaxies with marginally stable disks. The figure is adopted from Mikhailova *et al.* (2001, Fig. 2). In our notation $\langle h \rangle / L$ is z_0/h.

bending perturbations that increase stellar velocity dispersion along the z-coordinate. Eventually, the disk becomes marginally stable against both perturbations in its plane and the bending perturbations. See Khoperskov *et al.* (2001, 2002) for a detailed description of numerical simulations.

Figure 5 (same as Fig. 2 in the paper by Mikhailova *et al.* (2001)) compares the relative disk thickness and $M_s = (M_t - M_d)/M_d$—the spherical-to-disk mass ratio—based on the results obtained by constructing numerical models for galaxies with different component masses and different shapes of rotation curves corresponding to those actually observed $V(R)$ in real galaxies.

To compare the relations shown in Figs. 4 and 5, one must convert disk luminosities into disk masses. Assuming, like we did in the previous section, that $W_{50}^2 h/G$ determines the total mass of the galaxy within $R_m = 4h$, we can write the quantity M_s laid off along the horizontal axis in Fig. 5 as

$$M_s = \frac{W_{50}^2 h}{G \times M_d} - 1. \quad (6)$$

It follows from this equation that

$$\frac{W_{50}^2 h}{G \times L_d} = (M_s + 1)(M/L)_d, \quad (7)$$

where $(M/L)_d$ is the disk mass-to-luminosity ratio for the chosen spectral interval.

Figure 6 shows in a logarithmic scale the diagram given in Fig. 4 with the superimposed curve from Fig. 5 computed in accordance with equation (7) for three mass-to-K_s-band luminosity ratios ($M/L_d =$ 1, 2, and 3).

Evolutionary models yield for the stellar population of cosmological-age galaxies a mass-to-luminosity ratio of $(M/L)_{model} \approx 1$ for the photometric K band, which is close to K_s (Bell and de Jong 2001). This ratio remains somewhat uncertain due to the lack of data about the low-mass end of the stellar mass function. All galaxies are actually situated in the domain between the adopted ratios, which are quite reasonable for a stellar population. This shows that models of marginally stable disks agree well with observations. This leads us to conclude that for most of the galaxies the mechanisms of additional disk heating (scattering by massive clouds, tidal perturbation of the disk) are not crucial for the formation of the vertical structure. The approximately constant disk thickness along the radius is then a result of two opposite tendencies—the radial decrease of both the surface disk density and the velocity dispersion at which the disk reaches stable equilibrium—almost exactly cancelling each other.

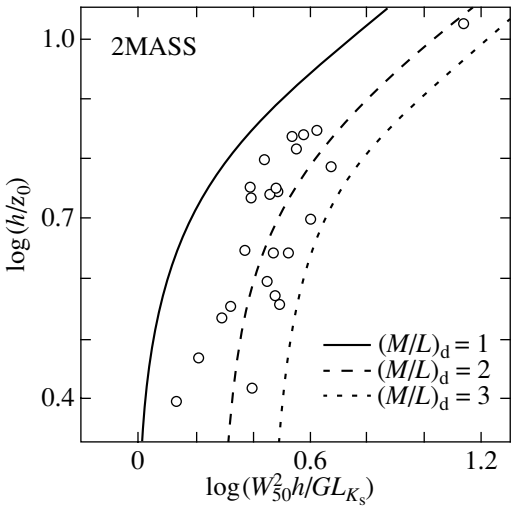


Fig. 6. The same diagram as in Fig. 4b with superimposed collissionless-disk relation inferred from numerical simulations (the line in Fig. 5) recalculated with three disk mass-to-luminosity ratios (in solar units): $(M/L)_d = 1$, 2, and 3.

Figure 7 compares the relative disk thickness with the face-on central surface brightness S_0 (in magnitudes) reduced to the face-on position using model R- and K_s-band brightness distributions. The correlation between these parameters is even more conspicuous than that between z_0/h and $(M/L)_t$, although the existence of such a relation is nothing unexpected: indeed, "normal" and low surface brightness galaxies were already shown to exhibit a close relation between S_0 and integrated $(M/L)_t$, which characterizes the dark halo mass fraction (MacGaugh and de Block 1998). The lower the central surface brightness (and, consequently, the surface density), the higher the dark halo mass fraction within the chosen $R = R_m$. The correlation between these two quantities implies, in particular, the existence of a single linear (in the logarithmic terms) Tully–Fisher relation (luminosity or mass of the disk-rotation velocity) for galaxies with different surface brightness S_0 (MacGaugh and de Block 1998).

This relation manifests itself most conspicuously in the diagram shown by Bizyaev and Mitronova (2002), which is based on an analysis of a 153-galaxy sample from the 2MASS survey. This relation appears to be more scattered at longer wavelength bands (as is evident from a comparison of diagrams a and b in Fig. 7); Bizyaev and Kaisin (in preparation) and Bizyaev (2000) came to the same conclusion based on R- and I-band photometry, respectively. The corresponding dependence in the B band is the least conspicuous (see Fig. 9 in the paper of de Grijs (1998)). The differences between the correlation coefficients and the slopes of the relations shown in Fig. 7 are

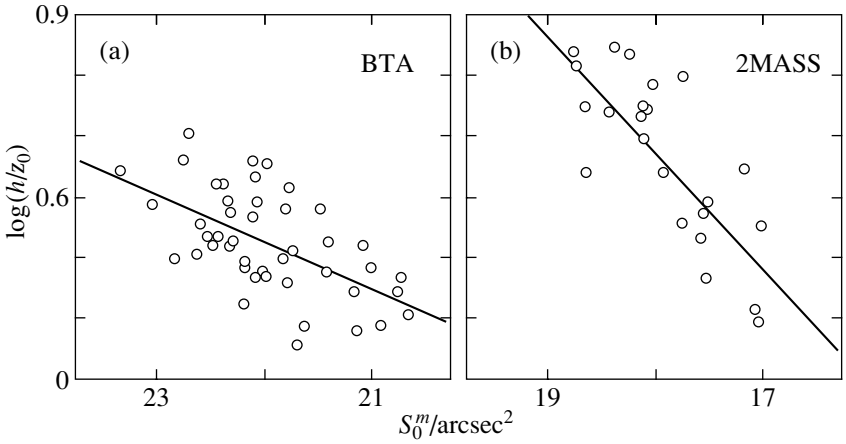


Fig. 7. Radial-to-vertical disk scale length ratio as a function of deprojected central disk brightness in magnitudes in the (a) R and (b) K_s band.

evidently due to selective internal extinction, which is strongest in galaxies with thin disks and becomes more important at shorter wavelengths. Underestimated brightness (or overestimated S_0, which is expressed in magnitudes) for galaxies with "thin" disks results in an underestimated slope of the relation in Fig. 7. The relation is more conspicuous in the K_s band where the flux decrease due to internal extinction amounts only to several tenths of a magnitude.

Thus the relations obtained lead us to conclude that the thinnest edge-on galaxies must be (after deprojection to a face-on orientation) low surface brightness spirals whose observed brightness is enhanced by the projection effect and whose dark-halo mass significantly exceeds the mass of the disk.

Note, however, that not all disk galaxies appear to obey the "dark-halo mass–disk thickness" relation. Relatively thick disks are observed not only in interacting systems (Reshetnikov and Combes, 1997) but also in some Irr galaxies, for which some authors pointed out a deficit of systems with strong apparent flattening (Hodge and Hitchcock 1966; Thuan and Seitzer 1979; van den Bergh 1988), and all this in spite of the fact that the dark-halo-mass fraction in low luminosity galaxies is, on the average, higher than in galaxies with high luminosity (Persik and Salucci 1996 ; Ashman 1992; Cote *et al.* 2000). Relatively thick disks observed in Irr galaxies may be a result of, among other things, a certain threshold level of stellar velocity dispersion, which cannot be lower than the velocity dispersion of gaseous clouds (usually $\approx$10 km s^{-1}). Unfortunately, the strong contribution of young stars to the disk luminosity in Irr galaxies and their nonuniform distribution within the galaxy complicate the vertical photometric disk structure and increase the uncertainty of the photometric tilt estimates in comparison to what we have in the case of spiral galaxies, thus preventing any direct comparison of the stellar disk thickness of these two types of objects.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 01-02-17597). We wish to thank I.D. Karachentsev for providing original data on galactic photometry.

REFERENCES

1. K. M. Ashman, Publ. Astron. Soc. Pac. **104**, 1109 (1992).
2. D. Barnaby and H. A. Thronson, Astron. J. **103**, 41 (1992).
3. E. F. Bell and R. S. de Jong, Astrophys. J. **550**, 212 (2001).
4. J. J. Binney, Astron. Soc. Pac. Conf. Ser. **197**, 107 (2000).
5. D. V. Bizyaev, Pis'ma Astron. Zh. **26**, 266 (2000) [Astron. Lett. **26**, 219 (2000)].
6. D. V. Bizyaev and S. S. Kaĭsin (in preparation).
7. D. V. Bizyaev and S. N. Mitronova, Astron. Astrophys. (submitted).
8. R. Bottema, Astron. Astrophys. **275**, 16 (1993).
9. S. Cote, C. Carignan, and K. C. Freeman, Astron. J. **120**, 3027 (2000).
10. R. de Grijs, Mon. Not. R. Astron. Soc. **299**, 595 (1998).
11. R. de Grijs and R. F. Peletier, Astron. Astrophys. **320**, L21 (1997).
12. R. de Grijs and P. C. van der Kruit, Astron. Astrophys., Suppl. Ser. **117**, 19 (1996).
13. R. de Grijs, M. Kregel, and K. H. Wesson, Mon. Not. R. Astron. Soc. **324**, 1074 (2001).
14. R. S. de Jong, Astron. Astrophys. **313**, 377 (1996).
15. J. Gerssen, K. Kuijken, and M. R. Merrifield, Mon. Not. R. Astron. Soc. **317**, 545 (2000).

16. P. W. Hodge and J. L. Hitchcock, Publ. Astron. Soc. Pac. **78**, 79 (1966).
17. I. D. Karachentsev, Ts. B. Georgiev, S. S. Kajsin, *et al.*, Astron. Astrophys. Trans. **2**, 265 (1992).
18. I. D. Karachentsev, V. E. Karachentseva, Yu. N. Kudrya, and S. L. Parnovskiĭ, Pis'ma Astron. Zh. **23**, 652 (1997) [Astron. Lett. **23**, 573 (1997)].
19. I. D. Karachentsev, V. E. Karachentseva, Y. N. Kudrya, *et al.*, Bull. SAO **47**, 5 (1999).
20. A. V. Khoperskov, A. V. Zasov, and N. V. Tyurina, Astron. Zh. **78**, 213 (2001) [Astron. Rep. **45**, 180 (2001)].
21. Yu. N. Kudrya, I. D. Karachentsev, V. E. Karachentseva, and S. L. Parnovskiĭ, Pis'ma Astron. Zh. **20**, 13 (1994) [Astron. Lett. **20**, 8 (1994)].
22. J. Ma, J. L. Zhao, C. G. Shu, and Q. H. Peng, Astron. Astrophys. **350**, 31 (1999).
23. J. Ma, Q.-H. Peng, R. Chen, and Z.-H. Ji, Astron. Astrophys., Suppl. Ser. **126**, 503 (1997).
24. S. McGaugh and W. de Block, Astrophys. J. **499**, 41 (1998).
25. S. McGaugh, J. M. Schombert, G. D. Bothun, and W. J. G. de Blok, Astrophys. J. **533**, L99 (2000).
26. E. A. Mikhailova, A. V. Khoperskov, and S. S. Sharpak, in *Stellar Dynamics: from Classic to Modern,* Ed. by L. P. Ossipkov and I. I. Nikiforov (St. Petersburg State University, St. Petersburg, 2001), p. 147.
27. M. Persic and P. Salucci, Mon. Not. R. Astron. Soc. **281**, 27 (1996).
28. V. L. Polyachenko and I. G. Shukhman, Pis'ma Astron. Zh. **3**, 254 (1977) [Sov. Astron. Lett. **3**, 134 (1977)].
29. V. Reshetnikov and F. Combes, Astron. Astrophys. **324**, 80 (1997).
30. T. X. Thuan and P. O. Seitzer, Astrophys. J. **231**, 680 (1979).
31. R. B. Tully, M. J. Pierce, Jia-Sheng Huang, *et al.*, Astron. J. **115**, 2264 (1998).
32. P. C. van der Kruit and L. Searle, Astron. Astrophys. **95**, 105 (1981b).
33. P. C. van der Kruit and L. Searle, Astron. Astrophys. **95**, 116 (1981a).
34. A. V. Zasov and A. G. Morozov, Astron. Zh. **62**, 475 (1985) [Sov. Astron. **29**, 277 (1985)].
35. A. V. Zasov, D. I. Makarov, and E. A. Mikhaĭlova, Pis'ma Astron. Zh. **61**, 282 (1991) [Sov. Astron. Lett. **17**, 374 (1991)].
36. A. V. Zasov, A. V. Khoperskov, and N. V. Tyurina, in *Stellar Dynamics: from Classic to Modern,* Ed. by L. P. Ossipkov and I. I. Nikiforov (St. Petersburg State University, St.-Pb., 2001), p. 95.

Translated by A. Dambis

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 536–542. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 609–615.

Nearby Young Single Black Holes

M. E. Prokhorov and S. B. Popov*

Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 199992 Russia

Received February 15, 2002

Abstract—We consider nearby young black holes formed after supernova explosions in close binaries whose secondary components are currently observed as the so-called runaway stars. Using data on runaway stars and making reasonable assumptions about the mechanisms of supernova explosion and binary breakup, we estimate the present positions of nearby young black holes. For two objects, we obtained relatively small error regions ($\sim$50–100 deg^2). The possibility of detecting these nearby young black holes is discussed. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *pulsars, neutron stars, and black holes; X-ray and gamma-ray sources; supernovae and supernova remnants; star clusters and associations, stellar dynamics*

INTRODUCTION

To date, stellar-mass black holes (BHs) have been discovered in close binaries (Cherepashchuk 1996) and supermassive BHs have been discovered in galactic nuclei (Kormendy 2001). It would be of great interest to find a single stellar-mass BH, but this is technically very difficult to do. Therefore, nearby single BHs are of considerable interest. To detect such objects, it would be desirable to reduce the search area, i.e., to estimate the positions of possible sources in advance. Below, we suggest a method of such estimation and use specific examples to illustrate it.

Popov *et al.* (2002) briefly discussed nearby young compact objects (neutron stars and BHs) and assumed that radio-quiet neutron stars in the solar neighborhood were associated with recent supernova explosions that produced various structures in the local interstellar medium (Local Bubble, Loop I, etc.). Here, we analyze nearby young BHs in more detail.

The main idea of our study is as follows. We estimate the present positions of nearby ($r < 1$ kpc) young (<6 Myr) BHs formed in close binaries with massive secondary components that broke up after the first supernova explosion. The so-called runaway star (Blaau 1961) appears after binary breakup. Knowing the present position and velocity of the runaway star and specifying certain parameters for the binary and supernova explosion (see, e.g., Lipunov *et al.* (1996) about the evolution of binary stars), we can estimate the present-day position of a black hole.

*E-mail: polar@sai.msu.ru

YOUNG MASSIVE STARS IN THE SOLAR NEIGHBORHOOD

The Galactic region where the Sun is located has some peculiarities. The so-called Gould Belt (Pöppel 1997) dominates in the solar neighborhood. This is a disk-like structure, $\sim 750 - -1000$ pc in size, whose center is at 150–250 pc from the Sun. The plane of the Gould Belt is inclined $\sim 18°$ with respect to the Galactic plane. The age of the Gould Belt is estimated to be 30–70 Myr; i.e., the life of the most numerous stars among those that can produce supernova explosions ($M \approx 8$–$10 M_\odot$) has come to an end there. Single radio-quiet neutron stars discovered by the ROSAT satellite (Popov *et al.* 2002) and some of the unidentified EGRET sources (Grenier and Perrot 2001) are probably associated with the Gould Belt.

Fifty-six runaway stars are known within $\sim$700 pc of the Sun (Hoogerwerf *et al.* 2001). They were formed either during the dynamical evolution of the clusters and associations where they were born (the most likely cause is a close encounter of binaries) or through the binary breakup during a supernova explosion. Four stars from this group have masses larger than $\sim 30 M_\odot$ (since these stars are single and massive, the accuracy of determining their masses is not very high).

Table 1 gives data [parameters from Hoogerwerf *et al.* (2001)] on the runaway stars considered here. Hoogerwerf *et al.* (2001) investigated all 56 nearby runaway stars in detail. These are nearby stars in that they were studied by the HIPPARCOS satellite and their sky positions, proper motions, and parallaxes are known within milliarcsecond accuracy (here, we ignore the errors in the velocities and other parameters

1063-7737/02/2808-0536$22.00

of the runaway stars). The authors traced the motion of these stars in the Galaxy and for most of them (including the four massive stars), they found when and from which association they escaped and which of the two possible ejection mechanisms operated for each particular star.

The four massive runaway stars are most likely to have acquired their high space velocities through binary breakups after supernova explosions (to all appearances, the fifth massive star, ι Ori, was ejected from its parent association through dynamical interaction; see Hoogerwerf *et al.* 2001). Several arguments may be advanced in support of this conclusion:

(1) These stars are very massive. To be ejected from the cluster (association), they had to pass near stars of comparable mass. Otherwise, according to the law of momentum conservation, less massive stars would be ejected from the cluster, whereas such massive stars are very few for any reasonable mass function. Close encounters of several massive stars turn out to be extremely rare events compared with rare close triple encounters of low-mass stars.

(2) Massive stars live only several Myr. This imposes an additional constraint on the rare events described above: the encounter must take place until the massive star explodes as a supernova.

(3) Finally, all these stars move at velocities that are several times higher than the velocity dispersion of their parent associations. This fact does not contradict anything; after a successful close encounter, the stars can acquire high velocities. However, this occurs only in rare cases; the mean velocity acquired in such processes is much lower.

More detailed arguments for each of the four stars from this group can be found in Hoogerwerf *et al.* (2001).

Thus, to all appearances, each of these four stars was a member of a binary in which its neighbor exploded some time ago. The exploded star traversed its entire evolutionary path faster; i.e., it was even more massive than the observed runaway star. Such massive stars ($M > 30$–$40M_\odot$) are currently believed to collapse not into neutron stars but into BHs (White and van Paradijs 1996; Fryer 1999). Moreover, the cores in stars with slightly higher masses ($M \gtrsim 40$–$50M_\odot$) are most likely to collapse directly into BHs without going through the intermediate stage of a hot neutron star (see, e.g., Bisnovatyĭ-Kogan 1968).

Table 1. Parameters of the four most massive runaway stars in the solar neighborhood (Hoogerwerf *et al.* 2001)

Star	Mass, $M_\odot$	Velocity, km s^{-1}	Kinematic age, Myr
ξ Per	33	65	1
HD 64760	25–35	31	6
ζ Pup	67	62	2
λ Cep	40–65	74	4.5

BINARY BREAKUP AFTER SUPERNOVA EXPLOSION

If a supernova explodes symmetrically in a binary with a circular orbit, then *at least half of the binary mass* must be ejected for the binary to break up [all aspects of binary breakup during mass ejection were considered in detail by Hills (1983)]. For example, if the mass of the runaway star is $M_{\rm opt} = 30M_\odot$ and if it did not change significantly since the binary breakup, while the BH mass is $M_{\rm BH} = 10M_\odot$, then the mass of the ejected envelope must be no less than $\Delta M \geq M_{\rm opt} + M_{\rm BH} = 40M_\odot$ and the mass of the exploded presupernova is $M_{\rm SN} = M_{\rm BH} + \Delta M \geq 50M_\odot$. Since the mass loss from such massive stars over their lifetimes is large (at least 30% of the initial mass), each of the stars under consideration was a member of an *extremely massive* binary. The presupernova mass for ζ Pup that follows from such reasoning is $87.5M_\odot$; i.e., either this was a particularly massive star ($>100M_\odot$ during its birth) or the mass loss was much lower than that expected.

We consider only binaries with two massive stars and assume that none of their components filled their Roche lobe before a supernova explosion. Note that it is unlikely that these systems passed through the stage of mass transfer. However, if such a process takes place, then for stable and unstable (with a common envelope) mass transfer, the primary component will lose part of its mass and the mass of the secondary component will be constant or increase. As a result, the binary-component mass ratio decreases and a symmetric supernova explosion will most likely be unable to tear the binary apart. Our second condition, a circular orbit, is guaranteed to be satisfied after the stage of mass transfer.

Since the binaries under consideration are close systems (the current velocities of the runaway stars are on the order of their orbital velocities in binaries), the assumption of circular orbits appears acceptable and the high presupernova mass makes probable the direct collapse of the supernova core into a BH (White

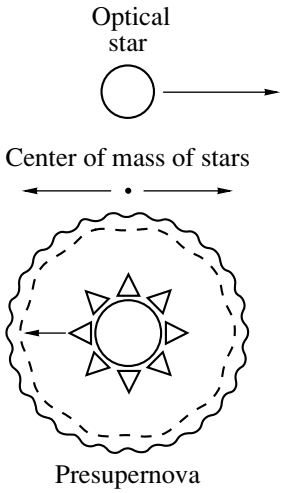


Fig. 1. A scheme for binary breakup after a supernova explosion.

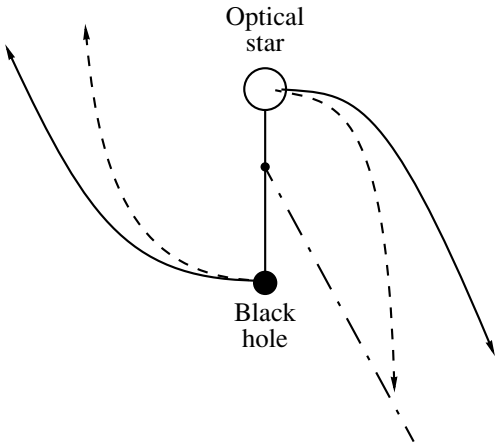


Fig. 2. Star separation after a supernova explosion in the presupernova center-of-mass frame of reference.

and van Paradijs 1996). Such collapse is generally believed to be symmetric and without recoil (i.e., the BH velocity is the same as that of the presupernova velocity before the explosion). This is in contrast to the formation of neutron stars, which are born with space velocities of several hundred kilometers per second (Lyne and Lorimer 1994).

Binary breakups through supernova explosions were considered by several authors (see, e.g., Tauris and Takens 1998; Hills 1983). However, since the above two conditions are most likely satisfied, the breakup proceeds in a simple way (see Fig. 1). The envelope is ejected symmetrically about the presupernova center and is carried away in a straight line in the direction and with the velocity of its orbital motion at the explosion time. The motion refers to the center of the envelope and is unaffected by its symmetric expansion. The center of mass of the two stars (the BH and the binary's secondary component, which became a runaway star) moves in the opposite direction but at a higher velocity, because the mass of the ejected envelope exceeds the total mass of the remaining stars.

In the center-of-mass frame of reference of the two stars (without the ejected envelope), the star velocities immediately after an explosion are directed perpendicularly to the line that connects them and the relative velocity of the star and the BH is equal to the relative orbital velocity of the stars before the explosion (see Fig. 2). The runaway star and the BH move along similar hyperbolas with the eccentricity $e = \Delta M/(M_{\rm opt} + M_{\rm BH}) \geq 1$. As the two stars move apart, the vectors of their velocities turn through angle φ: $\sin\varphi = 1/e$. In the limiting case where the ejected mass is exactly equal to half the binary mass, the stars move along parabolas ($e = 1$) and the direction of their velocities change by 90° in the separation time. The parabolic trajectories are indicated by the dashed lines in Fig. 2.

In the presupernova center-of-mass system (Fig. 3), the hyperbolic or parabolic star separation is supplemented with the uniform motion of their center of mass. As a result, both the runaway star and the BH move in the direction opposite to the motion of the ejected envelope.

CALCULATING THE BH POSITIONS

The errors in the proper motions and parallaxes of stars affect the *relative* positions of the BH and the runaway stars only slightly. The contribution of these errors to the BH localization is less significant than the uncertainties in the remaining parameters. Given the sky position of each of the stars, their distance, and velocity component, we can integrate a star's motion in the Galactic gravitational field back in time. We took the kinematic age (the time elapsed since the supernova explosion and binary breakup) from Hoogerwerf *et al.* (2001). Therefore, we can determine the relative velocity $v_{\rm opt}$ with which each of the runaway stars escaped from its parent association and its direction. The BH velocity $v_{\rm BH}$ must be determined from $v_{\rm opt}$. The problem has a unique solution if ΔM and $M_{\rm BH}$ are known, and we can find the velocity $v_{\rm BH}$ and the angle ψ that it makes with $v_{\rm opt}$: $\psi(v_{\rm BH}, v_{\rm opt}) = \widehat{\mathbf{v}_{\rm BH}\mathbf{v}_{\rm opt}}$. The center of mass of the envelope, the BH, and the runaway star moves in the orbital plane of the binary whose orientation is unknown. Thus, the velocity $\mathbf{v}_{\rm BH}$ is directed along the side surface of the cone whose axis coincides with $\mathbf{v}_{\rm opt}$ and whose half-angle is ψ. We characterize the specific position of vector $\mathbf{v}_{\rm BH}$ on the cone by an

Table 2. Parameters of the error regions for the BHs associated with massive runaway stars

Name	Distance, pc	Velocity, km s^{-1}	Error region	N_{EGRET}
ξ Per	537–611	19–70	$\sim 7^\circ \times 7^\circ$	1
HD 64760	263–645	11–59	$\sim 45^\circ \times 50^\circ$	12
ζ Pup	404–519	33–58	$\sim 12^\circ \times 12^\circ$	1
λ Cep	223–534	19–70	$\sim 45^\circ \times 45^\circ$	6

azimuthal angle ϕ (ϕ is related to the orientation of the binary orbital plane; the choice of the zero point from which it is counted off is of no importance for subsequent analysis). Since we cannot determine the specific position of $\mathbf{v}_{BH}$ on the cone surface (i.e., ϕ) from observations, this parameter must be varied.

After specifying the binary breakup parameters, we must integrate the motion of the BH from its birth and to the present time. To integrate the motion in the Galactic potential, we used the same code and constants specifying the Galactic potential as in our previous computations of the motion of single neutron stars (Popov *et al.* 2000).

Here, we make three simplifying assumptions, which are discussed below:

—the supernova explosion is symmetric, i.e., the space velocity of the remnant (BH) does not vary during the explosion;

—the association moves in a circular orbit in the Galactic disk;

—the binary velocity inside the association is disregarded.

These assumptions allow us to use the above relation between the velocities of the runaway star and BH at the point of binary breakup. For each set of parameters ϕ, ΔM, and M_{BH}, we obtain the vector $v_{BH}(\phi, \Delta M, M_{BH})$. Integrating the BH motion from the supernova explosion to the present time, we find its sky position. When exhausting the admissible values of the parameters, these points sweep the sky region where the BH must be searched for.

Table 2 gives the following data on BHs: the heliocentric distance; the BH velocity relative to the interstellar medium (i.e., relative to the circular Galactic rotation at a given point); the size of the error region; and the number of unidentified EGRET sources in this region. Despite the simplifying assumptions, we obtained large regions in the sky for λ Cep and HD 64760 in which the search was not promising. Figure 4 shows the trajectory of the optical star and a number of possible BH trajectories for the binary that produced ζ Pup. Figure 5 shows the possible BH error region for this system (both figures are in Galactic coordinates). Figures 6 and 7 show the same results for ξ Per. Since we obtained large error regions for other stars, no similar figures are given here for them.

The mass of the ejected envelope ΔM, the presupernova mass M_{SN}, and the BH velocity relative to the interstellar medium for the rings is shown in Figs. 5 and 7.

The largest masses are given for illustrative purposes. It should be noted, however, that a reduction in the upper limit of the presupernova mass to $100M_\odot$ for ξ Per and to $120M_\odot$ for ζ Pup changes the BH error regions only slightly.

DISCUSSION AND CONCLUSIONS

The errors in the proper motions and parallaxes affect only slightly the relative positions of the BHs and runaway stars. The contribution of these errors to the BH localization error is less significant than that of the uncertainty in other parameters ($\Delta M, M_{BH}, \phi$).

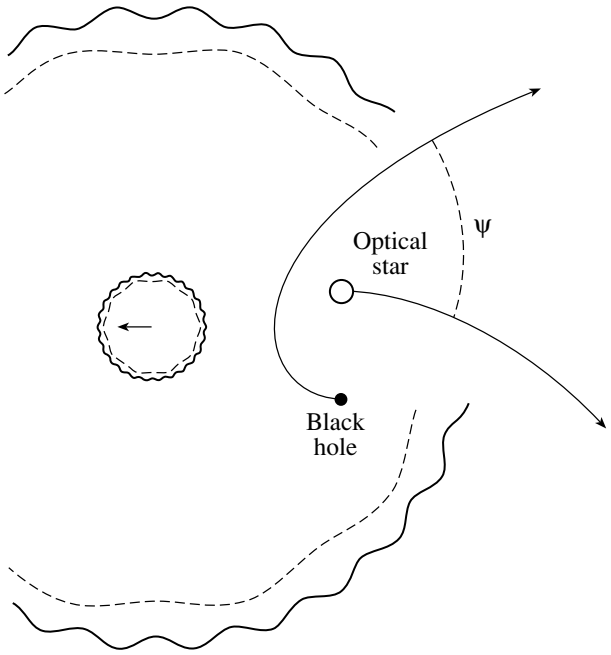


Fig. 3. Star separation after a supernova explosion in the presupernova center-of-mass system.

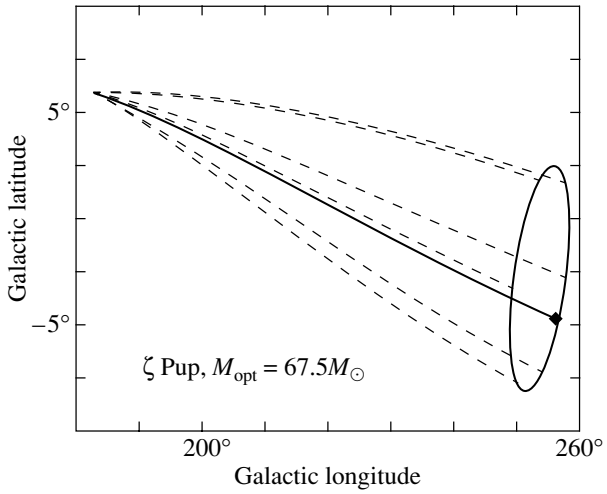


Fig. 4. The sky trajectory of the runaway star ζ Pup (solid line) and four possible BH trajectories (dashed lines). The BH mass was set equal to $M_{BH} = 10M_\odot$.

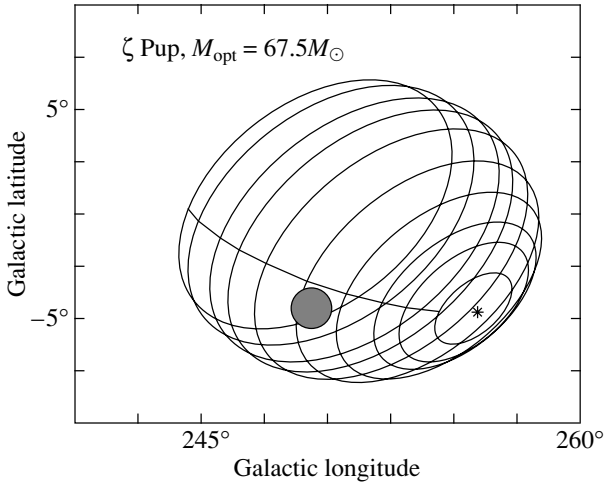


Fig. 5. The possible error region of the BH that originated from the same binary breakup as the runaway star ζ Pup. The rings correspond to different ejected masses ΔM and presupernova orbital orientations ϕ. The asterisk and the circle mark the positions of the runaway star and the unidentified EGRET source (3EG J0747−3412), respectively. The BH mass is set equal to $M_{BH} = 10M_\odot$. The smallest ΔM corresponds to the ring nearest the runaway star.

ζ Pup

$\Delta M, M_\odot$	78	79	80	82	85	90	95	100	110	120
$M_{SN}, M_\odot$	88	89	90	92	95	100	105	110	120	130
v, km s^{-1}	57–58	56–57	55–56	53–55	51–52	47–49	44–46	41–43	37–38	33–35

ξ Per

$\Delta M, M_\odot$	44	45	47	50	55	60	70	80	100	120
$M_{SN}, M_\odot$	54	55	57	60	65	70	80	90	110	130
v, km s^{-1}	69–70	66–68	62–63	56–58	49–51	44–46	33–35	31–32	24–25	19–20

From our assumptions about the velocities, the first assumption (about zero recoil during the BH formation) seems most uncertain. If we draw an analogy with neutron stars (i.e., if we scale the velocity in ac-

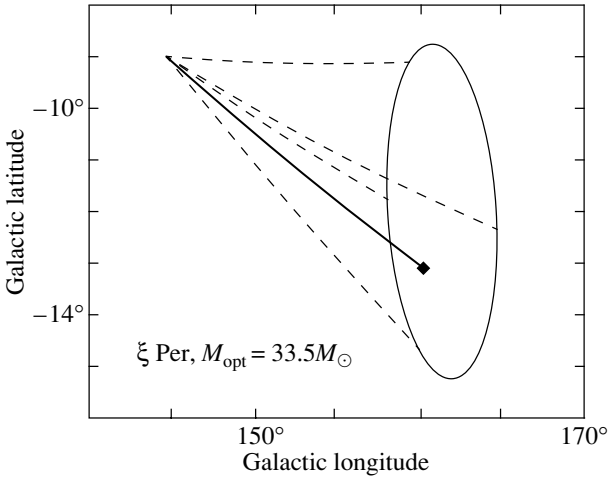


Fig. 6. Same as Fig. 4 for the runaway star ξ Per.

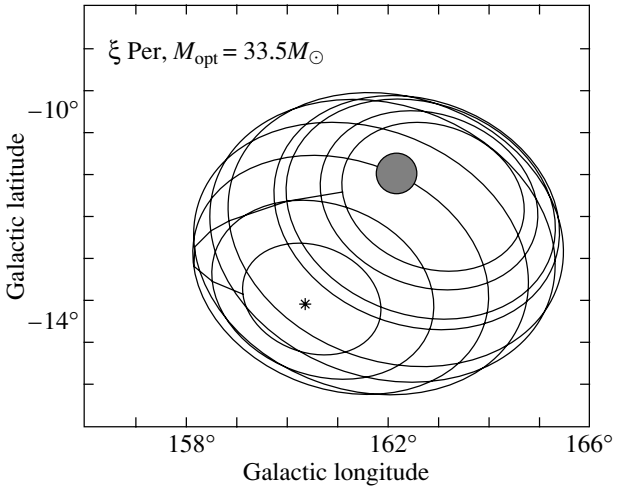


Fig. 7. Same as Fig. 5 for the runaway star ξ Per. The circle marks the position of the unidentified EGRET source (3EG J0416+3650).

cordance with an increase in the mass of the compact object and with changes in other parameters), then the BH could gain an additional velocity of up to several km s^{-1} during its birth, which would completely change the inferred error region. However, as yet, no compelling experimental evidence is available for a low or high BH recoil velocity.

The assumption about the circular motion of young stellar associations in the Galactic disk appears plausible enough. Moreover, this motion can, in principle, be measured. The motion of a binary inside an association can be taken into account in calculations (by adding a randomly oriented velocity on the order of the velocity dispersion inside the association to the velocity vector of its center). These velocities are low and their allowance causes a small increase in the error regions (by 15–20 %, according to our estimates). We ignore the small corrections here.

The probability of finding a BH increases as the corresponding runaway star is approached. This is because closer sky positions of the two components correspond to closer pre-explosion stellar masses. The close positions of the BH and the runaway star prove to be more probable for the mass functions that fall off toward more massive stars. However, the distribution function is wide enough and there is no sharp maximum at the current position of the runaway star. The situation is even more complicated observationally, because farther relative positions of the BH and the runaway star correspond to a lower BH space velocity, i.e., to a higher accretion rate at the same interstellar-gas density, which should facilitate the detection of such an object.

The BH activity in the hard energy band may result from the accretion of turbulized interstellar matter (Shvartsman 1971). Such matter has a nonzero angular momentum and cannot immediately fall to the BH. Instead, it forms an accretion ring near the latter, which, due to viscosity, transforms into an accretion disk. If the matter from such a ring does not fall completely to the BH in the time it takes for it to cross an interstellar turbulence cell, then a new ring of matter with a different orientation will begin to form near the BH. These rings "annihilate" (i.e., mutually destroy each other), thereby increasing the accretion rate. The accretion rate will vary greatly on the turbulent-cell crossing time scales (from several days to several years, depending on the BH velocity relative to the interstellar medium). The upper limits on the BH velocity (see Table 2) are large and no significant accretion rate should be expected in this case. However, the lower limits on the velocities make the $\dot{M}$ estimate optimistic.

Many authors considered the rate of accretion onto a single BH [see, e.g., Gruzinov (1998) and references therein]. However, of particular interest in our case is nonstationary BH activity (Gruzinov 1999). At a low mean luminosity and at relatively large heliocentric distances (hundreds of pc, see Table 2), the source can be open during a short-term increase in flux.

Since we obtained relatively small possible BH error regions for ξ Per and ζ Pup, only one candidate can be found in the third EGRET catalog for each of these stars. These sources are 3EG J0747−3412 (for ζ Pup) and 3EG J0416+3650 (for ξ Per). For λ Cep and HD 64760, for which our computations yielded large BH error regions, we found 6 and 12 sources, respectively, in the EGRET catalog. However, in the latter case, of particular interest may be the sources

that are especially close to the observed runaway star. These are 3EG J2227+6122 for λ Cep and 3EG J0724−4713, 3EG J0725−5140, 3EG J0828−4954, and 3EG J0903−3531 for HD 64760.

Note that analysis of massive runaway stars can shed additional light on the explosion mechanism of massive stars. According to the currently most popular supernova explosion mechanism (Fryer 1999), the collapse of stars with masses $>40M_\odot$ proceeds with no mass ejection and gives rise to the most massive BHs. However, in this case, it is difficult to explain the breakups of binaries in which the secondary components are heavier than $\sim 30M_\odot$. Prokhorov and Postnov (2001) considered various supernova explosion mechanisms and concluded that the magnetorotational mechanism best explains the observed mass distribution of compact objects. In this mechanism, the recoil for BHs is much weaker than that for neutron stars and, furthermore, the envelope is ejected even if a BH is formed. The study of the breakup products of close binaries may give additional arguments for a particular supernova explosion mechanism.

Apart from the BHs formed in massive close binaries, there must be about 20 more BHs younger than 10 Myr in the solar neighborhood. This follows from the supernova rate in the Gould Belt, which is about 20–30 per Myr (Grenier 2000), and the ratio of the number of neutron stars to that of BHs (on the order of 10 : 1). In addition, one might expect a large number of old BHs to exist within 1 kpc of the Sun. However, these objects are difficult to identify without some a priori knowledge about their positions and other parameters (space velocity, heliocentric distance). That is why we attempted to show how these parameters can be determined from data on runaway stars.

ACKNOWLEDGMENTS

We are grateful to Profs. K.V. Kuimov and A.S. Rastorguev for their advice and to the referees for their remarks that improved this paper. S.B. Popov thanks Monica Colpi, Aldo Treves, Roberto Turolla, and Luca Zampieri. We are also grateful to the organizers and participants of the HEA-2001 conference (Space Research Institute of the Russian Academy of Sciences) for providing a forum for the presentation and fruitful discussion of this work. This work was supported by the Russian Foundation for Basic Research (project no. 00-02-17164).

REFERENCES

1. G. S. Bisnovatyĭ-Kogan, Astrofizika **4**, 221 (1968).
2. A. Blaauw, Bull. Astron. Inst. Netherlands **15**, 265 (1961).
3. A. M. Cherepashchuk, Usp. Fiz. Nauk **166**, 809 (1996) [Phys. Usp. **39**, 759 (1996)].
4. C. L. Fryer, Astrophys. J. **522**, 413 (1999).
5. I. A. Grenier, Astron. Astrophys. Lett. **364**, L93 (2000).
6. I. A. Grenier and C. A. Perrot, in *Gamma 2001,* Ed. by S. Ritz, N. Gehrels, and C. R. Shrader, AIP Conf. Proc. **587**, 649 (2001).
7. A. Gruzinov, Astrophys. J. **501**, 787 (1998).
8. A. Gruzinov, astro-ph/9908101 (1999).
9. J. G. Hills, Astrophys. J. **267**, 322 (1983).
10. R. Hoogerwerf, J. H. J. de Bruijne, and P. T. de Zeeuw, Astron. Astrophys. **365**, 49 (2001).
11. J. Kormendy, Rev. Mex. Astron. Astrofis. **10**, 69 (2001).
12. V. M. Lipunov, K. A. Postnov, and M. E. Prokhorov, Astrophys. Space Phys. Rev. **9**, part 4 (1996).
13. A. G. Lyne and D. R. Lorimer, Nature **369**, 127 (1994).
14. S. B. Popov, M. Colpi, A. Treves, *et al.*, Astrophys. J. **530**, 896 (2000).
15. S. B. Popov, M. E. Prokhorov, M. Colpi, *et al.*, Gravitation and Cosmology (in press); astro-ph/0201030 (2002).
16. W. Pöppel, Fundam. Cosm. Phys. **18**, 1 (1997).
17. M. E. Prokhorov and K. A. Postnov, Odessa Astron. Publ. **14**, 78 (2001); astro-ph/0110176.
18. V. F. Shvartsman, Astron. Zh. **48**, 479 (1971) [Sov. Astron. **15**, 377 (1971)].
19. T. M. Tauris and R. J. Takens, Astron. Astrophys. **330**, 1047 (1998).
20. N. E. White and J. van Paradijs, Astrophys. J. Lett. **473**, L25 (1996).

Translated by A. Dambis

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 543–552. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 616–625.

The Two-Dimensional Structure of Thin Accretion Disks

V. S. Beskin[1*], R. Yu. Kompaneetz[2], and A. D. Tchekhovskoy[2]

[1] *Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 119991 Russia*

[2] *Moscow Institute of Physics and Technology, Institutskiĭ per. 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*

Received February 14, 2002

Abstract—The two-dimensional structure of a thin accretion disk in the vicinity of a Schwarzschild black hole after passing a marginally stable orbit ($r < 3r_g$) is discussed in terms of the Grad–Shafranov hydrodynamic equation. The accretion disk is shown to be sharply compressed as the sonic surface is approached, so the mass flow here is no longer radial. As a result, the dynamic forces $\rho[(\mathbf{v}\nabla)\mathbf{v}]_\theta$, which are equal in magnitude to the pressure gradient $\nabla_\theta P$ on the sonic surface, become significant in vertical balance. Therefore, the disk thickness in the supersonic region (and, in particular, near the black-hole horizon) may be assumed to be determined not by the pressure gradient but by the shape of ballistic trajectories. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *pulsars, neutron stars, and black holes*

INTRODUCTION

Investigation of accretion flows near black holes (BHs) is undoubtedly of considerable astrophysical interest. Substantial energy release must take place near BHs and general-relativity effects attributable to a strong gravitational field must show up here. Depending on external conditions, both quasi-spherical and disk accretion flows can be realized.

The structure of thin accretion disks has been the subject of many papers. Many results were included in textbooks (Shapiro and Teukolsky 1983; Lipunov 1987). Lynden-Bell (1969) was the first to point out that supermassive BHs surrounded by accretion disks could exist in galactic nuclei. Subsequently, Shakura (1972), Shakura and Sunyaev (1973), and Novikov and Thorne (1973) laid the foundations of the theory for such disks that is now called the standard model or the model of an α-disk. This model introduces a phenomenological proportionality coefficient $\alpha_{SS} < 1$ between the viscous stress tensor t^r_φ and pressure P: $t^r_\varphi = \alpha_{SS} P$. According to this model, under certain conditions, matter forms a thin equilibrium disk and moves in virtually circular orbits with the Keplerian velocity

$$v_K(r) = (GM/r)^{1/2}. \quad (1)$$

Indeed, if the accreted-gas temperature is low enough for the speed of sound c_s to be much lower than the Keplerian rotation velocity v_K, then the pressure forces are much weaker than the gravitational and centrifugal forces, which balance each other out. In this case, the disk is thin and its thickness is determined by the accreted-matter pressure

$$H \approx r\frac{c_s}{v_K}. \quad (2)$$

Let there now be a small viscosity in the matter. Minor friction between the rotating gas layers will then lead to the loss of angular momentum, so the accreted matter will gradually approach the compact object. This motion disturbs the vertical balance only slightly and is commonly considered as a small perturbation. As a result, the radial velocity can be written as

$$\frac{v_r}{v_K} \approx \alpha_{SS}\frac{c_s^2}{v_K^2}, \quad (3)$$

so the radial velocity at $c_s \ll v_K$ is much lower than the Keplerian velocity v_K and the speed of sound c_s.

The foregoing is valid far from the compact object where the relativistic effects are unimportant. As for the inner regions of the accretion disk, the general-relativity effects here lead to at least two new qualitative phenomena. First, there are known to be no stable circular orbits near BHs. Thus, for a Schwarzschild (nonrotating) BH, the radius of the marginally stable orbit is $r_0 = 3r_g$, where $r_g = 2GM/c^2$ is the gravitational radius. This implies that the accreted matter that fell into the region $r < r_0$ will approach the BH horizon rapidly (to be more precise, in a dynamic time $\tau_d \sim [v_r(r_0)/c]^{-1/3} r_g/c$); importantly, such a motion will also take place in the absence of viscosity. Second, the fact that the accretion onto a BH must be

*E-mail: beskin@lpi.ru

transonic in nature is of fundamental importance. This conclusion follows from the well-known theorem stating that the flow must be supersonic on the BH horizon (see, e.g., Beskin 1997). Indeed, as we see from relation (3), the condition for the radial velocity being small, $v_r \ll c_s$, remains valid up to the marginally stable orbit. Therefore, the sonic surface, on which, by definition, the poloidal velocity is equal to the speed of sound, must be located somewhere between the horizon and the marginally stable orbit. Thus, the question of what is the structure of the inner accretion-disk regions requires a consistent analysis of the transonic flow.

Transonic accretion onto BHs has been the subject of much research. Of crucial importance was the paper of Paczynski and Bisnovatyi-Kogan (1981), who formulated equations with the derivatives with respect to the radius alone; they used the quantities averaged over the disk thickness. Subsequently, virtually all authors considered only such one-dimensional models. In most cases, the model potential of Paczynski and Wiita (1980), $\varphi_g = -GM/(r - r_g)$, which simulates the properties of the strong gravitational field of a centrally symmetric BH, was used (Abramowicz *et al.* 1988; Papaloizou and Szuszkiewicz 1994; Chen *et al.* 1997; Narayan *et al.* 1997; Artemova *et al.* 2001). The Schwarzschild and Kerr metrics that describe the field of an axisymmetric BH were used appreciably more rarely (Riffert and Herold 1995; Peitz and Appl 1997; Beloborodov 1998; Gammie and Popham 1998a, 1998b). As for the two-dimensional structure, it was investigated mostly only numerically and only for thick disks (see, e.g., Igumenshchev and Beloborodov 1997).

Thus, for thin accretion disks, basic results were obtained in terms of the one-dimensional problem. However, although the procedure for averaging over the disk thickness seems obvious enough at first glance (it is undoubtedly valid in the region of stable orbits), it actually requires a more serious analysis. In our view, the assumption that the transverse velocity may always be disregarded in thin accretion disks is most debatable. This assumption is widely used, explicitly or implicitly, virtually in all the papers devoted to thin accretion disks (see the papers cited above and Abramowicz and Zurek 1981; Chakrabarti 1996).

The importance of allowing for the transverse velocity in the inner accretion-disk regions was pointed out by Abramowicz *et al.* (1997). This is because the transverse component of $[(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}]_\theta$ (as well as its relativistic generalization; see below) that appears in the Euler equation under the natural assumption of $v_\theta \propto \cos\theta$ as $\cos\theta \to 0$ tends to zero on the equator in the same way as the components of the gravitational force and the pressure gradient. Therefore, the fact that the transverse velocity becomes zero on the equator by no means implies that the contribution of $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ in the vertical balance of forces may be ignored compared to the gravitational force and the gas pressure. Nevertheless, at the end of the paper, we conclude that for thin disks, $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ may still be disregarded, so the accretion-disk thickness can be determined from the balance of the gravitational force and pressure gradient up to the BH horizon.

As we show below, this conclusion appears to be untrue. Here, we only note that the above suggestion also disagrees with the well-known pattern of motion during the accretion of Bondi (1952), a transonic spherically symmetric ideal flow. Recall that in such a flow, the term $[(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}]_r$ at large distances compared to the radius of the sonic surface is actually negligible compared to the pressure and gravitational force, which balance each other out almost completely. However, all terms are of the same order of magnitude near the sonic surface. Moreover, after passing the sonic surface, the motion of the matter differs little from free fall, so the pressure gradient, as well as all the remaining thermodynamic quantities, no longer affect the dynamics of the accreted matter.

In this paper, we discuss the two-dimensional structure of a thin accretion disk immediately after passing the marginally stable orbit ($r < 3r_g$). We show that the flow ceases to be radial before passing the $r = 3r_g$ surface. It turns out that the components of $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ actually become significant near the sonic surface. Consequently, we may assume that, as in the case of Bondi accretion, the pressure gradient in the supersonic region must be of minor importance. Therefore, the disk thickness here is determined not by the pressure gradient but by the shape of ballistic trajectories.

BASIC EQUATIONS

Let us consider the gas motion in the gravitational field of a BH. As was said above, after passing the marginally stable orbit, the effect of viscosity is no longer crucial. Therefore, the approximation of ideal hydrodynamics may be assumed to be good enough in describing the flows in the inner accretion-disk regions.

Exact equations of motion for an ideal medium in the Kerr metric were written out in terms of the so-called $(3 + 1)$-split by Beskin and Par'ev (1993) (see also Beskin 1997). The essence of this approach is that the physical processes for a stationary gravitational field can be described in terms of three-dimensional vectors measured by locally nonrotating observers (Thorne *et al.* 1986). Here, for simplicity, we formulate basic equations only for a nonrotating BH, i.e., in the Schwarzschild metric.

The Schwarzschild metric in the Boyer–Lindquist coordinates r, θ, and φ is known to be (Landau and Lifshitz 1973)

$$ds^2 = -\alpha^2 dt^2 + g_{ik}dx^i dx^k, \tag{4}$$

where

$$\alpha^2 = 1 - 2M/r, \quad g_{rr} = \alpha^{-2}, \quad g_{\theta\theta} = r^2, \tag{5}$$
$$g_{\varphi\varphi} = \varpi^2 = r^2 \sin^2\theta.$$

Here, α is the gravitational redshift, which is zero on the horizon $\alpha(r_g) = 0$. The indices with and without a hat denote the vector components relative to the coordinate basis $\partial/\partial r$, $\partial/\partial\theta$, $\partial/\partial\varphi$ in absolute three-dimensional space and their physical components, respectively. The symbol ∇_k always denotes a covariant differentiation in absolute three-dimensional space with metric g_{ik} (5). Finally, with the exception of specially specified cases, we use the system of units $c = G = 1$.

When investigating axisymmetric steady flows, it is convenient to introduce the stream function $\Phi(r,\theta)$ that defines the physically poloidal component of the 4-velocity $\mathbf{u}_p$

$$\alpha n \mathbf{u}_p = \frac{1}{2\pi\varpi}(\nabla\Phi \times \mathbf{e}_{,\hat{\varphi}}), \tag{6}$$

where n is the particle number density in a comoving coordinate system. The $\Phi(r,\theta) = \text{const}$ lines precisely determine the streamlines of the matter.

Introducing the potential Φ proves to be convenient when considering the flows in which the integrals of motion are different on different streamlines. There are known to be three integrals of motion for an ideal flow that are conserved on the $\Phi(r,\theta) = \text{const}$ surfaces. These are the energy flux and the z component of the angular momentum

$$E(\Phi) = \mu\alpha\gamma, \tag{7}$$
$$L(\Phi) = \mu\varpi u_{\hat{\varphi}}, \tag{8}$$

as well as the entropy $S = S(\Phi)$. Here, $\mu = (\rho_m + P)/n$ (ρ_m is the internal energy density, $P = nT$ is the pressure) is the relativistic enthalpy. Below, for simplicity, we use the polytropic equation of state

$$P = k(S)n^\Gamma, \tag{9}$$

so the temperature and speed of sound can be written as (Shapiro and Teukolsky 1983)

$$T = k(S)n^{\Gamma-1}; \qquad c_s^2 = \frac{\Gamma}{\mu}k(S)n^{\Gamma-1}. \tag{10}$$

In particular, for the case of nonrelativistic temperatures, $c_s \ll 1$, we can write $\mu = m_p + m_p W$, where $W = c_s^2/(\Gamma-1)$ is the nonrelativistic enthalpy and m_p is the particle mass. In this case, the function $k(S)$ for $\Gamma = \text{const}$ has a definite form:

$$k(S) = k_0 \exp[(\Gamma-1)S], \tag{11}$$

which follows from expressions (9) and (10) and the thermodynamic relation $dP = nd\mu - nTdS$.

As a result, the relativistic Euler equation (Frolov and Novikov 1998)

$$nu^b\nabla_b(\mu u_a) + \nabla_a P - \mu n (u_{\hat{\varphi}})^2 \frac{1}{\varpi}\nabla_a\varpi \tag{12}$$
$$+ \mu n\gamma^2\frac{1}{\alpha}\nabla_a\alpha = 0,$$

in which the indices a and b take on the values of r and θ, can be rewritten as a scalar Grad–Shafranov equation, i.e., as an equation of the second order for the stream function $\Phi(r,\theta)$ containing the three integrals $E(\Phi)$, $L(\Phi)$, and $S(\Phi)$ (Beskin 1997):

$$-M^2\left[\frac{1}{\alpha}\nabla_k\left(\frac{1}{\alpha\varpi^2}\nabla^k\Phi\right) + \frac{1}{\alpha^2\varpi^2(\nabla\Phi)^2}\right. \tag{13}$$
$$\left.\times\ \frac{\nabla^i\Phi\nabla^k\Phi\nabla_i\nabla_k\Phi}{D}\right] + \frac{M^2\nabla'_k F\nabla^k\Phi}{2\alpha^2\varpi^2(\nabla\Phi)^2 D}$$
$$+\frac{64\pi^4}{\alpha^2\varpi^2M^2}\left(\varpi^2 E\frac{dE}{d\Phi} - \alpha^2 L\frac{dL}{d\Phi}\right) - 16\pi^3 nT\frac{dS}{d\Phi} = 0,$$

where the derivative ∇'_k acts on all variables except the quantity M. Here,

$$D = -1 + \frac{1}{u_p^2}\frac{c_s^2}{1-c_s^2}, \tag{14}$$

$$F = \frac{64\pi^4}{M^4}\left[\varpi^2(E-\omega L)^2 - \alpha^2L^2 - \varpi^2\alpha^2\mu^2\right], \tag{15}$$

and the thermodynamic function M^2 is defined as

$$M^2 = \frac{4\pi\mu}{n}. \tag{16}$$

Equation (13) contains only one singular surface, the sonic surface determined from the $D = 0$ condition, on which the type of equation changes from elliptic to hyperbolic.

To close the system, the Grad–Shafranov equations must be supplemented with the relativistic Bernoulli equation $u_p^2 = \gamma^2 - u_{\hat{\varphi}}^2 - 1$. Using definitions (7) and (8), the latter can be rewritten as

$$u_p^2 = \frac{E^2 - \alpha^2L^2/\varpi^2 - \alpha^2\mu^2}{\alpha^2\mu^2}. \tag{17}$$

Below, we will need this equation more than once. Finally, using definition (6), we obtain

$$E^2 = \alpha^2\mu^2 + \frac{\alpha^2}{\varpi^2}L^2 + \frac{M^4}{64\pi^4\varpi^2}(\nabla\Phi)^2. \tag{18}$$

Relation (18) allows M, together with the remaining thermodynamic quantities, to be defined via the potential $\Phi(r,\theta)$ and the three invariants of motion.

FORMULATION OF THE PROBLEM

In this paper, our goal is to make only the first step in determining the two-dimensional flow structure or, more specifically, to show that the role of the dynamic terms becomes crucial when the sonic surface $r = r_*$ is approached. Therefore, we restrict our analysis here to the subsonic flow region near the marginally stable orbit, where the poloidal velocity u_p is still much lower than the speed of sound c_s. In other words, we ignore the terms of smallness $u_\mathrm{p}^2/c_\mathrm{s}^2$ everywhere. As we see from estimate (3), the condition $u_\mathrm{p}^2/c_\mathrm{s}^2 \ll 1$ is satisfied in the marginally stable orbit with a comfortable margin. In contrast, the passage of the sonic surface (as well as the supersonic flow structure) requires special analysis and is beyond the scope of this paper.

For $u_\mathrm{p} \ll c_\mathrm{s}$, the complete equation (13) can be significantly simplified. Indeed, we see from definition (14) that the terms proportional to D^{-1} are of smallness $u_\mathrm{p}^2/c_\mathrm{s}^2$ and, hence, may be discarded. As a result, we have

$$-M^2\frac{1}{\alpha}\nabla_k\left(\frac{1}{\alpha\varpi^2}\nabla^k\Phi\right) \tag{19}$$
$$+\frac{64\pi^4}{\alpha^2\varpi^2M^2}\left(\varpi^2E\frac{dE}{d\Phi}-\alpha^2L\frac{dL}{d\Phi}\right)-16\pi^3nT\frac{dS}{d\Phi}=0.$$

As would be expected, this equation for the subsonic flow is elliptic.

On the other hand, yet another significant simplification is related to the subsonic flow. The point is that there is no critical surface in the subsonic region, which makes it necessary to impose an additional constraint on the flow parameters. Therefore, our problem requires all five boundary conditions: three conditions determined the integrals of motion and the two remaining conditions are the boundary conditions for the second-order equation (19).

Thus, to determine the two-dimensional flow structure, we must specify five quantities (three velocity components and any two thermodynamic functions) on some $r = r_0(\theta)$ surface. It would be natural to choose the surface of the marginally stable orbit, $r_0 = 3r_\mathrm{g}$, as this surface. Here, as we know (Landau and Lifshitz 1973), $\alpha_0 = \alpha(r_0) = \sqrt{2/3}$, $u_{\hat\varphi}(r_0) = 1/\sqrt{3}$, and $\gamma_0 = \gamma(r_0) = \sqrt{4/3}$. Below, for simplicity, we consider the case where the radial velocity is constant on the $r = r_0$ surface and the toroidal velocity exactly corresponds to $u_{\hat\varphi}(r_0)$:

$$u_{\hat r} = -u_0, \tag{20}$$
$$u_{\hat\Theta} = \Theta u_0, \tag{21}$$
$$u_{\hat\varphi} = (1/3)^{1/2}, \tag{22}$$

where the small quantity $u_{\hat\Theta}$ corresponds to a plane-parallel flow in the region of stable orbits and we passed to a new angular variable $\Theta = \pi/2 - \theta$ counted off from the equator in the vertical direction.

Next, we assume that the speed of sound is also constant on the $r = r_0$ surface:

$$c_\mathrm{s}(r_0,\theta) = c_0. \tag{23}$$

In our model of the polytropic equation of state (9), this implies that both the temperature $T_0 = T(r_0)$ and the relativistic enthalpy $\mu_0 = \mu(r_0)$ are also constant on this surface. If we now define E_0 and L_0 as $E_0 = \alpha_0\gamma_0 = \sqrt{8/9}$ and $L_0 = u_{\hat\varphi}(r_0)r_0 = \sqrt{3}r_\mathrm{g}$ (to within the discarded factor μ, they correspond to E and L for a free particle in the equatorial plane), then the invariants $E(\Phi)$ and $L(\Phi)$ can be written as

$$E(\Phi) = \mu_0E_0, \tag{24}$$
$$L(\Phi) = \mu_0L_0\cos\Theta_m, \tag{25}$$

and, hence, they are directly specified by the boundary conditions. Below, $\Theta_m = \Theta_m(\Phi)$ is the angle for which $\Phi(r,\Theta) = \Phi(r_0,\Theta_m)$. In other words, the function $\Theta_m(r,\Theta)$ has the meaning of angle Θ on the surface of the marginally stable orbit, which is connected by the $\Phi(r,\Theta) = \mathrm{const}$ streamline with a given point with coordinates r and Θ. It is therefore clear that $\Theta_m(r_0,\Theta) = \Theta$. As a result, by definition (6),

$$d\Phi = 2\pi\alpha_0r_0^2n(r_0,\Theta_m)u_0d\Theta_m. \tag{26}$$

When deriving relation (26), we again disregarded the quantities of order $u_\mathrm{p}^2/c_\mathrm{s}^2$ related to the low poloidal velocity.

As we see, for the boundary conditions we chose, the Bernoulli integral is the same on all streamlines: $E(\Phi) = \mathrm{const}$, implying that Eq. (19) can be rewritten in an even simpler form,

$$\frac{\partial^2\Phi}{\partial r^2}+\frac{\cos\Theta}{\alpha^2r^2}\frac{\partial}{\partial\Theta}\left(\frac{1}{\cos\Theta}\frac{\partial\Phi}{\partial\Theta}\right) \tag{27}$$
$$=-4\pi^2n^2\frac{L}{\mu^2}\frac{dL}{d\Phi}-4\pi^2n^2r^2\cos^2\Theta\frac{T}{\mu}\frac{dS}{d\Phi}.$$

As for the fifth boundary condition, it is convenient to choose the entropy $S(\Phi)$. Indeed, as we show in the Appendix, in the marginally stable orbit, the right-hand side of Eq. (27) describes the transverse balance of pressure force and effective potential, whereas the

left-hand side of Eq. (27) corresponding to the dynamic contribution $(\mathbf{v}\nabla)\mathbf{v}$ has the order of smallness u_0^2/c_0^2 and may be discarded here. Therefore, it would be natural to choose the entropy $S(\Phi)$ from the condition for the transverse balance of forces on the $r = r_0$ surface,

$$r_0^2 \cos^2\Theta_m \frac{dS}{d\Theta_m} = -\frac{\Gamma}{c_0^2}\frac{L}{\mu_0^2}\frac{dL}{d\Theta_m}, \qquad (28)$$

where $L(\Theta_m)$ can be determined from the boundary condition (25). As a result, we have

$$S(\Theta_m) = S(0) - \frac{\Gamma}{3c_0^2}\ln(\cos\Theta_m), \qquad (29)$$

and, hence, owing to (11) and (23), the density $n(r,\Theta)$ at $r = r_0$ has a standard profile:

$$n(r_0,\Theta) \approx n_0 \exp\left(-\frac{\Gamma}{6c_0^2}\Theta^2\right) \qquad (30)$$

(to be more precise, $n(r_0,\Theta) = n_0(\cos\Theta)^{\Gamma/3c_0^2}$).

Equation (27) together with the boundary condition (26) and relation (21), which specifies the derivative $\partial\Phi/\partial r$, and with expressions (25) and (29), which are needed to specify the invariants L and S, define the structure of an inviscid flow in the inner accretion-disk region. Actually, however, the flow depends only on two parameters: the speed of sound c_0 and the initial poloidal velocity u_0. If the stream function $\Phi(r,\Theta)$ is known, then all the remaining quantities describing the flow can be easily calculated by using the algebraic relations (6), (11), and (18).

Thus, for example, for nonrelativistic temperatures $c_s \ll 1$ considered below, we obtain from (10) and (17)

$$u_p^2 = u_0^2 + w^2 + \frac{1}{3}\left(\Theta_m^2 - \Theta^2\right) + \frac{2}{\Gamma-1}(c_0^2 - c_s^2). \qquad (31)$$

Here, the quantity

$$w^2(r) = \frac{E_0^2 - \alpha^2 L_0^2/r^2 - \alpha^2}{\alpha^2}, \qquad (32)$$

which depends on the radius r alone, is the poloidal 4-velocity of a free particle with a zero poloidal velocity in the marginally stable orbit. In the region $r \approx 3r_g$, it is

$$w^2 = \frac{1}{6}\left(\frac{r_0 - r}{r_0}\right)^3 \qquad (33)$$

and, hence, slowly increases as one recedes from the marginally stable orbit. Therefore, as we show below, the contribution of w^2 proves to be negligible and it may be disregarded in many cases.

DISCUSSION

Before presenting the results of our numerical calculations, let us consider the flow of gas near the equatorial plane itself, where the basic flow properties can be analyzed analytically. First, a crucially important conclusion can be drawn directly from the algebraic relation (31), in which $\Theta_m = \Theta = 0$ on the equator. Assuming now that $u_p = c_s = c_*$ and disregarding the quantity w^2, we immediately conclude that the speed of sound c_* on the sonic surface $r = r_*$ must be equal in order of magnitude to the speed of sound in the marginally stable orbit c_0. Since in our case of an ideal flow, the entropy S is constant along streamlines, the gas density n_* on the sonic surface will differ little from its value in the marginally stable orbit. This conclusion corresponds to Bondi accretion, where the gas density on the sonic surface is known to differ from its value at infinity only by a numerical coefficient (Shapiro and Teukolsky 1983). In other words, in a zero approximation, the subsonic flow may be assumed to be incompressible. Importantly, this conclusion is also valid far from the equator, because the additional term $1/3(\Theta_m^2 - \Theta^2)$ for the angles $\Theta \sim c_0$ concerned, which correspond to the characteristic disk thickness in the region of stable orbits, is also c_0^2 in order of magnitude. Of course, the density in this case differs on different streamlines.

Even on this basis, we can make a key assertion of our study. Indeed, if the gas density is essentially constant and if the poloidal velocity increases from u_0 to $c_* \sim c_0$, i.e., it changes by several orders of magnitude, then the accretion-disk thickness must also change in the same proportion because of the continuity equation

$$H(r_*) \approx \frac{u_0}{c_0} H(r_0). \qquad (34)$$

Thus, we conclude that the approximation of purely radial motion near the sonic surface is inapplicable. Accordingly, a sharp decrease in disk thickness must also be accompanied by the appearance of a vertical velocity component, which must be included in the Euler equation (12).

Let us now consider the region near the equator in more detail. First, as we show below, the scale of variation in quantities with radius Δr is appreciably smaller than the disk thickness $H(r_0)$. Therefore, when the flow far from the sonic surface is analyzed qualitatively, the derivatives with respect to Θ on the left-hand side of Eq. (27) may be discarded. In addition, the quantities r and $\alpha(r)$ may be assumed to be constant and equal to their values in the marginally stable orbit.

As a result, given the boundary condition (28), Eq. (27) takes the form

$$\frac{\partial^2\Phi}{\partial r^2} = -\frac{2\pi}{3\alpha}\frac{n^2}{n_0}\frac{\Theta_m}{u_0}\left(\frac{r_0^2}{r^2} - \frac{c_s^2}{c_0^2}\right), \qquad (35)$$

where only the leading terms were retained on the right-hand side.

Since all quantities in Eq. (35) are odd functions of Θ, it is convenient to define the new dimensionless variables $X(t)$ and $Y(t)$ as

$$\frac{\partial\Phi}{\partial r} = -\Theta A\frac{c_0}{r_0 u_0^2}X(t), \qquad (36)$$

$$\Theta_m = \Theta\frac{c_0}{u_0}Y(t). \qquad (37)$$

Here, $A = 2\pi r_0^2\alpha_0 u_0 n_0$ and the new radial variable

$$t = \frac{r_0 - r}{r_0 u_0} \qquad (38)$$

is counted off from the marginally stable orbit toward the BH. In this case, the quantity $Y(t) \propto \Theta_m/\Theta$ describes the streamline compression. The choice of normalization will become clear below.

Since, by definition, Θ_m is only a function of Φ, we obtain from the obvious relation $(\partial\Phi/\partial r)\,/\,(\partial\Phi/\partial\theta) = (\partial\Theta_m/\partial r)\,/\,(\partial\Theta_m/\partial\theta)$

$$\frac{dY}{dt} = X. \qquad (39)$$

Next, assuming that $n = n_0$ in (35) and using the definition of A and the fact that $\alpha_0^2 = 2/3$, we have

$$\frac{dX}{dt} = \frac{1}{2}Y\left(2u_0 t + 1 - Z\right), \qquad (40)$$

where $Z = c_s^2/c_0^2$. As for the quantity Z, it is also convenient to use a differential equation for its definition. Differentiating the algebraic equation (31) with allowance for the asymptotics w^2 (33) yields

$$\frac{dZ}{dt} = -(\Gamma - 1)XY + \frac{1}{4}(\Gamma - 1)\frac{u_0^3}{c_0^2}t^2. \qquad (41)$$

The system of ordinary differential equations (39)–(41) together with the boundary conditions on the $r = r_0$ surface $\partial\Phi/\partial r = 0$, $\Theta_m(r,\theta) = \Theta$, and $c_s(r,\theta) = c_0$, i.e.,

$$X(0) = 0, \qquad Y(0) = u_0/c_0, \qquad Z(0) = 1 \quad (42)$$

defines the flow structure near the equatorial plane.

Note once again that the two-dimensional problem can be reduced to a system of ordinary differential equations, because we disregard the transverse derivatives $\partial/\partial\theta$ here. A more careful analysis shows that the terms discarded in (40) are generally of the same order of smallness as the right-hand side of Eq. (40). This fact is attributable to the sharp compression of streamlines near the equator; far from the equator, the transverse derivatives no longer play a crucial role. That is why the solution of system (39)–(41) can be used only to qualitatively estimate the flow parameters near the equatorial plane. Importantly, however, as is easy to verify, the discarded terms would only increase the right-hand side of Eq. (40), so including the transverse derivatives must lead to an even sharper compression of the accretion disk.

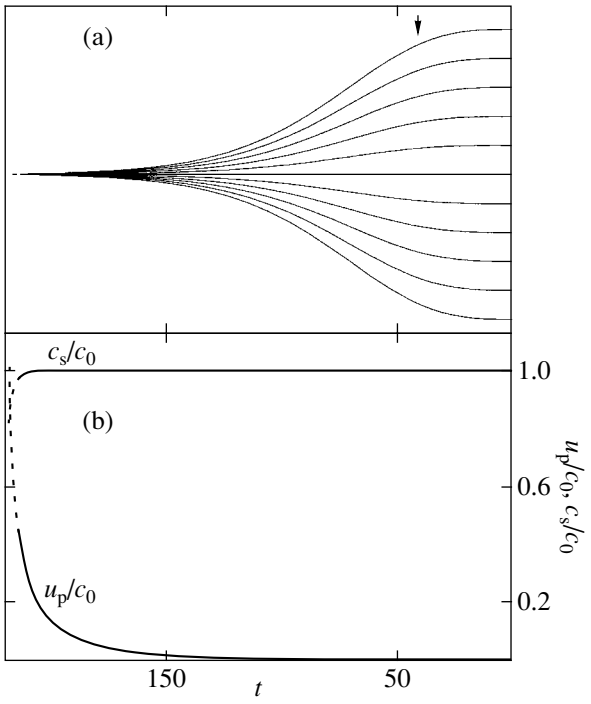


Fig. 1. (a) The pattern of streamlines on the actual scale and (b) the profiles of poloidal velocity u_p and speed of sound c_s on the equator for $c_0 = 10^{-2}$ and $u_0 = 10^{-5}$. We see that near the sonic surface, the radial scale of variation in all quantities, δt, is of the order of unity. The arrow indicates the scale $(\Delta r)_1$ (43) from which the accretion-disk compression must begin. The solid lines correspond to the range of parameters $u_p^2/c_s^2 < 0.2$, where the solution to the approximate equation (27) should not differ greatly from the solution to the complete equation (13). The dashed line indicates an extrapolation of the solution to the sonic-surface region.

Figure 1 shows the pattern of streamlines and the pattern of variations in the speed of sound and in the poloidal flow velocity near the equator on the actual scale for $c_0 = 10^{-2}$ and $u_0 = 10^{-5}$ ($\alpha_{SS} \approx 0.1$) obtained by integrating system (39)–(41). As we see, the accretion-disk thickness actually begins to decrease sharply as the sonic surface is approached.

Importantly, the basic flow parameters can be calculated analytically by analyzing the asymptotics for the system of equations (39)–(41). First, for $t < t_1$,

where $t_1 = u_0^{-1/3}$ and, hence, within small distances

$$(\Delta r)_1 = u_0^{2/3} r_0 \tag{43}$$

from the marginally stable orbit, the perturbation of $Y(t)$, which determines the streamline compression, remains insignificant:

$$Y(t) \approx \frac{u_0}{c_0}\left(1 + \frac{1}{6}u_0 t^3\right). \tag{44}$$

This must be the case, because here the poloidal velocity determined from relation (31),

$$u_\mathrm{p}^2 \approx u_0^2\left(1 + \frac{1}{3}u_0 t^3\right), \tag{45}$$

has no time to change significantly either. Since the contribution from the change in the speed of sound and the contribution given by w^2 are identical, the scale (43) indicated by the arrow in Fig. 1 exactly corresponds to the longitudinal length on which the poloidal velocity of a free particle (32) is constant. Therefore, it comes as no surprise that the disk thickness also remains constant within this scale.

Next, for $t > t_1$, where the quantity $1 - Z$ on the right-hand side of Eq. (40) may still be disregarded, Eqs. (39) and (40) reduce to the Airy equation $d^2Y/dt^2 = u_0 Y t$, so the streamlines in this range are compressed rapidly:

$$Y \approx \frac{u_0}{c_0}\frac{1}{u_0^{1/12} t^{1/4}} \exp\left(\frac{2}{3}u_0^{1/2} t^{3/2}\right). \tag{46}$$

Here, the contribution of w^2 may already be ignored. Finally, when approaching the sonic surface where $1 - Z$ is of the order of unity, the quantities X and Y and the longitudinal scale δt, which determines the radial derivatives, must also be of the order of unity. This is the reason for our choice of normalization in the definition of X, Y, and t. Of course, Eq. (27) and, hence, Eq. (40) are no longer applicable here. Nevertheless, we can use the solution of this equation to roughly estimate the flow behavior near the sonic surface.

First, Eq. (46) allows the location of the sonic surface to be estimated. We obtain from the condition $Y(t_*) = 1$

$$(\Delta r)_* = \Lambda u_0^{2/3} r_0, \tag{47}$$

where the logarithmic factor

$$\Lambda = \left(\frac{3}{2}\right)^{2/3}\left[\ln\left(\frac{c_0}{u_0}\right)\right]^{2/3} \approx 5\text{–}7 \tag{48}$$

depends weakly on the problem parameters. As we see, the sonic surface must be reached at distances that exceed the scale (43) on which the accretion-disk compression begins by only several times. Using

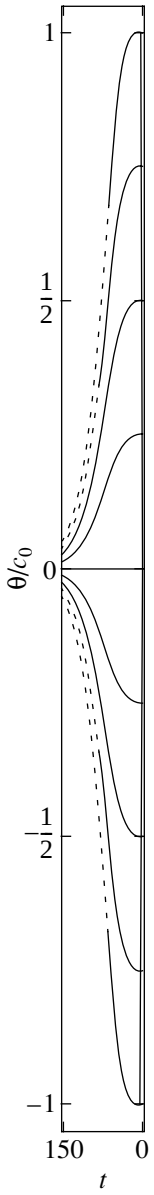


Fig. 2. The pattern of streamlines (actual scale) obtained by numerically solving Eq. (27) for the same parameters as in Fig. 1. The solid lines correspond to $u_\mathrm{p}^2/c_\mathrm{s}^2 < 0.2$. The dashed lines indicate an extrapolation of the solution to the sonic-surface region.

now relation (3), we obtain an estimate for the ratio of the longitudinal and transverse sizes of the transition region,

$$\frac{(\Delta r)_*}{H(r_0)} \sim \alpha_\mathrm{SS}^{2/3} c_0^{1/3}. \tag{49}$$

For the standard $\alpha_\mathrm{SS} \ll 1$, the longitudinal size $(\Delta r)_*$ is always much smaller than the disk thickness $H(r_0)$ in the marginally stable orbit.

Next, as was said above, our analysis of Eqs. (39)–(41) shows that $Y(t)$ becomes of the order of unity as

the sonic surface is approached. Using definitions (6) and (36), we find that, for $r \to r_*$,

$$u_{\hat{\Theta}} \to -\frac{c_0}{u_0}\Theta, \tag{50}$$

and, hence,

$$\frac{u_{\hat{\Theta}}}{r}\frac{\partial u_{\hat{\Theta}}}{\partial \Theta} \to \frac{c_0^2}{u_0^2}\frac{\Theta}{r}. \tag{51}$$

On the other hand, the approximate constancy of the accreted-matter density and the speed of sound along each streamline allows the vertical pressure profile to be estimated as

$$P \approx n(\Theta_m)T_0 \approx n_0 T_0 \exp\left(-\frac{\Theta^2}{\Theta_*^2}\right), \tag{52}$$

where $\Theta_* \sim H_*/r_0 \sim u_0$ is the angular size of the disk near the sonic surface. As would be expected, a sharp compression of the accretion disk causes the transverse component of the pressure gradient to increase,

$$(\boldsymbol{\nabla}_{\hat{\theta}}P)_* \to \frac{c_0^2}{u_0^2}(\boldsymbol{\nabla}_{\hat{\theta}}P)_0, \tag{53}$$

with

$$-\frac{\boldsymbol{\nabla}_{\hat{\theta}}P}{\mu} \to \frac{c_0^2}{u_0^2}\frac{\Theta}{r}. \tag{54}$$

The dynamic term (51) is actually on the same order of magnitude as the pressure gradient. Finally, since $v_{\hat{r}} \to c_0$ near the sonic surface and since the characteristic scale of variation in all quantities is $\delta r \approx u_0 r_0$ (see Fig. 1), we can estimate the other component of the dynamic term as

$$u_{\hat{r}}\frac{\partial u_{\hat{\Theta}}}{\partial r} \to \frac{c_0^2}{u_0^2}\frac{\Theta}{r}. \tag{55}$$

Consequently, this component near the sonic surface is also on the order of the pressure gradient.

All of the basic flow properties obtained from a qualitative analysis—the onset of compression for $(\Delta r)_1 \sim u_0^{2/3} r_0$ (43), the longitudinal scale $(\Delta r)_* \sim \Lambda(\Delta r)_1$ (47), and a sharp disk compression near the sonic surface—are also clearly seen from the results of our numerical solution of Eq. (27), which are shown in Fig. 2. Here, the solid lines correspond to the range of parameters $u_{\rm p}^2/c_{\rm s}^2 < 0, 2$, where the solution to the approximate equation (27) should not differ greatly from the solution to the complete equation (13). Thus, our numerical calculations completely confirm the above qualitative analysis of the basic accretion-flow parameters before passing the sonic surface.

As was said above, analysis of the flow near the sonic surface and in supersonic regime is beyond the scope of this paper. Nevertheless, we make several preliminary remarks concerning the flow properties in the supersonic region. As can be obtained from estimate (50), the slope of the streamlines near the sonic surface

$$\frac{u_{\hat{\Theta}}}{u_{\hat{r}}} \sim -\frac{(c_0/u_0)\Theta}{c_0} \tag{56}$$

for a characteristic disk thickness $\Theta_* \sim u_0$ is close to unity. In other words, a large compression of the accretion disk virtually ceases here. This is also clearly seen from the results of our numerical calculations shown in Figs. 1 and 2. Therefore, we may assume that the disk thickness in the supersonic region (and, in particular, near the BH horizon) will be no less than the disk thickness near the sonic surface: $H(r_{\rm g}) \gtrsim (u_0/c_0)H(r_0)$. We emphasize once again that the latter conclusion is preliminary and it requires additional analysis for its justification.

CONCLUSIONS

Thus, we have shown that after passing the marginally stable orbit, a sharp transverse compression of the accretion flow must take place. As a result, the disk thickness near the sonic surface decreases by several orders of magnitude: $H(r_*) \approx (u_0/c_0)H(3r_{\rm g})$. In fact, the compression is attributable to a sharp increase in the radial velocity and to the fact that the matter density changes only slightly in subsonic regime. Note that such a compression takes place only near a BH and only for transonic flow. In the region of stable orbits, the disk thickness is given, to a good accuracy, by the standard relation (2).

Next, we showed that a sharp compression of the accretion disk must inevitably be accompanied by the appearance of a vertical velocity component, which was generally disregarded. As a result, the dynamic terms $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ in the vertical balance of forces near the sonic surface cannot be discarded. In this sense, the situation is quite similar to a Bondi spherically symmetric flow, in which the contribution of the dynamic term is substantial near the sonic surface and dominant for supersonic flow.

At the same time, it should be emphasized that this case differs significantly from Bondi spherically symmetric accretion. The point is that the dynamic term $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ in the radial flow has only one component, $v_r \partial v_r/\partial r$. This component is of the same order of magnitude as the pressure gradient and as the gradient in effective gravitational potential near the sonic surface. In our case, however, the dynamic term $[(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}]_\theta$ has two components, (51) and (55); they both become equal to the pressure gradient. On the other hand, the effective-potential gradient becomes unimportant here, because it is of the order of $\nabla_\theta(\varphi_{\rm eff}) \sim \Theta/r$, i.e., a factor of u_0^2/c_0^2 smaller than

the leading terms. Therefore, we may assume that in the supersonic region, where the pressure gradient must be of minor importance, the two components of the dynamic force $(\mathbf{v}\boldsymbol{\nabla})\mathbf{v}$ must become dominant in vertical balance. A detailed analysis of this issue is beyond the scope of our paper.

In our view, of crucial importance is also the conclusion that a new small scale, $\delta r \approx u_0 r_0$, arises near the sonic surface, which causes the spatial derivatives to increase sharply. Consequently, additional energy release due to viscous stresses is possible here. However, this question also requires special analysis.

ACKNOWLEDGMENTS

We wish to thank A.V. Gurevich for his interest in our study, as well as for helpful discussions, and support. We are also grateful to R. Narayan and I.D. Novikov for fruitful discussions. This work was supported in part by the Russian Foundation for Basic Research (project nos. 00-15-96594 and 02-02-16762).

APPENDIX

Below, we show how Eq. (27) can be derived directly from the Euler equation. For simplicity, we restrict ourselves to the nonrelativistic case and to the range of small angles $\Theta = \pi/2 - \theta$ near the equatorial plane.

Let us write the θ component of the Euler equation (Landau and Lifshitz 1986) as

$$v_r\frac{\partial v_\theta}{\partial r} + v_\theta\frac{\partial v_\theta}{r\partial\theta} + \frac{v_r v_\theta}{r} - \frac{v_\varphi^2}{r}\cot\theta = -\frac{\boldsymbol{\nabla}_\theta P}{\rho} - \boldsymbol{\nabla}_\theta(\varphi_\mathrm{g}). \quad (57)$$

Here, the last term is written out formally, because $\boldsymbol{\nabla}_\theta(\varphi_\mathrm{g}) = 0$ for any spherically symmetric gravitational potential. Adding $v_\varphi\partial v_\varphi/r\partial\theta$ to both sides and adding and subtracting $v_r\partial v_r/r\partial\theta$ on the left-hand side, we have

$$v_r\frac{\partial v_\theta}{\partial r} - v_r\frac{\partial v_r}{r\partial\theta} + \boldsymbol{\nabla}_\theta\left(\frac{v^2}{2}\right) + \frac{v_r v_\theta}{r} = \frac{v_\varphi^2}{r}\cot\theta + v_\varphi\frac{\partial v_\varphi}{r\partial\theta} - \frac{\boldsymbol{\nabla}_\theta P}{\rho} - \boldsymbol{\nabla}_\theta(\varphi_\mathrm{g}). \quad (58)$$

Using the definition of the Bernoulli integral $E = v^2/2 + W + \varphi_\mathrm{g}$ and the thermodynamic equality $dP = \rho dW - \rho T dS$, we obtain for our case of $E =$ const

$$v_r\frac{\partial v_\theta}{\partial r} - v_r\frac{\partial v_r}{r\partial\theta} + \frac{v_r v_\theta}{r} = \frac{r v_\varphi\sin\theta}{\varpi^2} \times \frac{\partial}{r\partial\theta}(r v_\varphi\sin\theta) + T\frac{\partial}{r\partial\theta}S. \quad (59)$$

Taking into account definitions (6),

$$v_r = \frac{1}{2\pi n r^2\sin\theta}\frac{\partial\Phi}{\partial\theta}, \quad v_\theta = -\frac{1}{2\pi n r\sin\theta}\frac{\partial\Phi}{\partial r}, \quad (60)$$

and for $\rho = \mu n \approx$ const (which exactly corresponds to the subsonic flow), we have

$$-\frac{1}{4\pi^2 n^2}\left(\frac{\partial\Phi}{\partial\theta}\right)\left[\frac{\partial^2\Phi}{\partial r^2} + \frac{\sin\theta}{r^2}\frac{\partial}{\partial\theta}\left(\frac{1}{\sin\theta}\frac{\partial\Phi}{\partial\theta}\right)\right] = r v_\varphi\sin\theta\frac{\partial}{\partial\theta}(r v_\varphi\sin\theta) + \varpi^2 T\frac{\partial}{\partial\theta}S, \quad (61)$$

which after the division by $-(\partial\Phi/\partial\theta)$ leads to Eq. (27).

Thus, we see that, whereas the first term on the left-hand side of Eq. (27) actually corresponds to the component $v_r\partial v_\theta/\partial r$ and the last term on the right-hand side (on $c_\mathrm{s} =$ const) corresponds to the pressure gradient, the role of the term $\propto L\partial L/\partial\Phi$ proves to be less trivial. It contains both the effective-potential gradient $(v_\varphi^2/\varpi)\boldsymbol{\nabla}_\theta\varpi$ and, in fact, the component $v_\theta\partial v_\theta/\partial\theta$. In this case, the effective-potential gradient turns out to be dominant only near the boundary $r \approx 3r_\mathrm{g}$ itself, while the dynamic term $v_\theta\partial v_\theta/\partial\theta$ gives a major contribution as the sonic surface is approached.

REFERENCES

1. M. A. Abramowicz and W. Zurek, Astrophys. J. **246**, 314 (1981).
2. M. A. Abramowicz, B. Czerny, J.-P. Lasota, and E. Szuszkiewicz, Astrophys. J. **332**, 646 (1988).
3. M. A. Abramowicz, A. Lanza, and M. J. Percival, Astrophys. J. **479**, 179 (1997).
4. Yu. V. Artemova, G. S. Bisnovatyi-Kogan, I. V. Igumenshchev, and I. D. Novikov, Astrophys. J. **549**, 1050 (2001).
5. A. M. Beloborodov, Mon. Not. R. Astron. Soc. **297**, 739 (1998).
6. V. S. Beskin, Usp. Fiz. Nauk **167**, 689 (1997) [Phys. Usp. **40**, 659 (1997)].
7. V. S. Beskin and V. I. Par'ev, Usp. Fiz. Nauk **163** (6), 95 (1993) [Phys. Usp. **36**, 529 (1993)].
8. H. Bondi, Mon. Not. R. Astron. Soc. **112**, 195 (1952).
9. S. Chakrabarti, Astrophys. J. **471**, 237 (1996).
10. X. Chen, M. A. Abramowicz, and J.-P. Lasota, Astrophys. J. **476**, 61 (1997).
11. V. P. Frolov and I. D. Novikov, *Black Hole Physics* (Kluwer, Dordrecht, 1998).
12. C. F. Gammie and R. Popham, Astrophys. J. **498**, 313 (1999a).
13. C. F. Gammie and R. Popham, Astrophys. J. **504**, 419 (1998b).
14. I. V. Igumenshchev and A. M. Beloborodov, Mon. Not. R. Astron. Soc. **284**, 767 (1997).

15. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics,* Vol. 2: *The Classical Theory of Fields* (Nauka, Moscow, 1973; Pergamon, Oxford, 1975).
16. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics,* Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
17. V. M. Lipunov, *Astrophysics of Neutron Stars* (Nauka, Moscow, 1987).
18. D. Lynden-Bell, Nature **223**, 690 (1969).
19. R. Narayan, S. Kato, and F. Honma, Astrophys. J. **476**, 49 (1997).
20. I. D. Novikov and K. S. Thorne, in *Black Holes,* Ed. by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1973), p. 343.
21. B. Paczynski and G. S. Bisnovatyi-Kogan, Acta Astron. **31**, 283 (1981).
22. B. Paczynski and P. J. Wiita, Astron. Astrophys. **88**, 23 (1980).
23. J. Papaloizou and E. Szuszkiewicz, Mon. Not. R. Astron. Soc. **268**, 29 (1994).
24. J. Peitz and S. Appl, Mon. Not. R. Astron. Soc. **286**, 681 (1997).
25. H. Riffert and H. Herold, Astrophys. J. **450**, 508 (1995).
26. N. I. Shakura, Astron. Zh. **49**, 921 (1972) [Sov. Astron. **16**, 756 (1972)].
27. N. I. Shakura and R. A. Sunyaev, Astron. Astrophys. **24**, 337 (1973).
28. S. L. Shapiro and S. A. Teukolsky, *Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects* (Wiley, New York, 1983; Mir, Moscow, 1985), Vol. 2.
29. K. S. Thorne, R. H. Price, and D. A. Macdonald, *Black Holes: The Membrane Paradigm* (Yale Univ. Press, New Haven, 1986; Mir, Moscow, 1988).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 553–559. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 626–632.

Electron–Positron and Photon Wind from the Surface of a Strange Star

A. G. Aksenov*

Institute for Theoretical and Experimental Physics,
ul. Bol'shaya Cheremushkinskaya 25, Moscow, 117259 Russia

Received March 23, 2002

Abstract—We consider the problem of strange-star (SS) radiation. The bare quark SS surface and electrons on the stellar surface generate an electric field that is strong enough for electron–positron pairs to be produced from a vacuum at a nonzero temperature. The luminosity in pairs is assumed to be within $\lesssim 10^{49}$ erg s^{-1} from a surface with a characteristic radius of 10 km. We consider the energy transfer from pairs to photons by taking into account the well-studied reactions between e, e^+, γ and obtain a change in the photon spectrum with luminosity. Our analysis is restricted to the spherically symmetric case. The magnetic field is disregarded. To solve the problem, we developed a new numerical method of integrating the Boltzmann kinetic equations for pairs and photons. This method is used to calculate the problem up to a luminosity of 10^{42} erg s^{-1}. This region is difficult to investigate when the optical path for pairs or photons is considerably larger than unity but the two optical depths are not simultaneously much larger than unity (when hydrodynamics with heat conduction is applicable). It turns out that the mean photon energy is approximately equal to $\bar{\epsilon}_\gamma \approx m_e c^2$ (the annihilation line for pairs) at a modest luminosity, $L \lesssim 1 \times 10^{37}$ erg s^{-1}, and decreases to $\sim$210 keV at $L \gtrsim 10^{38}$ erg s^{-1}. Hydrodynamic estimates point to an increase in the mean energy $\bar{\epsilon}_\gamma$ to 1 MeV as the luminosity further increases to $L \gtrsim 10^{49}$ erg s^{-1}. Our calculations may prove to be useful in interpreting soft gamma repeaters (SGRs) and are of methodological interest. © *2002 MAIK "Nauka/Interperiodica".*

Key words: *neutron stars, γ-ray radiation, radiation mechanism*

INTRODUCTION

The method of distinguishing strange stars (SSs) from neutron stars (NSs) still remains to be found, as does the answer to the question of whether SSs made of free quarks (Glendennig 1996) actually exist in the Universe. Despite their different properties (for example, different equations of state, different mass–radius relations), SSs and NSs have similar external parameters: masses, radii, angular momenta, and periods for typical masses $\sim M_\odot$ (Haensel *et al.* 1986; Zdunik and Haensel 1990; Øvergård and Østgaard 1991; Colpi and Miller 1992). Thus, the question of what compact object is observed as a pulsar is nontrivial.

In this paper, we investigate the photon radiation from a hot SS so as to compare it subsequently, for example, with the spectrum of Soft Gamma Repeaters (SGRs). Below, we give well-known SS properties. The characteristic scale of strong interaction is 1 Fm. The quark surface thickness is on the order of this size. At the same time, electrons are also distributed outside the quark surface at a distance of $\sim 10^3$ Fm. Estimates yield an outward-directed electric field of $\gtrsim 10^{16}$ V cm^{-1} (Alcock *et al.* 1986). This field exceeds the critical field $E_{\rm cr} = 1.3 \times 10^{15}$ V cm^{-1} and is strong enough to produce e, e^+ pairs from a vacuum (Schwinger 1951). In this case, it is necessary to overcome the potential barrier and the final states of the pairs must be free, i.e., the pair production is possible at a temperature $T > 0$. A nonzero temperature can be achieved, for example, during mass accretion onto an SS; SGRs are possible and observable. The temperature dependence of a steady flux of pairs, as well as for the time scale of the emergence and continuation of radiation during short SS heating after the infall of matter, was calculated by Usov (2001). It turns out that the radiation in pairs on the stellar surface is much stronger than the photon radiation at $T \lesssim 10^{10}$ K, while at $T > 10^{11}$ K the blackbody photon radiation becomes dominant. This is a result of the dispersion relation for a SS: radiation below a certain frequency does not propagate in the star. Even if the radiation in pairs dominates, the luminosity is significant.

We have the following problem. We specify a steady energy flux in electron–positron pairs on the SS surface. This is a function of the surface temperature. The total intensity of the radiation is also known, because the SS radius is known. It is of interest to

*E-mail: alexei.aksenov@itep.ru

find out what this flux is transformed into far from the stellar surface, i.e., we must calculate the transfer of $e^{\mp}$, γ above the star for various luminosities, and to obtain the spectral photon light curves. At low luminosities, a small number of pairs are formed. The photons produced by annihilation will escape without interaction with a high energy, $\sim m_e c^2 = 511$ keV. At the same time, for a high luminosity, the particle flux may prove to be significant and the energy exchange between particles is possible until the particle flux becomes small after expansion far from the stellar surface. In this case, the photon energy may decrease. The question is how large this decrease is and at what luminosities. Can this be consistent with observations of SGRs with an inverse energy dependence of the spectrum?

The reaction rates for pairs and photons (Compton scattering, annihilation, scattering of $e^{\mp}$ between themselves and by one another, bremsstrahlung) are well known (Berestetskiĭ *et al.* 1982). We set the stellar radius, which weakly depends on the equation of state for typical masses $\sim M_{\odot}$, equal to 10 km. The only parameter to be varied is the luminosity. The need to perform simultaneous calculations in optically thick (for pairs) and optically thin (for photons) spatial regions poses the greatest difficulty. Statistical modeling (Pilla and Shaham 1997; Beloborodov 1999), which is universal for including the reactions under consideration, is hardly applicable to our case of arbitrary optical depths. Unfortunately, we failed to find an appropriate method in the literature and it had to be developed to solve the problem. This is the principal difficulty of the proposed direct approach to investigating SSs.

FORMULATION OF THE PROBLEM

We solve the Boltzmann kinetic equations for the distribution functions $f_\nu(\epsilon_\nu, \mu, r, t)$ of a ν particle ($\nu = \gamma$ or e, $f_{\mp} = f_e/2$) for the spherically-symmetric case in special relativity in the laboratory frame of reference (Mihalas 1984),

$$\frac{\partial f_\nu}{\partial t} + c\beta_\nu \left(\mu \frac{\partial f_\nu}{\partial r} + \frac{1-\mu^2}{r}\frac{\partial f_\nu}{\partial \mu}\right) = \eta_\nu^0 - \chi_\nu f_\nu,$$

until a steady-state solution is obtained. Here, r is the radius and μ is the cosine of the angle between the radius vector and the particle momentum $\beta_\nu = \sqrt{1-(m_\nu c^2/\epsilon_\nu)^2}$ ($\beta_\gamma = 1$ for a photon). The general-relativity effects are disregarded; η_ν^0 and χ_ν are the absorption and emission coefficients, respectively. We consider two-particle processes with conservation of the number of particles: the Compton scattering of γ by $e^{\mp}$, γ-annihilation, and the scattering of $e^{\mp}$ by one another. From multiparticle processes (at the exit of a 3-particle), we qualitatively take into account bremsstrahlung. In our calculations, it is more convenient to use the transfer equations for the spectral energy density $E_\nu \equiv \frac{4\pi g_\nu}{(2\pi\hbar c)^3}\epsilon_\nu^3 \beta_\nu f_\nu \equiv \frac{4\pi}{\beta_\nu c} I_\nu$ ($g_\nu = 2$ is the statistical weight, I_ν is the spectral intensity of the radiation from a ν particle):

$$\frac{1}{c}\frac{\partial E_\nu}{\partial t} + \mu\frac{\partial r^2 \mathcal{I}}{\partial r^3/3} + \frac{1}{r}\frac{\partial}{\partial \mu}\left[(1-\mu^2)\mathcal{I}\right] = \frac{4\pi g_\nu(\eta_\nu^0 - \chi_\nu f_\nu)}{(2\pi\hbar c)^3 c}\epsilon_\nu^3 \beta_\nu = \eta_\nu - \chi_\nu E_\nu,$$

where $\mathcal{I} = \beta_\nu E_\nu = \frac{4\pi}{c} I_\nu$, $\eta_\nu = \frac{4\pi g_\nu \eta_\nu^0}{(2\pi\hbar c)^3 c}$. Using the energy densities in the conservative scheme instead of the occupation numbers allows us to consider highly opaque regions without loss of the approximation.

In equilibrium,

$$f_\nu^{\mathrm{eq}} = \frac{1}{1 \mp \exp\frac{\epsilon_\nu - \mu_\nu}{kT}}, \qquad (1)$$

where μ_ν are the chemical potentials and T is the temperature.

The mean photon energy $\bar{\epsilon}_\nu \equiv \int d\epsilon_\nu do / \int d\epsilon do \times E_\nu \frac{E_\nu}{\epsilon_\nu}$ as a function of the power $L_e = 2\pi R^2 \times \int d\epsilon d\mu \mu I_e$ at the outer boundary is an important parameter of the solution (o is the solid angle).

Boundary Conditions

The inner boundary is a reflecting wall at $r = R$ for $\mathcal{I}$ (the same quantity is added to the computational region and leaves it only at the mirror angle); for pairs, such a value of $\mathcal{I}_e$ is added so that $\int d\epsilon_e \frac{d\mu}{2} c\mu \mathcal{I}_e(r_s, \mu, \epsilon, t) \equiv \int d\epsilon_e do \mu I_e = \epsilon_{\mp}\dot{n}_{\mp}$, where $\epsilon_{\mp} = m_e c^2 + kT$, and the pair production rate is

$$\dot{n}_{\mp} = 10^{39}\left(\frac{T}{10^9\ \mathrm{K}}\right)^3 \exp\left[-11.9\left(\frac{T}{10^9\ \mathrm{K}}\right)^{-1}\right] \times \left(\frac{1}{3}\frac{\zeta^3 \ln(1+2\zeta^{-1})}{(1+0.074\zeta)^3} + \frac{\pi^5}{6}\frac{\zeta^4}{(13.9+\zeta)^4}\right)\ \mathrm{s}^{-1},$$

where $\zeta = 2\sqrt{\frac{e^2}{\pi\hbar c}\frac{\epsilon_{\mathrm{F}}}{kT}}$, $\epsilon_{\mathrm{F}} = \hbar c(\pi^2 n_e)^{1/3} \approx 18$ MeV, $R = 10^6$ cm, and $T \sim m_e c^2/k$ (Usov 1998, 2001). For $T_s = m_e c^2/k$, the luminosity in pairs is $L_{\mp} = 8.9 \times 10^{48}$ erg s^{-1}. The luminosity is varied. At the inner boundary, e, e^+ with the equilibrium spectrum (1)

and with a zero chemical potential μ_e are added to the computational region. The choice of a spectrum should not affect the solution in the optically thick case where the details of the initial distribution f_e are forgotten. Nor does this choice strongly affect the photon spectrum in the optically thin case where the annihilation line is mainly present in the spectrum. At the outer boundary, there is no particle flux into the computational region: $\mathcal{I}_e(r_s,\,\mu,\,\epsilon,\,t)=0$ for $\mu<0$.

THE NUMERICAL METHOD

We introduce a grid $r_{j+1/2}$, $\epsilon_{\nu,\,\omega+1/2}$, $\mu_{k+1/2}$. The derivatives with respect to space, angle, and energy are substituted with finite differences. For the spectral energy densities

$$\Delta E_{\nu,\omega,k,j}=\frac{2}{\Delta\mu_k}\int_{\Delta\epsilon_\omega,\Delta\mu_k}d\epsilon d\mu\frac{E_\nu}{2}$$
$$=\frac{2}{\Delta\mu_k}\int_{\Delta\epsilon_\omega,\Delta\mu_k}d\epsilon do\frac{g_\nu\epsilon^3\beta_\nu f_\nu}{(2\pi\hbar c)^3}$$

(we use the notation $\Delta r_j^3=r_{j+1/2}^3-r_{j-1/2}^3$, $\Delta\epsilon_{\nu,\omega}=\epsilon_{\nu,\omega+1/2}-\epsilon_{\nu,\omega-1/2}$, $\Delta\mu_k=\mu_{k+1/2}-\mu_{k-1/2}$ and the brief designations $\Delta\epsilon_\omega$, $\Delta\mu_k$ to indicate the integration limits under the integral instead of $\epsilon\in(\epsilon_{\omega-1/2},\epsilon_{\omega+1/2})$, $\mu\in(\mu_{k-1/2},\mu_{k+1/2})$); we obtain a system of ordinary differential equations (ODEs)

$$\frac{1}{c}\frac{\partial\Delta E_{\nu,\omega,k,j}}{\partial t}+\frac{\Delta(r^2\bar\mu_k\Delta\mathcal{I}_{\nu,\omega,k})_j}{\Delta V_j}\qquad(2)$$
$$+\left(\frac{\bar1}{r}\right)_j\frac{\Delta\left[(1-\mu^2)\Delta\mathcal{I}_{\nu,\omega,j}\right]_k}{\Delta\mu_k}$$
$$=\Delta\eta_{\nu,\omega,k,j}-\Delta(\chi E)_{\nu,\omega,k,j},$$

where $\Delta\eta_{\nu,\omega,k}=\int_{\Delta\epsilon_\omega,\Delta\mu_k}\frac{d\epsilon do}{4\pi}\eta_\nu$, $\Delta(\chi E)_{\nu,\omega,k}=\int_{\Delta\epsilon_\omega,\Delta\mu_k}\frac{d\epsilon do}{4\pi}\chi_\nu E_\nu$, $\Delta\mathcal{I}_{\nu,\omega,k,j}=\Delta E_{\nu,\omega,k,j}\beta_{\nu,\omega,j}=\Delta E_{\nu,\omega,k,j}\sqrt{1-\frac{\epsilon_{\nu,0}^2}{\epsilon_{\nu,\omega}^2}}$, $\epsilon_{\nu,\omega}=\frac{\epsilon_{\nu,\omega-1/2}+\epsilon_{\nu,\omega+1/2}}{2}$,

$$(\bar\mu_k\Delta\mathcal{I}_{\nu,\omega,k})_{j+\frac12}=(1-\tilde\chi_{\nu,\omega,k,j+\frac12})$$
$$\times\left(\frac{\bar\mu_k+|\bar\mu_k|}{2}\Delta\mathcal{I}_{\nu,\omega,k,j}+\frac{\bar\mu_k-|\bar\mu_k|}{2}\Delta\mathcal{I}_{\nu,\omega,k,j+1}\right)$$
$$+\tilde\chi_{\nu,\omega,k,j+\frac12}\bar\mu_k\frac{\Delta\mathcal{I}_{\nu,\omega,k,j}+\Delta\mathcal{I}_{\nu,\omega,k,j+1}}{2},$$

$$\tilde\chi_{\nu,\omega,k,j+\frac12}=\frac{\frac{\chi_{\nu,\omega,k,j}\Delta r_j\chi_{\nu,\omega,k,j+1}\Delta r_{j+1}}{\chi_{\nu,\omega,k,j}\Delta r_j+\chi_{\nu,\omega,k,j+1}\Delta r_{j+1}}}{1+\frac{\chi_{\nu,\omega,k,j}\Delta r_j\chi_{\nu,\omega,k,j+1}\Delta r_{j+1}}{\chi_{\nu,\omega,k,j}\Delta r_j+\chi_{\nu,\omega,k,j+1}\Delta r_{j+1}}},$$

$$\Delta\mathcal{I}_{\nu,k+\frac12,\omega,j}=\Delta\mathcal{I}_{\nu,\omega,k,j}$$
$$+\frac{\Delta\mu_k(\Delta\mathcal{I}_{\nu,\omega,k,j}-\Delta\mathcal{I}_{\nu,k-1,\omega,j})}{\Delta\mu_k+\Delta\mu_{k-1}}.$$

The coefficients $\tilde\chi_{\nu,\omega,k,j+\frac12}$ are defined in such a way that in an opaque region the differential equations are approximated by the central difference with the second order of accuracy necessary for the correct passage to the heat conduction equations, while in an optically transparent region the approximation is reduced to the first order to avoid false numerical oscillations [such an approach was used by Mezzacappa and Bruenn (1993) and Aksenov (1998) to solve the hydrodynamic equations together with neutrino kinetics when analyzing collapse].

To simplify our calculations, we averaged the emission and absorption coefficients η_ν and χ_ν over their angle.

The derived system of ODEs (2) can be solved by the Gear method (Hall and Watt 1976), which is efficient when integrating a stiff system of ODEs (distinctly different real parts of the eigenvalues of the Jacobian).

As an example, below we give the absorption and emission coefficients for the Compton scattering of a photon by an electron. The evolution of the photon and electron distribution functions in this process is (Ochelkov *et al.* 1979)

$$\left(\frac{\partial f_{\mathbf{p}}}{\partial t}\right)_{\rm Cs}=\int\frac{d\mathbf{k}d\mathbf{p}'d\mathbf{k}'}{(2\pi\hbar)^3}w_{\mathbf{p}',\mathbf{k}';\mathbf{p},\mathbf{k}}(f_{\mathbf{k}'}f_{\mathbf{p}'}-f_{\mathbf{k}}f_{\mathbf{p}}),$$
$$\left(\frac{\partial f_{\mathbf{k}}}{\partial t}\right)_{\rm Cs}=\int\frac{d\mathbf{p}d\mathbf{p}'d\mathbf{k}'}{(2\pi\hbar)^3}w_{\mathbf{p}',\mathbf{k}';\mathbf{p},\mathbf{k}}(f_{\mathbf{k}'}f_{\mathbf{p}'}-f_{\mathbf{k}}f_{\mathbf{p}}),$$

where (Berestetskiĭ *et al.* 1982) the probability of the process is

$$w_{\mathbf{p}',\mathbf{k}';\mathbf{p},\mathbf{k}}=\frac{c\delta(\epsilon_e+\epsilon_\gamma-\epsilon_e'-\epsilon_\gamma')}{(2\pi\hbar)^2}$$
$$\times\delta(\mathbf{p}+\mathbf{k}-\mathbf{p}'-\mathbf{k}')\frac{|M_{fi}^\gamma|^2}{4\epsilon_e\epsilon_\gamma\cdot4\epsilon_e'\epsilon_\gamma'},$$

and the square of the matrix element $|M_{fi}|^2$ can be expressed in terms of the electron radius $r_e=\frac{e^2}{m_ec^2}$ and the invariants $s=(\mathfrak{p}+\mathfrak{k})^2=(\mathfrak{p}'+\mathfrak{k}')^2$, $t=(\mathfrak{p}-\mathfrak{p}')^2=(\mathfrak{k}'-\mathfrak{k})^2$, $u=(\mathfrak{p}-\mathfrak{k}')^2=(\mathfrak{p}'-\mathfrak{k})^2$, $(s+t+u=2m_e^2c^2)$, which are functions of the 4-vectors of energy–momentum for electrons and photons $\mathfrak{p}=(\epsilon_e/c)(1,\beta\mathbf{e}_e)$, $\mathfrak{k}=(\epsilon_\gamma/c)(1,\mathbf{e}_\gamma)$, and $d\mathbf{p}=(\epsilon^2\beta_\nu)/(c^3)d\epsilon d\mu d\phi$. Integration leads to the

following expressions for the coefficients (in the difference-scheme notation):

$$\Delta\eta^{\rm Cs}_{e,\omega,k}=\frac{2}{\Delta\mu_k}\int_{\substack{\epsilon'_e\in\Delta\epsilon_{e,\omega}\\ \mu'_e\in\Delta\mu_k}} d\epsilon_e do_e d\epsilon_\gamma do_\gamma do'_\gamma \epsilon'_e \times\frac{\epsilon'^2_\gamma\dfrac{|M^\gamma_{fi}|^2}{(2\pi\hbar)^2 2c^3}}{4\epsilon_e\epsilon_\gamma\cdot 4\epsilon'_e\epsilon'_\gamma}\frac{g_e\epsilon_e^2\beta_e f_e}{(2\pi\hbar c)^3}\frac{g_\gamma\epsilon_\gamma^2 f_\gamma}{(2\pi\hbar c)^3},$$

$$\Delta(\chi E)^{\rm Cs}_{e,\omega,k}=\frac{2}{\Delta\mu_k}\int_{\substack{\epsilon_e\in\Delta\epsilon_{e,\omega}\\ \mu_e\in\Delta\mu_k}} d\epsilon_e do_e d\epsilon_\gamma do_\gamma do'_\gamma \epsilon_e \times\frac{\epsilon'^2_\gamma\dfrac{|M^\gamma_{fi}|^2}{(2\pi\hbar)^2 2c^3}}{4\epsilon_e\epsilon_\gamma\cdot 4\epsilon'_e\epsilon'_\gamma}\frac{g_e\epsilon_e^2\beta_e f_e}{(2\pi\hbar c)^3}\frac{g_\gamma\epsilon_\gamma^2 f_\gamma}{(2\pi\hbar c)^3},$$

$$\Delta\eta^{\rm Cs}_{\gamma,\omega,k}=\frac{2}{\Delta\mu_k}\int_{\substack{\epsilon'_\gamma\in\Delta\epsilon_{\gamma,\omega}\\ \mu'_\gamma\in\Delta\mu_k}} d\epsilon_e do_e d\epsilon_\gamma do_\gamma do'_\gamma \epsilon'_\gamma \times\frac{\epsilon'^2_\gamma\dfrac{|M^\gamma_{fi}|^2}{(2\pi\hbar)^2 2c^3}}{4\epsilon_e\epsilon_\gamma\cdot 4\epsilon'_e\epsilon'_\gamma}\frac{g_e\epsilon_e^2\beta_e f_e}{(2\pi\hbar c)^3}\frac{g_\gamma\epsilon_\gamma^2 f_\gamma}{(2\pi\hbar c)^3},$$

$$\Delta(\chi E)^{\rm Cs}_{\gamma,\omega,k}=\frac{2}{\Delta\mu_k}\int_{\substack{\epsilon_\gamma\in\Delta\epsilon_{\gamma,\omega}\\ \mu_\gamma\in\Delta\mu_k}} d\epsilon_e do_e d\epsilon_\gamma do_\gamma do'_\gamma \epsilon_\gamma \times\frac{\epsilon'^2_\gamma\dfrac{|M^\gamma_{fi}|^2}{(2\pi\hbar)^2 2c^3}}{4\epsilon_e\epsilon_\gamma\cdot 4\epsilon'_e\epsilon'_\gamma}\frac{g_e\epsilon_e^2\beta_e f_e}{(2\pi\hbar c)^3}\frac{g_\gamma\epsilon_\gamma^2 f_\gamma}{(2\pi\hbar c)^3},$$

where $\epsilon'_\gamma=\dfrac{\epsilon_e\epsilon_\gamma(1-\beta\mathbf{e}_e\mathbf{e}_\gamma)}{\epsilon_e(1-\beta\mathbf{e}_e\mathbf{e}'_\gamma)+\epsilon_\gamma(1-\mathbf{e}_\gamma\mathbf{e}'_\gamma)}$, $\epsilon'_e=\epsilon_e+\epsilon_\gamma-\epsilon'_\gamma$. The goal of this transformation is to ensure that the energy conservation law $\sum_{\nu,\omega,k,}\dfrac{\mu_k}{2}\Delta\epsilon_{\nu,k}\times\Delta E_{\nu,\omega,k,j}=0$ holds with computer accuracy, which is important in calculating optically thick regions. Our calculations show that the conservation of the number of particles in an opaque region is also important. This property of the scheme can be achieved by averaging the absorption and emission coefficients over the neighboring energy intervals.

The remaining absorption and emission coefficients are obtained in a similar way. We did not use detailed calculations for bremsstrahlung. First, this is too time-consuming, because three particles are produced at the exit. Second, this effect is important only in the opaque case for photons and pairs, i.e., when the distribution functions are close to equilibrium ones. Bremsstrahlung was taken into account qualitatively by using the spectral power of the energy release for e, e^+ scattering in an equilibrium plasma from Haug (1985) with a temperature that ensures equality of the energy of the equilibrium distribution function to the energy of electrons and positrons.

RESULTS OF THE SOLUTION

The problem was computed for various luminosities in electron–positron pairs specified at the inner boundary until steady-state solutions were obtained. We used nonuniform grids of 15 intervals in energies ($0\le\epsilon_\nu-m_\nu c^2\lesssim m_e c^2$), 4 intervals in angles ($-1\le\mu<1$), and 110 intervals in radius ($0<r-R\lesssim 10^9$ cm). The CPU time is mostly spent computing the Jacobian and can be estimated as $\propto j_{\max}(\omega_{\max}k_{\max})^5$. This accounts for the small number of intervals in energies and angles. However, our calculations indicate that the solution is weakly sensitive to the number of chosen intervals.

The integrated parameters of the solution (mean energies) and the values of some quantities at the outer boundary are given in the table and shown in the figure. At the inner boundary, we specify the energy and particle flux transported by electron–positron pairs alone. Pair annihilation produces photons, which interact with pairs. At the boundary of the computational region, which is sufficiently far from the star, this interaction becomes weak, which leads to energies that no longer depend on radius.

For low luminosities, $L\lesssim 10^{37}$ erg s^{-1}, the photons produced by pair annihilation escape from the computational region without interacting with $e^\mp$. Because of the low number density of pairs, the photon optical path is $\tau_\gamma<1$. The bulk of the energy flux is carried away by the noninteracting pairs, and the mean photon energy is close to the electron rest mass $\bar\epsilon_\gamma\approx m_e c^2=511$ keV. A small difference between the photon energy and this value (table) is attributable to a nonzero electron energy. As the luminosity increases to $L\gtrsim 10^{38}$ erg s^{-1}, the electron number density becomes high enough to change the photon spectra; the photon energy falls below $\bar\epsilon_\gamma\lesssim 210$ keV at a luminosity $L\gtrsim 10^{39}$ erg s^{-1}. In this case, most of the energy is carried away by photons. We can write out an approximation formula for the mean photon energy,

$$\bar\epsilon_\gamma=\frac{490\ \text{keV}-210\quad\text{keV}}{1+(L/2.5\times10^{37}\quad\text{erg s}^{-1})^{0.9}}+210\quad\text{keV},$$

for the total luminosity (in pairs and photons) of $10^{34}\lesssim L\lesssim 10^{42}$ erg s^{-1}, whose accuracy is seen from the figure. It should be noted that a thermodynamic equilibrium at a high luminosity is not yet achieved, because the mean energy ($\approx 3kT$) for such a luminosity is lower than the values obtained.

Mean particle energies $\bar{\epsilon}_\nu = \int f_\nu \epsilon^3 d\epsilon do / \int f_\nu \epsilon^2 d\epsilon do$ and optical paths $\tau_\nu = \int_0^{r_{\max}} dr \left[\int d\epsilon\ \chi_\nu E_\nu / \int d\epsilon E_\nu\right]$ versus luminosity

L, erg s^{-1}	L_e, erg s^{-1}	$\bar{\epsilon}_\gamma$, 10^2 keV	$\bar{\epsilon}_e - m_e c^2$, 10^2 keV	τ_γ	τ_e
1×10^{34}	8.6×10^{33}	5.0	0.18	1.5×10^{-2}	3.4×10^{-2}
1×10^{35}	2.4×10^{34}	4.8	0.48	7.4×10^{-2}	2.3
1×10^{36}	9.3×10^{34}	4.6	2.1	0.15	5.9
1×10^{37}	3.7×10^{35}	4.1	2.3	0.54	1.9×10^{1}
1×10^{38}	1.0×10^{36}	2.7	1.9	3.9	2.4×10^{2}
1×10^{39}	2.9×10^{36}	2.1	1.5	81	1.3×10^{4}
1×10^{40}	$\approx 2.3 \times 10^{37}$	2.1	≈ 1.5	5.8×10^{2}	9.8×10^{4}
1×10^{41}	$\approx 2.2 \times 10^{38}$	2.1	≈ 1.5	9.8×10^{2}	1.7×10^{5}
1×10^{42}	$\approx 5.3 \times 10^{39}$	≈ 2.1	≈ 1.5	1.3×10^{3}	2.4×10^{5}

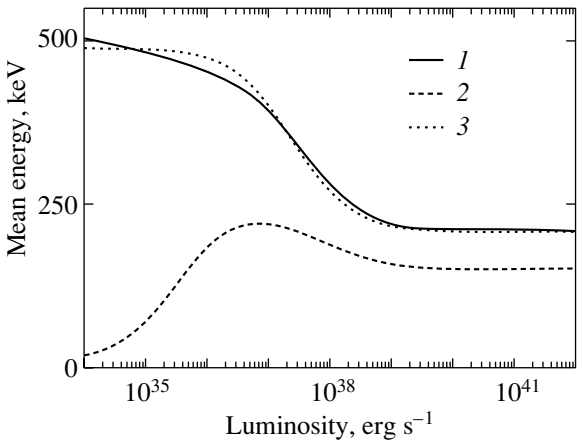
Note: For $L \geq 1 \times 10^{40}$ erg s^{-1} in the last computational zone $\tau_e > 1$, the quantities L_e, $\bar{\epsilon}_\nu$ were determined approximately.

The effect of an increase in the particle flux during bremsstrahlung begins to show up only at luminosities $L > 10^{38}$ erg s^{-1}. Although the process is weak compared to two-particle reactions, the effect is qualitatively important. The local parameters of the matter will be functions of not only the temperature but also of the total particle number density only when two-particle reactions are admitted in a thermodynamic equilibrium. Quantitatively, the additional formation of photons during bremsstrahlung causes the mean photon energy to decrease by a factor of almost 1.5.

As for the mean electron energy, it has a small local maximum $\bar{\epsilon}_e(L) \approx 230$ keV at luminosity $L \approx 10^{37}$ erg s^{-1} (the photon optical depth is $\tau_\gamma \gtrsim 1$). The pairs take energy away from the photons during Compton scattering. At high luminosities, $\bar{\epsilon}_e(L)$ decreases with $\bar{\epsilon}_\gamma(L)$ and remains smaller than the mean photon energy.

In our calculations, we can note the presence of an optically thick region, $L \gtrsim 10^{38}$ erg s^{-1}, both for pairs and for photons. However, only at luminosities above $L \gtrsim 10^{43}$ erg s^{-1} does the equilibrium blackbody radiation with mean energy $3kT_{\rm bb}$ become close to the results of our calculations. In this case, the photons may be considered to be in a local thermodynamic equilibrium with pairs and the hydrodynamic approximation with heat conduction is applicable (Paczyński 1990). For this case, there are estimates of the increase in the mean photon energy (Usov 1994; Lyutikov and Usov 2000) from several hundred keV to 1 MeV at luminosities $L \gtrsim 10^{47}$ erg s^{-1} due to a significant increase in the Lorentz factor of pairs, $\sim$10, at the photosphere. In these estimates, apart from local thermodynamic equilibrium, the characteristic temperature at the photosphere is used, $\sim 10^{8.5}$ K. Unfortunately, we failed to verify such suggestions by this method. First, the optical path becomes too large for this method. Second, to take into account large Lorentz factors, we must use a different energy grid: the range $\epsilon \gtrsim m_e c^2$ must be considered.

To estimate the effect of ignoring the gravitational acceleration in the Boltzmann equations for pairs, we should compare this acceleration, GMm_e/R^2, with the fluid pressure gradient $(m_e/\rho)(\partial P/\partial r) \sim kT_e(1 + P_\gamma/P_e)/\Delta r$, where Δr is the characteristic scale of pressure variation. If we take $\Delta r = R$ and



Mean energies $\bar{\epsilon}_\nu - m_\nu c^2 = \int f_\nu \epsilon^3 d\epsilon do / \int f_\nu \epsilon^2 d\epsilon do - m_\nu c^2$ of photons (*1*) and electrons (*2*) at the outer boundary of the computational region versus luminosity L; (*3*) the approximation of the photon energy with the function $\bar{\epsilon}_\gamma = (490 - 210\ \text{keV})/(1 + (L/2.5 \times 10^{37}\ \text{erg s}^{-1})^{0.9}) + 210$ keV.

$T_e \approx 100$ keV, $GMm_e/R = 1.7 \times 10^{-7}$ erg, $kT = 1.4 \times 10^{-16}$ erg, then we find it necessary to take into account the gravity for low luminosities, $L \lesssim 10^{37}$ erg s^{-1}. Further, for typical SS masses $M = 1.4M_\odot$, the gravitational radius is $r_{\rm gr} = 2GM/c^2 = 4.1$ km. It is close to the surface radius $R = 10$ km. General relativity affects the solution near the stellar surface at $r \sim 10$ km. The characteristic radius of the photosphere is found to be on the order of several stellar radii. The general-relativity factor $\Gamma = \sqrt{1 + (v/c)^2 - r/r_{\rm gr}}$ differs from unity by several tens of percent. This apparently causes a decrease in the mean energies. Unfortunately, in the case under consideration, it is difficult to introduce comoving coordinates, because the inner boundary is fixed. Such an approach can be used for a hydrodynamic analysis with heat conduction (Lindquist 1966).

CONCLUSIONS

(1) The main result of our calculations is the derivation of the luminosity dependence of the mean photon energy, $\bar{\epsilon}_\gamma(L)$ (see the figure). Near the point with the optical path $\tau_\gamma = 1$ ($L \approx 1 \times 10^{37}$ erg s^{-1}), the situation is qualitatively different. The radiation spectrum is softer at high luminosities, $L \gtrsim 1 \times 10^{38}$ erg s^{-1}, and harder at low luminosities, $L \lesssim 1 \times 10^{37}$ erg s^{-1}. At significant luminosities $L \gtrsim 10^{47}$ erg s^{-1} possible for SSs, our hydrodynamic estimates point to an increase in the mean energy $\epsilon_\gamma(L)$ to 1 MeV for $L \sim 10^{49}$ erg s^{-1}. Our calculations reveal a minimum of the mean energy $\bar{\epsilon}_\gamma \lesssim 200$ keV in the luminosity range $10^{42} \lesssim L \lesssim 10^{47}$ erg s^{-1}.

(2) Taking into consideration our calculations, we can attempt to explain the existence of abnormal SGRs with a high super-Eddington luminosity, for example, $\gtrsim 10^{38}$–10^{42} erg s^{-1}, and with a soft photon spectrum (Fenimore *et al.* 1996; Mazets *et al.* 1999) compared to normal SGRs. Clearly, an SS can emit photons with the required intensity with a decrease in the mean energy as L increases. Our calculations yielded a small decrease in the mean photon energy to $\lesssim$200 keV with increasing luminosity. This may be the result of ignoring various three-particle reactions. Therefore, it is still too early to conclude which radiation mechanism is simpler: an SS or a normal NS but with a strong magnetic and anisotropic radiation?

(3) In our calculations, we neglected the ordinary matter, which can be above the quark surface. The density of this matter is below the formation point of free neutrons, which freely penetrate into the star through a strong electric field: $\rho < 4 \times 10^{11}$ g cm^{-3}, and the mass of the matter is $\sim 10^{-5} M_\odot$. This neglect is most likely admissible for a high luminosity, $\gtrsim 10^{38}$ erg s^{-1}. The predicted disappearance of the annihilation line, which is of greatest interest in observations, occurs at the same or slightly lower luminosity, and the problem of taking into account the SS atmosphere arises. To determine whether there is an annihilation line in the radiation, we must also estimate the SS magnetic field, in which photons with energies different than 511 keV can be produced by pair annihilation.

(4) We presented a new numerical method that allows us to solve a wide range of evolutionary problems on the transport of electron–positron plasma to large optical depths. A weak point of the method—stringent constraints on the number of energy and angle intervals—is mainly attributable to the necessity of calculating the Jacobian when solving the system of ODEs by an implicit method. For a high luminosity, $L \gtrsim 10^{43}$ erg s^{-1}, it is appropriate to replace the Boltzmann equations by hydrodynamics with the calculation of photon energy transfer in the heat conduction approximation. Such a problem can be solved in general relativity.

(5) Our calculation using the absorption and emission coefficients not averaged over the angles suggests that the averaging procedure may be used at this stage of our study. This approximation allowed us to achieve an acceptable energy resolution (without averaging, we were able to use up to seven energy zones) while reducing computational time considerably (one day instead of one or two weeks). When averaging the coefficients, the opaque stage begins at approximately the same total luminosity. Thus, for $L = 10^{37}$ erg s^{-1}, our calculations yield $\tau_\gamma = 0.54$, whereas the correct calculation of the coefficients yields 0.57. At high luminosities, the averaged coefficients are highly overestimated; for example, for $L = 10^{38}$ erg s^{-1}, we obtain $\tau_\gamma = 3.9$ instead of 1.5, and as the luminosity increases further, the difference in the photon optical depths is several orders of magnitude. Consequently, the passage to the opaque stage was quantitatively determined correctly, while the dependence of the mean energy on L at high luminosities is correctly described only qualitatively, although this luminosity dependence is weak. An increase in the number of computational intervals in the angle grid by a factor of 1.5 did not cause any appreciable change in the mean photon energy (for example, the difference was 3% at $L = 10^{39}$ erg s^{-1}, the agreement is better at lower L, and the agreement becomes poorer but is still within 7% as the luminosity increases).

(6) We did not investigate the accuracy of allowance for bremsstrahlung. Initially, we assumed

this effect to be small compared to annihilation and scattering. This was actually the case, particularly at low luminosities. Nevertheless, allowance for this radiation caused an increase in the particle flux and a decrease in the mean photon energy by a factor of almost 1.5. A more accurate allowance for this process, as well as including other three-particle reactions (three-photon annihilation, double Compton scattering), will cause a change in the photon energy at luminosities of 10^{42} erg s^{-1}. At the initial stage of our study of the SS hypothesis, it would have been desirable to take into account bremsstrahlung in our calculations with a large error. If refinement will cause a decrease in the photon energy (for example, to 100 keV), it will be easier to relate SSs to SGR observations.

ACKNOWLEDGMENTS

The statement of the problem belongs to V.V. Usov. The main calculations were performed under the leadership of V.V. Usov and M. Milgrom during my visit to the Weizmann Institute (Israel). I am grateful to my instructors for numerous remarks on the calculations. In particular, they pointed out the necessity of allowing for bremsstrahlung, a weak but important effect at high luminosities.

REFERENCES

1. A. G. Aksenov, Pis'ma Astron. Zh. **24**, 563 (1998) [Astron. Lett. **24**, 482 (1998)].
2. C. Alcock, E. Farhi, and A. Olinto, Astrophys. J. **310**, 261 (1986).
3. A. M. Beloborodov, Mon. Not. R. Astron. Soc. **305**, 181 (1999).
4. V. B. Berestetskii, E. M. Lifshitz, and L. P Pitaevskiĭ, *Course of Theoretical Physics,* Vol. 4: *Quantum Electrodynamics* (Pergamon, New York, 1982; Nauka, Moscow, 1989).
5. E. E. Fenimore, R. W. Klebesadel, and J. G. Laros, Astrophys. J. **460**, 964 (1996).
6. M. Colpi and J. C. Miller, Astrophys. J. **388**, 513 (1992).
7. N. K. Glendennig, *Compact Stars: Nuclear Physics, Particle Physics, and General Relativity* (Springer-Verlag, New York, 1996).
8. P. Haensel, J. L. Zdunek, and R. Scaeffer, Astron. Astrophys. **160**, 121 (1986).
9. G. Hall and J. M. Watt, *Modern Numerical Methods for Ordinary Differential Equations* (Clarendon, Oxford, 1976).
10. E. Haug, Astron. Astrophys. **148**, 386 (1985).
11. R. W. Lindquist, Ann. Phys. **37**, 487 (1966).
12. M. Lyutikov and V. V. Usov, Astrophys. J. **543**, L129 (2000).
13. E. P. Mazets, T. L. Klaĭn, R. L. Aptekar', *et al.*, Pis'ma Astron. Zh. **25**, 735 (1999).
14. A. Mezzacappa and S. W. Bruenn, Astrophys. J. **405**, 669 (1993).
15. D. Mihalas, *Foundations of Radiation Hydrodynamics* (Oxford Univ. Press, Oxford, 1984).
16. Yu. P. Ochelkov, O. F. Prilutskiĭ, I. L. Rozental', and V. V. Usov, *Relativistic Kinetics and Hydrodynamics* (Atomizdat, Moscow, 1979).
17. T. Øvergård and E. Østgaard, Astron. Astrophys. **243**, 412 (1991).
18. B. Paczyński, Astrophys. J. **363**, 218 (1990).
19. R. P. Pilla and J. Shaham, Astrophys. J. **486**, 903 (1997).
20. J. Schwinger, Phys. Rev. **82**, 664 (1951).
21. V. V. Usov, Mon. Not. R. Astron. Soc. **267**, 1035 (1994).
22. V. V. Usov, Phys. Rev. Lett. **80**, 230 (1998).
23. V. V. Usov, Astrophys. J. Lett. **550**, L179 (2001).
24. J. L. Zdunek and P. Haensel, Phys. Rev. D **42**, 710 (1990).

Translated by G. Rudnitskii

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 560–567. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 28, No. 8, 2002, pp. 633–640.

Determining the Rotation Law of the Solar Convection Zone

Yu. V. Vandakurov*

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

Received February 13, 2002

Abstract—The condition of minimum total dissipation is used to derive stationary rotation and azimuthal magnetic field distributions in the bulk of the solar convection zone with an upper boundary at which the relative radius is $r/R = 0.95$. General equilibrium configurations with symmetric and antisymmetric (about the equator) angular-velocity and field components are determined. The calculated rotation law matches the observed one in general parameters, but the decrease in angular velocity at high latitudes in theory is larger than that in observations. Besides, there are additional sharp variations in the rotation and field distributions in the theoretical curves near the generation zone of solar torsional waves. The possible cause of the latter discrepancy is discussed. The change in equilibrium distributions due to the presence of an inverse molecular-weight gradient at the base of the convection zone is also studied. This gradient is known to be produced by accelerated gravitational helium settling in the convection zone.

Key words: *Sun, magnetic field, convection, rotation, torsional waves*

INTRODUCTION

Based on exact nonlinear equations, Elliott *et al.* (2000) numerically analyzed the rotation distribution in a nonmagnetic solar convection zone. They obtained several models with a latitudinal rotation distribution of the type observed on the Sun. Such a rotation distribution arises when the constancy condition for the emergent heat flux is satisfied over the entire upper boundary of the convection zone. Although this boundary condition is confirmed by observations, its validity still needs to be proved. If, for example, the formation of the observed differential rotation was determined by other factors, then the constancy condition for the emergent heat flux under discussion could be the effect rather than the cause of the differential rotation.

In addition, the assumption mentioned by Elliott *et al.* (2000) that there is no stationary magnetic field is important. An equilibrium with an antisymmetric (about the equatorial plane) azimuthal magnetic field was studied previously (Vandakurov 2001a). For example, if we are concerned with an adiabatically stratified convection zone and if the latitudinal differential rotation is negligible, then Eq. (42) from the cited paper holds. It follows from this equation that in the absence of a magnetic field, the angular velocity does not depend on radius. This conclusion is also valid when the latitudinal differential rotation is taken into account. Thus, the assumption of Elliott *et al.* (2000) about a nonmagnetic zone is in conflict with helioseismic data. The latter suggest the existence of a radial gradient in angular velocity (see, e.g., Schou *et al.* 1998).

An important conclusion that follows from the numerical studies of Elliott *et al.* (2000) is the negligible effect of turbulent viscosity on the formation of differential rotation in the convection zone when the rotation distribution is similar to that observed on the Sun. This conclusion is in agreement with the law of irreversible thermodynamics on minimum entropy production in nearly equilibrium systems (Prigogine 1980), because the generation of additional dissipative forces that produce or maintain nonuniform rotation of the medium would imply an increase in entropy production.

The generation models for the solar magnetic field in which the latter is generated by dissipative forces seem unrealistic for similar reasons. This question has already been discussed previously (Vandakurov 1999b, 2001a, 2001b). In the last two papers, we showed that a longitude-dependent magnetic field may well be generated in the nondissipative regime if the radial angular-velocity gradient is nonzero and if there is meridional circulation of matter. In the case under consideration, we are concerned with an oscillatory instability whose frequency is comparable to the rotation frequency. Such a field generation mechanism can be realized if, apart from the generation zone mentioned above, there is an overlying zone

*E-mail: yv.mhd@pop.ioffe.rssi.ru

where the field of the preceding cycle is transformed (and subsequently rises to the surface). Below, we use vector-field expansions in terms of orthogonal vector spherical harmonics considered previously (Vandakurov 1999a, 1999b).

Thus, the condition for the functional that determines total dissipation in the convection zone being at a minimum may serve as an equation for deriving nearly stationary distributions of the rotation, circulation, and magnetic fields formed in convection zones. The models obtained in this way are called minimized models. Clearly, the conditions at the convection-zone boundaries in the approximation under consideration are simply disregarded. Note, incidentally, that using the principle of minimum dissipation to study the structure of rotating stars was also discussed by Tassoul (1979).

In the next section, we consider a general statement of the problem of finding a minimized equilibrium configuration. We take into account the fact that, in general, rotation with asymmetric (about the equator) corrections can be formed spontaneously. Similar problems attributable to velocity asphericity of the acoustic waves generated on the Sun arise when analyzing the solar oscillation frequency splittings. These problems were discussed by Woodard and Libbrecht (1993), Dziembowski *et al.* (2000), and Antia *et al.* (2001). Below, we assume that, apart from the formation of configurations with a normal symmetry (for which the angular velocity is symmetric and the magnetic field is antisymmetric about the equator), structures with an additional (opposite) field symmetry can be formed.

It turns out that difficulties in constructing models arise either from the presence of many shallow local minima of total dissipation or from the tendency for complex rotational structures with sharp angular-velocity and magnetic-field variations to be formed. These factors cause serious difficulties when upper layers of the convection zone are taken into account. Since these layers do not significantly affect the equilibrium of the denser lower levels, we consider only the zone with a relative radius $r/R \leq 0.95$.

It is also difficult to choose the initial state. We assume that in the rough approximation, our exact solutions (Vandakurov 2001a) can be used to estimate the principal (first) coefficients in the angular-velocity and field expansions in terms of vector spherical harmonics. The shallow minima correspond to a decrease in the initial total viscous dissipation determined by the dimensionless parameter $\varepsilon^{1/2}$ by several times. The solutions that we consider are characterized by the deep minima corresponding to a decrease in dissipation compared to the initial one by more than two orders of magnitude. Our studies show that such solutions are difficult to find.

Complicated problems arise because there are generally sharp radial velocity and field variations (see the next section) in minimized models. Interestingly, the localization zone of the most abrupt changes in angular velocity is close to that in which torsional waves, i.e., drifting waves of faster or slower rotation of the medium, are observed on the Sun (very sharp variations apparently arise in the uppermost layers, which we do not consider here). These waves were studied by LaBonte and Howard (1982), Kosovichev and Schou (1997), Schou *et al.* (1998), Howe *et al.* (2000b), and Antia and Basu (2000); it is not yet clear why the waves appear. The hypothesis that, because of the slow latitudinal drift of matter, the formation of the abrupt changes in angular velocity mentioned above is accompanied by a replacement of the matter in which these abrupt changes are formed deserves a study. As a result, instead of sharp spatial variations in angular velocity, we obtain a smooth velocity distribution and drifting torsional waves.

An additional molecular-weight gradient is known to be present at the base of the convection zone due to the accelerated (compared with deeper layers) gravitational settling of helium and heavy elements in the convection zone (Kosovichev *et al.* 1997; Elliott and Gough 1999). We discuss the changes in the equilibrium model that are attributable to such a molecular-weight gradient. In the case of a modest change in the depth of the convection zone with time, a transition from the model with an enhanced density gradient at the lower boundary to another model without such a gradient is possible.

MODELS FOR THE DIFFERENTIAL ROTATION

Let us consider the equilibrium rotation distribution for a magnetized solar convection zone under the assumption of minimum total dissipation. We assume that there are no meridional flow and no poloidal magnetic field and that the rotation of the medium and the azimuthal magnetic field have an arbitrary symmetry about the equatorial plane. The presence of a magnetic field is required if the angular velocity depends on radius. A representation of the fields under consideration as series in orthogonal vector spherical harmonics has already been discussed previously (Vandakurov 1999a, 1999b). In our case, the angular velocity and magnetic field have only toroidal components with the coefficients of the series $v_{J0}^{(0)}$ and $B_{J_b0}^{(0)}$, respectively, where $J_b = J + 1$; $J = 1, 3, 5, \ldots$ if the symmetry is normal, i.e., the rotation is symmetric, and $J = 2, 4, 6, \ldots$ otherwise.

Assuming the stratification of the convection zone to be adiabatic, we find that the rotation and field distributions are determined by the balance of vortical

forces. The latter can be written as formula (14) from Vandakurov (1999b):

$$\left[r/v_0\right]^4 \sum_J R_{J0}^{(0)} R_{J0}^{(0)*} = \varepsilon, \quad J = 1, 2, 3 \ldots, \quad (1)$$

where

$$R_{J0}^{(0)} = \frac{1}{ir^2} \sum_{J_1 J_2} G_{J_1 J_2}^J \Bigg\{ J(J+1) \quad (2)$$

$$\times Z_{J_1 J_2}^J \left[v_{J_2 0}^{(0)} \frac{\partial}{\partial r}\left(r v_{J_1 0}^{(0)} \right) - \frac{1}{4\pi\rho} B_{J_2 0}^{(0)} \frac{\partial}{\partial r}\left(r B_{J_1 0}^{(0)} \right) \right]$$

$$+ J_1(J_1+1) Z_{J J_2}^{J_1} r \frac{\partial}{\partial r} \left[v_{J_1 0}^{(0)} v_{J_2 0}^{(0)} - \frac{1}{4\pi\rho} B_{J_1 0}^{(0)} B_{J_2 0}^{(0)} \right] \Bigg\}.$$

The quantities G_{jl}^k and Z_{jl}^k depend on the indices j, l, k alone and are given by formulas (10) and (11) from the above paper; v_0 is identical to the coefficient $v_{10}^{(0)}$ for the lower boundary of the convection zone; ρ is the density, which may roughly be considered as an equilibrium one; r is the radius; ε is a dimensionless coefficient determined by turbulent viscous forces; and the asterisk denotes complex conjugation. We do not write out an expression for ε, because our objective below is only to find minimum values for the left-hand side of Eq. (1).

The models discussed below were obtained by examining equilibrium conditions in the solar convection zone divided into 12 layers along the radius from the lowest boundary at $r/R = 0.713$ to the upper boundary at $r/R = 0.95$, where R is the solar radius. The number of latitudinal modes in the representations for the angular velocity and magnetic field is taken to be five; i.e., the largest value of J in Eq. (1) is 10.

Characteristics of the medium (its density and density gradient) were chosen in accordance with the numerical solar model of Guenther *et al.* (1992). We determined the radial gradients of the coefficients $v_{J0}^{(0)}$ and $B_{J_b 0}^{(0)}$ by using cubic splines, provided that the function being fitted and its first three derivatives were continuous. The theory of such splines was considered by Zav'yalov *et al.* (1980).

We use the lowest layer ($0.713 \le r/R \le 0.733$) to study the influence of an additional radial density gradient at the base of the solar convection zone. Because of the accelerated gravitational settling of helium and heavy elements in the convection zone, a layer with an inverse molecular-weight gradient is known to be formed at its base (Elliott and Gough 1999; Schou *et al.* 1998). As a result, the increase in density with depth in this layer will greatly slow down. To qualitatively analyze this effect, below we consider two models. In one (C) model, the dimensionless derivative $-d\ln\rho/d\ln r$ varies continuously in accordance with the model of Guenther *et al.* (1992), while in the other (gradient G) model, this derivative in the lowest layer is assumed to be small.

The total dissipation was minimized, in turn, for each latitudinal mode by successively determining all those velocity and field coefficients that characterized the radial dependence of the mode under discussion from the condition of minimum dissipation. However, no variations were made for the normalization coefficient $v_{10}^{(0)}$ that corresponds to the lower boundary of the convection zone (and equal to v_0). Subsequently, the calculations were repeated until an absolute minimum was found. Of course, we first calculated the coefficients that corresponded to the field with a normal symmetry. We determined a total of 239 coefficients $v_{J0}^{(0)}(r)$ and $B_{J_b 0}^{(0)}(r)$, where $J_b = J+1$ and r corresponds to the radial-layer number.

Note that when we pass from varying one coefficient to varying another coefficient, the total dissipation can increase. The entire procedure under consideration then may diverge. Therefore, the question of whether the solution is unique is of current interest. At this preliminary stage, we restricted our analysis to models with sufficiently deep minima. For the two types of G-models considered below, the parameter ε in Eq. (1) is approximately equal to 4×10^{-4}. This parameter is almost an order of magnitude smaller for the C-model with a normal field symmetry.

Below, we mark the calculated angular velocity (or Alfvén velocity) by the subscript 0 or 1, depending on whether the field under consideration has a normal or additional symmetry. The total field with an arbitrary symmetry is produced by the summation of the components with the subscripts 0 and 1 [given that for an antisymmetric (about the equator) field, the contribution of the latter for the southern hemisphere enters with the opposite sign]. Because of the large difference in the orders of magnitude of the fields, it seems more convenient not to perform this summation.

Figures 1 and 2 refer to the fields with a normal symmetry where the index J in Eq. (1) is even. Figures 1a and 1b show the derived radial distributions of the relative angular velocity Ω_0, respectively, for the G- and C-models at latitudes 0°, 15°, ..., and 60°. Here, we consider the equatorial angular velocity at the lower boundary of the convection zone as the normalization velocity. For comparison, Fig. 1 also shows the helioseismic data from Howe *et al.* (2000a) in our normalization. In this case, the triangles, pluses, and crosses refer to latitudes of 0°, 30°, and 60°, respectively. Next, Fig. 2 shows the distributions of the relative Alfvén velocity A_0 (which

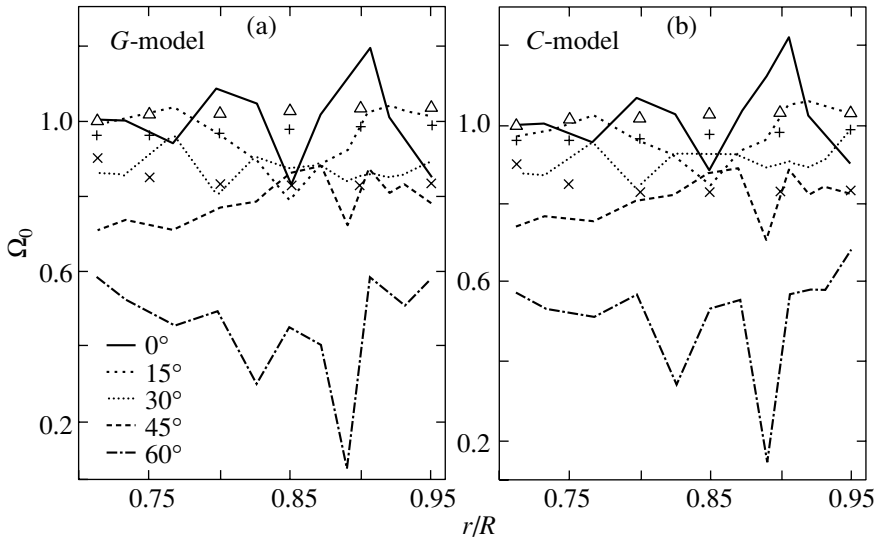


Fig. 1. Relative angular velocity Ω_0 versus the relative radius for the (a) G- and (b) C-models with a normal symmetry at latitudes of 0°, 15°, 30°, 45°, and 60°. The triangles, pluses, and crosses represent the data from Howe *et al.* (2000a) for latitudes of 0°, 30°, and 60°, respectively. The angular velocity is normalized in such a way that $\Omega_0 = 1$ on the equator at the lower boundary of the convection zone.

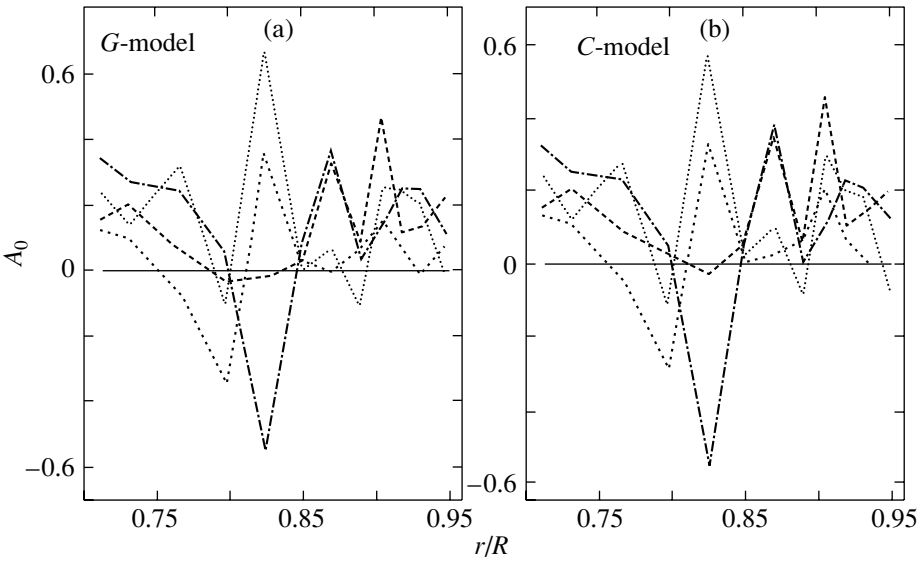


Fig. 2. Relative Alfvén velocities A_0 of the same models and at the same latitudes as in Fig. 1. The normalization velocity is equal to the equatorial angular velocity of the lower boundary of the convection zone.

is antisymmetric about the equator) for the same G- and C-models and at the same latitudes as in Fig. 1. Here, the Alfvén velocity is related to the velocity v_0. The G-model with an arbitrary symmetry is shown in Figs. 3 and 4.

It should be emphasized that when constructing the curves shown in Figs. 1–4, we used only the values of Ω_l and A_l (where $l = 0, 1$) that refer to the boundaries of the radial layers under consideration. Therefore, the radial dependence of these velocities is reproduced with distortions. However, the other extremely sharp radial variations discussed below are of greater importance.

If we are concerned with the mean rotation distributions, then our data confirm the very fact that differential rotation with more rapid rotation of equatorial layers is spontaneously formed in the solar convection zone. An additional reduction in the theoretical angular velocity of the lower layers at high latitudes stems from the fact that our calculations disregarded the relationship of the angular velocity with the rotation of the underlying layers. Under ac-

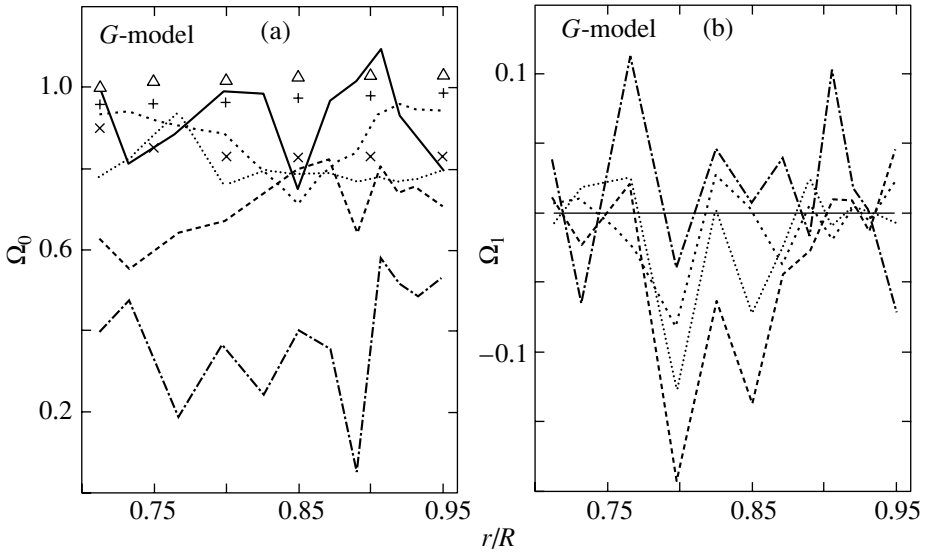


Fig. 3. Radial distributions of the (a) symmetric and (b) antisymmetric (about the equator) relative angular velocities for the G-model with an arbitrary symmetry. The notation is the same as in Fig 1a.

tual conditions, the rotation must be nearly rigid at depth $r/R \sim 0.6$. However, there are also more significant discrepancies between theory and observations. These discrepancies are attributable both to a more rapid (compared with observations) decrease in the angular velocity at high latitudes and to the presence of sharp peaks and variations in theoretical angular velocity, particularly near and slightly below the layer at $r/R \approx 0.9$. Of course, for such sharp variations in equilibrium quantities to be accurately described, we must obtain a solution with a sufficiently high accuracy, because the contribution of high-order latitudinal modes can be significant.

Mean coefficients of the angular (u_{lk}) and Alfvén (b_{lk}) velocities for the G-model with a normal symmetry (rows 1–2), the C-model with the same symmetry (rows 3–4), and the G-model with an arbitrary symmetry (rows 5–8). Here, $l = 0, 1$ and k is the latitudinal-mode number

k	1	2	3	4	5
u_{0k}	0.9966	−0.1546	−0.0588	−0.0538	−0.0163
b_{0k}	0.1399	−0.0199	−0.0458	−0.0116	−0.0397
u_{0k}	1.0223	−0.1407	−0.0501	−0.0356	−0.0064
b_{0k}	0.1348	−0.0137	−0.0354	−0.0067	−0.0211
u_{0k}	0.9789	−0.1669	−0.0612	−0.0533	−0.0137
b_{0k}	0.1343	−0.0245	−0.0425	−0.0150	−0.0246
u_{1k}	−0.0244	0.0054	−0.0058	0.0035	0.0011
b_{1k}	−0.0141	0.0057	0.0070	−0.0011	−0.0077

The table gives the coefficients averaged over the 12 radial layers under consideration: $u_{0k} = v^{(0)}_{2k-1,0}/v_0$, $b_{0k} = B^{(0)}_{2k,0}/[v_0(4\pi\rho)^{1/2}]$, $u_{1k} = v^{(0)}_{2k,0}/v_0$, and $b_{1k} = B^{(0)}_{2k-1,0}/[v_0(4\pi\rho)^{1/2}]$, where $k = 1, 2, \ldots, 5$; and v_0 is equal to $v^{(0)}_{1,0} = v^{(0)}_{10}$, which refers to the lower radial boundary of the convection zone. We see that the amplitudes of the high-order modes do not decrease with increasing mode number k, although these amplitudes are an order of magnitude smaller than those for the principal first terms of the series.

Nevertheless, an increase in the number of included latitudinal modes does not cause the situation to change qualitatively, because the sharp variations in angular velocity under discussion are preserved. Importantly, the sharpest variations take place near the zone where solar torsional waves are generated. According to the latest helioseismic data from Antia and Basu (2000) and Antia *et al.* (2001), the wave-generation zone lies in the range $0.9 \le r/R \le 1$. The nature of these solar waves is unknown. Since the zones of the observed torsional waves are located near regions with sharp theoretical variations in angular velocity, we cannot rule out the possibility that the sharp spatial variations in angular velocity under discussion are actually formed. However, a slow latitudunal drift of matter appears to arise simultaneously; this drift contributes to the replacement of the accelerated (or decelerated) matter with other matter from the neighboring layers. As a result, sharp variations in angular velocity characteristic of the theoretical steady-state equilibrium model have no time to be formed; drifting torsional waves appear instead. Of

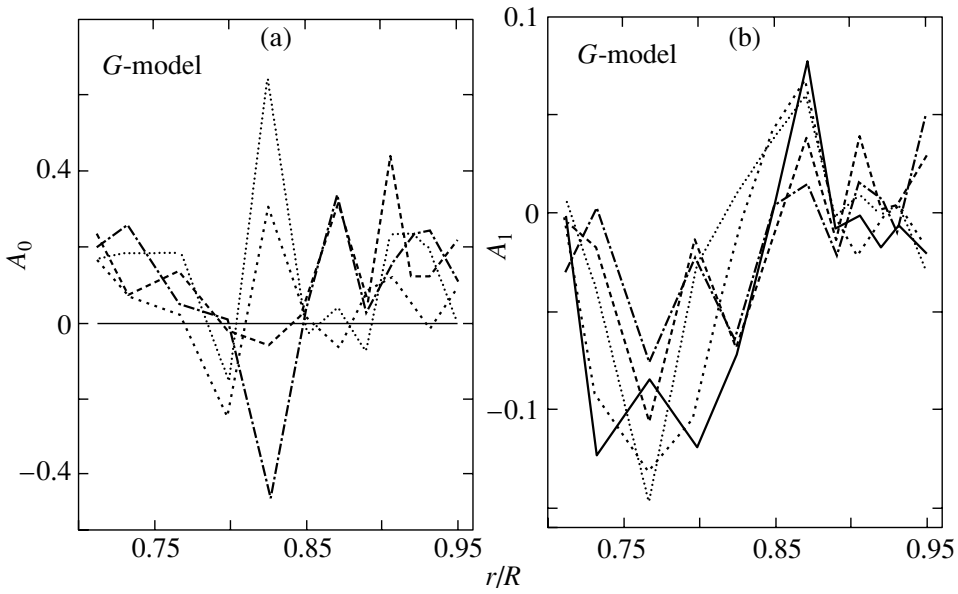


Fig. 4. Same as Fig. 3 for the relative (a) antisymmetric and (b) symmetric (about the equator) Alfvén velocities.

course, the above reasoning still needs to be proved and we plan to return to this problem in the future.

It is not yet clear why the reduction in angular velocity at high latitudes in the theoretical model is larger than that in observations. It may well be that the rotation distribution under discussion will change significantly once the above effect produced by wavelike motions of the medium has been taken into account. We must also ascertain whether other minimized models of the solar convection zone are possible.

The processes that contribute to the formation of asymmetric (about the equator) rotation of the medium are also of considerable interest. For the most realistic G-model with a μ-gradient at the base of the convection zone, the asymmetric angular velocity under discussion has a number of peak-shaped maxima or minima; the largest of these have the following parameters (we give r/R, latitude, and Ω_1): 0.77, 67°, 0.135; 0.80, 43°, −0.196; 0.85, 43°, −0.142; 0.91, 69°, 0.140; and 0.95, 77°, −0.210. In addition, there are extrema at the pole, for example, at r/R close to 0.73, 0.83, and 0.93; the angular velocity Ω_1 at a latitude of 90° is 0.28, −0.32, and −0.35, respectively.

Note that, according to the helioseismic data from Antia *et al.* (2001), there is an excess in the time-averaged speed of sound relative to the spherically symmetric velocity component that corresponds to the mean frequencies of acoustic waves at $r/R \approx 0.92$ and at 60° latitude. In this case, the excess in the speed of sound is symmetric about the equator. In our case, there is an excess in the asymmetric (about the equator) angular velocity at $r/R \approx 0.91$ at latitude 69°. It is not yet clear whether the described phenomena are related. For the problem to be considered in detail, we must study the distribution of the potential forces that determine the behavior of the pressure and density of the medium.

In conclusion, it should be mentioned that the azimuthal magnetic field corresponding to the leftmost peak in Fig. 4a is ∼60 kG. This field estimate matches that obtained by Dziembowski *et al.* (2000), who considered the observed frequency splitting under the effect of a variable stochastic magnetic field. For the upper boundary of our convection zone, the field is a factor of 5 weaker than the above field.

DISCUSSION

The results obtained suggest that our method for finding nearly equilibrium, rotating magnetized convective structures by minimizing the total dissipation is rather efficient. The possible formation of a new type of astrophysical structure is also of considerable interest. Here, we bear in mind that the theoretical steady-state configuration corresponding to minimum total dissipation can contain sharp spatial variations in angular velocity and magnetic field in some regions. As a result, the transition to another quasi-equilibrium state in which there are no sharp variations mentioned above but where wavelike motions of the medium are generated becomes probable. In fact, the waves are generated as if through a continuous transition to a state with minimum total dissipation; no actual transition of this type occurs because of the replacement of one matter with other. As a result, structures with wavelike motions of the

medium and with smooth rotation and magnetic-field distributions can be produced.

For the solar convection zone, we identify the above wavelike motions with torsional waves. The fact that the localization region of the observed torsional waves is comparatively close to the zone where the angular velocity varies greatly in steady-state theoretical models argues for the validity of this conclusion. As a result, it becomes possible to elucidate the nature of the torsional waves on the Sun. Of course, a final conclusion can be reached only after a detailed analysis of the models that take into account the very presence of torsional waves.

The general pattern of our rotation law for the convection zone is the same as that observed on the Sun with faster rotation of the equatorial layers. However, the decrease in angular velocity with increasing latitude is larger than the observed one. It may well be that the situation will change if the contribution from torsional waves is taken into account. It is also unclear how a smooth transition to the rotation of the radiative zone that must take place at the lower boundary of the convection zone can be taken into account. Since the general pattern of radial variations in angular velocity in our approximation is the same at different latitudes, a difference between the convection-zone rotation and the rotation of the underlying layers that follows from helioseismic data will arise.

We reached the unexpected conclusion that there are small but nonzero corrections corresponding to asymmetric (about the equator) rotation of the medium in minimized models. It is not yet clear how significant these corrections are. Note that the presence of components with different symmetries about the equator in the global axisymmetric solar magnetic field is confirmed by the data from Stenflo and Güdel (1988).

The question of what role the enhanced density gradient at the base of the convection zone discussed above can play is also of considerable interest. If the depth of the convection zone could vary with time for some reason or another (by several percent), then an additional density gradient during the periods of a shallow zone could be outside the convection zone. The emergence of the gradient under discussion in our designations is equivalent to a transition from the C-model to the G-model. Incidentally, we constructed the latter model by considering a spontaneous transition from the derived C-model to a new model through the emergence of an additional density gradient in the lowest radial layer. These calculations confirm that an equilibrium in the models under discussion is established spontaneously, although the problem of whether the minimized model is unique is yet to be solved.

The possible relationship between variations in the average level of solar activity and transitions from the C-model to the G-model and vice versa deserves attention. In general, the basic parameters of these models are similar but the differences between the corrections attributable to the presence of fields with an additional symmetry can be more significant. Observations suggest that the aforementioned slow variations in solar activity actually take place on the Sun. For example, Makarov *et al.* (2001) provide evidence for an increase in the past 120 years in the area of the solar polar zone occupied by a magnetic field of one polarity under minimum solar activity. Since the theoretical models under discussion are preliminary and rough, a detailed analysis of this question seems premature. On the whole, the equilibrium state of the solar convection zone still needs to be studied further.

ACKNOWLEDGMENTS

This study was supported by the Integration Foundation (contract no. KO854) and the Russian Foundation for Basic Research (project no. 00-02-16939).

REFERENCES

1. H. M. Antia and S. Basu, Astrophys. J. **541**, 442 (2000).
2. H. M. Antia, S. Basu, F. Hill, *et al.*, Mon. Not. R. Astron. Soc. **327**, 1029 (2001).
3. W. A. Dziembowski, P. R. Goode, A. G. Kosovichev, and J. Schou, Astrophys. J. **537**, 1026 (2000).
4. J. R. Elliott and D. O. Gough, Astrophys. J. **516**, 475 (1999).
5. J. R. Elliott, M. S. Miesch, and J. Toomre, Astrophys. J. **533**, 546 (2000).
6. D. B. Guenther, P. Demarque, Y.-C. Kim, *et al.*, Astrophys. J. **387**, 372 (1992).
7. R. Howe, J. Christensen-Dalsgaard, F. Hill, *et al.*, Science **287**, 2456 (2000a).
8. R. Howe, J. Christensen-Dalsgaard, F. Hill, *et al.*, Astrophys. J. Lett. **533**, L163 (2000b).
9. A. G. Kosovichev and J. Schou, Astrophys. J. **482**, L207 (1997).
10. A. G. Kosovichev, J. Schou, P. H. Scherrer, *et al.*, in *Proceedings of the 181st Symposium of the International Astronomical Union "Sounding Solar and Stellar Interiors,"* Ed. by J. Provost and F.-X. Schmider (Kluwer, Dordrecht, 1997), p. 203.
11. B. J. LaBonte and R. Howard, Sol. Phys. **75**, 161 (1982).
12. V. I. Makarov, V. N. Obridko, and A. G. Tlatov, Astron. Zh. **78**, 859 (2001) [Astron. Rep. **45**, 746 (2001)].
13. I. Prigogine, *From Being to Becoming: Time and Complexity in the Physical Sciences* (Freeman, San Francisco, 1980; Nauka, Moscow, 1985).

14. J. Schou, H. M. Antia, S. Basu, *et al.*, Astrophys. J. **505**, 390 (1998).
15. J. O. Stenflo and M. Güdel, Astron. Astrophys. **191**, 137 (1988).
16. J.-L. Tassoul, *Theory of Rotating Stars* (Princeton Univ. Press, Princeton, 1978; Mir, Moscow, 1982).
17. Yu. V. Vandakurov, Astron. Zh. **76**, 29 (1999a) [Astron. Rep. **43**, 24 (1999a)].
18. Yu. V. Vandakurov, Pis'ma Astron. Zh. **25**, 868 (1999b) [Astron. Lett. **25**, 758 (1999b)].
19. Yu. V. Vandakurov, Pis'ma Astron. Zh. **27**, 700 (2001a) [Astron. Lett. **27**, 596 (2001a)].
20. Yu. V. Vandakurov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **44**, 735 (2001b) [Radiophys. Quant. Electron. **44**, 678 (2001)].
21. M. F. Woodard and K. G. Libbrecht, Astrophys. J. Lett. **402**, L77 (1993).
22. Yu. S. Zav'yalov, B. I. Kvasov, and V. L. Miroshnichenko, *Spline Function Methods* (Nauka, Moscow, 1980).

Translated by V. Astakhov

Astronomy Letters, Vol. 28, No. 8, 2002, pp. 568–577.

ERRATA

Erratum: Comprehensive Studies of Solar Activity on the CORONAS-F Satellite*
[*Astronomy Letters* 28, 401–410 (2002)]

V. N. Oraevsky[1] and I. I. Sobelman[2]

[1] *Institute of Terrestrial Magnetism, Ionosphere, and Radio-Wave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142090 Russia; E-mail: oraevsky@izmiran.rssi.ru*

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

Received December 27, 2001

Abstract—The first results of comprehensive CORONAS-F observations of solar activity are presented. The CORONAS-F instrumentation and principal scientific objectives are briefly described and examples of the first results of data reduction are given. © *2002 MAIK "Nauka/Interperiodica"*.

Key words: *Sun; astronomical observing techniques, equipment and instruments*

THE CORONAS PROGRAM AND THE CORONAS-F PROJECT

The CORONAS (Comprehensive Orbital Near-Earth Solar Activity Observations) International Program, as part of which the CORONAS-F near-Earth spaceborne solar observatory was launched, is intended for studies of the Sun at various phases of the 11-year solar cycle. On the previous CORONAS-I satellite (launched in 1994), the Sun was observed near the minimum of its activity. The CORONAS-F will study the solar activity near the maximum of the current cycle 23. On July 31, 2001, CORONAS-F was placed in an orbit with the following parameters: an orbital inclination of 82.49°, a minimum altitude of 500.9 km, a maximum altitude of 548.5 km, and a revolution period of 94.859 min. Such an orbit provides repetitive ~20 day-long periods of continuous solar observations, which is of particular importance in helioseismology and in patrolling solar flares. The actually achieved spacecraft attitude stabilization turned out to be a factor of 3 to 5 better than the projected one (a few arcseconds per second), which allows the spatial resolution of solar observations to be appreciably improved.

SCIENTIFIC OBJECTIVES OF THE CORONAS-F PROJECT

The principal scientific objectives of the CORONAS-F project are observations of global solar oscillations, investigation of the seismology of the solar interior and internal structure, comprehensive studies of powerful dynamical processes on the active Sun (active regions, flares, plasma ejections) over a wide wavelength range, from optical to gamma rays, including the radiation from solar cosmic-ray particles accelerated during active solar events, conditions for their escape, propagation in the interplanetary magnetic field, and impact on the Earth's magnetosphere.

SCIENTIFIC INSTRUMENTATION AND THE FIRST OBSERVATIONAL RESULTS

In accordance with the CORONAS-F objectives, the scientific instrumentation of the spacecraft includes four major groups of instruments: an instrument to detect global solar oscillations; X-ray instruments to image active regions on the Sun with a spatial resolution of ~2″–3″ instruments to measure the fluxes of electromagnetic radiation from active

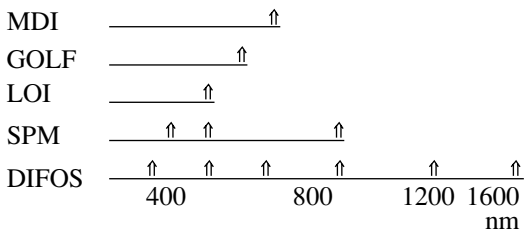


Fig. 1. Comparison of the DIFOS (CORONAS-F) channels with the SOHO instruments (MDI—Michelson Doppler Imaging of solar oscillations, GOLF—Global low-degree velocity, LOI—Luminosity Oscillation Imager, SPM—SunPhotoMeter).

*This article was republished due to two color figures (nos. 4 and 5) missed in the original translation. Our apologies to the authors and the readers of the journal.

1063-7737/02/2808-0568$22.00

Table 1. The CORONAS-F instruments

Instrument	Purpose	Designer	Principal investigators
	Helioseismology		
DIFOS spectrophotometer	Helioseismological monitoring	IZMIRAN	V.N. Oraevsky
	High-angular-resolution monochromatic imaging		
SRT-C X-ray telescope	Studying the spatial structure and dynamics of the upper solar atmosphere using narrow-band XUV images	FIAN	I.I. Sobelman I.A. Zhitnik
RES-C X-ray spectroheliograph	Diagnosing the hot solar atmospheric plasma using images in X and XUV spectral lines	FIAN	I.I. Sobelman I.A. Zhitnik
DIOGENESS spectrophotometer	Studying the X-ray radiation from solar active regions and flares	SRC PAS*	J. Silvestr
	Measuring the fluxes and polarization of electromagnetic radiation (from UV to γ rays)		
RESIK X-ray spectrometer	Studying the solar X-ray radiation with a high spectral resolution	SRC PAS*	J. Silvestr
SPR-N solar spectropolarimeter	Studying the polarization of X-ray radiation from solar flares	FIAN NIIYaF	I.I. Sobelman I.P. Tindo S.I. Svertilov
IRIS flare spectrometer	Studying the X-ray solar flare activity	FTI	G.E. Kocharov
GELIKON gamma-ray spectrometer	Studying the X-ray and gamma-ray solar flare activity	FTI	E.P. Masetz
RPS X-ray spectrometer	Studying the X-ray radiation from solar flares and their precursors	IKI RAS MIPhI	V.M. Pankov Yu.D. Kotov
ATS amplitude–time spectrometer	Studying the X-ray and gamma-ray radiation from solar flares	MIPhI	Yu.D. Kotov
SUVR-Sp-C solar ultraviolet radiometer	Studying the variations in total UV solar flux	IPG	T.V. Kazachevskaya
VUSS-L ultraviolet solar spectrophotometer	Studying the solar UV radiation near the L_α resonance line	IPG	A.A. Nusinov
	Studying solar corpuscular fluxes		
Instrumentation to study SCRs	Studying solar cosmic rays	NIIYaPh MGU	S.N. Kuznetsov
	Gathering and studying of scientific information		
Scientific Information Acquisition System (SSNI)	Controlling the scientific instrumentation and the modes of its operation	IZMIRAN	V.N. Oraevsky A.I. Stepanov

* Space Research Center of the Polish Academy of Sciences.

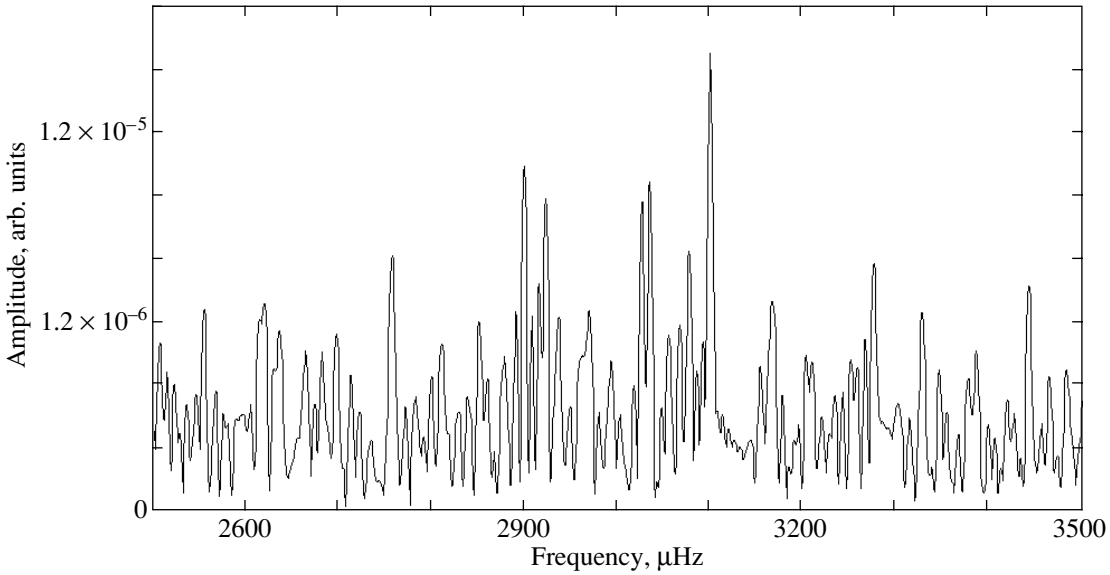


Fig. 2. The relative amplitudes of 5-min solar oscillation modes at a wavelength of 350 nm as inferred from the results of the DIFOS data reduction.

regions and flares; and instruments to study solar corpuscular fluxes. The wide range of measurements of the electromagnetic spectrum and solar cosmic-ray particle [both neutral (neutrons) and charged (electrons, protons, nuclei)] fluxes yields the fullest picture of the physical processes in active regions on the Sun. The CORONAS-F instruments and their main purpose are given in Table 1.

The main operational control of the scientific instrumentation is performed from the Flight Control Center located at the Cosmic Information Technology Center, Institute of Terrestrial Magnetism, Ionosphere, and Radio-Wave Propagation (IZMIRAN), Troitsk. Commands are sent to the spacecraft on a daily basis. Up to 24 kbytes of commands are sent simultaneously. Apart from controlling the modes of operation, the controllers of the scientific instruments are dynamically reprogrammed in flight, which allows their software to be customized for current observations. Requests for control through the Flight Control Center of the Cosmic Information Technology Center (IZMIRAN) are collected via E-mail with no less than 10 min before the control session, which ensures the required speed and flexibility in controlling the scientific instrumetation.

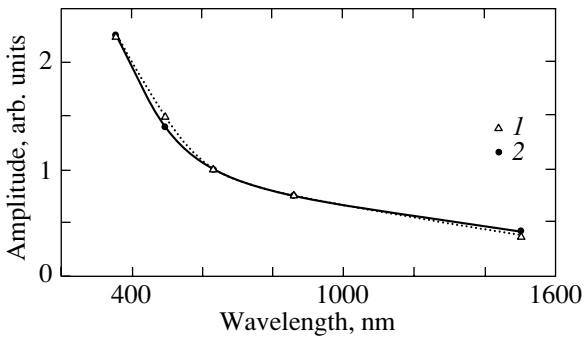


Fig. 3. The relative amplitude of free solar oscillations versus wavelength.

THE DIFOS SPECTROPHOTOMETER

The DIFOS spectrophotometer is designed to measure fluctuations in the intensity of solar optical radiation in an effort to obtain the spectrum of free oscillations. The intensity is simultaneously measured in six optical spectral bands: 350, 500, 650, 850, 1100, and 1500 nm with a band width equal to 10% of the central frequency. In comparison with the CORONAS-I project (Lebedev *et al.* 1995), the photometer was significantly upgraded: the detector sensitivity increased by more than an order of magnitude and the spectral range of observations was extended by almost a factor of 2; concurrently, the number of spectral channels increased from three to six. Global solar oscillations are being observed over a wide spectral range, including the most informative UV line. The intensity of the radiation from global solar oscillations in this line appreciably exceeds the radiation intensity in other lines (Oraevsky *et al.* 1999). Figure 1 compares the DIFOS channels with the SOHO instruments (1988).

Figure 2 shows the relative amplitudes of 5-min solar oscillation modes at a wavelength of 350 nm during November 28–30, 2001 as inferred from the

Table 2. Characteristics of the SRT-C telescope

Channel	Spectral range, Å	Ion lines	Field of view, arcmin	Angular scale per cell	Mode of operation
BP1	171	Fe IX–Fe X	42 × 48	2″.5	Full disk
	195	Fe XII			
	284	Fe XV			
	304	He II			
BP2	175	Fe IX–Fe XI	45 × 50	2″.6	Full disk, pointing at corona up to several solar radii
	304	HeII + SiXI	45 × 50	2″.6	
Optical sensors OS1 and OS2	4000–8000		Circular 11°	1′ × 2′	Stars $> 12^m$

results of the DIFOS data reduction. Similar results were obtained for the other channels. After averaging over the modes, the mean relative amplitudes were determined for each of the six channels and plotted against wavelength (see Fig. 3). All data points fall on a smooth curve fitted by a $\lambda^{-1.2}$ law (*1*). A similar result was established for another interval of observations, November 30–December 3, 2001 (*2*). A comparison of the DIFOS result with the results in the range of ground-based observations (Jimenez *et al.* 1998) gives close agreement. According to ground-based measurements, the amplitude ratio of global oscillations at wavelengths of 500 and 680 nm was 1.6, while for the DIFOS measurements, the amplitude ratio at 500 and 650 nm was 1.5. For a different pair of wavelengths, the amplitude ratio at 500 and 870 nm was 2.2 for ground-based observations, while the amplitude ratio at 500 and 850 nm for the DIFOS data was 2.

THE SPIRIT (SRT–RES–SPR) ASTROPHYSICAL COMPLEX TO STUDY THE SOLAR X-RAY RADIATION

The SPIRIT astrophysical complex consists of the SRT-C solar X-ray telescope, the RES-C X-ray spectroheliograph, and the SPR-N spectropolarimeter. It is designed to solve a wide variety of problems in solar physics based on XUV imaging spectroscopy of the Sun (soft X-ray and extreme UV solar spectroscopy). This method consists in obtaining simultaneously solar images and spectra with high angular and spectral resolutions in narrow portions of the XUV range that characterize the various temperature layers of solar plasma. It was applied in the CORONAS-I project, in which the TEREK-C multichannel telescope and the RES-C spectroheliometer in the wavelength range 0.18–30.4 nm designed at the Lebedev Physical Institute (Russian Academy of Science) were used (Sobelman *et al.* 1996; Zhitnik *et al.* 1998).

Observations with the SPIRIT astrophysical complex onboard the CORONAS-F satellite will allow the following processes to be investigated: the formation and development of solar flares from temporal variations and changes in the structure, spectra, and polarization of the flare regions; nonstationary phenomena in solar plasma (hot points, coronal mass ejections, and others)—the frequency of their occurrence, conditions for their emergence, dynamics of their development, and residual phenomena; and the helium distribution in the transition layer and in the near and far coronas. They also allow us to search for the possible sources of the solar wind, to study the processes of its generation, to elucidate the relationships of the observed solar-wind parameters to solar processes, and to diagnose the structural elements of active regions, areas of the quiet Sun, and coronal holes on time scales from 0.1 s to tens of days.

The SPIRIT complex is distinguished by a wide spectral range (0.04–100 keV) and by high spatial (up to 3″), spectral (up to 2×10^3), and temporal (0.01 s) resolutions intended for observations near maximum solar activity. The SRT-C telescope is a new instrument with significantly improved characteristics compared to the TEREK-C telescope that successfully operated on the CORONAS-I satellite. The working aperture of the optical channels increased from 30 to 100 mm with the simultaneous increase in actual angular resolution. The number of spectral subbands also increased. The RES-C-F spectroheliometer is a modification of the RES-C-I spectroheliometer that successfully operated on the CORONAS-I satellite with an extended spectral range and improved characteristics. In the SRT-C

Table 3. Characteristics of the RES-C X-ray spectroheliograph

Channel	Spectral range, Å	Ion lines	Spectral resolution, Å per cell	Angular scale per cell	Mode of operation
Fe XXV	1.85–1.87	Fe XXIV, FeXXV	1.5×10^{-4}	1′	Image and spectrum from the entire disk
Mg XII (two crossed subchannels)	8.41–8.43	Mg XII doublet	3×10^{-3}	4.1″	Same
XUV (two crossed subchannels)	176–205	Fe VIII, X, XI XII, XIII, XXIV, O VI S IX	0.03	6″ (perpendicular to dispersion)	Spectral images of the disk in all lines
	285–335	He II, Si XI			

and RES-C-F instruments, the resolution and sensitivity of the imaging detectors were increased, their dynamical range was expanded, and a powerful high-performance computer with a large amount of random access memory was used in the electronic system. The new instruments provide an increase in temporal resolution and sensitivity by up to a factor of 10 compared to the TEREK and RES instruments on the CORONAS-I satellite and by a factor of 3 compared to the EIT/SOHO telescope.

The SRT-C Solar X-ray Telescope

The SRT-C solar X-ray telescope is designed for X-ray imaging of the Sun with a high spatial resolution. It includes XUV channels, BP1 and BP2, and two optical sensors for the telescope attitude control using stars. The telescope characteristics are listed in Table 2.

X-ray mirrors with multilayer Mo–Si coating are used for imaging in the X-ray channels of the instrument: in channel BP1, a Ritchey–Chretien objective with a diameter of 120 mm and a focal length of 1660 mm that has four sectors with coverages at wavelengths of 171, 195, 284, and 304 Å (an analog of the EIT/SOHO telescope objective); and in channel BP2, off-axis paraboloids with a diameter of 100 mm and a focal length of 1500 mm.

The imaging detectors in all channels are CCD arrays; image intensifiers, which also act as electronic shutters, are used in the X-ray channels. The brightness dynamic range in the X-ray channels is of the order of 10^5. The exposure times can vary between 0.01 and 600 s.

The mirrors of channel BP2 are equipped with angular alignment and focusing systems to correct for possible changes in alignment when placing the satellite in orbit. In addition, the channel is supplied with artificial moons blocked by commands and with additional optical windows, which in combination with mirror deflection makes it possible to observe portions of the X-ray corona within several solar radii.

The RES-C X-ray Spectroheliograph

The RES-C spectroheliograph is designed to image the Sun in spectral lines of highly ionized atoms of abundant elements (helium, magnesium, iron, silicon) with a high spatial resolution and to measure the linear polarization in spectral lines. The instrument includes three XUV channels: Fe, Mg, and XUV. The RES-C characteristics are listed in Table 3.

X-ray optics is used in the RES-C instrument: in the Fe XXV channel, a ring Bragg mirror made of crystalline quartz with a sphere radius of 196.6 mm operating at a Bragg angle of 52°; in the Mg XII channel, two spherical Bragg mirrors made of crystalline quartz with a sphere radius of 1286 mm operating at angle 81.7°; in the XUV channels, two holographic diffraction gratings with a period of 3600 lines per mm whose spectra are focused on the detectors using multilayer Mo–Si-coated mirrors with a radius of 1600 mm were used: spherical and parabolic ones for the wavelength ranges 180–205 Å and 285–335 Å, respectively.

The imaging detectors in the XUV channels, just as those in the SRT-C instrument, are CCD arrays in combination with image intensifiers. The dynamic range is of the order of 10^5 and the exposure time ranges from 0.01 to 600 s. In the Mg XII and Fe XXV channels, X-ray radiation is recorded by the open CCD arrays.

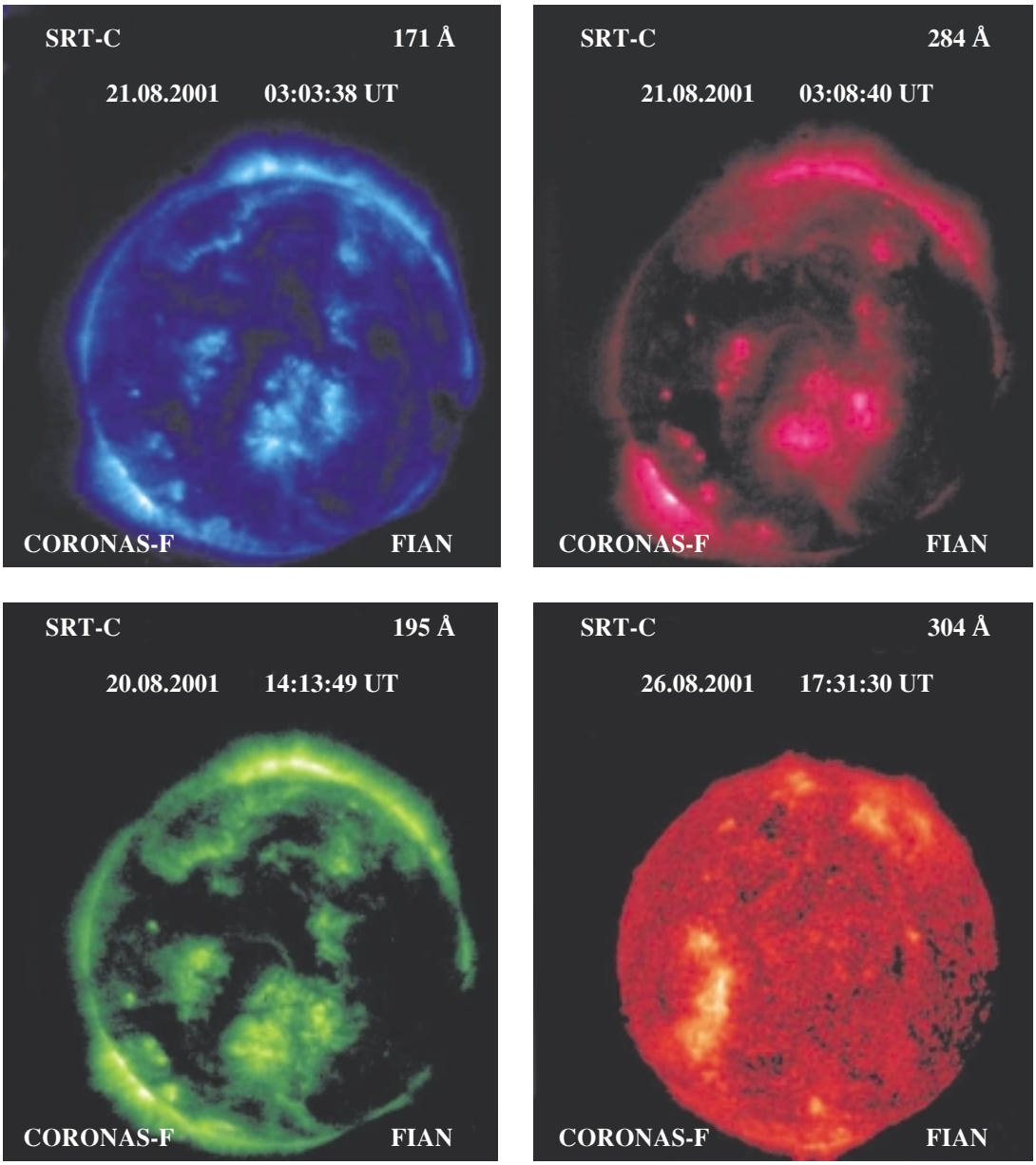


Fig. 4. The solar images at wavelengths of 171, 195, 284, and 304 Å obtained in channel BP1 of the SRT-C instrument.

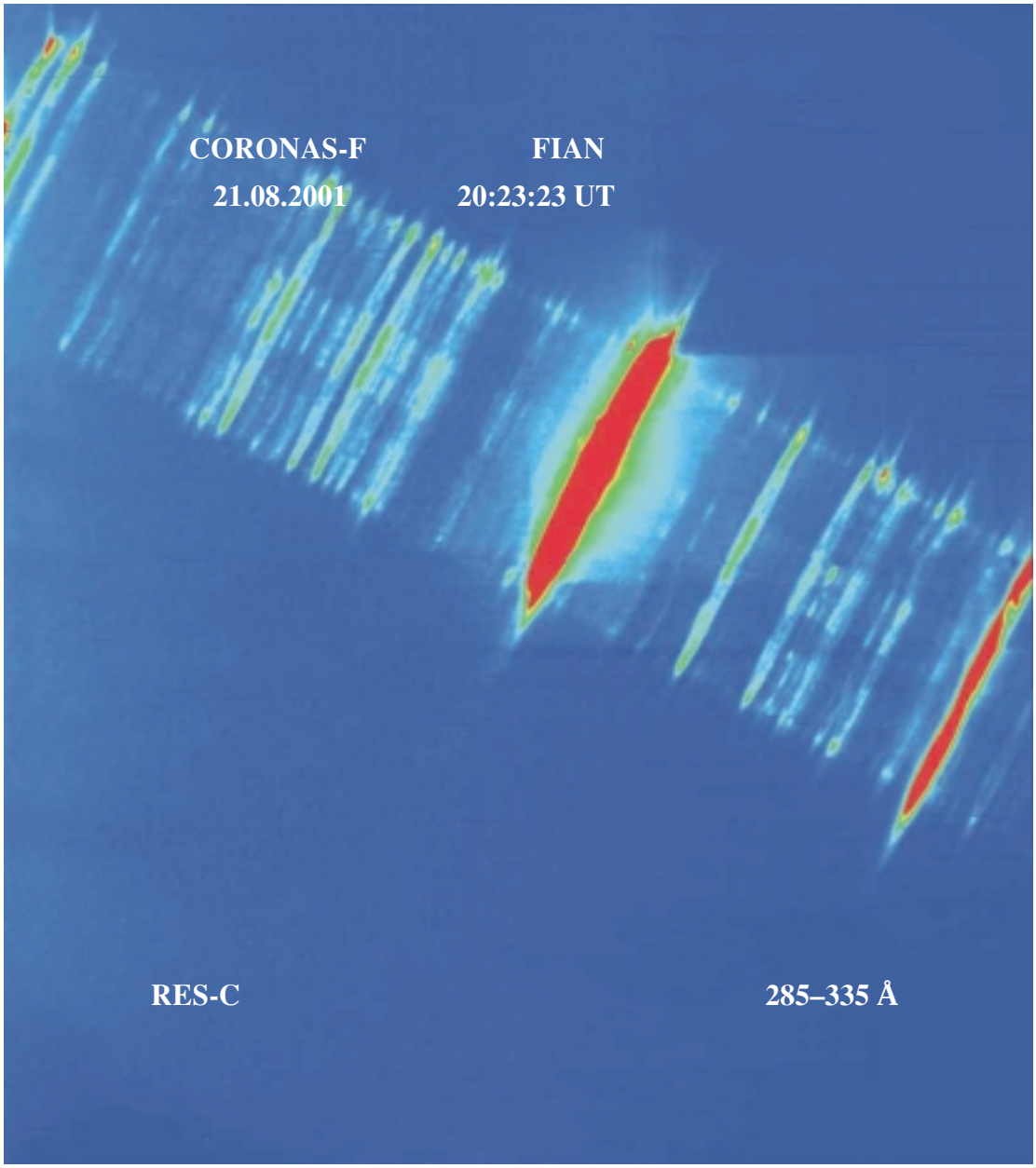


Fig. 5. The spectroheliogram in the range 285–335 Å taken in the XUV channel of the RES-C instrument.

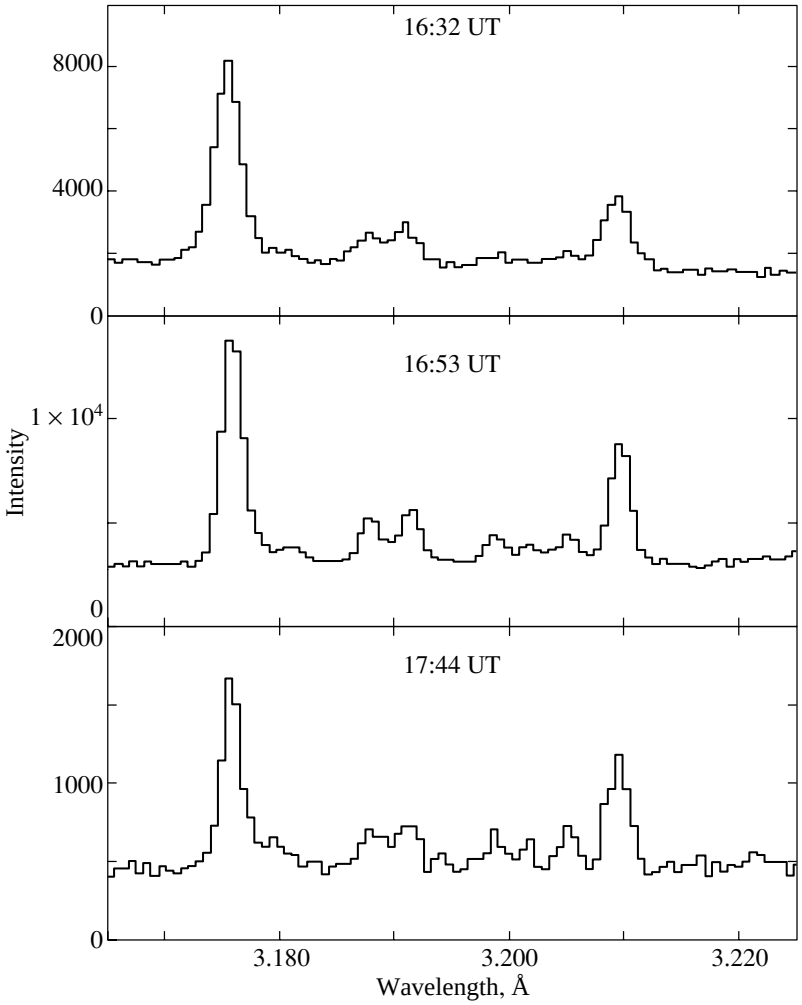


Fig. 6. Ca XIX line spectra with a high spatial resolution ($\sim 5''$) at the growth phase, intermediate phase, and the decay phase of the strongest flare in the current solar cycle (August 25, 2001).

The SPR-N Solar Spectropolarimeter

The SPR-N solar X-ray spectropolarimeter measures the polarization of continuum X-ray radiation from solar flares in the spectral range 20–100 keV (nonthermal radiation from the accelerated electrons generated in flares). The instrument has the following characteristics:

Spectral range	20–100 keV
Subbands	20–40 keV
	40–60 keV
	60–100 keV
Polarization detection sensitivity	~5% at flux 10^{-6} erg cm^{-2} s^{-1}
Temporal resolution	4–16 s

The SPR-N instrument has two detection channels: a polarization detector and a patrol detector. The polarization detector contains an X-ray radiation scatterer (a hexahedral Be prism) and a system of three pairs of scintillation detectors (SSD), which measure the intensity of the scattered radiation. CsI(Na) crystal scintillators with photomultipliers are used to record X-ray radiation. To eliminate the systematic errors during polarization measurements attributable to the possible sensitivity drift in individual channels, the polarization detector was mounted on a turning drive. For typical solar flares of importance I, II, and III, the expected sensitivity of polarization measurement is ~5% at a flux of $\sim 10^6$ phot. cm^{-2} s^{-1} ($E = 20$–100 keV) and an integration time of ~8 s. The patrol detector is designed to measure the total intensity of the X-ray radiation from solar flares at energies 20–100 keV. It consists of a CsI(Na) crystal placed behind an anticoincidence shield and a photomultiplier.

Apart from the general scientific program of the

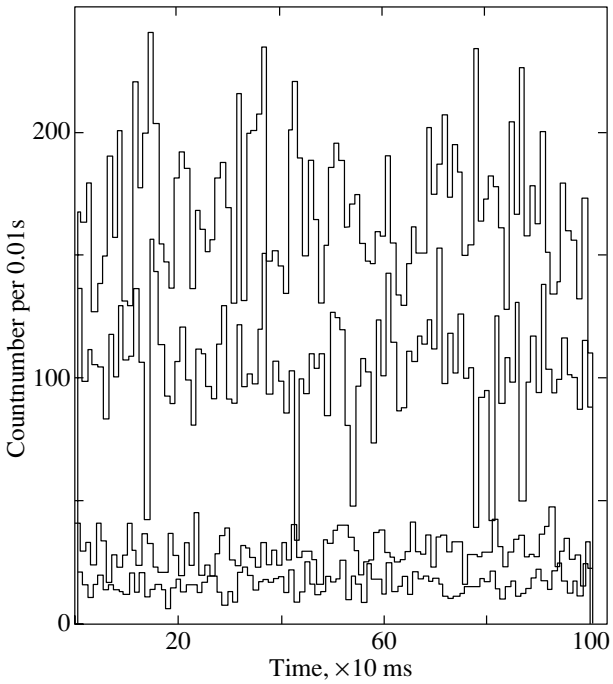


Fig. 7. An intensity record of hard X-ray radiation on September 27, 2001, 18:26 UT in the energy range 20–100 keV with a high temporal resolution in four energy channels (from top to bottom: 20–25, 25–35, 35–100, and 100–200 keV).

CORONAS-F project, the programs of joint observations with instruments on other spacecraft that will operate simultaneously with CORONAS-F, primarily with instruments on the SOHO international spaceborne observatory, the TRACE telescope, YOHKOH, and others, are envisaged. It is planned that the CORONAS-F observations will be coordinated with observations at ground-based observatories, in particular, with the Institute of Solar–Terrestrial Physics (Itkutsk), the Crimean Astrophysical Observatory, and other ground-based observatories that carry out solar observations.

Test experiments and preliminary scientific results obtained during the first three months of observations show that the entire instrumentation of the SRT–RES–SPR operates normally. Figure 4 shows the solar images at 171, 195, 284, and 304 Å obtained in channel BP1 of the SRT-C instrument. Figure 5 shows the spectroheliogram in the range 285–335 Å taken in the XUV channel of the RES-C instrument.

THE DIOGENESS SPECTROPHOTOMETER AND THE RESIK X-RAY SPECTROMETER

The RESIK and DIOGENESS instruments measure the solar spectra in the wavelength range 3–7 Å. They are designed to investigate the X-ray radiation from active regions and flares with a high spectral resolution. Their solar X-ray spectra are comparable in spectral and temporal resolution to the best observations performed to date. Because of the CORONAS-F polar orbit and the complete coverage of a wide spectral range, they are a valuable complement to the Yohkoh spectra. Since the detectors of the RESIK instrument and the Yohkoh spectrometers are saturated at high X-ray fluxes, the DIOGENESS instrument is currently the world's only operating spectrometer that takes the spectra of intense flares of importance above M2 (as exemplified by the X5.3 flare on August 25, 2001).

Figure 6 shows Ca XIX line spectra with a high spatial resolution ($\sim 5''$) at the growth phase, intermediate phase, and the decay phase of the strongest flare in the current solar cycle (August 25, 2001). The temperature of the emitting plasma decreases from 2.5×10^6 K at the growth phase to 1.2×10^6 K at the decay phase. The spectra exhibit changes in the line width, which points to the role of plasma turbulence during the flare evolution. A detailed analysis of the relative intensities will allow one to study the energy balance in flares and the role of non-Maxwellian and nonequilibrium processes in the flare energy release region. In the spectral range in question (3–7 Å), such spectra for an X flare were obtained with such a high spatial resolution for the first time. They provide information on the many emission lines produced by the collisional excitation of atoms, the excitation of inner atomic shells, and dielecton recombination.

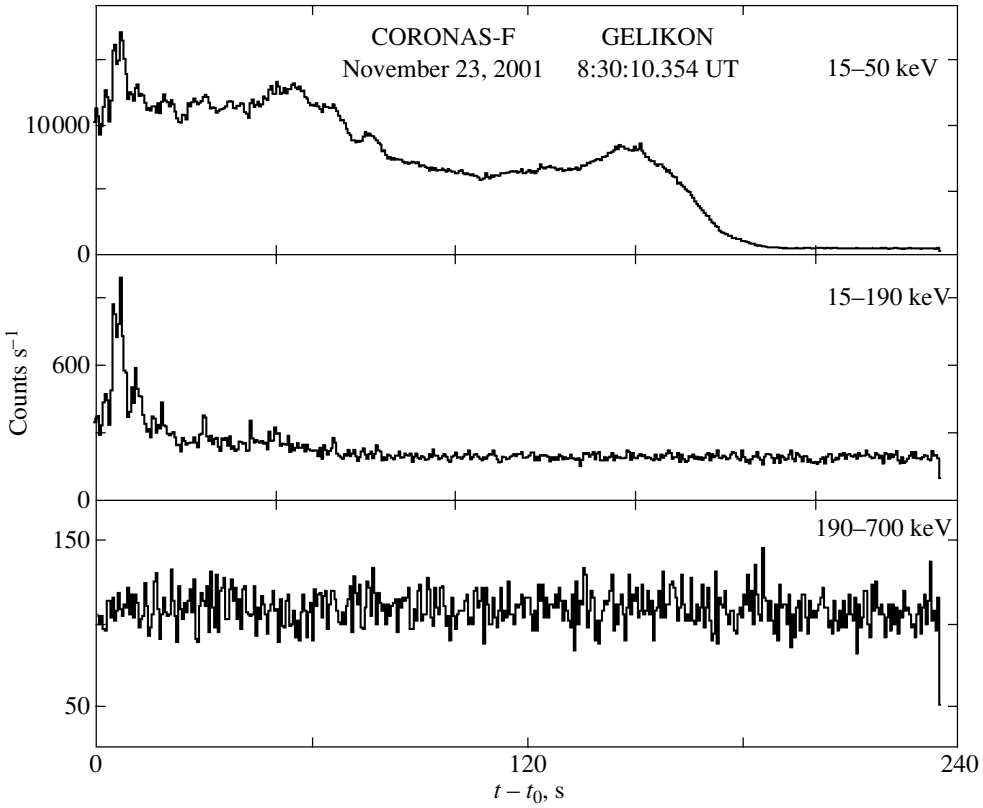


Fig. 8. The time profile for the intense solar flare of November 23, 2001, in three energy bands with a high temporal resolution, from 2 to 256 ms.

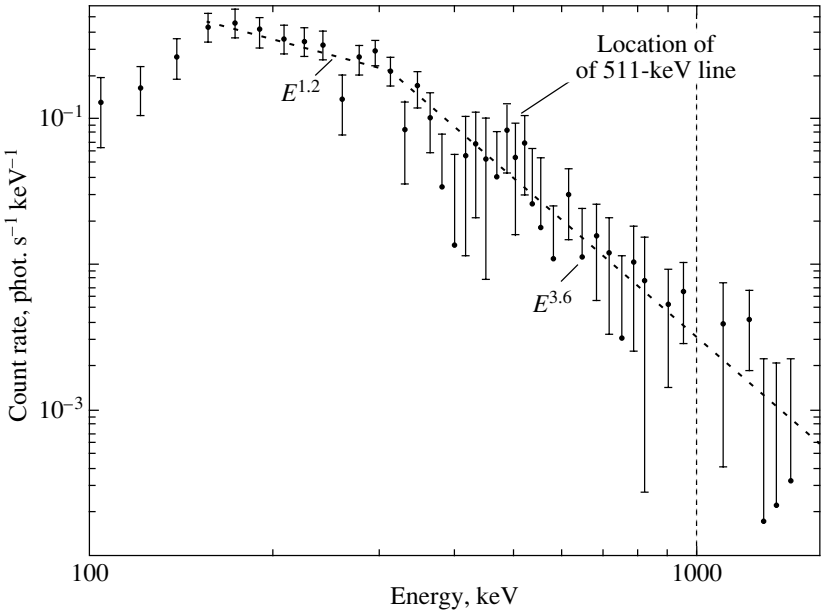


Fig. 9. The differential energy spectrum for the September 18, 2001 event.

THE IRIS FLARE SPECTROMETER

The instrument is designed to investigate the solar flare activity in the X-ray spectral range 2–200 keV. It has 12 energy channels with a temporal resolution of 2.5 s in patrol mode and 64 energy channels with a temporal resolution of 1 s and four channels with a 10 ms resolution in burst mode. As yet no spectral measurements have been carried out with such a temporal resolution in any experiment aimed at investigating the solar X-ray radiation. Observations

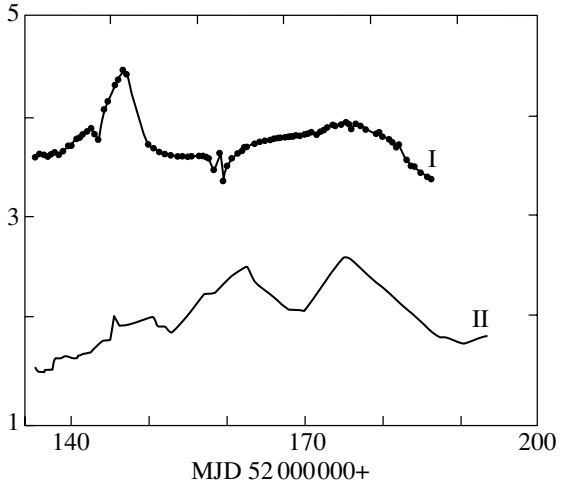


Fig. 10. The VUSS-L temporal record of a signal proportional to the solar UV flux (I) and the solar radio flux at a wavelength of 10.7 cm (II).

of the fine temporal structure in hard X-ray fluxes will provide direct information on the development of the energy release at the explosive phase of flares. Figure 7 shows an intensity record of hard X-ray radiation in four energy channels in the range 20–100 keV with a high temporal resolution for the event of September 27, 2001, 18:26 UT.

THE GELIKON GAMMA-RAY SPECTROMETER

The instrument measures the temporal and spectral characteristics of the hard electromagnetic radiation from solar flares over a wide energy range, from X-rays to gamma-rays (10 keV–10 MeV). It is used to monitor the radiation conditions and soft solar flares and to detect and record in detail hard ($E\gamma > 50$ keV) flares and gamma-ray bursts. Over the observing period from November 14 through December 14, 2001, the temporal profiles and spectra were measured for 134 solar flares. Most of them had soft spectra and were recorded in background mode with a temporal resolution of 1 s in eight adjacent energy bands from 15 to 200 keV. Figure 8 shows the temporal profile for the intense solar flare of November 23, 2001, detected in trigger mode in three energy bands with a high temporal resolution from 2 to 256 ms.

THE RPS-1 X-RAY SPECTROMETER

The instrument is designed to investigate solar flares and their precursors in the X-ray range 3–30 keV. Over the period of measurements, the instrument calibration parameters were refined, the background conditions were mapped, and the experimental data statistics was accumulated. No SOHO measurements are carried out in this energy range, while the SXS/Yohkoh results have not yet been published.

THE ATS AMPLITUDE–TIME SPECTROMETER

The instrument is designed to study the X-ray and gamma-ray radiation from solar flares, in particular, to analyze solar-flare or gamma-ray burst events in the energy bands 3–30 keV, 0.1–8.0 MeV, and 2.0–80.0 MeV. Over the observing period, several tens of candidates for solar-flare or gamma-ray burst events were recorded and their energy spectra were obtained for various phases of the event (before, during, and after the event). The duration of the candidate events ranges from 16 s to 3–5 min. In Fig. 9, the differential energy spectrum for the event of September 18, 2001, was reconstructed from measurements. It exhibits a kink near energy 300 keV and a possible spectral feature that can be identified with the 511-keV annihilation line.

THE SUVR-Sp-C SOLAR ULTRAVIOLET RADIOMETER

The SUVR-Sp-C instrument measures the fluxes from the Sun as a star in several spectral bands, from 1 to 130 nm. In contrast to SOHO, it has a *wider* spectral range. Apart from the scientific objectives, the radiometer monitors one of the most important elements of cosmic *weather*—the geoeffective solar radiation. the SUVR obtained data on the intense X5.3B flare of August 25 at 16:23 UT at X-ray wavelengths $\lambda < 12$ nm.

THE VUSS-L ULTRAVIOLET SOLAR SPECTROPHOTOMETER

The VUSS-L instrument carries out measurements in a band near 120 nm and is designed to investigate the solar ultraviolet radiation near the L_α resonance line. The measurements yield numerous data on the UV flux and intensity in the L_α line, which was tentatively estimated to be $\sim$10 erg cm^{-2} s^{-1}. The VUSS-L temporal record of the UV flux is shown in Fig. 10.

THE SCR INSTRUMENTATION TO STUDY SOLAR COSMIC RAYS

The SCR scientific instrumentation consists of three instruments: SONG (gamma-ray and neutron spectrometer, principal investigator S.P. Ryumin), CRM (cosmic-ray monitor, principal investigator S.N. Kuznetsov), and CRS-3 (cosmic-radiation spectrometer, principal investigator A.F. Podorol'skii).

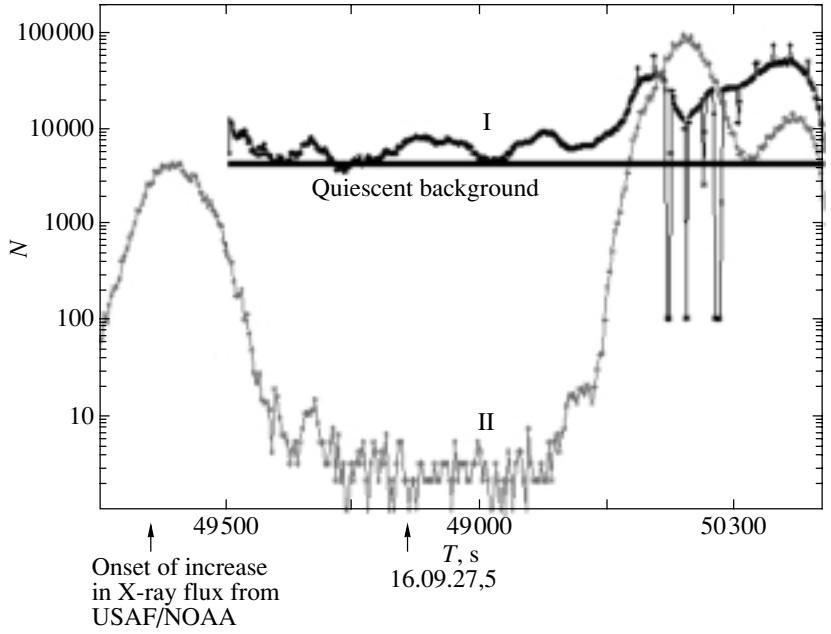


Fig. 11. The recording of gamma rays with energy 60–150 keV (I) and electrons with energy 0.3–0.6 MeV (II) by the SCR complex in the November 4, 2001 event.

It is designed for comprehensive studies of solar cosmic rays. The SONG instrument records X-ray and gamma-ray spectra in the energy range 0.03–100 MeV, detailed gamma-ray line spectra in the range 0.3–20 MeV, neutrons with energies >20 MeV, and fluxes of charged cosmic-ray particles—protons with energies >70 MeV and electrons with energies >50 MeV. The SRM instrument measures the fluxes and spectra of protons with energies 1–200 MeV and electrons with energies 0.5–12 MeV. The CRS-3 measures the chemical composition and spectra of ions in the range $Z = 1$–10 and in the energy range 1.5–20 MeV per nucleon for He and 4–40 MeV per nucleon for Ne. In comparison with the SOHO and Yohkoh instruments, the SONG spectrometer is capable of recording energetic (with energies up to 100 MeV) gamma-rays, which, in turn, makes it possible to observe gamma-rays from the decay of the π^0 mesons produced in interactions of high-energy protons. The CORONAS-F measurements show that the SONG background for recording solar neutrons is at least a factor of 5 to 7 lower than that of GRS (SMM) — the only instrument that recorded solar neutrons in the same energy range. Figure 11 shows the SONG recording of gamma-rays with energy 60–150 keV (I) and electrons with energy 0.3–0.6 MeV (II). The first increase in gamma-ray flux correlates with the electron flux and can be explained in part as electron bremsstrahlung. The two remaining increases represent solar flare radiation. Since the flare occurred in the western hemisphere, the recorded increase in relativistic particle flux began almost immediately.

ACKNOWLEDGMENTS

V.D. Kuznetsov and I.A. Zhitnik played a major role in designing the CORONAS-F complex. They also helped in writing this paper, for which we thank them. We are also grateful to T.V. Kazachevskaya, Yu.D. Kotov, G.E. Kocharov, S.N. Kuznetsov, N.I. Lebedev, E.P. Masetz, A.A. Nusinov, V.M. Pankov, A.F. Podorol'skii, S.P. Ryumin, and J. Silvestr, who provided the reduced data from the CORONAS-F satellite.

REFERENCES

1. A. Jiménez, P. L. Palle, T. Roca Cortes, and V. Domingo, Astron. Astrophys. **193**, 298 (1988).
2. N. I. Lebedev, V. E. Oraevsky, Y. D. Zhugzda, *et al.*, Astron. Astrophys. **296**, L25 (1995).
3. V. N. Oraevsky, N. I. Lebedev, I. M. Kopaev, *et al.*, oral communication (1999).
4. I. I. Sobelman, I. A. Zhitnik, and A. P. Ignat'ev, Pis'ma Astron. Zh. **22**, 604 (1996) [Astron. Lett. **22**, 539 (1996)].
5. *The SOHO Mission: Scientific and Technical Aspects of the Instruments* (ESA, Paris, 1988), ESA SP-1104.
6. I. Zhitnik, A. Ignatiev, V. Korneev, *et al.*, Proc. SPIE **3406**, 1 (1998).

Translated by V. Astakhov