Constraints on the Photon Charge from Observations of Extragalactic Sources

V. V. Kobychev[1]* and S. B. Popov[2,3]**

[1]*Institute for Nuclear Research, National Academy of Sciences of Ukraine, Kiev, Ukraine*
[2]*University of Padova, Padova, 35122 Italy*
[3]*Sternberg Astronomical Institute, Universitetskiĭ pr. 13, Moscow, 119992 Russia*
Received January 26, 2004

Abstract—Having analyzed high-resolution observations of extragalactic compact radio sources with modern systems of radio telescopes, we obtained an estimate of the upper limit for the photon electric charge, $e_\gamma \lesssim 3 \times 10^{-33}$ of the elementary charge (assuming the photon charge to be energy independent). This is three orders of magnitude better than the limit obtained from radio pulsar timing. We also set a limit on the charge of a gamma-ray (energy ~ 0.1 MeV) photon. In the future, the estimate based on extragalactic sources can be improved significantly. © *2005 Pleiades Publishing, Inc.*

Key words: *fundamental constants, radio sources, extragalactic magnetic fields.*

INTRODUCTION

At present, the most stringent upper limit on the photon electric charge is obtained from the timing of Galactic millisecond radio pulsars. Radio pulses are smeared due to the dispersion of charged photons as they move in the interstellar magnetic field (Cocconi 1988; the result was subsequently refined by Raffelt 1994): $e_\gamma/e < 5 \times 10^{-30}$.

Slightly weaker constraints on the photon charge were discussed by Cocconi (1992) using a different approach based on the angular spread of photons from distant extragalactic sources (which arises from the deviation of a photon with a hypothetical small charge in a magnetic field from rectilinear motion). This method was used to obtain an estimate of $e_\gamma/e < 10^{-27.7} \approx 2 \times 10^{-28}$ by analyzing the motion of photons in the Galactic magnetic field ($B \sim 10^{-6}$ G, a path length of $l \sim 10$ kpc). This bound on the photon charge can be significantly improved primarily by increasing the path length, i.e., when considering the effect in intergalactic fields, and by extending the bandwidth (the Cocconi's limit (1992) is based on observations in a rather narrow bandwidth of ~ 2 MHz).

Another constraint has recently been obtained by studying the properties of the cosmic microwave background (CMB). The existence of a small photon charge would result in charge asymmetry of the Universe and would contribute to the observed CMB anisotropy. Quantitative analysis (Caprini *et al.* 2003) yields a very stringent upper limit, $e_\gamma/e < 10^{-38}$, but only if the charge asymmetries produced by different types of particles are not anticorrelated and, more importantly, if the photons have charges of only one sign. These assumptions make this limit model-dependent.[1]

Note that the best laboratory limit, $e_\gamma/e < 8.5 \times 10^{-17}$ (Semertzidis *et al.* 2003), is much worse than the astrophysical bounds.

CALCULATIONS AND ESTIMATES

Recent observations have revealed that the magnetic fields in many clusters of galaxies reach significant strengths (several microgauss) with a characteristic autocorrelation length of several kpc (see the review by Carilli and Taylor (2002) and references therein). If the photon had a small, but finite electric charge (with all photons having charges of the same sign), then photons with different energies would move along different trajectories in the intracluster magnetic field. This would cause the angular size of a source to increase (of course, an identical but weaker effect must also exist for intercluster fields). In addition, during observations at two different frequencies, the image centroids would be shifted rela-

---

*E-mail: kobychev@kinr.kiev.ua
**E-mail: polar@sai.msu.ru

[1]See also the previous paper by Sivaram (1994) who discussed the limit based on CMB data.

tive to one another,[2] and the photons emitted simultaneously with different energies would reach an observer at different times. If, alternatively, the emitted photons had identical (in absolute value) charges of different signs, the images would be smeared even for monoenergetic radiation. However, the position of the image centroid does not depend on the photon energy in this case.

An ultrarelativistic particle (a photon) with an electric charge e_γ and a momentum $p = h\nu/c$ in a magnetic field with a component B_Y perpendicular to the momentum moves along a trajectory with a radius of curvature r_H (the Larmor radius) equal to $cp/(e_\gamma B_Y)$. The deviation is thus seen to be largest for low-energy radio photons; furthermore, the angular resolution of VLBI radio observations is much better than that in other parts of the electromagnetic spectrum. Thus, one would expect the best limit on the photon charge to be obtainable in the radio frequency range. However, since the effective photon charge can originate from Lorentz invariance violation, we will also discuss the bounds on the charge for high-energy (of the order of MeV) photons.

A photon that travels along an arc of a circumference with a radius r_H is deflected through an angle dl/r_H (in radians) when traversing a distance dl. Therefore, a source at a distance of $l_\star$ from the detector will be seen shifted along the X axis perpendicular to the line of sight (the Z axis) by the angle

$$\varphi_X = \int_0^{l_\star} \frac{dl}{r_H(l)} = \frac{e_\gamma}{h} \int_0^{l_\star} \frac{B_Y(l)dl}{\nu(l)}. \quad (1)$$

The l dependence of the frequency appears for cosmological distances due to the redshift: $\nu(z) = (1+z)\nu_0$.

Two photons with different energies will diverge by the angle

$$\Delta\varphi = \varphi_1 - \varphi_2 = \frac{e_\gamma}{h} \int_0^{l_\star} B_Y \left(\frac{1}{\nu_1} - \frac{1}{\nu_2} \right) dl \quad (2)$$

(assuming the photon charge to be energy independent). Thus, the following constraint on the photon charge can be obtained from observations of a source with an angular diameter $\Delta\varphi$ in a band $\Delta\nu$ (with $\Delta\nu \ll \nu$):

$$e_\gamma/e \lesssim \frac{\Delta\varphi h}{e} \left(\int_0^{l_\star} \frac{B_Y \Delta\nu dl}{\nu^2} \right)^{-1}. \quad (3)$$

---

[2]In this paper, we assume the angles of deviation to be small.

On the other hand, during observations in two widely separated frequency bands ($\nu_1 \ll \nu_2$), we may assume that $\Delta\nu/\nu^2 \approx 1/\nu_1$:

$$e_\gamma/e \lesssim \frac{\Delta\varphi h}{e} \left(\int_0^{l_\star} \frac{B_Y dl}{\nu_1} \right)^{-1}. \quad (4)$$

Here, $\Delta\varphi$ should be considered not as the angular diameter of the source, but as the angular distance between the apparent positions of the source in two different bands. The integrals in Eqs. (3) and (4) can be estimated from observational data on the Faraday rotation of the polarization plane of radio waves that propagate in a magnetized plasma. This quantity can be expressed in terms of the rotation measure, which is defined as the rotation angle of the polarization divided by the wavelength squared. The rotation measure RM can be expressed in terms of the electron density n_e on the line of sight and the longitudinal projection of the magnetic field as follows (Clarke et al. 2001):

$$RM = 8.12 \times 10^5 \int_0^{l_\star} n_e B_Z dl. \quad (5)$$

Here, the distance is in Mpc, n_e is in cm^{-3}, and B_Z is the longitudinal projection of the magnetic field (in microgauss). Assuming the field to be distributed isotropically, without a preferential direction, the longitudinal field projection under the integral can be substituted with the projection onto any other axis, for example, B_Y. Taking this assumption, below we omit the subscript on B.

Above, we disregarded the redshift dependence of the quantities. It would be reasonable to rewrite the formulas using the redshift z rather than the distance as the variable. In addition, the cosmological effects should be taken into account.

The integration element can be written as

$$dl = -\frac{c}{H_0}(1+z)^{-3/2}dz$$

(this equation corresponds to a flat Universe without any dark energy contribution; including this contribution would make the limit more stringent; H_0 is the Hubble constant), $B(z) = B_0(1+z)^2$, $\nu(z) = \nu_0(1+z)$ (Ryu et al. 1998).

Equation (3) then transforms to

$$e_\gamma/e < \frac{\Delta\varphi h}{e} \left[\int_0^{z_\star} \frac{\Delta\nu_0(1+z)}{\nu_0^2(1+z)^2} \right. \quad (6)$$

$$\left. \times B_0(1+z)^2 \frac{c}{H_0}(1+z)^{-3/2}dz \right]^{-1}$$

$$= \frac{\Delta\varphi h}{e} \frac{H_0 \nu_0^2}{c B_0 \Delta\nu_0} \frac{1}{2(\sqrt{1+z_\star}-1)}.$$

For $z_\star \ll 1$, the last fraction in Eq. (6) turns into $1/z_\star$.

In most cases, the line-of-sight magnetic field for extragalactic sources is difficult (or impossible) to estimate with a sufficient accuracy. As an illustration, let us obtain a constraint on the photon electric charge using existing estimates of large-scale extragalactic magnetic fields. When considering the effect of a large-scale magnetic field (with scales larger than the size of a cluster of galaxies), one can use the estimate by Kronberg (1994) for the upper limit on the "cosmologically aligned" magnetic field, $B_0 < 10^{-11}$ G (these data were obtained from the upper bound of 5 rad m^{-2} on any systematic growth of the rotation measure with distance for $z = 2.5$), and the upper limit $B_0 < 10^{-9}$ G for a changing field with a correlation length of ~ 1 Mpc. Widrow (2002) gave the upper limit on the uniform cosmological field component, $B_0 < 6 \times 10^{-12}$ G $(n_e/10^{-5}\text{cm}^{-3})^{-1}$, which agrees with the above estimate by Kronberg. Various studies (see the review by Widrow 2002) indicate that the actual rotational measure cannot be less than the current upper limit by two or three orders of magnitude. Therefore, for our illustrative estimate, we can conservatively take $B_0 > 6 \times 10^{-15}$ G as the lower limit for the "noncompensated" cosmological field.

To estimate what constraints on the charge can be obtained for a photon that travels in such a weak field, we will use actual observations. As part of the VSOP (the VLBI Space Observatory Program), Lobanov et al. (2001) observed the quasar PKS 2215+020 at $\nu_0 = 1.6$ GHz in a bandwidth of $\Delta\nu_0 = 32$ MHz. The angular resolution was about 1 mas. The redshift of the source, $z_\star = 3.57$, corresponds to a distance of $l_\star = 4700$ Mpc. Substituting these values into Eq. (6) and taking $B_0 = 6 \times 10^{-15}$ G and $H_0 = 70$ km s^{-1} Mpc$^{-1} = 2.3 \times 10^{-18}$ s^{-1}, we obtain a bound for the limit on the photon charge,

$$e_\gamma/e \lesssim 6 \times 10^{-29},$$

which is only an order of magnitude worse than the constraint of Raffelt (1994). However, we used a very conservative lower limit for the noncompensated intergalactic magnetic field on the line of sight. Therefore, as another example, let us consider (at the same frequencies and angular resolution) how this limit would improve if the source were observed through a typical cluster of galaxies (relatively close to us, $z \ll 1$). When the cosmological effects are disregarded, the integral in Eq. (3) transforms into $(\Delta\nu/\nu^2)\int B dl$. If the value of the latter integral for our estimate is taken to be $Bl = 1\,\mu$G Mpc (the product of the typical

values of the intracluster field and the size of the central part of the cluster), we then obtain

$$e_\gamma/e < \frac{\Delta\varphi h}{e} \frac{\nu^2}{B l \Delta\nu} = 2 \times 10^{-33}.$$

Thus, observing a source through relatively strong intracluster fields (which, in addition, are known with a better accuracy that the fields outside clusters of galaxies) allows us to improve significantly the limit on the photon electric charge at the same angular resolution, although the path length in the field decreases.

A SPECIFIC EXAMPLE

In the example considered below, the cluster of galaxies used as a scattering screen on the line of sight has $z \ll 1$ (we ignore the influence of the intercluster field). Since the distance dependence of the frequency in Eqs. (3) and (4) may be disregarded at low z, we can directly express the limit on the photon charge in terms of the observed rotation measure by eliminating the distribution of the magnetic field and the electron density along the line of sight from the formulas:

$$e_\gamma/e \lesssim 3.2 \times 10^{-19} \frac{\Delta\varphi h}{e} f(\nu)^{-1} \frac{812 h_{70}^{1/2}}{RM}, \quad (7)$$

where $f(\nu) = \Delta\nu/\nu^2$ or $1/\nu_1$, depending on the relationships between the frequencies. We used Eq. (5) and expressed the electron density as $n_e = 10^{-3} h_{70}^{1/2}$ cm^{-3} (Clarke et al. 2001).[3]

For observations in widely separated frequency bands ($\nu_1 \ll \nu_2$), this formula can be rewritten as

$$e_\gamma/e \lesssim 1.8 \times 10^{-32} h_{70}^{1/2} \left(\frac{\Delta\varphi}{0\rlap{.}''001}\right) \quad (8)$$

$$\times \left(\frac{\nu_1}{1\,\text{GHz}}\right) \left(\frac{RM}{1\,\text{rad/m}^2}\right)^{-1}.$$

Let us consider the compact radio source 3C 84 in the galaxy NGC 1275 located near the center of the Perseus cluster (Abell 426, $z = 0.018$). This source was observed by Scott et al. (2004), who surveyed 102 active galactic nuclei at 5 GHz as part of the VSOP project (the VLBI network of telescopes, including the HALCA satellite antenna with an apogee of 21 400 km). The smallest (in angular size) of the six observed components of 3C 84 has an isophotal diameter at half maximum (FWHM) of 0.8 mas.

---

[3]Note that the Hubble constant h_{70} appears in the formulas only with the subscript that denotes the normalization to 70 km s^{-1} Mpc^{-1}, while the Planck constant h is written everywhere without a subscript.

Since these authors provided the measurement error of the angular diameter, 10%, we conservatively take 0.9 mas as a more reliable estimate of the angular diameter. The central frequency and the bandwidth are 4.8 GHz and 32 MHz, respectively.

The rotation measure for 3C 84 was measured by Rusk (1988) (cited by the review of Aller *et al.* 2003): $RM = +76$ rad m^{-2}. In addition, the Perseus cluster is a source of polarized dispersed radio emission at 350 MHz (Brentjens and de Bruyn 2003) with $RM \sim 25-90$ rad m^{-2}, including the cluster outskirts. Therefore, although part of the rotation measure on the line of sight to 3C 84 was gained in regions with a high electron density at the cluster center, we can safely assume that at least 25 rad m^{-2} was gained in the outer cluster regions. According to the model of the radial n_e distribution in the Perseus cluster (formula (4) from Churazov *et al.* 2003), the electron density outside the central sphere with a radius of 0.3 Mpc is low ($\lesssim 10^{-3}$ cm^{-3}) and depends weakly on the distance from the cluster center.

Let us first consider the case where the photons have charges of opposite signs. In this case, a point source will be smeared even during observations at a single frequency, so we can use Eq. (8). Substituting $\nu_1 = 4.8$ GHz, $RM = 25$ rad m^{-2}, and $\Delta\varphi = 0.9$ mas, we obtain a limit on the absolute value of the photon charge,

$$e_\gamma/e \lesssim 3 \times 10^{-33}.$$

For photon charges of the same sign, the broadening of the source results from the photon energy difference, i.e., the finite receiver bandwidth ($\Delta\nu = 32$ MHz), and the effect will be weaker. For this case, using $f(\nu) = \Delta\nu/\nu^2$, we obtain the following limit from Eq. (7):

$$e_\gamma/e \lesssim 4 \times 10^{-31}.$$

DISCUSSION

Various methods can be used to set limits on the photon charge. They can be associated with different techniques and observations in different spectral ranges.

VLBI observations of close pairs of sources at several frequencies can give a stringent constraint. In this case, the angular distance between the sources can be measured with a high accuracy: as high as several tens of μas (Bartel 2003). Multifrequency observations of two sources with different redshifts (such as those performed by Rioja and Porcas (2000)) can yield important upper limits.

Cocconi (1992) also obtained fairly weak constraints from optical and X-ray data on the angular

dispersion: $e_\gamma/e < 10^{-25.4}$. This author used data on the galactic magnetic fields. The limit can be improved significantly by using currently available data on the intergalactic magnetic fields.

Observations of cosmic gamma-ray bursts (GRBs) with known redshifts (these data were lacking at the time the papers by Cocconi (1988, 1992) and Raffelt (1994) were published) cannot seriously compete with pulsar timing data and data on the smearing of radio sources. However, as we noted above, the possible energy dependence of the photon charge makes obtaining limits on the charge over a wide energy range justifiable.

The time delay (the dispersion for gamma-ray photons in the interstellar medium may be disregarded) can be written as (Barbiellini and Cocconi 1987)

$$\Delta t = \frac{e_\gamma^2 B^2 l_\star^3}{24cE^2}.$$

Here, the delay is calculated relative to the arrival time of photons with energies much higher than E, since the delay for them, which is inversely proportional to the square of the energy, is negligible. If the observations are performed in a narrow energy range, $\Delta E \ll E$, then a different formula is applicable:

$$\Delta t = \frac{e_\gamma^2 B^2 l_\star^3}{12cE^2}\frac{\Delta E}{E}.$$

Both formulas are applicable to photons with charges of the same and different signs.

The duration of the GRB leading edge, which can be shorter than 1 ms ($\sim 200-250$ μs; Schaefer and Walker 1999), may be taken as an estimate of the maximum time delay.

We then obtain (for $\Delta E/E = 0.5$ with the cosmological effects disregarded)

$$e_\gamma/e < 5.6 \times 10^{-21}$$
$$\times \left(\frac{E}{100\text{ keV}}\right)\left(\frac{B}{6 \times 10^{-15}\text{ G}}\right)^{-1}$$
$$\times \left(\frac{\Delta t}{0.1\text{ ms}}\right)^{1/2}\left(\frac{l_\star}{1000\text{ Mpc}}\right)^{-3/2}.$$

Here, we normalize the magnetic field to the lower limit on the uniform component of the extragalactic field without including the chaotic field component, for which there are no reliable estimates. As in the case considered above, the constraint can be improved significantly if a GRB is observed through a cluster with a known magnetic field.

CONCLUSIONS

To summarize, we can say that, at present, observations of extragalactic radio sources yield the most stringent constraints on the absolute value of the photon electric charge, $e_\gamma/e \lesssim 3 \times 10^{-33}$ (assuming the photons with unlike charges to be emitted with an equal probability and the photon charge to be energy independent). These constraints can be improved by VLBI observations of close pairs of compact sources at widely separated frequencies through a cluster of galaxies with a known magnetic field, since the accuracy of measuring the angular distance between close sources can reach 10 μas. In addition, it is highly desirable to use data on several sources; this would allow a limit on the photon charge to be set from statistical analyses.[4] In the future, with the launch of new spaceborne radio antennas, the angular resolution of observations is likely to increase (Bartel 2003; Fomalont and Reid 2004). Therefore, the most stringent upper limits on the photon charge might be expected precisely from highly accurate observations of extragalactic radio sources.

ACKNOWLEDGMENTS

REFERENCES

1. M. F. Aller, H. D. Aller, and P. A. Hughes, Astrophys. J. **586**, 33 (2003); astro-ph/0211265.
2. G. Barbiellini and G. Cocconi, Nature **329**, 21 (1987).
3. N. Bartel, Astron. Latin America, ADeLA Publ. Ser. **1**, 35 (2003); astro-ph/0303342.
4. M. A. Brentjens and A. G. de Bruyn, *The Riddle of Cooling Flows in Galaxies and Clusters of Galaxies*, *Charlottesville, VA, USA, May 31–June 4, 2003*, Ed. by T. H. Reiprich, J. C. Kempner, and N. Socker, Publ. electronically at http://www.astro.virginia.edu/coolflow/proc.php.
5. C. Caprini, S. Biller, and P. G. Ferreira, hep-ph/0310066 (2003).
6. C. L. Carilli and G. B. Taylor, Annu. Rev. Astron. Astrophys. **40**, 319 (2002); astro-ph/0110655.
7. E. Churazov, W. Forman, C. Jones, and H. Böhringer, Astrophys. J. **590**, 225 (2003).
8. T. E. Clarke, P. P. Kronberg, and H. Böhringer, Astrophys. J. **547**, L111 (2001).
9. G. Cocconi, Phys. Lett. B **206**, 705 (1988).
10. G. Cocconi, Am. J. Phys. **60**, 750 (1992).
11. E. Fomalont and M. Reid, astro-ph/0409611 (2004).
12. P. P. Kronberg, Rep. Prog. Phys. **57**, 325 (1994).
13. A. P. Lobanov, L. I. Gurvits, S. Frey, *et al.*, Astrophys. J. **547**, 714 (2001).
14. G. Raffelt, Phys. Rev. D **50**, 7729 (1994); hep-ph/9409461.
15. M. J. Rioja and R. W. Porcas, Astron. Astrophys. **355**, 552 (2000); astro-ph/0002097.
16. R. E. Rusk, Ph. D. Thesis (Univ. Toronto, 1988).
17. D. Ryu, H. Kang, and P. L. Biermann, Astron. Astrophys. **335**, 19 (1998).
18. B. E. Schaefer and K. C. Walker, Astrophys. J. Lett. **511**, L89 (1999).
19. W. K. Scott, E. B. Fomalont, S. Horiuchi, *et al.*, Astrophys. J., Supp. Ser. (2004) (in press); astro-ph/0407041.
20. Y. K. Semertzidis, G. T. Danby, and D. M. Lazarus, Phys. Rev. D **67**, 017701 (2003).
21. C. Sivaram, Am. J. Phys. **63**, 1473 (1994).
22. L. M. Widrow, Rev. Mod. Phys. **74**, 775 (2002); astro-ph/0207240.

Translated by A. Astakhov

---

[4] The importance of statistics was pointed out by an anonymous referee.

Global Properties of Nearby Galaxies in Various Environments

I. D. Karachentsev[1*] and A. V. Kasparova[2]

[1]*Special Astrophysical Observatory, Russian Academy of Sciences, Nizhniĭ Arkhyz,*
357147 Karachai-Cherkessian Republic, Russia

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

Received August 4, 2004

Abstract—We analyze a sample of the Local Volume that contains 451 galaxies within 10 Mpc. We compare the various global parameters of these galaxies with their tidal index that characterizes the local density of the environment. The closest correlation is observed between the density of the galaxy's environment and its morphological type. The abundance of neutral hydrogen in the members of close groups was found to be, on average, a factor of 3 lower than that in isolated galaxies. However, much of this difference is attributable to different morphological composition for the group members and field galaxies. The total mass-to-luminosity ratio is virtually independent of the tidal index of the galaxy, which indirectly indicates a low percentage of tidal systems among dwarf galaxies. All of the galaxies with three or more companions in the Local Volume are shown to have masses above the threshold value of $10^{10} M_\odot$.
© *2005 Pleiades Publishing, Inc.*

Key words: *dwarf galaxies, galaxy evolution, interstellar gas in galaxies.*

INTRODUCTION

Elliptical and spheroidal galaxies are known to have distinctly different distributions in space: galaxies of early morphological types are more commonly encountered in rich clusters and compact groups, while late-type objects dominate in the general field. This morphological segregation is obviously caused by the collective interaction of galaxies, including their mergers, which gives rise to spherical subsystems (bulges) of galaxies in dense environments (Dressler *et al.* 1984). Apart from normal and giant galaxies, dwarf spheroidal (dSph) galaxies, which are encountered almost exclusively in clusters and groups, exhibit strong morphological segregation (Karachentseva and Sharina 1987), whereas dwarf irregular galaxies are found in both dense and rarefied regions. It is commonly assumed that low-mass galaxies in groups and clusters easily lose their gas, exhausting their resources for new star formation, and transform from irregulars into spheroidals. Another viewpoint assumes the formation of new dwarf galaxies from tidal tails and bars. In this picture, the fraction of tidal dwarfs must increase with density of the environment.

According to Giovanelli and Haynes (1991) and many other authors, the abundance of neutral hydrogen (H I) in the disks of galaxies depends on their environment: spiral galaxies in the central regions of rich clusters show a high H I underabundance compared to field galaxies of the same morphological type. However, the situation with a H I deficiency in groups of galaxies so far appears less certain.

The above examples show that an analysis of the dependence of various global parameters of galaxies on their environment can serve as a useful tool for studying their dynamical evolution. This requires that various selection effects be taken into account; ignoring them can lead to false conclusions about the pattern of evolution of galaxies.

THE TIDAL INDEX AND THE ENVIRONMENT OF GALAXIES

A sample of galaxies in a fixed volume of space is best suited to properly comparing the global parameters of galaxies in various environments. Using a distance-limited rather than magnitude-limited sample of galaxies minimizes the various selection effects. The recently published CNG catalog of nearby galaxies (Karachentsev *et al.* 2004), which contains observational data for 451 galaxies within 10 Mpc, is most suitable in this sense.

The commonly used descriptive scale—a field galaxy, a group member group, and a cluster member—is simplified and subjective. Therefore, Karachentsev and Makarov (1999) introduced the tidal index to quantify the environment of a galaxy i:

$$\mathrm{TI}_i = \max\{\log(M_k/D_{ik}^3)\} + C, \quad k = 1, 2, \ldots, N,$$

where M_k and D_{ik} are the mass and spatial distance of the neighboring galaxy, respectively. In this case, for each galaxy, we can find the most significant galaxy that exerts the strongest tidal effect on the galaxy considered. The arbitrary constant C was chosen in such a way that at TI = 0 the cyclic Keplerian period of the galaxy with respect to its significant neighbor

$$t_{ik} = D_{ik}^{2/3} \times G^{-1/2} \times (M_i + M_k)^{-1/2}$$

was equal to the age of the Universe, $1/H_0$, where G is the gravitational constant, and H_0 is the Hubble parameter. It can be easily seen that TI also characterizes the overdensity produced by the significant neighbor, in units of the critical density $\rho_c = 3H_0^2/8\pi G$. For this definition of the tidal index, galaxies with TI < 0 may be treated as isolated (field) galaxies that have not interacted with their significant neighbors over the cosmological time.

The tidal index for each galaxy with $D < 10$ Mpc and the name of its significant neighbor are given in the CNG catalog. The TI range in the catalog, from -3 to $+7$, corresponds to a local density difference of about 10 orders of magnitude, which is important in studying the effects of the environment.

THE HYDROGEN DEFICIENCY

The abundance of neutral hydrogen in galactic disks can be characterized by various parameters. Roberts and Haynes (1994) showed that the hydrogen mass in normal and dwarf galaxies, M_{HI}, is proportional to the square of their linear optical diameter, A_{25}, so the mean surface brightness $S_{HI} = 4M_{HI}/\pi A_{25}^2$ is approximately constant. The mean hydrogen surface brightness for nearby galaxies is plotted against the tidal index in Fig. 1. The solid and two dashed lines correspond to the sliding median and sliding quartiles of the sample, respectively. Several early-type dwarf galaxies in the Local Group (Fornax, Sculptor, NGC 185, NGC 205) have weak but detectable HI flows, thereby introducing a certain selection effect in the $\{S_{HI}, \mathrm{TI}\}$ diagram. Using the median instead of the mean partially suppresses the contribution of these nearby dwarfs with positive TI. As we see from Fig. 1, the mean surface brightness S_{HI} decreases smoothly from isolated galaxies to group members. The mean HI brightness for galaxies with the largest tidal indices is a factor of 3 lower than that for isolated galaxies. An identical but slightly less pronounced trend can be seen in Fig. 2, where the hydrogen mass-to-luminosity ratio (in solar units) was taken as a measure of the hydrogen abundance. High $M_{HI}/L > 3M_\odot/L_\odot$ are usually encountered in isolated galaxies. The isolated dwarf

system KKs40, which was recently studied in detail by Warren $et\ al.$ (2004), has a record high hydrogen mass-to-luminosity ratio. Another dwarf galaxy, Ho IX, a companion of M 81, may have a falsely high M_{HI}/L due to the confusion of the H I flows from it and from M 81. As we see from the figure, when the mean density of the galaxy environment increases by seven orders of magnitude, its mean hydrogen mass-to-luminosity ratio decreases by a factor of about 2.

The hydrogen deficiency in a galaxy can also be characterized by yet another parameter: the hydrogen-to-total mass ratio within its standard radius, M_{HI}/M_{25}, which can be determined from the rotation curve or from the HI line width. Figure 3 shows the distribution of 300 nearest galaxies in this dimensionless parameter and tidal index. The M_{HI}/M_{25} ratio for half of the galaxies lies within the range 1 to 1/10 with a median of 0.4. However, there are galaxies in which the hydrogen mass exceeds the total (indicative) mass within the standard radius, i.e., within the 25^m arcsec^{-2} isophote, by an order of magnitude. Such "fertile" galaxies are grouped in the range $[-2 < \mathrm{TI} < 2]$. It can be concluded from Fig. 3 that the difference between galaxies in M_{HI}/M_{25} correlates weakly with the density of their environment; i.e., the conditions for their initial formation rather than tidal interactions are mainly responsible for the observed variety of galaxies in terms of the hydrogen mass fraction.

MORPHOLOGICAL SEGREGATION

Figure 4 shows the distribution of 451 nearby galaxies in morphological type and tidal index. The sliding median shows the systematic decrease in tidal index from TI $\simeq +2$ for elliptical galaxies to TI $\simeq -0.5$ for Sd, Sm, and Ir galaxies; in other words, the characteristic density of the environment for elliptical galaxies is a factor of about 300 higher than that for the latest-type galaxies.

The table gives an idea of the statistical significance of the dependences of a particular parameter for a galaxy on its tidal index. The first two columns list the variables Y and X the relationship between which is expressed by the linear regression $Y = A + B \times X$. The third column gives the number of galaxies in the sample; the fourth column lists the correlation coefficients; and the fifth and sixth columns give the slope and standard error of the regression, respectively. We see from these data that the most significant correlation is observed between the tidal index and the morphological type of a galaxy. A further analysis shows that the dependence of the type of a galaxy on its environment is mainly responsible for the HI deficiency in groups that we pointed out above.
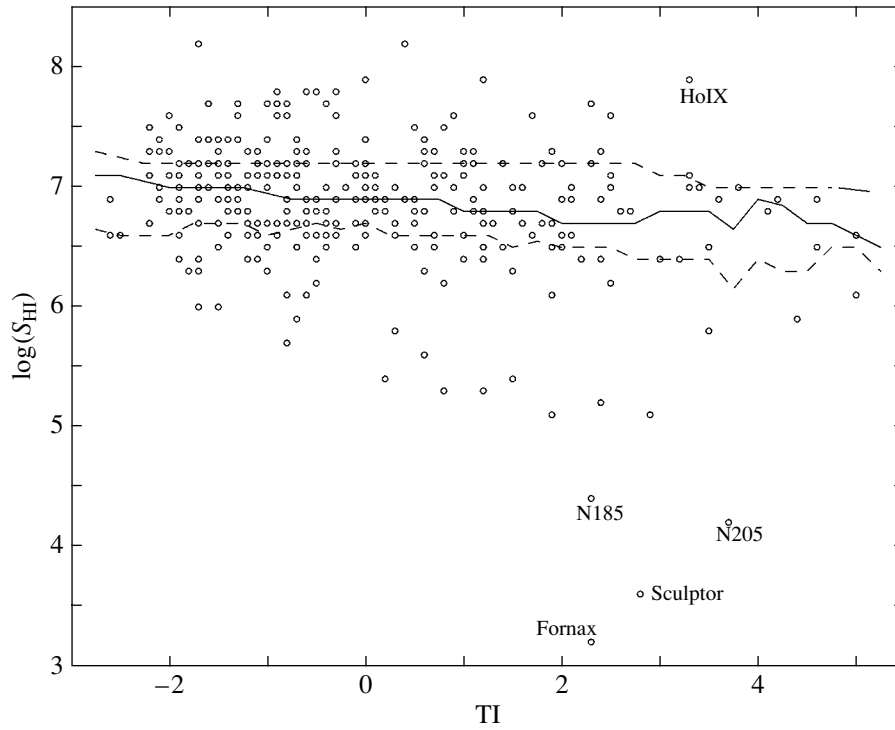
Fig. 1. The distribution of 316 nearby ($D < 10$ Mpc) galaxies in hydrogen surface brightness and tidal index. The solid and two dashed lines correspond to the sliding median and quartiles of the sample.
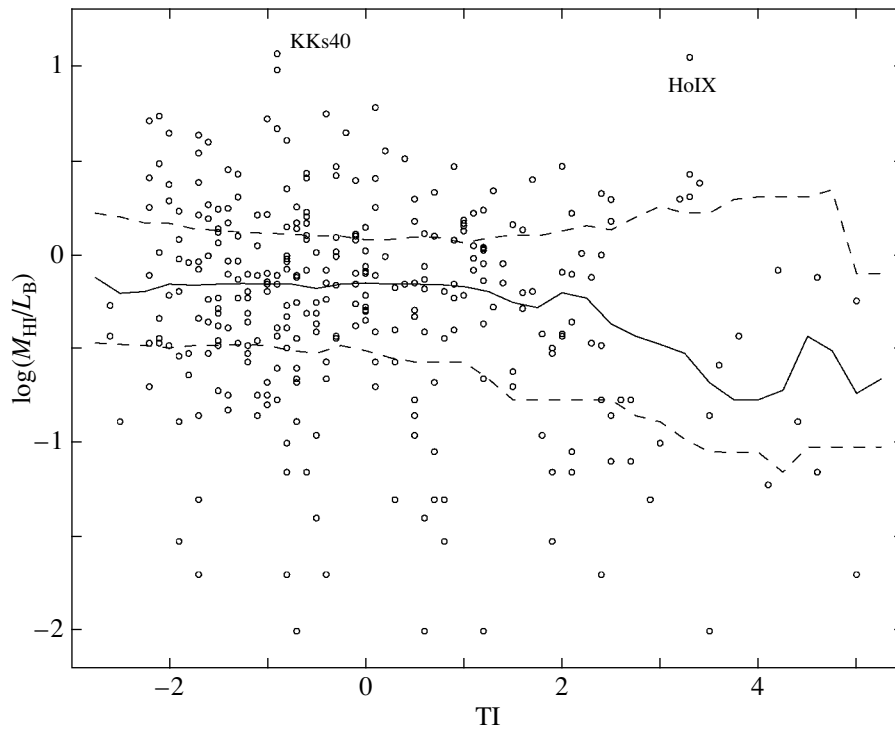
Fig. 2. Ratio of the hydrogen mass to the B luminosity (in solar units) versus tidal index for nearby galaxies. The solid and dashed lines indicate the sliding median and quartiles of the sample.
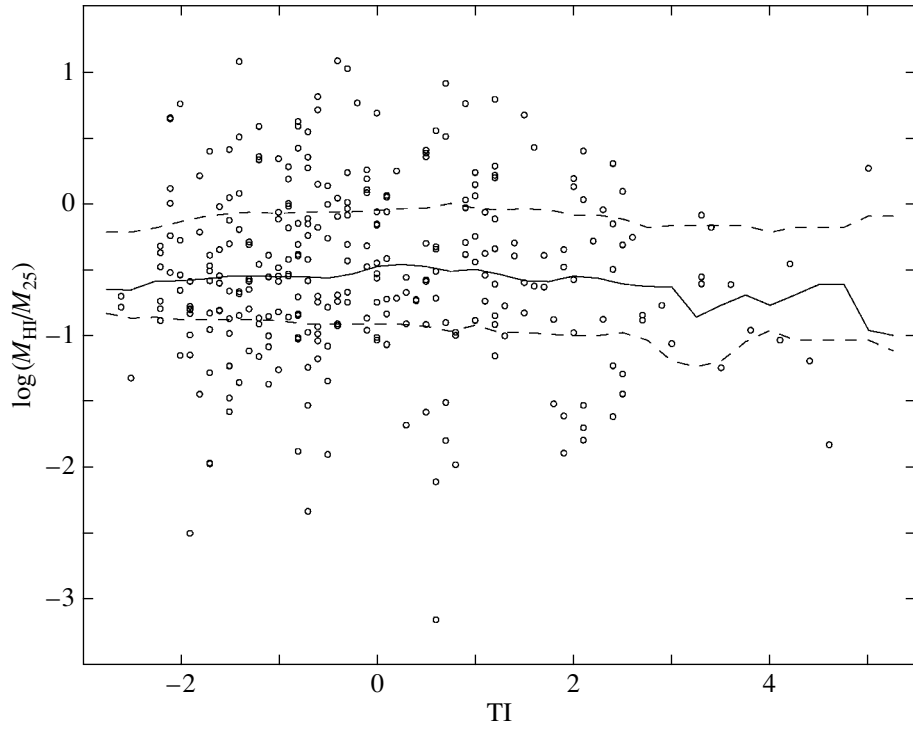
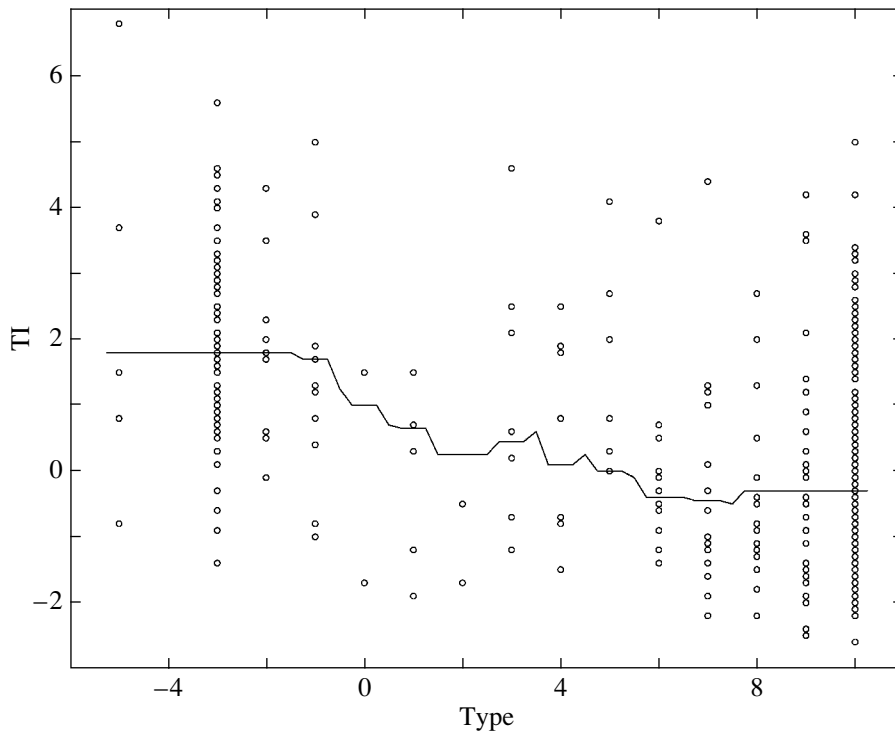**Fig. 3.** Hydrogen-to-total mass ratio (within the standard radius) versus tidal index for Local-Volume galaxies.

Fig. 4. Tidal index versus morphological type for 451 nearby galaxies. The broken line indicates the sliding median.
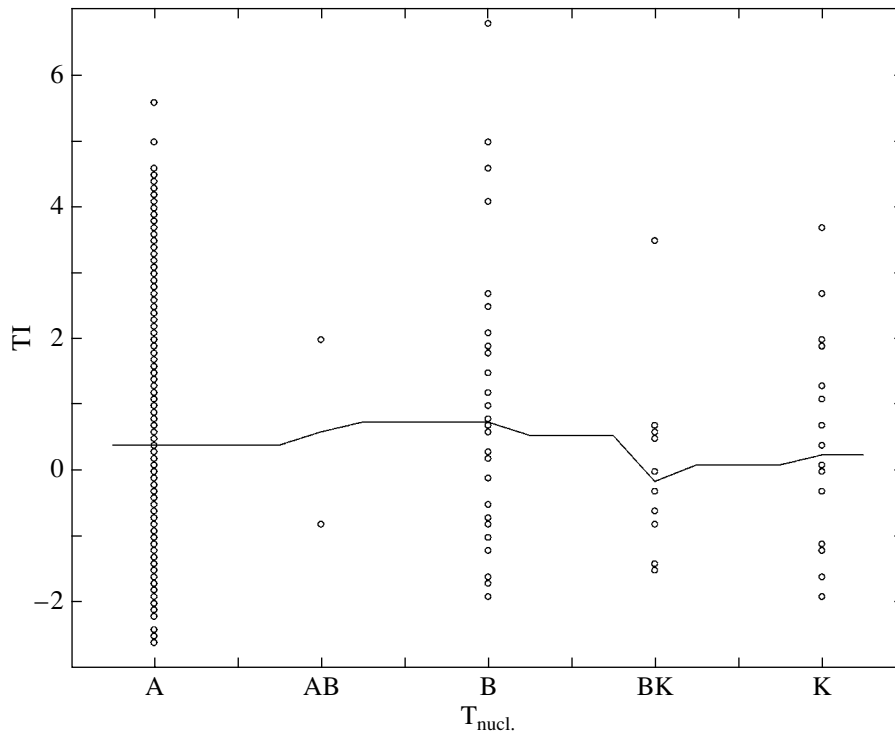
Fig. 5. Tidal indices for galaxies with various types of central circumnuclear region.

Apart from the morphological type, the CNG catalog contains a characteristic of the central region of each galaxy according to the following scheme: K, a kern, or starlike nucleus; B, a bulge where a starlike nucleus could be present, but it is unseen due to a steep brightness gradient or dust clouds; and A, the definite absence of a starlike nucleus. Figure 5 shows the lack of any significant correlation between the type of the central region of a galaxy and its tidal

Parameters of the linear regression between the global parameters of nearby galaxies and their tidal index

Y	X	N	R	B	σ_B
$\log(S_{HI})$, $M_\odot/\mathrm{kpc}^2$	TI	316	-0.304	-0.112	0.020
$\log(M_{HI}/L_b)$, $M_\odot/L_\odot$	TI	308	-0.174	-0.061	0.020
$\log(M_{HI}/M_{25})$	TI	300	-0.036	-0.016	0.025
TI	Type	451	-0.433	-0.140	0.014
TI	$T_{nucl.}$	451	0.025	0.042	0.081
$\log(M_{25}/L_b)$, $M_\odot/L_\odot$	TI	308	-0.103	-0.037	0.020
SB, mag. arcsec^{-2}	TI	451	0.206	0.147	0.033
M_b	TI	451	0.140	0.238	0.080

index. Consequently, the presence or absence of a starlike nucleus in a galaxy is an intrinsic property that depends weakly on the degree of interaction with its neighbors.
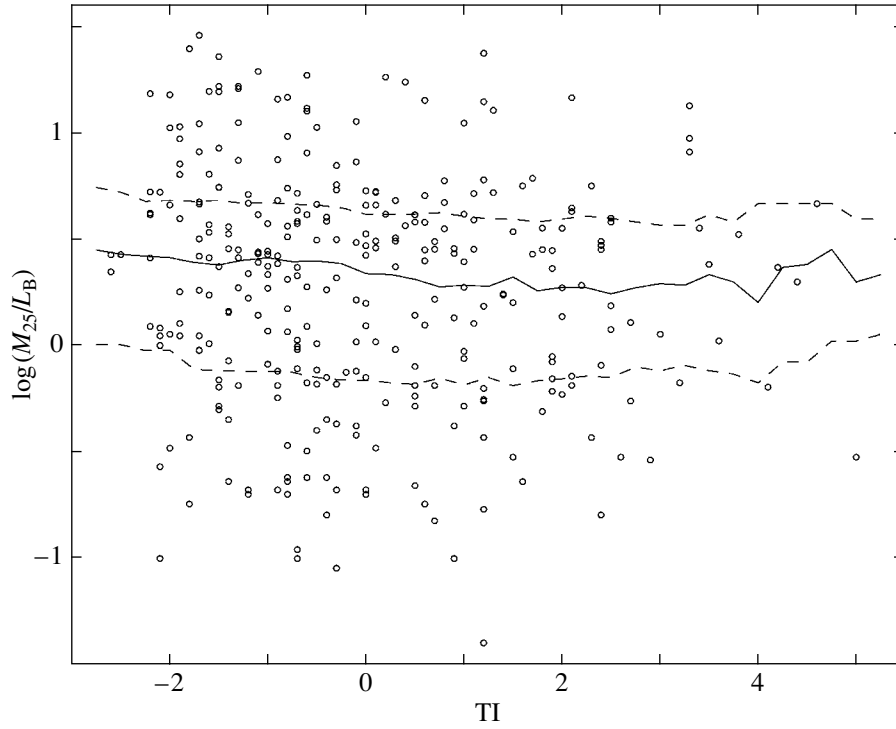
OTHER PARAMETERS OF GALAXIES AND THE TIDAL INDEX

In recent years, the hypothesis about the existence of a special category of dwarf galaxies that were formed from the tails and bars observed in interacting galaxies (Duc and Mirabel 1999) has been discussed extensively. Such tidal dwarfs must consist of a relatively young stellar population and have the low mass-to-luminosity ratios typical of the disks. In the neighboring group around M 81, Makarova *et al.* (2002) found dwarf systems that contained no red giant branch stars, i.e., old population stars. If the population of tidal dwarfs accounts for a significant fraction of nearby galaxies with positive tidal indices, then one could expect a negative correlation between the mass-to-luminosity ratios and the tidal index of galaxies. Figure 6 show that such a weak trend does exist, but the slope of the linear regression, -0.037, is only slightly larger than its standard error, 0.020. Therefore, the fraction of tidal dwarfs in our sample must be small.
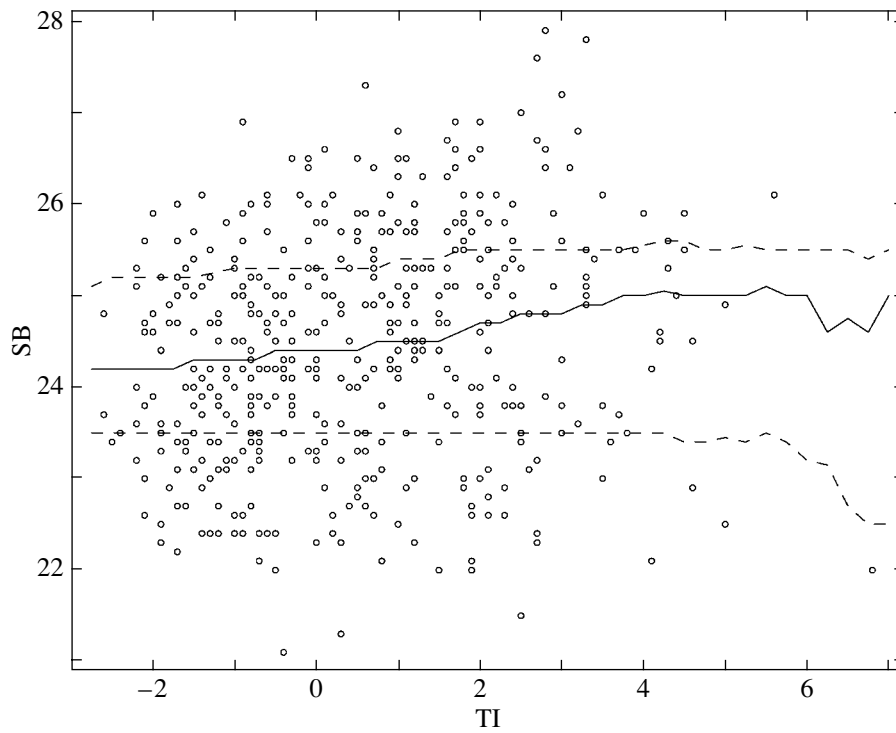
The tidal index of galaxies also reveals a weak correlation with absolute magnitude M_b and mean optical surface brightness (SB). The parameters of these

Fig. 6. Ratio of the total mass (within the standard radius) to the luminosity (in solar units) versus tidal index for nearby galaxies. The lines correspond to the median and quartiles.

Fig. 7. The distribution of nearby galaxies in mean optical surface brightness (in magnitude per square arcsec) and tidal index.
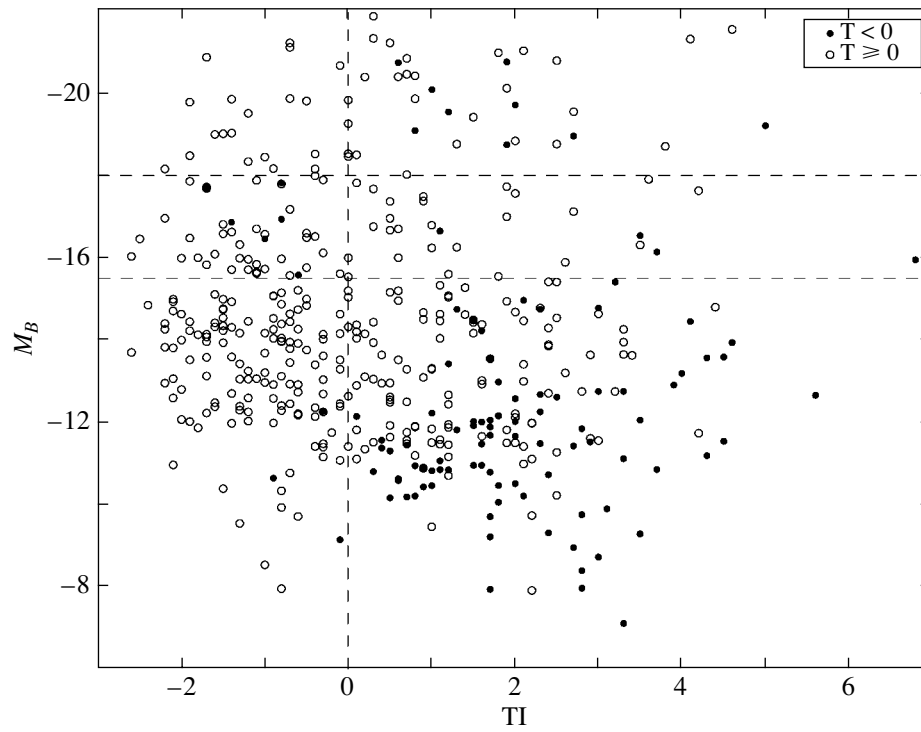
Fig. 8. Absolute B-band magnitude versus tidal index for Local Volume galaxies. The filled circles indicate early-type (E, S0, dSph) galaxies.
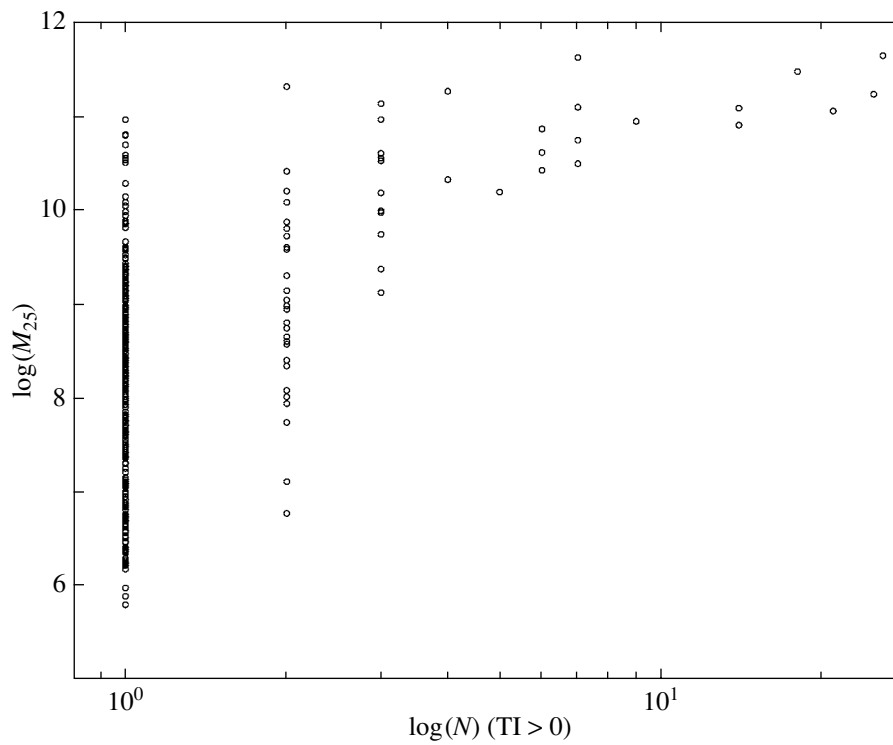
Fig. 9. Mass of the main galaxy in nearby groups (in solar masses) and the group population (only the members with positive tidal indices).

regressions are listed in the last rows of the table. The range of surface brightnesses for nearby galaxies is very wide, reaching seven magnitudes. Galaxies with extremely low surface brightnesses similar to Draco, UMin, IKN, and KK221 with SB $> 27.0^m$ arcsec^{-2} are encountered only in the immediate neighborhood of massive galaxies (Fig. 7). This may be attributable to the peculiarities of the dynamical evolution of dwarf companions.

As was pointed out above, the closest correlation is observed between the tidal index and the morphological type. However, the morphological segregation effect shows up differently in giant, normal, and dwarf galaxies. Figure 8 shows the distribution of nearby galaxies in an absolute magnitude and tidal index. The filled circles indicate early-type objects (T $<$ 0). Among the high-luminosity galaxies with $M_b < -18^m$, all elliptical and lenticular galaxies are within the range of positive TI, i.e., in groups. Intermediate-luminosity ($-15.5 > M_b > -18.0$) E and S0 galaxies are present both in groups and in the general field. NGC 404 with TI $= -1.0$ and $M_b = -16.49$ is an example of a well isolated galaxy. Recent detailed HI observations of this galaxy (del Rio et al. 2004) revealed an unusually extended hydrogen envelope in it that stretches to eight standard radii. Among dwarf galaxies with $M_b > -15.5$, the overwhelming majority of early-type (dSph) objects are in the region TI > 0. Only three dwarf spheroidal galaxies, Tucana (TI $= -0.1$), KKs3 (TI $= -0.3$), and KK 258 (TI $= -0.9$), are located outside well-known groups. The distances and, hence, the tidal indices for the last two objects have so far been estimated unreliably. The peculiarities of the distribution of galaxies in the {TI, M_b, Type} diagram require a more detailed interpretation in terms of various scenarios for the dynamical evolution of galaxies.

CONCLUSIONS

To summarize, let us point out several quantities that play a key role in characterizing the distribution of galaxies in the Local Volume. Among 451 galaxies with $D < 10$ Mpc, 197 objects, or 44%, are located in regions of locally low densities and have tidal indices TI < 0. Such objects can be called isolated galaxies or a general field population.

Early-type (E, S0, and dSpH) galaxies with (T $<$ 0) account for 23% of the sample of Local Volume galaxies, whereas their fraction among field galaxies (TI < 0) decreases to 5%, which represents the effect of galaxy segregation in morphological types.

Many Local Volume galaxies are associated with one another in the sense that they are located in the zone of predominant gravitational influence of the same galaxy, their significant neighbor. Clearly,

such significant neighbors are generally fairly massive galaxies. In our sample, 112 galaxies or 25% of their total number in the Local Volume belong to the category of significant galaxies. Only 78 significant galaxies (17%) have companions with positive indices in their retinue. Such significant galaxies together with their close (TI > 0) companions correspond to the typical definition of a group of galaxies. Figure 9 shows the distribution of these groups in the number of members and the mass of the main galaxy. As would be expected, there is a clear correlation between the mass of the galaxy and the number of its companions. For example, only galaxies with masses M_{25} larger than the threshold value of $10^9 M_\odot$ have two or more companions, while the threshold mass of the main galaxy for systems with $N > 3$ members is $10^{10} M_\odot$.

The numerical characteristics of the distribution of galaxies in the Local Volume given above are only slightly affected by various selection effects and can be of use in testing various scenarios for the formation of a small-scale structure in models of the hierarchical clustering of galaxies.

REFERENCES

1. M. S. del Rio, E. Brinks, and J. Cepa, Astron. J. (2004) (in press); astro-ph/0403467.
2. A. Dressler, Annu. Rev. Astron. Astrophys. **22**, 185 (1984).
3. P. A. Duc and F. Mirabel, *IAU Symp. No. 186: Galaxy Interactions at Low and High Redshift*, Ed. by J. E. Barnes and D. B. Sanders (Kyoto, 1999), p. 61.
4. R. Giovanelli and M. P. Haynes, Annu. Rev. Astron. Astrophys. **29**, 499 (1991).
5. I. D. Karachentsev, V. E. Karachentseva, W. K. Huchtmeier, and D. I. Makarov, Astron. J. **127**, 2031 (2004).
6. I. D. Karachentsev and D. I. Makarov, *IAU Symp. No. 186: Galaxy Interactions at Low and High Redshift*, Ed. by J. E. Barnes and D. B. Sanders (Kyoto, 1999), p. 109.
7. V. E. Karachentseva and M. E. Sharina, Publ. Spec. Astrophys. Observ. **57**, 1 (1987).
8. L. N. Makarova, E. K. Grebel, I. D. Karachentsev, et al., Astron. Astrophys. **396**, 473 (2002).
9. M. S. Roberts and M. P. Haynes, Annu. Rev. Astron. Astrophys. **32**, 115 (1994).
10. B. E. Warren, H. Jerjen, and B. S. Koribalski, Astron. J. (2004) (in press); astro-ph/0406010.

Translated by A. Dambis

The Gas Content in Galactic Disks: Correlation with Kinematics

A. V. Zasov[1*] and A. A. Smirnova[2]

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

[2]*Special Astrophysical Observatory, Russian Academy of Sciences, Nizhnii Arkhyz,
357147 Karachai-Cherkessian Republic, Russia*

Received June 22, 2004

Abstract—We consider the relationship between the total HI mass in late-type galaxies and the kinematic properties of their disks. The mass M_{HI} for galaxies with a wide variety of properties, from dwarf dIrr galaxies with active star formation to giant low-brightness galaxies, is shown to correlate with the product $V_c R_0$ (V_c is the rotational velocity, and R_0 is the radial photometric disk scale length), which characterizes the specific angular momentum of the disk. This correlation, along with the decrease in the relative mass of the gas in a galaxy with increasing V_c, can be explained in terms of the previous assumption that the gas density in the disks of most galaxies is maintained at a level close to the threshold (marginal) stability of a gaseous layer to local gravitational perturbations. In this case, the regulation mechanism of the star formation rate associated with the growth of local gravitational instability in the gaseous layer must play a crucial role in the evolution of the gas content in the galactic disk. © *2005 Pleiades Publishing, Inc.*

Key words: *galaxies, groups and clusters of galaxies, intergalactic gas, interstellar gas in galaxies, galaxy evolution.*

INTRODUCTION

Elucidating the mechanisms that determine the current gas content in galactic disks is an outstanding problem. Several processes that are capable of significantly changing the total mass of the interstellar medium in galaxies over their lifetimes are known. These include star formation, galactic wind, the ejection of matter by evolved stars, the accretion of intergalactic gas, and the absorption of gas-containing companion galaxies. Undoubtedly, the efficiency of these processes and the extent to which they are balanced vary greatly from galaxy to galaxy, so the total mass of the gas in a galaxy can both decrease and increase with time during certain evolutionary periods. Still, the main mechanism that determines the current mass of the gas is its consumption for star formation: if we proceed from the current star formation rates (SFRs), then, in most cases, the gas depletion time scale t_c proves to be much smaller than the Hubble age of galaxies (see, e.g., Devereux and Hameed 1997; Bendo *et al.* 2002; Zasov and Bizyaev 1994; Zasov 1995).

As was first shown by Larson and Tinsley (1978), the color differences between galaxies can be well explained in terms of a difference in the SFR decay rates at the same age of the systems: in redder galaxies, star formation was more intense and has

been almost completed. From this point of view, the relative amount of the gas left in the galactic disk must be determined by the star formation efficiency in them (the current SFR per unit gas mass), which is higher in systems with more favorable star formation conditions. Undoubtedly, the current SFR must be affected by the rotational velocity of the gas via both the angular velocity of spiral density waves in the disk and the formation conditions of large-scale gas condensations, which are definitely different in rapidly and slowly rotating disks.

Actually, however, the situation proves to be not so simple. The current star formation efficiency (or its reciprocal, the gas depletion time scale) was found to correlate rather weakly with other parameters of galaxies. It correlates weakly with the gas mass-to-luminosity ratio for galaxies (Zasov 1995) and shows no clear correlation with the morphological type, the luminosity (Boselli *et al.* 2002), and the rotational velocity or color of spiral galaxies (Boissier *et al.* 2001). Both the observed star formation efficiency and the relative gas content in the disk, which is defined as the ratio M_{HI}/L, can differ in galaxies of the same morphological type or in galaxies with the same color index by more than an order of magnitude (Verheijen and Sancisi 2001; Boissier *et al.* 2001).

It would be natural to expect the total mass of the gas in the disk, which is a derivative of many factors and primarily of the SFR history, to be insensitive also

---

[*]E-mail: `zasov@sai.msu.ru`

to other parameters of disk galaxies. Nevertheless, several simple relationships exist between the gas mass and galactic disk parameters.

(1) *The relationship between M_{HI} and the galaxy size.* Although the mass fraction of the gas in a galaxy (M_{gas}/M_{tot}) does not correlate with the linear size of the galaxy (McGaugh and de Blok 1997), observations reveal a close correlation between the total gas mass M_{HI} and the diameter D exhibited by all types of disk galaxies, except the earliest types (S0−Sab) (Hewitt *et al.* 1983; Broeils and Rhee 1997; Becker *et al.* 1988; Martin 1998). The form of the relationship is $M_{HI} \sim D^n$, where $n \approx 1.8-2$, implying that the mean neutral hydrogen surface density $\langle \sigma_{HI} \rangle$ is approximately constant in various galaxies. If the early-type galaxies are excluded, then $\langle \sigma_{HI} \rangle$ undergoes virtually no systematic changes with galaxy morphological type T and rotational velocity (Karachentsev *et al.* 1999a, 2004). Low- and high-surface-brightness galaxies also lie on the same $M_{HI}(D)$ relationship (Verheijen and Sancisi 2001).

Shaya and Federman (1987) suggested the transition of the gas to a molecular state when the ionizing radiation is screened by an H I layer where its surface density exceeds a certain threshold value as a mechanism that could lead to an approximately constant H I surface density. This mechanism can indeed partly or completely explain the slower decrease in the azimuthally averaged surface density $\sigma_{HI}(R)$ compared to the molecular gas surface density, but the constancy of the mean gas surface density in galaxies by no means follows from it. Moreover, if the total atomic and molecular gas surface density $\langle \sigma_{HI+H_2} \rangle$ is taken instead of $\langle \sigma_{HI} \rangle$, it also changes little along the sequence of morphological types, which by no means follows from the suggested mechanism. This is clearly illustrated by Fig. 1, in which the total gas mass (including the H I, H_2, and He masses) is plotted against the optical galaxy diameter for late-type ($T > 3$) galaxies using data from the catalog by Bettoni *et al.* (2003) (below designated as CISM). Including earlier-type galaxies increases the scatter of points without changing significantly the slope of the relationship.

The correlation of the H I mass with the disk size is particularly pronounced when the radial disk scale length R_0 is substituted for the isophotal diameter (Swaters *et al.* 2002) or when D_{HI} ($=2R_{HI}$), the size of the gaseous disk bounded by a fairly low threshold gas surface density ($\sigma_{HI} = 1 M_\odot \ pc^{-2}$), within which almost all of the H I mass is contained (Verheijen and Sancisi 2001), is used as the diameter.

(2) *The decrease in the relative gas content M_{HI}/L with increasing surface brightness of*
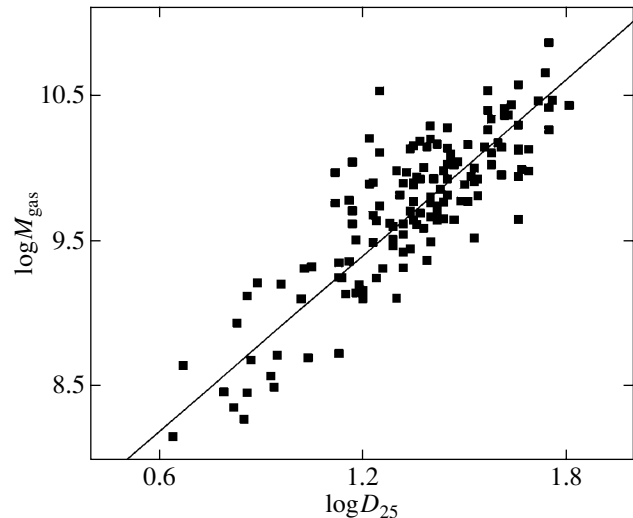
Fig. 1. Total mass of the interstellar gas (in solar units) versus optical diameter D_{25} (in kpc) for late-type ($T > 3$) galaxies (CISM data.)

late-type galaxies (Karachentsev *et al.* 1999a; McGaugh and de Blok 1997; Swaters *et al.* 2002). This decrease can be easily shown to be a direct result of the relationship $M_{HI} \sim R_0^2$. Indeed, since the disk luminosity L is proportional to $I_0 R_0^2$, where I_0 is the disk surface brightness extrapolated to the center, the relationship $M_{HI} \sim R_0^2$ leads to inverse proportionality between M_{HI}/L and I_0. For galaxies whose brightnesses differ by five magnitudes, the M_{HI}/L ratios must differ by two orders of magnitude; this in excellent agreement with the observations of nearby late-type galaxies (see Fig. 5 from Karachentsev *et al.* 1999a).

(3) *The relationship between the galaxy rotational velocity V_c and the relative gas mass M_{HI}/M_{25}.* Here, M_{25} is the indicative mass of the galaxy within its optical radius equal to $V_c^2 D/2G$, where V_c can be determined from the H I line width corrected for the disk inclination. The faster the galaxy's rotation, the lower, on average, its relative gas content (Karachentsev *et al.* 1999a, 2004; Boissier *et al.* 2001). As an illustration, Fig. 2 shows the $\log(M_{25}/M_{HI})-\log(V_c)$ diagram constructed using data from the catalog of nearby galaxies by Karachentsev *et al.* (2004) (where the observational selection effects are probably at a minimum) for Sbc and later-type galaxies. Such a relationship cannot be a simple reflection of different compression ratios of the gas when it passes through spiral density waves, since irregular (including dwarf) galaxies, which constitute a majority among the galaxies of the catalog, obey it.
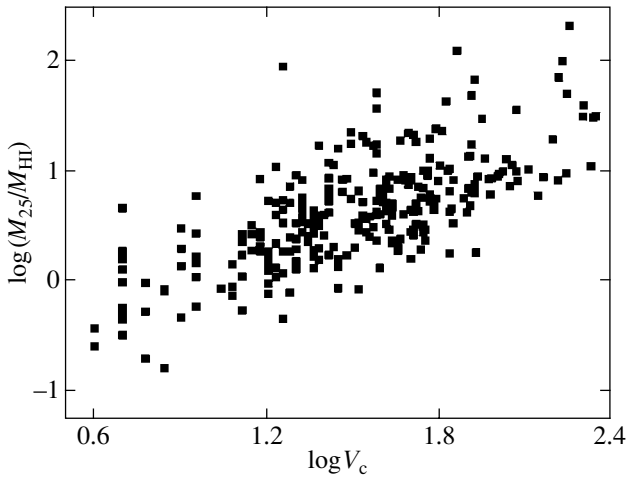
Fig. 2. Ratio of the total (indicative) galaxy mass to the total H I mass (in solar units) versus rotational velocity, as constructed for later-type ($T > 3$) galaxies using data from the catalog of nearby galaxies by Karachentsev *et al.* (2004).

(4) *The relationship between* M_{HI} *and the disk specific angular momentum* $V_c D$ (Zasov 1974; Zasov and Rubtsova 1989). The higher the specific angular momentum, the larger the amount of gas in the galaxy. As we show below, this relationship, which has a simple physical interpretation, is probably the key one.

In this paper, we restrict our analysis to the total mass of the gas in later-type galaxies. We show that the relationships listed above can be explained by the fact that the bulk of the gas in galaxies have a surface density close to its threshold value $\langle \sigma_c \rangle$ for a gravitationally stable gaseous layer, and that the galaxy diameter correlates with the disk rotational velocity.

THE GAS MASS−SPECIFIC ANGULAR MOMENTUM RELATIONSHIP FOR A MARGINALLY STABLE GASEOUS LAYER

Beginning with the classic paper by Quirk (1972), various authors have considered the relationship between the radial gas density distribution in late-type galaxies and the distribution of the critical density σ_c for the growth of local gravitational perturbations. In the simplest case of a thin gaseous disk in the gravitational field of a galaxy, it is defined by the equation

$$\sigma_c = c_g \kappa / Q_T \pi G. \qquad (1)$$

Here, c_g is the one-dimensional velocity dispersion of gaseous clouds (6–8 km s^{-1}), which is commonly assumed to be constant and independent of the morphological type and the galactocentric distance, consistent with direct H I and CO measurements (except for the galactic circumnuclear regions) (see, e.g.,

Lewis 1984; Combes *et al.* 1997); κ is the epicyclic frequency; and $Q_T > 1$ is the dimensionless (Toomre) parameter of stability to arbitrary perturbations in the plane of the disk. In general, its value depends on the shape of the rotation curve and the mass distribution in the disk and can be determined analytically or by numerically simulating the growth of disk instability. For purely radial perturbations of a thin disk, $Q_T = 1$. The existing theoretical Q_T estimates that include nonradial perturbations were obtained only for simplified models of gaseous disks, and their values lie within the range $Q_T = 1.2$–1.7 (see Morozov 1985; Polyachenko *et al.* 1997; Kim and Ostriker 2001; and references therein). The latter authors took into account the magnetic field of the interstellar medium, which does not change greatly the result.[1]

If the gravitational stability of the outer gaseous disk determined for the azimuthally averaged gas density is assumed to be responsible for the sharp reduction in SFR starting from a certain galactocentric distance R_c, then a comparison of R_c with the location of the outer boundary of the distribution of H II regions in the disks of spiral galaxies corresponds most closely to $c_g/Q_T \approx 4$ km s^{-1} (Martin and Kennicutt 2001); it thus follows that $Q_T \approx 1.5$–2 at $c_g \approx 6$–8 km s^{-1}.

Although the question of how closely the boundary of active star formation in galactic disks corresponds to the radius R_c at which the gas density becomes equal to σ_c is debatable, and this condition is definitely not satisfied in all of the galaxies studied; the azimuthally averaged density $\sigma_{HI}(R)$ or $\sigma_{HI+H_2}(R)$ in most of the S and Irr galaxies over a wide range of galactocentric distances is indeed close to $\sigma_c(R)$, differing from it by no more than a factor of 2[2] (Zasov and Simakov 1989; Martin and Kennicutt 2001; Wong and Blitz 2002; Boissier *et al.* 2001; Hunter *et al.* 1998).

Including the molecular gas poses a serious problem in estimating the gas mass in the disk. This is because, first, the dependence of the CO−H_2 conversion factor on chemical abundances is known poorly,

---

[1] The critical density (1) was determined under the assumption of a one-component gaseous disk. The presence of a stellar disk makes the gaseous disk slightly less stable. However, its effect is quantitatively small, since the velocity dispersion of the old disk stars is much higher than that of the gas and since the density of the stellar disk is very low in the outer regions where their values could be comparable. Boissier *et al.* (2003) showed for several spiral galaxies that including the stellar disk reduces the local critical mass by 10–15%; as would be natural to expect, the effect of the stellar disk decreases with increasing galactocentric distance.

[2] The gas density is most often higher than its critical value in the inner galactic region and lower in the outer disk regions.

and, second, direct CO measurements are generally restricted only to the inner galactic region, so when the total mass of the molecular gas is estimated, one has to extrapolate its density to large galactocentric distances. However, the latter is unlikely to play a significant role, since the H_2 fraction decreases with R (Wong and Blitz 2002).

The situation is alleviated by the fact that the bulk of the observed gas is in the form of H I in most galaxies (at least late-type galaxies): the molecular hydrogen mass M_{H_2}, on average, accounts for about 15% of the mass M_{HI} (Casoli *et al.* 1998; Boselli *et al.* 2002). Note, however, that this value may prove to be slightly underestimated: according to the CISM, which combines the H_2 estimates obtained by various authors by assuming the conversion factor to be constant (2.3×10^{20} mol K^{-1} km s^{-1}), the molecular gas (including helium) accounts for a slightly higher mass fraction ($\sim 40\%$ of the H I mass), except for the Sd−Irr galaxies that contain a very small amount of molecular gas (Bettoni *et al.* 2003).

In any case, including the molecular gas increases the total mass of the cold gas in spiral galaxies, on average, by no more than a factor of 1.5 (although this factor could be much larger in some galaxies); this justifies using the H I mass estimate to characterize the total amount of the gas. The predominance of the atomic gas makes it easier to analyze the evolution of the gaseous component of the disk, since the mass of the molecular gas is known for a much smaller number of galaxies than the H I mass. Here, we do not consider the possibility that very cold and, hence, unobservable H_2 cloudlets in galactic disks might form a layer with such a high total surface density that they produce the effect of hidden mass by contributing significantly to the disk mass (Combes and Pfenniger 1997). This hypothesis encounters serious difficulties in analyzing the gravitational stability of a gaseous layer (Elmegreen 1995) or in dynamical mass estimations for the disks of spiral galaxies (Kranz *et al.* 2003; Zasov *et al.* 2004), which do not leave much space for dark matter in the disks.

Let us estimate the gas mass M_c in a disk if the gas density is close to the critical value over its entire length:

Let the gas density everywhere from the center ($R = 0$) to the radius R_{HI} within which almost all of the H I mass is contained be defined by Eq. (1). We will approximate the circular velocity by a simple function, $V(R) = V_c(R/R_{HI})^n$, where V_c is the rotational velocity on the periphery of the galaxy ($R \approx R_{HI}$), and the constant n can have values for different galaxies starting from zero (a plateau on the rotation curve) to unity (rigid rotation). In the inner regions of spiral galaxies, the rotation curve is generally more

complex in shape, but the bulk of the H I is located in regions far from the center, where $V_c(R) \approx$ const ($n \approx 0$). In irregular galaxies, $0 < n < 1$, while for the least massive systems, the parameter n could be close to unity (rigid rotation).

For the epicyclic frequency, we obtain

$$\kappa = 2\Omega[1 + (n - 1)/2]^{1/2}, \qquad (2)$$

where Ω is the angular velocity of circular rotation of the disk, whence it follows that the critical gas mass is

$$M_{gas}^c = \int_0^{R_{HI}} 2\pi R \sigma_c(R) dR \qquad (3)$$

$$= \frac{2^{3/2}}{G} \frac{c_g}{Q_T} (1 + n)^{1/2} V_c R_{HI}.$$

The total mass of the gas is related to the masses of the atomic and molecular gases by

$$M_{gas} \approx 1.4(M_{HI} + M_{H2}) = \eta^{-1} M_{HI}, \qquad (4)$$

where $\eta^{-1} \approx 1.4-2$ is a coefficient that characterizes the fraction of the molecular gas, helium, and heavier elements in the total mass of the gas (the first value corresponds to a negligible mass of the molecular gas).

It follows from (3) and (4) that

$$M_{HI}^c = \eta M_{gas}^c = \eta 2^{3/2} K(1 + n)^{1/2} V_c R_{HI}/G, \quad (5)$$

where $K \equiv c_g/Q_T$. Given the uncertainties in the Q_T and c_g estimates (see above), K may be assumed to lie within the range 3.5−6.5 km s^{-1}.

Since the resulting estimate depends weakly on the parameter n (the largest uncertainty is associated with the ratio c_g/Q_T and with the assumption of its constancy along the radius), we assume below that $n = 0$, which corresponds to the same rotational velocity at all R. In this case,

$$M_{HI}^c = 2^{3/2} \eta \frac{K}{G} V_c R_{HI}. \qquad (6)$$

A relation similar to (5) (but with a different numerical value of the proportionality coefficient) can also be written for the case where the photometric radius R_{25}, which, on average, is a factor of 1.7−1.8 smaller than R_{HI} for both spiral and irregular galaxies (Broeils and Rhee 1997; Swaters *et al.* 2002), is used instead of R_{HI} as the upper integration limit in Eq. (3). Therefore, Eq. (6), to within numerical coefficients, describes the previously reached conclusion (Zasov 1974; Zasov and Rubtsova 1989) that the total mass of the gas at its threshold density is proportional to its specific angular momentum DV_c, where D is the optical diameter of the galaxy. This relationship is actually in good agreement with the

observations, both for single galaxies and for galaxies in pairs (Zasov 1974; Zasov and Rubtsova 1989; Zasov and Sulentic 1994; Karachentsev *et al.* 1999a, 2004).

To check how universal this conclusion is for various galaxies and to quantitatively compare M_{HI} and M_{HI}^c, we consider below several samples of late-type galaxies with widely differing parameters—from clumpy irregular galaxies with bright sites of star formation to Malin 1 galaxies of extremely low disk surface brightness with very low SFRs.

The optical isophotal diameter is of little use for this purpose, since it offers no possibility of comparing galaxies with different surface brightnesses: the lower the surface brightness of the disk with the same radial scale length, the smaller its isophotal diameter. Therefore, it would be more appropriate to compare galaxies by using the radial disk scale length R_0 rather than the optical diameter. On average, $R_{HI} \approx 5.4 R_0$ for both irregular and spiral galaxies (Swaters *et al.* 2002). Below, we use this relationship to estimate M_{HI}^c.

SAMPLES OF GALAXIES

In this paper, we consider several different samples of late-type galaxies with known hydrogen masses M_{HI}, rotational velocities V_c (determined in most cases from the H I line width), and photometric radial scale lengths R_0: dwarf irregular (dIrr) galaxies, H I-rich late-type galaxies, UMa cluster galaxies, clumpy irregular (cIrr) galaxies (which were included in the atlases of interacting systems by Vorontsov-Vel'yaminov (1959, 1977) due to their peculiar appearance), edge-on late-type galaxies, and low-surface-brightness galaxies to which three objects with extreme Malin-1-type characteristics belong. If required, the distances to galaxies with significant systemic velocities were reduced to the Hubble constant $H_0 = 75$ km s^{-1} Mpc^{-1}. For nearby dwarf galaxies, we used distance estimates from original works.

Let us consider the samples separately:

(1) Galaxies that morphologically belong to late-type spirals or irregulars with an absolute magnitude of -18 or fainter and line-of-sight velocities lower than 3000 km s^{-1} constituted the sample of dwarf galaxies. These are noninteracting systems: they have no apparent companions within $30'$ whose velocities differ by less than 500 km s^{-1} from the velocity of the galaxy. All of the parameters, except the optical isophotal radius R_{25}, were taken from the paper by van Zee (2001). The radius was calculated from the assumed distance and the angular diameter D_{25} (the latter was taken from the LEDA database).

(2) Late-type galaxies with an absolute magnitude of -17 or fainter and a high flux density in the H I line ($S_{HI} > 200$ mJy) constituted the second sample of dwarf galaxies. Data for these were taken from the paper by Swaters *et al.* (2002). With the exception of several galaxies, all of them lie at distances less than 25 Mpc.

(3) Galaxies from the catalog by Vorontsov-Vel'yaminov (1959, 1977) whose peculiar shapes are very likely attributable not to the system multiplicity but to the clumpy distribution of bright star-forming regions, constituted the sample of clumpy irregular (cIrr) galaxies. Several dozen such galaxies were identified, but the radial disk scale length could be estimated only for several objects: due to the presence of bright condensations, the surface brightness of the disks in such galaxies can by no means always be described by an exponential law. Data for these galaxies were taken from the literature (Patterson and Thuan 1996; Yasuda *et al.* 1997; Martin 1998; Bremnes *et al.* 1999; Iglesias-Paramo and Vilchez 1999; Makarova 1999; Thuan *et al.* 1999; Barazza *et al.* 2001; Cairos *et al.* 2001a, 2001b; Shapley *et al.* 2001; Pustilnik *et al.* 2003) and from the HyperLeda database (http://www-obs.univ-lyon1.fr/hypercat/).

(4) Late-type spirals from the UMa open cluster that differ widely in surface brightness constituted the sample of normal spiral galaxies. Data for these galaxies were taken from the paper by Verheijen and Sancisi (2001). Only the galaxies with the most reliable measurements that were identified by the authors as galaxies with fully analyzed data were used.

(5) The sample of low-surface-brightness galaxies was taken from the paper by de Block (1996). The central brightness of the disks in these galaxies is $\mu_0(B) \geq 23$. The sample was supplemented by three objects with extremely low surface brightnesses: Malin 1, F568−6, and 1226+0105. All of the data for the latter were taken from the paper by Sprayberry *et al.* (1993) and reduced to $H_0 = 75$ km s^{-1} Mpc^{-1}.

(6) The sample of edge-on spiral galaxies was drawn from the RFGC catalog (Karachentsev *et al.* 1999b). The latest version of this catalog contains 4236 galaxies distributed over the entire sky with apparent axial ratios $a/b \geq 7$ and angular diameters $a \geq 0'.6$. Sc−Sd galaxies with small bulges constitute the bulk of the catalog. The RFGC objects are rich in gas and are easily detectable in the 21-cm H I line. No correction for the projection is required for these when estimating the rotational velocity of the outer disk. More importantly, the selection criteria make the sample homogeneous in the structure of its constituent galaxies. H I data for the galaxies are given in the paper by Karachentsev and Smirnova (2002). The

radial disk scale lengths for the edge-on galaxies were taken from the papers by Bizyaev (2000), Bizyaev and Mitronova (2002), and Kregel *et al.* (2002). Since R_0 determined for the B band was used in other samples, the disk scale length for the edge-on galaxies obtained in the near infrared was also reduced to the B band using the relations

$$R_{0B} = 1.44R_{0I}, \quad R_{0B} = 1.65R_{0K}, \quad (7)$$

where R_{0B}, R_{0K}, and R_{0I} are the radial disk scale lengths in the B, K, and I bands, respectively (de Grijs 1998).

THE UNIVERSALITY OF THE GAS MASS−SPECIFIC ANGULAR MOMENTUM RELATIONSHIP

Since the samples are heterogeneous, the M_{HI} and R_0 estimates of interest may have various systematic errors. Nevertheless, all of the samples show similar relationships, although some of them are noticeably displaced from one another in the diagrams.

Figure 3 shows the relationships between the neutral hydrogen mass and the products $R_{25}V_c$ (Fig. 3a) and R_0V_c (Fig. 3b) whose existence follows from the assumption that the mass of the gas is close (or proportional) to the critical value for gravitational stability (see the previous section). Both these products characterize the specific angular momentum of rotating disks. The diagrams reveal a correlation between the quantities being compared, which, as would be expected, becomes more pronounced when using the radial scale length R_0.

If these relationships were a simple reflection of the already discussed size−H I mass relation, then including the rotational velocity would blur them appreciably. In fact, although M_{HI} correlates both with R_0 and (slightly worse) with V_c (Figs. 4a, 4b), the correlation coefficient between the H I mass and the product of these quantities is as high as (if not higher than) that between M_{HI} and R_0. It thus follows that the observed relationship between the H I mass and the angular momentum of the disk can not be reduced to the combination of two simpler relationships, but reflects the actually existing correlation between the total gas mass and the kinematic parameters of the gaseous disk.

The correlation coefficients r and the parameters of the linear dependences $\log Y = a + b \log X$ considered in this section are given in the table. The coefficient b_c is the mean slope Y/X of two regression lines: $Y(X)$ and $X(Y)$.

The band bounded by the parallel lines in Fig. 3b describes the expected relationship between the critical mass M_{HI}^c and $V_c R_0$ for the case where the total mass of the gas in the disk (which was assumed

here to be proportional to M_{HI}) is close to its critical value: $M_{gas} = M_{gas}^c$ (see (6)). The band width characterizes the uncertainty in the coefficients in the equation (see the section "The Gas Mass−Specific Angular Momentum Relationship...") and primarily in the coefficient K, the ratio of the gas velocity dispersion to the Toomre parameter. The most probable values of M_{HI}^c correspond to the dashed line drawn through the center of the band. As we see from Fig. 3b, the overwhelming majority of points in the diagram lie within or slightly below the uncertainty band. The relatively large scatter in the positions of cIrr galaxies (triangles) and edge-on galaxies (filled circles) in the $\log M_{HI} - \log(V_{rot}R_0)$ diagram is probably attributable to the larger errors in the radial scale lengths of their disks. The H I mass also deviates significantly from the expected value for several of the most slowly rotating gas-poor dwarf galaxies (in the lower part of the diagram).

Some of the gas-rich dwarf galaxies chosen for their high flux density in the H I line (open stars) are located slightly above, but also parallel to the relationship for M_{HI}^c. The mean gas surface density in them slightly exceeds the critical value calculated with the assumed values of c_g/Q_T (see the section "The Gas Mass−Specific Angular Momentum Relationship...").

The estimates of the observed hydrogen mass and the most probable critical mass M_{HI}^c (corresponding to the central line of the band in Fig. 3b) for the galaxies under consideration are compared in Fig. 5. Like the diagram in Fig. 3b, the diagram in Fig. 5 clearly reveals a close correlation between the observed gas content in the galaxies and the kinematic parameters of their disks.

DISCUSSION AND CONCLUSIONS

The agreement between the theoretically expected (for a marginally stable disk) and observed H I masses (Figs. 3b and 5), along with the relatively small number of galaxies located above the highlighted band in Fig. 3b, in which the hydrogen mass probably exceeds M_c, suggest that the gravitational stability condition for a gaseous layer is an important factor that determines the amount of gas at the current epoch and, hence, regulates the star formation efficiency and the gas depletion rate. If we exclude the galaxies chosen for their high H I flux density (open asterisks), then, on average, the total H I mass in the galaxies under consideration proves to be slightly lower than M_{HI}^c (by a factor of 1.5−2). This is probably the result of a low (compared to the critical value) gas density in the outer disk regions, beyond D_{25}, where a considerable fraction of the total H I mass is contained (Broeils and Rhee 1997).
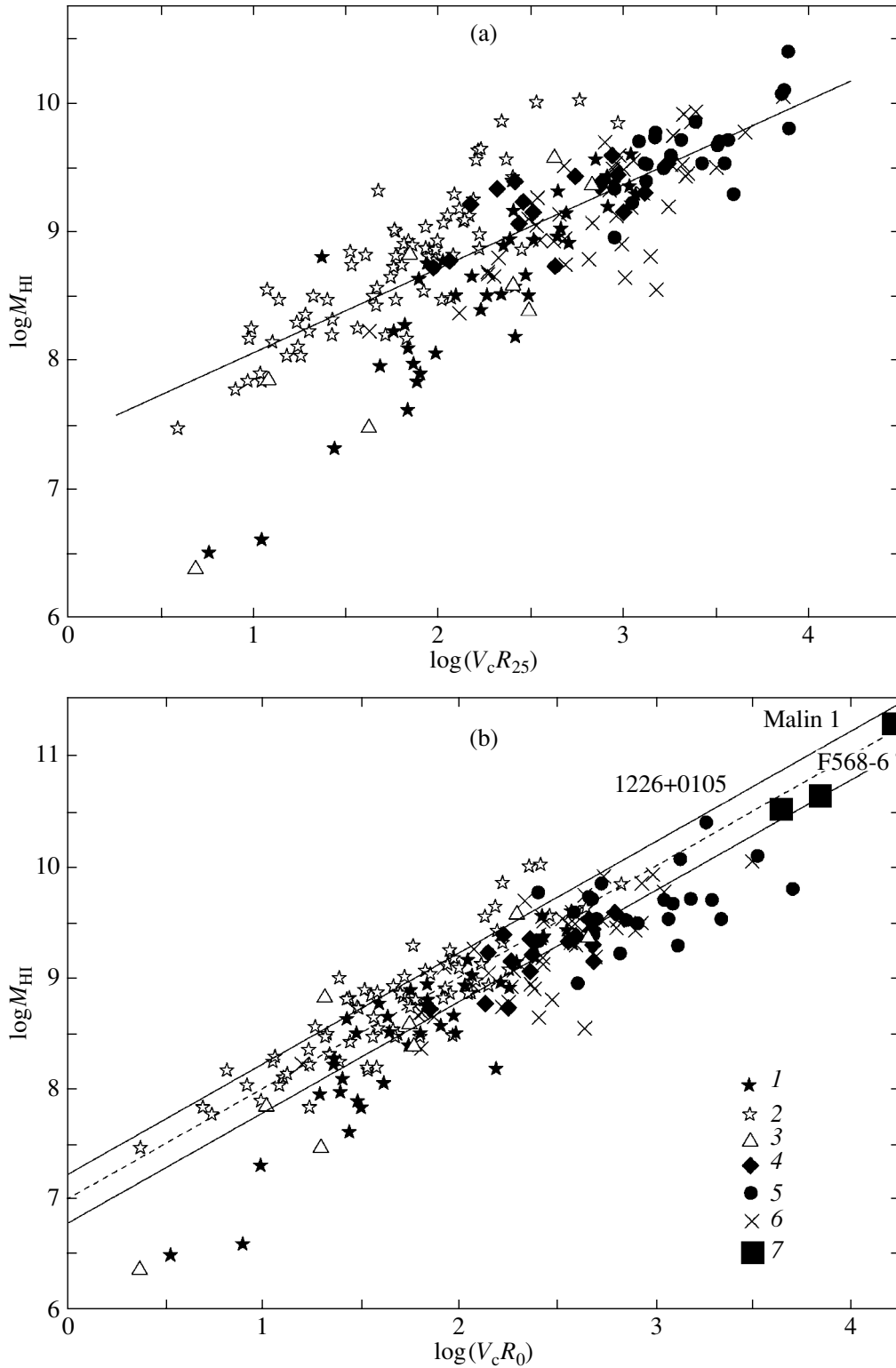
Fig. 3. H I mass–specific angular momentum diagram for galaxies of different samples. The logarithm of the product $V_c R_{25}$ (a) or $V_c R_0$ (b), where V_c is the rotational velocity (in km s^{-1}), R_{25} is the optical radius of the galaxy (in kpc), and R_0 is the photometric radial disk scale length, is along the horizontal axis; the logarithm of the H I mass (in solar units) is along the vertical axis: (*1*) dwarf galaxies with quiescent star formation, (*2*) hydrogen-rich galaxies, (*3*) clumpy irregular (cIrr) galaxies, (*4*) low-surface-brightness galaxies, (*5*) edge-on late-type spirals, (*6*) UMa cluster spirals, and (*7*) three Malin-1 type galaxies.
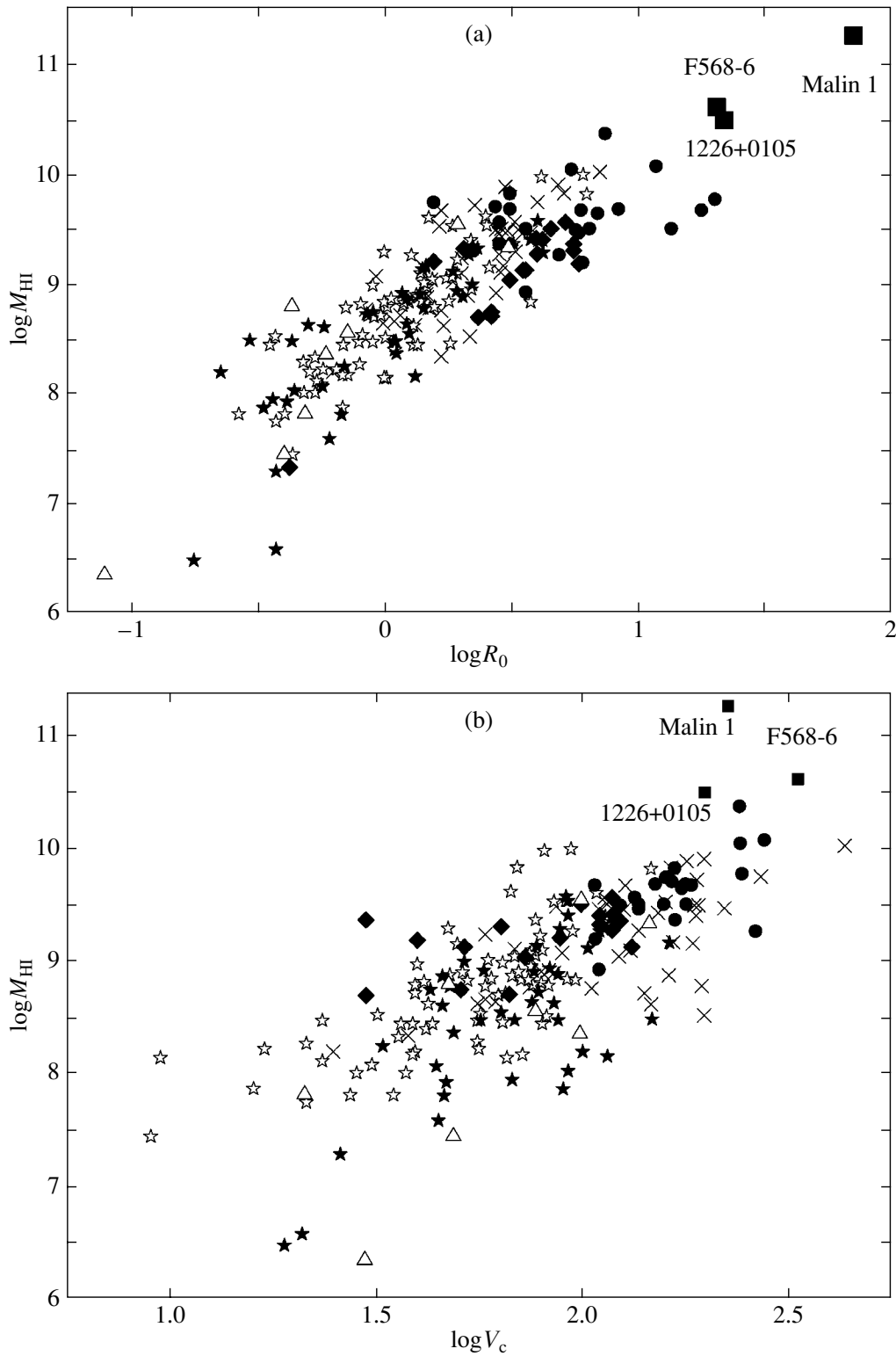
Fig. 4. (a) Radial disk scale length R_0 (in kpc)−H I mass (in solar units) diagram; (b) rotational velocity V_c (km s^{-1})−H I mass (in solar units) diagram.

The correspondence of the H I mass to its expected critical value of M_{HI}^c (and of the total gas mass to its critical value of M_{gas}^c) for late-type galaxies can be naturally explained by assuming that during the evolution of the galaxy, when the mean gas density decreased to a level close to that critical for the
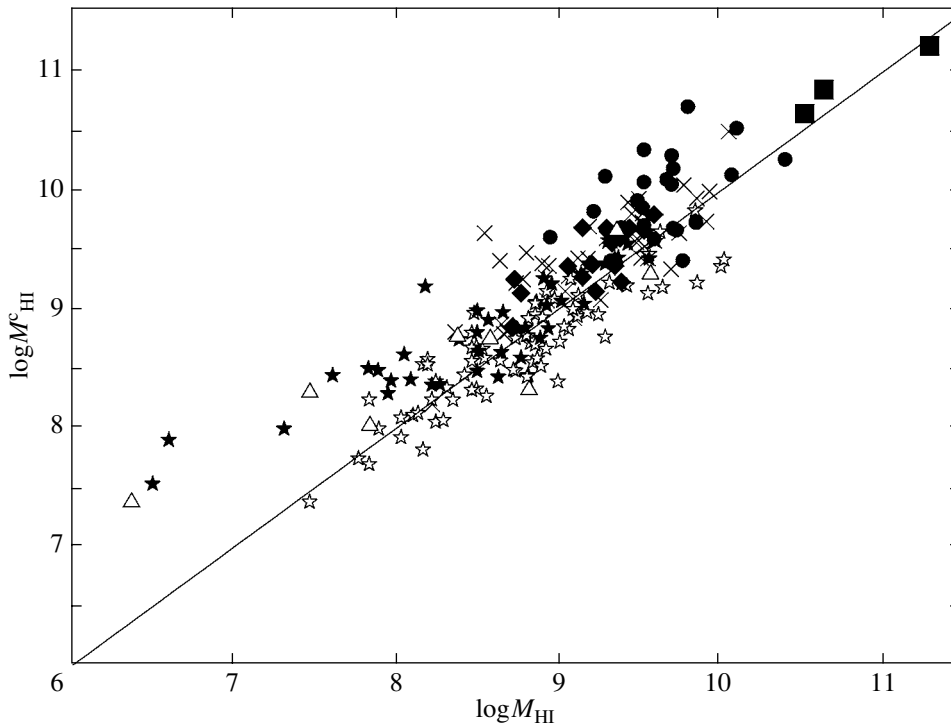
Fig. 5. Logarithmic H I mass–critical mass M_{HI}^c (in solar units) dependence. The straight line corresponds to the equality between M_{HI} and M_{HI}^c.

gravitational stability of the gaseous layer, the gas depletion slowed down significantly, and, as a result, most galaxies were able to preserve an amount of gas close to (or slightly smaller than) M_{gas}^c up to the present time. This conclusion applies to most of the galaxies from all of the samples considered, including extremely-low-brightness Malin 1 galaxies, since they lie on the general relationship (Fig. 3b). The gas content in these unusual galaxies proved to be close to the expected values for disks with enormous angular momentum. The star formation rates and efficiency in their disks, as well as in the disks of normal-brightness galaxies, must have been decreased when the gas density reduced below the critical level. However, this does not rule out the possibility that different initial formation conditions or a lower initial density of the gaseous disk were true of them.

The approximate proportionality between the rotational velocity V_c and the size of galaxies that has been pointed out by several authors (Tully and Fisher 1977; Karachentsev *et al.* 1999a) could explain why the gas mass proportional to $V_c R_{25}$ changes linearly with the square of the galactic disk size (the mean H I surface density is almost constant). A clear and almost linear relationship between the rotational velocity and the optical size of galaxies was found for a homogeneous sample of nearby galaxies: $\log D \sim (0.99 \pm 0.06) \log V_c$ (Karachentsev *et al.* 1999a).

Another important conclusion also follows from the existence of a linear relationship between the observed mass M_{HI} and M_{HI}^c: galaxies with slowly rotating disks must, on average, possess a higher relative gas mass, i.e., have a lower ratio of the total (indicative) mass M_{25} within the photometric radius to the total H I mass. Indeed, since the photometric radius is proportional to R_0 and since the total mass $M_{25} \sim V_c^2 R_0$, the relation $M_{25}/M_{HI} \sim V_c$ follows from the condition $M_{HI} \sim V_c R_0$. This conclusion is in good agreement with the observational data (Fig. 2).

Thus, the observed gas content in late-type galaxies reflects a similar (for most) pattern of evolutionary change of the gas mass in the disk. The gravitational instability of the gaseous layer must play a crucial role in this evolution. The growth of instability probably facilitated the enhancement of star formation and to the fast gas depletion at the initial (violent) evolutionary stage of the galactic disk. At that time, the density of the gaseous disk exceeded significantly the critical value calculated for the current gas velocity dispersion that reflects the quiescent pattern of star formation. This conclusion is valid for galaxies with various diameters, rotational velocities, current SFR rates, and disk surface brightnesses.

In this paper, we have not considered early-type (S0–Sab) disk galaxies, which contain little gas

Correlation coefficients and parameters of the linear dependences

X, Y	r	a	b	b_{c}
$\log V_c R_{25}, \log M_{\mathrm{HI}}$	0.78	7.25 ± 0.10	0.71 ± 0.04	0.94 ± 0.07
$\log V_c R_0, \log M_{\mathrm{HI}}$	0.88	7.05 ± 0.07	0.91 ± 0.03	1.05 ± 0.04
$\log R_0, \log M_{\mathrm{HI}}$	0.86	8.63 ± 0.03	1.45 ± 0.06	1.69 ± 0.08
$\log V_c, \log M_{\mathrm{HI}}$	0.74	5.57 ± 0.22	1.79 ± 0.11	2.52 ± 0.24

at the same sizes and rotational velocities as those of late-type galaxies. These systems must have a slightly different history of star formation; they have lost a significant fraction of their gas either through external factors (e.g., due to the gas being swept up as it moved in the intergalactic medium of the cluster) or through internal processes that ensured active star formation even when the gas density decreased below the critical level for large-scale gravitational instability.

ACKNOWLEDGMENTS

REFERENCES

1. F. D. Barazza, B. Binggeli, and P. Prugniel, Astron. Astrophys. **373**, 12 (2001).
2. R. Becker, U. Mebold, K. Reif, and H. van Woerden, Astron. Astrophys. **203**, 21 (1988).
3. G. J. Bendo, R. D. Joseph, M. Wells, *et al.*, Astron. J. **124**, 1380 (2002).
4. D. Bettoni, G. Galletta, and S. Garcia-Burillo, Astron. Astrophys. **405**, 5 (2003).
5. D. Bizyaev, astro-ph/0007242 (2000).
6. D. Bizyaev and S. Mitronova, Astron. Astrophys. **389**, 795 (2002).
7. S. Boissier, A. Boselli, N. Prantzos, *et al.*, Mon. Not. R. Astron. Soc. **321**, 733 (2001).
8. S. Boissier, N. Prantzos, A. Boselli, and G. Gavazzi, Mon. Not. R. Astron. Soc. **346**, 1215 (2003).
9. A. Boselli, J. Lequeux, and G. Gavazzi, Astron. Astrophys. **384**, 33 (2002).
10. T. Bremnes, B. Binggeli, and P. Prugniel, Astron. Astrophys., Suppl. Ser. **137**, 337 (1999).
11. A. H. Broeils and M.-H. Rhee, Astron. Astrophys. **324**, 877 (1997).
12. L. Cairos, J. Vilchez, J. Gonzalez Perez, *et al.*, Astrophys. J., Suppl. Ser. **133**, 321 (2001a).
13. L. Cairos, N. Caon, J. Vilchez, *et al.*, Astrophys. J., Suppl. Ser. **136**, 393 (2001b).
14. F. Casoli, S. Sauty, M. Gerin, *et al.*, Astron. Astrophys. **331**, 451 (1998).
15. F. Combes and J.-F. Becquaert, Astron. Astrophys. **326**, 554 (1997).
16. F. Combes and D. Pfenniger, Astron. Astrophys. **327**, 453 (1997).
17. W. J. G. de Block, S. S. McGaugh, and J. M. van der Hulst, Mon. Not. R. Astron. Soc. **283**, 18 (1996).
18. R. de Grijs, Mon. Not. R. Astron. Soc. **299**, 595 (1998).
19. N. Devereux and S. Hameed, Astron. J. **113**, 599 (1997).
20. B. G. Elmegreen, Mon. Not. R. Astron. Soc. **275**, 944 (1995).
21. J. Hewitt, M. Haynes, and R. Giovanelli, Astron. J. **88**, 272 (1983).
22. D. A. Hunter, B. G. Elmegreen, and A. L. Baker, Astrophys. J. **493**, 595 (1998).
23. J. Iglesias-Paramo and J. M. Vilchez, Astrophys. J. **518**, 94 (1999).
24. I. D. Karachentsev, V. E. Karachentseva, W. K. Huchtmeier, and D. I. Makarov, Astron. J. **127**, 2031 (2004).
25. I. Karachentsev, V. Karachentseva, Y. Kudrya, *et al.*, Bull. SAO **47**, 5 (1999b).
26. I. D. Karachentsev, D. I. Makarov, and W. K. Huchtmeier, Astron. Astrophys., Suppl. Ser. **139**, 97 (1999a).
27. I. D. Karachentsev and A. V. Smirnova, Astrofiz. **45**, 448 (2002).
28. W.-T. Kim and E. Ostriker, Astrophys. J. **559**, 70 (2001).
29. T. Kranz, A. Slyz, and H.-W. Rix, Astrophys. J. **586**, 143 (2003).
30. M. Kregel, P. van der Kruit, and R. de Grijs, Mon. Not. R. Astron. Soc. **334**, 646 (2002).
31. R. Larson and B. Tinsley, Astrophys. J. **219**, 46 (1978).
32. B. Lewis, Astrophys. J. **285**, 453 (1984).
33. L. Makarova, Astron. Astrophys., Suppl. Ser. **139**, 491 (1999).
34. M. C. Martin, Astron. Astrophys., Suppl. Ser. **131**, 77 (1998).
35. C. Martin and R. Kennicutt, Astrophys. J. **555**, 301 (2001).
36. S. McGaugh and W. J. G. de Blok, Astrophys. J. **481**, 689 (1997).
37. A. G. Morozov, Astron. Zh. **62**, 209 (1985) [Sov. Astron. **29**, 120 (1985)].

38. R. Patterson and T. Thuan, Astrophys. J., Suppl. Ser. **107**, 103 (1996).

39. V. L. Polyachenko, E. V. Polyachenko, and A. V. Strel'nikov, Pis'ma Astron. Zh. **23**, 551 (1997) [Astron. Lett. **23**, 483 (1997)].

40. S. Pustilnik, A. Zasov, A. Kniazev, *et al.*, Astron. Astrophys. **400**, 841 (2003).

41. W. Quirk, Astrophys. J. Lett. **176**, L9 (1972).

42. A. Shapley, G. Fabbiano, and P. Eskridge, Astrophys. J., Suppl. Ser. **137**, 139 (2001).

43. E. J. Shaya and S. R. Federman, Astrophys. J. **319**, 76 (1987).

44. D. Sprayberry, C. D. Impey, M. J. Irwin, *et al.*, Astrophys. J. **417**, 114 (1993).

45. R. A. Swaters, T. S. van Albada, J. M. van der Hulst, and R. Sancisi, Astron. Astrophys. **390**, 829 (2002).

46. T. X. Thuan, V. A. Lipovetsky, J.-M. Martin, and S. A. Pustilnik, Astron. Astrophys., Suppl. Ser. **139**, 1 (1999).

47. R. B. Tully and J. R. Fisher, Astron. Astrophys. **54**, 661 (1977).

48. M. A. W. Verheijen and R. Sancisi, Astron. Astrophys. **370**, 765 (2001).

49. L. van Zee, Astron. J. **121**, 2003 (2001).

50. B. A. Vorontsov-Velyaminov, *Atlas and Catalogs of Interacting Galaxies* (Mosk. Gos. Univ., Moscow, 1959) [in Russian].

51. B. A. Vorontsov-Velyaminov, Astron. Astrophys. **28**, 1 (1977).

52. T. Wong and L. Blitz, Astrophys. J. **569**, 157 (2002).

53. N. Yasuda, M. Fukugita, and S. Okamura, Astrophys. J., Suppl. Ser. **108**, 417 (1997).

54. A. V. Zasov, Astron. Zh. **51**, 1225 (1974) [Sov. Astron. **18**, 730 (1974)].

55. A. V. Zasov, Pis'ma Astron. Zh. **21**, 730 (1995) [Astron. Lett. **21**, 652 (1995)].

56. A. V. Zasov and D. V. Bizyaev, Publ. Astron. Soc. Pac. **66**, 73 (1994).

57. A. V. Zasov, A. V. Khoperskov, and N. V. Tyurina, Pis'ma Astron. Zh. **30**, 653 (2004) [Astron. Lett. **30**, 593 (2004)].

58. A. V. Zasov and T. V. Rubtsova, Pis'ma Astron. Zh. **15**, 118 (1989) [Sov. Astron. Lett. **15**, 51 (1989)].

59. A. V. Zasov and S. G. Simakov, Astrofiz. **29**, 518 (1989).

60. A. V. Zasov and J. Sulentic, Astrophys. J. **430**, 179 (1994).

Translated by V. Astakhov

Parameters of the Spiral Structure of the Galaxy from Data on Open Star Clusters

M. E. Popova and A. V. Loktin[*]

Astronomical Observatory, Urals State University, pr. Lenina 51, Yekaterinburg, 620083 Russia
Received July 2, 2004

Abstract—We use data on open star clusters (OSCs) from the Homogeneous Catalog of OSC Parameters to determine some of the parameters of the spiral structure of our Galaxy: the pitch angle of the spiral arms $i = 21°.5$, the pattern speed $\Omega_P = 20.4 \pm 2.5$ km s^{-1} kpc^{-1}, and the initial phase of the spiral $\theta_0 = 206°$. The spiral pattern of the Galaxy proves to have been virtually unchanged over the last billion years, and signatures of the concentration of objects toward the spiral arms can be traced back to this age. However, the number of spiral arms in the structure cannot be determined from OSCs. © *2005 Pleiades Publishing, Inc.*

Key words: *Milky Way, spiral structure—parameters, open star clusters.*

INTRODUCTION

At present, the spiral structure of our Galaxy is generally believed to be a manifestation of a density wave propagating over its disk. In turn, the spiral density wave acts as a global trigger of star formation. Thus, the spiral structure is an important link between the dynamics of the Galaxy as a whole and the properties of the stellar and gaseous components of the Galactic disk. Therefore, it is important to determine the parameters of the spiral structure of the Galaxy. In this paper, by the parameters of the spiral structure we mean both the geometric (the number of spiral arms, the initial phase angle of the density wave, the pitch angle, and the positions of the Lindblad and corotation resonances) and kinematic (the pattern speed and the amplitude of the velocity field distortions of the stellar and gaseous disks) parameters.

The parameters of the spiral structure have been determined repeatedly and by many authors. Among the most recent works, we can mention the paper by Mishurov and Zenina (1999) and references therein. References to previous works can be found in the monograph by Marochnik and Suchkov (1984). The question of whether the spiral structure of our Galaxy is two- or four-armed has not yet been completely resolved. Thus, for example, Russeil (2003) and, previously, one of us (Loktin and Matkin 1992) found the spiral structure to be four-armed by statistically analyzing the positions of young objects in the Galactic disk. Amaral and Lepine (1997) suggested a

pattern of superimposed two- and four-armed structures. Note that the pitch angle of the spirals in the pattern suggested by these authors is $14°$. On the other hand, Efremov (1999) does not attribute the Orion arm to the system of large-scale spiral arms (the so-called grand design). A recent estimate of the pattern speed by Debattista *et al.* (2002), $\Omega_P = 59 \pm 5$ km s^{-1} kpc^{-1}, which is more than twice as high as that found by other authors (Loktin and Matkin 2002), gives an idea of the accuracy of determining this parameter. Yano *et al.* (2002) pointed out that the data currently available on the positions and motions of young objects are not enough for a reliable kinematic analysis of the spiral structure.

In this paper, we attempt to determine the parameters of the spiral structure of our Galaxy using data on the positions and ages of open star clusters (OSCs), traditional objects for our research team. Recall that the OSC positions and ages are related to the properties of the spiral density wave not directly but via star formation, which, for example, can lead to local deviations of the positions of groups of clusters from those dictated by the density wave. Therefore, in this paper, our conclusions are drawn from the spiral structure as a whole rather than from an analysis of individual structural elements.

THE SAMPLE

We used open star clusters as the objects of our study. OSCs are convenient to use for analyzing the parameters of the spiral structure because they have reliable estimates of their ages and heliocentric distances. We took OSC data from the current version

[*]E-mail: Alexhander.Loktin@usu.ru
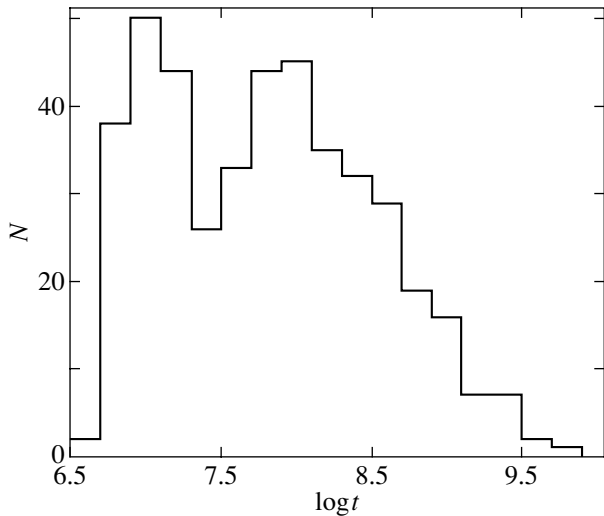
Fig. 1. The frequency distribution of cluster ages for our sample.

of the Homogeneous Catalog of OSC Parameters described by Loktin *et al.* (2001). The current version of the catalog contains the color excesses, ages, and heliocentric distances for 458 OSCs; the OSC distance scale was reconciled with the Hipparcos trigonometric parallaxes (see Beshenov and Loktin 2001).

OSC age estimates must be used to estimate the pattern speed and the possible evolution of the spiral pattern with time. Figure 1 shows the frequency distribution of cluster ages for our sample. This distribution indicates that clusters with ages up to $t \approx 10^9$ yr can be used for our analysis; the number of older clusters in the Galaxy is too small. In all of our calculations, we used weighted OSC parameters. The weights are given in the catalog and are proportional to the quality of the corresponding parameters.

PARAMETERS OF THE SPIRAL STRUCTURE

The youngest OSCs in projection onto the Galactic plane are known to lie mostly along several features extended in the direction of Galactic rotation that are usually identified with spiral-arm fragments. Intermediate-age and older clusters are believed to show no spiral structure. This can be seen from Fig. 2. Figures 2a and 2b show the spatial distributions of the youngest ($\log t < 7.0$) and intermediate-age ($7.8 < \log t < 8.4$) clusters from our sample in projection onto the Galactic plane, respectively. In these figures, the Galactic center is to the right, and the Galactocentric distances calculated with the assumed Galactocentric distance of the Sun $R_0 = 8.3$ kpc (Gerasimenko 2003) and the Galactocentric azimuthal an-

gle θ measured from the Sun−Galactic center direction in the sense of Galactic rotation are along the axes. We deliberately do not show the positions of the spiral arms to demonstrate clearly that even the youngest OSCs reliably outline only one fragment of the spiral structure, the Carina−Sagittarius arm, which also shows up in the right panel of the figure. The Orion and Perseus arms can be identified only in the case of great imagination. Clearly, an objective quantitative criterion is required to determine the positions of spiral-arm fragments. We chose a criterion based on the rms deviation from the straight line parallel to the spiral-arm fragments in the $\ln(R) - \theta$ diagram, where R is the Galactocentric distance of the clusters. We calculated the rms deviations using only those OSCs that deviated from the straight line by no more than a specified value, and added the squares of the deviations for specific OSCs with weights proportional to the quality of the observational data. This procedure is explained in Fig. 3. More specifically, we performed the following calculations: For each cluster, we calculated the deviation $\Delta \ln R_i$ from the moving straight line along the $\ln(R)$ axis. The straight line was chosen here because the logarithmic spiral, which is generally believed to fit the spiral arms of galaxies well, turns into a straight line in the coordinate axes used. As the straight line moves along the $\ln(R)$ axis, the rms deviation S is a measure of the number of clusters lying near the straight line. At each step in the motion of the straight line along the coordinate axis parallel to itself, clusters that deviated from the straight line by $\Delta \ln R \leq 0.1$ were included in the calculations of S.

The slope of the straight line to the vertical axis is the pitch angle of the spiral arms. A correct choice of the slope of the moving straight line ensures the largest difference between the maxima and minima of the function $S(\ln R)$. We determined the most suitable pitch angle of the spiral arms, $i = 22°1$, by changing the slope of the straight line. Similarly, we can shift the straight line in the vertical direction to obtain the function $S(\theta)$, which has the largest amplitude at $i = 20°9$. We finally took the mean of $i = 21°5 \pm 0°6$ for the pitch angle of the spiral arms in our Galaxy. The solid and dashed lines in Fig. 3 indicate the straight lines with the best slope in the amplitude of the functions $S(\ln R)$ and $S(\theta)$, respectively. A test calculation with the Galactocentric distance of the Sun $R_0 = 6$ kpc (a short Galactic distance scale) yielded a spiral-arm pitch angle of $18°8$, so the previous low values of this parameters obtained by other authors cannot be explained by the choice of a different Galactic distance scale. Naturally, we applied the above procedure for determining of the spiral arm pitch angle only to the youngest OSCs, since older clusters lose information about their birth
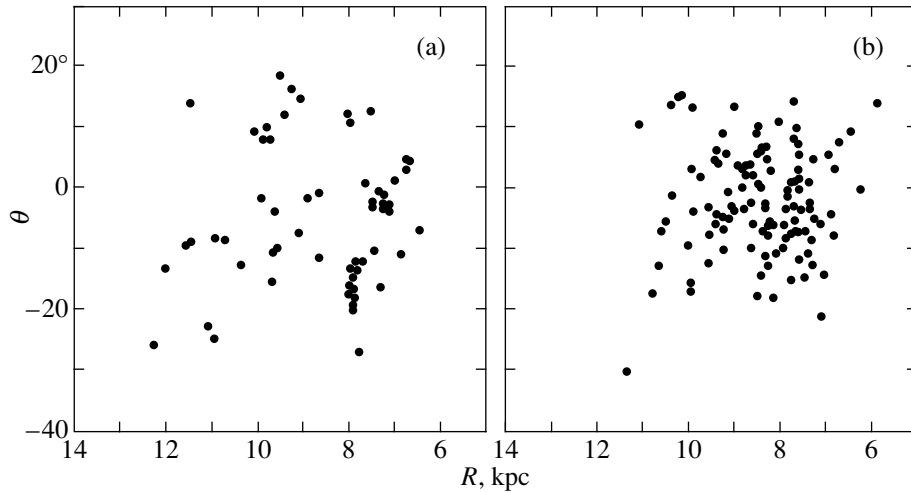
Fig. 2. The spatial distribution of young ($\log t < 7.0$) (a) and intermediate-age ($7.8 < \log t < 8.4$) (b) OSCs in projection onto the Galactic plane.
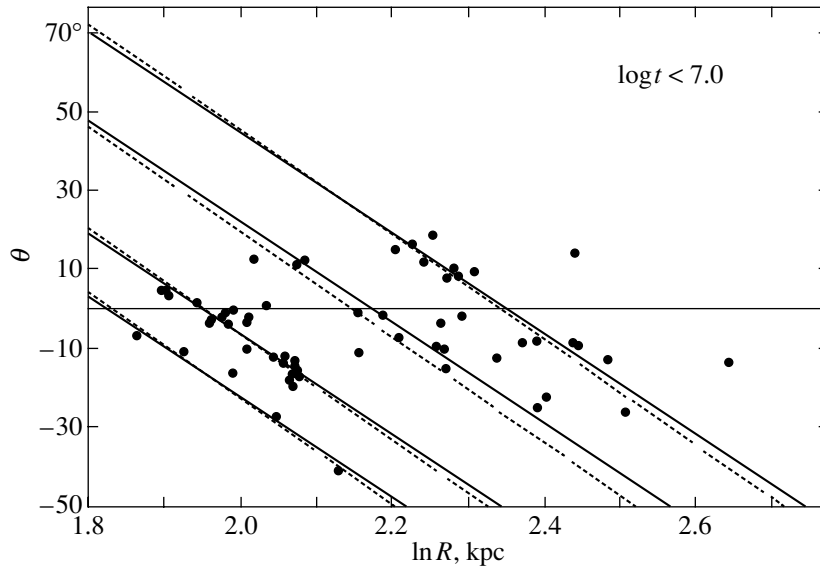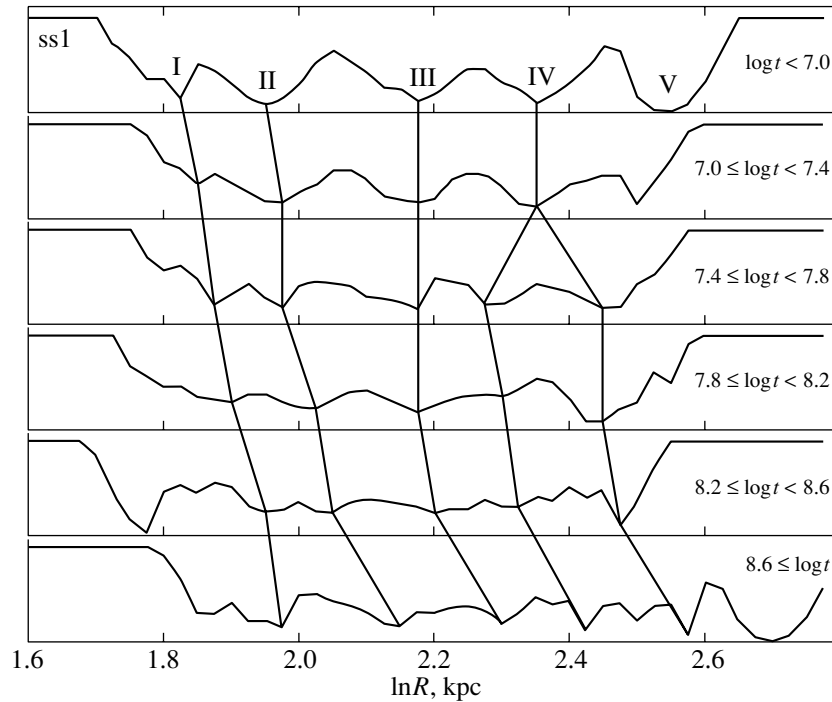

Fig. 3. Determining the positions of the spiral arms for one OSC age group.

in the spiral arm regions due to their peculiar motions and differential Galactic rotation.

Figure 4 plots the function of the sum of the rms deviations $S(\ln R)$ for six OSC age groups with ages younger than 10^9 yr. In this figure, the ages increase from the top downward, and the minima denoted by Roman numerals correspond to the well-known spiral arm fragments (see Figs. 2 and 3): II, the Carina–Sagittarius arm; III, the Orion arm; IV, the Perseus arm; and I and V, the outer and inner arms, respectively. Since the outer and inner arms for our sample are outlined by only a few clusters in each age interval, our attention is naturally focused on arms

II, III, and IV. We clearly see that the maxima and minima of the function $S(\ln R)$ are well defined for the youngest OSCs. However, the positions of the minima corresponding to the spiral arms identified by the youngest objects can be determined even for the oldest OSCs. The vertical broken lines in Fig. 4 connect the minima corresponding to the same elements of the spiral structure. Our first attempt to identify the minima on the $S(\ln R)$ curves for different age groups ran into difficulties, which we were able to overcome by using plots that corresponded to overlapping age intervals. However, even this approach did not allow us to clearly trace the minima corresponding to the

Fig. 4. Dependence of the rms deviations of the cluster positions from the moving straight line along the R axis for six OSC age groups. The broken lines running from top downward connect the minima corresponding to the spiral arms of the same name.

Perseus arm for all ages. We get the impression that in the past, there were two different elements of the spiral structure at the location of the Perseus arm.

We made similar constructions and determinations of the spiral arm locations for the plots of the function $S(\theta)$. The nonsmooth structure of the broken lines connecting the minima can be primarily explained by the fact that the spiral arm locations are determined not directly, but through the results of star formation triggered by the spiral density wave.

Table 1 gives the data obtained for overlapping age intervals. We used the overlapping age intervals to smooth the results under conditions of small sample sizes. Column 1 in Table 1 lists the intervals of age logarithms, column 2 lists the mean age logarithms for a given interval, column 3 gives the number of clusters in the interval, and columns 4−9 give the positions of the minima for the derived spiral arms on the straight line connecting the Sun and the Galactic center ($\theta = 0°$). The last three columns correspond to the split of the Perseus arm in the past. Since arm V (in our notation) contains only two or three OSCs in each age interval, we did not include the results for this arm in the table due to their low reliability.

Table 2 was compiled in a similar way. It gives the spiral arm locations inferred from the plots of the functions $S(\theta)$. The data in Table 2 clearly allow the pattern speed to be directly calculated, since they give the time (cluster age) dependence of the location of each observed spiral arm fragment. Similarly, the pattern speed can also be estimated from the data of Table 1 by performing obvious calculations (differentiation of the formula for the logarithmic spiral with respect to time).

Figures 5 and 6 show the spiral arm locations as a function of time (the mean age of cluster groups) based on the data of Tables 1 and 2. Galactocentric distances (in kpc) and Galactocentric azimuthal angles (in degrees) are along the vertical axes in Figs. 5 and 6, respectively. These figures show that the locations of the identified spiral arms as a function of time fall nicely on the segments of parallel straight lines, so the spiral pattern actually rotates rigidly and the pattern speed has not changed appreciably over the past billion years. Note that the data points corresponding to the split of the Perseus arm in the past fall nicely on the straight lines parallel to those corresponding to other spiral arms. Figure 6 does not show arm IVb, because no corresponding minima could be reliably identified in the plot of $S(\theta)$ against $\log R$. For unclear reasons, the function $S(\ln R)$ results in a smaller scatter of data points in the plots of the spiral arm locations against time than does the function $S(\theta)$.

Table 3 gives the pattern speeds estimated from the slopes of the straight lines that fit the positions of the data points in Figs. 5 and 6. The data in

Table 1

Intervals of $\log t$	Mean $\log t$	N	$\ln R\,(\theta = 0°)$					
			I	II	III	IV	IVa	IVb
6.4−6.8	6.75	15	−	1.950	−	2.325	−	−
6.5−7.0	6.87	62	1.825	1.950	2.175	2.350	−	−
6.8−7.2	7.00	102	1.850	1.950	2.150	2.350	−	−
7.0−7.4	7.17	83	1.850	1.975	2.175	2.350	−	−
7.2−7.6	7.40	64	1.875	1.975	2.175	2.350	−	−
7.4−7.8	7.62	73	1.875	1.975	2.170	−	2.275	2.450
7.6−8.0	7.81	86	1.875	1.975	2.150	−	2.275	2.450
7.8−8.2	8.00	90	1.900	2.025	2.175	−	2.300	2.450
8.0−8.4	8.17	71	1.900	2.000	2.200	−	2.300	2.475
8.2−8.6	8.42	65	1.950	2.050	2.200	−	2.325	2.475
8.4−8.8	8.56	56	1.950	2.075	2.200	−	2.325	−
8.6−9.0	8.76	36	1.975	2.150	2.300	−	−	−
8.8−9.2	8.98	33	1.975	2.140	−	−	−	−

Table 2

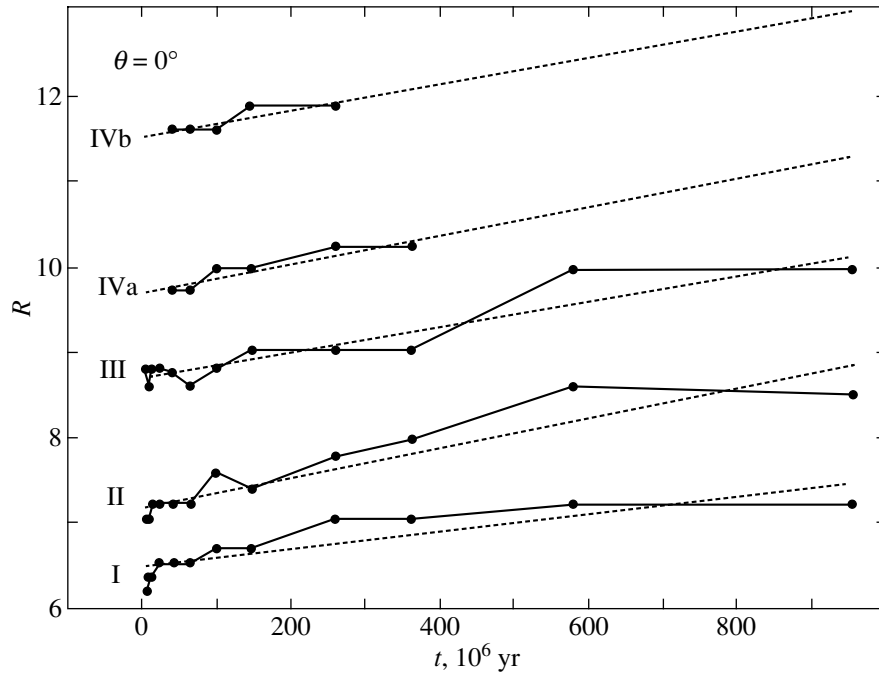| Intervals of $\log t$	$\theta°$(at $\ln R = 1.8$)					
	I	II	III	IV	IVa	IVb
6.4−6.8	−	20	−	72	−	−
6.5−7.0	4	20	46	72	−	−
6.8−7.2	6	22	48	74	−	−
7.0−7.4	−	24	50	74	−	−
7.2−7.6	−	24	46	74	−	−
7.4−7.8	−	22	42	−	64	86
7.6−8.0	−	22	44	−	64	85
7.8−8.2	14	30	52	−	66	84
8.0−8.4	14	28	52	−	66	−
8.2−8.6	20	32	50	−	66	−
8.4−8.8	−	32	50	−	70	−
8.6−9.0	26	42	−	−	−	−
8.8−9.2	25	46	64	−	−	−

Fig. 5. Spiral arm locations (along the Sun–Galactic center straight line) as a function of the mean ages of the cluster groups.
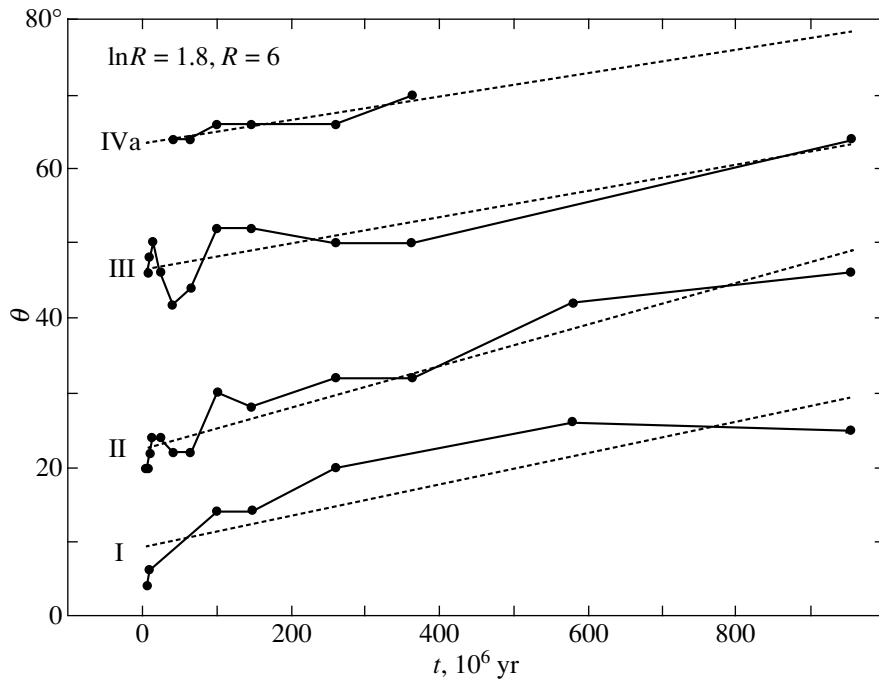

Fig. 6. Spiral arm locations (along the axis of azimuthal angles) as a function of the mean ages of the cluster groups.

Table 3 yielded a weighted mean pattern speed of $\Omega_P = 22.7 \pm 3.4$ km s^{-1} kpc^{-1}. Clearly, since the rotation of the spiral pattern is rigid and the rotation of the Galactic disk is differential, the OCS-producing

clouds of gas and dust and the OSCs born in them move with different angular velocities with respect to the spiral density wave. This effect is seen in the data of Table 3 despite the relatively large errors. The loca-

Table 3. Estimates of the pattern speed

Arm no.	Ω_P, km s^{-1} kpc^{-1} (inferred from R)	Ω_P, km s^{-1} kpc^{-1} (inferred from θ)	Number of OSCs in arm
I	22.0 ± 4.1	20.0 ± 5.6	48
II	33.0 ± 4.1	27.0 ± 2.6	147
III	24.0 ± 3.1	17.0 ± 3.5	156
IVa	24.0 ± 5.0	15.0 ± 3.7	67
IVb	19.0 ± 7.0	$-$	11

tions of the observed spiral arm fragments are symmetric about the Sun, which makes it possible to average the estimates in Table 3 and to obtain the above value of Ω_P. However, this effect increases appreciably the error in Ω_P. It is easy to apply corrections for the difference between the Galactic rotation rates at different Galactocentric distances. To this end, we used the Kamm function (which will be described in our next paper) constructed from the radial velocities and proper motions of OSCs. After applying these corrections, the weighted mean pattern speed was found to be $\Omega_P = 20.4 \pm 2.5$ km s^{-1} kpc^{-1}.

## GEOMETRY OF THE SPIRAL STRUCTURE

Unfortunately, OSCs yield only short fragments of three spiral arms and give only slight hints of two other arms. Therefore, such important parameters as the number of arms and the initial phase angle of the spiral structure are difficult to determine from objects of this class. Moreover, as we show below, it is difficult to solve the question of whether our Galaxy has a grand design or a flocculent structure. Let us show this with examples: As usual, we represent the spiral arms by logarithmic spirals of the form

$$R = \exp[(\vartheta - \vartheta_0)/\tan i],$$

where the phase angle ϑ of the spiral is shifted by 180° relative to the angle θ used above, and ϑ_0 denotes the initial phase angle of the spiral. Applying a nonlinear least-squares method to our data yielded $\vartheta_0 = 206°$ and $i = 16°8$ (we used only clusters with $\log t < 7.5$). In this case, as in the analyses of other authors (see references in the Introduction), the least-squares method yields smaller pitch angles for the spirals than the direct measurement performed above. Figure 7a shows three spiral arms with the parameters derived here. The structure can be either two- or four-armed, depending on whether we consider the Orion arm to be an element of the main structure.

If we individually fit all of the identified spiral-arm fragments by logarithmic spirals, the phase differences between the fitting spirals will be far from even

fractions of π. Figure 7b shows the resulting structure, and Table 4 gives the initial phase angles for each of the fitting logarithmic spirals. To avoid difficulties
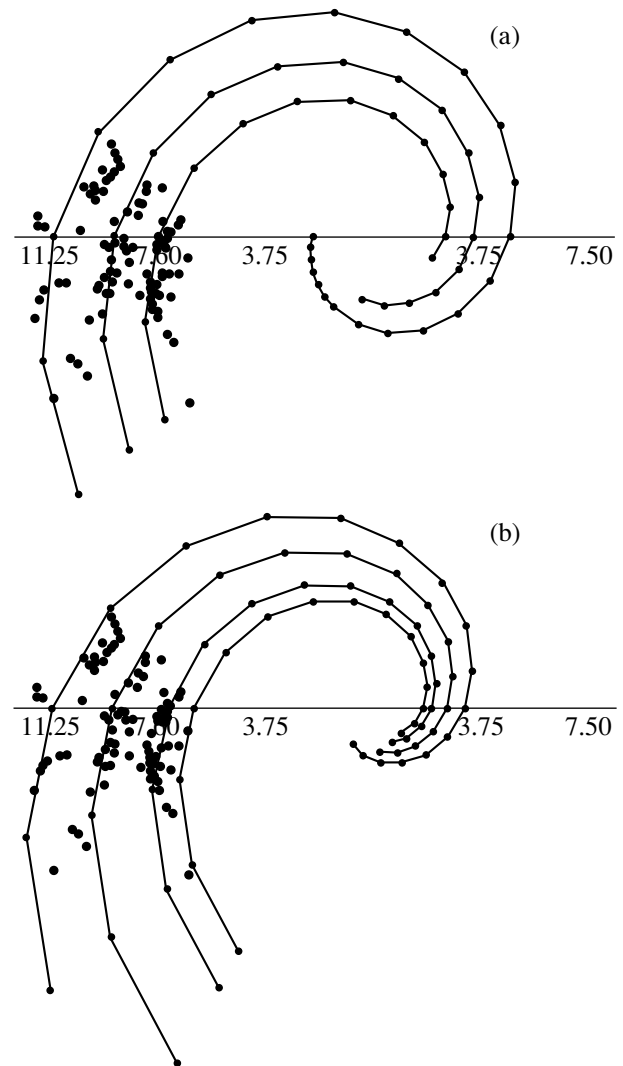
Fig. 7. Observed fragments of the Galactic spiral arms as inferred from the positions of clusters with ages $\log t < 7.2$ fitted individually by logarithmic spirals (a) and fitted by a four-armed pattern with the parameters estimated in this paper (b).

Table 4

Arm	ϑ_0, deg
II	255
III	220
IV	190

with the possible splitting of the Perseus arm in the past mentioned above, we used only clusters with ages $\log t < 7.5$ to fit the observed arms by segments of a logarithmic spiral.

CONCLUSIONS

As we see from the above discussion, the available data on the positions and ages of OSCs from the Homogeneous Catalog of OSC Parameters allow some of the parameters of the spiral structure of our Galaxy to be reliably determined. Thus, for example, the pitch angle and the pattern speed are $i = 21°5$ and $\Omega_{\mathrm{P}} = 20.4 \pm 2.5$ km s^{-1} kpc^{-1}, respectively. The spiral structure of our Galaxy has remained virtually unchanged over the past billion years, and signatures of the concentration of objects toward the spiral arms can be traced back to such ages. Therefore, it would be of interest to perform a similar analysis based on Cepheids. However, the number of arms in the structure cannot be determined from OSCs. So far, the positions of H II regions yield more reliable results.

The positions of young objects are related to the spiral pattern parameters not directly, but via the response of the gas−dust medium that produces these objects to a spiral density wave. This can lead to certain errors in the parameters of the spiral structure. In our next paper, we hope to test our conclusions by analyzing the distortions of the velocity field in the Galactic disk.

REFERENCES

1. L. H. Amaral and J. R. D. Lepine, Mon. Not. R. Astron. Soc. **286**, 885 (1997).
2. G. V. Beshenov and A. V. Loktin, Pis'ma Astron. Zh. **27**, 450 (2001) [Astron. Lett. **27**, 386 (2001)].
3. V. P. Debattista, O. Gerhard, and M. N. Sevenster, Mon. Not. R. Astron. Soc. **334**, 355 (2002).
4. Yu. N. Efremov, Astron. Astrophys. Trans. **18**, 321 (1999).
5. T. P. Gerasimenko, Astron. Rep. **48**, 103 (2004).
6. A. V. Loktin, T. P. Gerasimenko, and L. K. Malysheva, Astron. Astrophys. Trans. **20**, 607 (2001).
7. A. V. Loktin and N. V. Matkin, Astron. Astrophys. Trans. **3**, 169 (1992).
8. L. S. Marochnik and A. A. Suchkov, *Galaxy* (Nauka, Moscow, 1984).
9. Yu. N. Mishurov and I. A. Zenina, Astron. Astrophys. **341**, 81 (1999).
10. D. Russeil, Astron. Astrophys. **397**, 133 (2003).
11. T. Yano, M. Chiba, and N. Gouda, Astron. Astrophys. **389**, 143 (2002).

Translated by A. Dambis

Interaction of the Supernova Remnant HB3 with the Ambient Interstellar Gas

I. V. Gosachinskiĭ*

Special Astrophysical Observatory, St. Petersburg Branch, Russian Academy of Sciences, Pulkovo, St. Petersburg, 196140 Russia
Received August 4, 2004

Abstract—The well-known shell supernova remnant (SNR) HB3 is part of a feature-rich star-forming region together with the nebulae W3, W4, and W5. We study the H I structure around this SNR using five RATAN-600 drift curves obtained at a wavelength of 21 cm with an angular resolution of $2'$ in one coordinate over the radial-velocity range -183 to $+60$ km s^{-1} in a wider region of the sky and with a higher sensitivity than in previous works by other authors. The spatial–kinematic distribution of H I features around the SNR clearly shows two concentric expanding shells of gas that surround the SNR and coincide with it in all three coordinates (α, δ, and V). The outer shell has a radius of 133 pc, a thickness of 24 pc, and an expansion velocity of 48 km s^{-1}. The mass of the gas in it is $\approx 2.3 \times 10^5$ $M_\odot$. For the inner shell, these parameters are 78 pc, 36 pc, 24 km s^{-1}, and 0.9×10^5 $M_\odot$, respectively. The inner shell is immediately adjacent to the SNR. Assuming that the outer shell was produced by the stellar wind and the inner shell arose from the shock wave of the SNR proper, we estimated the age of the outer shell, $\approx 1.7 \times 10^6$ yr, and the mechanical luminosity of the stellar wind, 1.5×10^{38} erg s^{-1}. The inner shell has an age of $\approx 10^6$ yr and corresponds to a total supernova explosion energy of $\approx 10^{52}$ erg. © *2005 Pleiades Publishing, Inc.*

Key words: *interstellar medium, gaseous nebulae, neutral hydrogen, supernovae and supernova remnants.*

INTRODUCTION

The supernova remnant (SNR) HB3 (G132.6+1.5) is well known, since it is part of a rich star-forming region in the Cas OB6 stellar association together with the H II regions W3, W4, and W5. This SNR is commonly assumed to be physically associated with them. At radio wavelengths, it has a significant angular size ($\approx 80'$) and a shell structure, suggesting that its evolution time is fairly long. This SNR has been extensively studied in the continuum, infrared, X rays, and molecular lines (see Landecker *et al.* (1987) and references therein).

The neutral hydrogen in this region has been studied by several authors. The papers by Routledge *et al.* (1991) and Normandeau *et al.* (1997), who used the DRAO aperture-synthesis radio telescope with an angular resolution of $2'$ and $1'$, are particularly noteworthy. Note, however, that Routledge *et al.* (1991) restricted themselves to a field of view of one primary beam ($3°$ at the 10% level) with an angular resolution of $2'$ and a sensitivity of 1.34 K, so the sensitivity at the edges of the field deteriorated sharply during the data processing. Normandeau *et al.* (1997) had a rather

poor sensitivity ($3-15$ K) and found nothing new in this region, compared to Routledge *et al.* (1991). Note also that Routledge *et al.* (1991) gave a brief review of the then available observations of HB3 in the radio, X-ray, and optical ranges and in molecular lines.

In the period $2001-2004$, we obtained five drift curves through the SNR HB3 in the H I line on the southern sector of the RATAN-600 radio telescope in a region that was twice that in Routledge *et al.* (1991) and with a sensitivity that was a factor of 20 higher than that in Normandeau *et al.* (1997). We confirmed several conclusions of the former paper and found new features in the distribution of gas around the SNR.

Note also that the distribution of neutral hydrogen emission in the Galactic plane is highly nonuniform and rich in various features. It thus follows that searching for evidence of the physical association of H I gas with such objects as SNRs (i.e., evidence that this is not a chance projection) is not an easy task. In our case, however, the problem was simplified, since the SNR is an active expanding object, and this must also be reflected in the kinematics of the ambient gas. Unfortunately, many authors pay little attention to this and restrict their analysis to the distribution of gas only in the plane of the sky. The goal of our work

was to search for expanding H I shells around the SNR, which manifest themselves as ringlike structures in the coordinate−radial velocity plane.

INSTRUMENTATION AND TECHNIQUE

To study the distribution of interstellar neutral hydrogen in the region of the SNR HB3, we obtained five drift curves in right ascension at $0°6$ declination intervals in the range of declinations from $+61°25$ to $+63°75$ on the southern sector of RATAN-600 at lower culmination (an elevation of $17°3$). In this elevation range, the RATAN-600 antenna has an angular resolution of $2'6 \times 65'$ and an effective area of about 900 m^2 (Esepkina et al. 1979). An uncooled HEMT amplifier (Il'in et al. 1997) was used at the input. The system noise temperature was about 70 K, the 39-channel filter-bank spectral analyzer had a channel width of 30 kHz (6.3 km s^{-1}), and the channel spacing was also 30 kHz (Venger et al. 1982). The system control, data acquisition, and primary data processing were performed with an IBM PC computer (Alferova et al. 1986).

The drift curve at each declination consisted of two series with three observations each obtained with a receiver tuning shift by half the channel width, so each drift curve had 78 spectral channels following at 3.15 km s^{-1} intervals. This technique also allowed us to effectively eliminate interference. On the averaged record, the mean square of the antenna temperature fluctuations in spectral channels was 0.2 K. In each series of observations, the antenna and instrumentation parameters were controlled using measurements of a series of reference sources (Venger et al. 1981). For the southern sector, we used 3C 147 and 3C 430.

We determined the parameters of features in each channel using a Gaussian fitting program and then attempted to establish a relationship between H I emission features at different radial velocities and at different declinations. This is the only procedure where a significant subjective factor could be introduced.

The measured parameters of H I features have the following errors. The radial velocity of an isolated medium-brightness H I feature is measured with an accuracy of no lower than 1 km s^{-1}. In certain cases, the accuracy deteriorates due to the difficulties in separating an object from the background or from adjacent features. Given the antenna calibration errors, the measurement error of the brightness temperature of the H I line is about 0.3 K, and the error in the estimated angular sizes in right ascension is $0°1$. In declination, the antenna resolution is much lower, and, therefore, the accuracy of measuring the angular sizes is lower. The accuracy of estimating the distances depends on the method of their determination

and, in each case, must be considered separately. As a result, the accuracy of estimating the H I mass in an isolated cloud is no better than 0.5−1 order of magnitude.

RESULTS OF THE OBSERVATIONS

Figures 1−5 show the $\alpha−V$ diagrams for each drift curve without elimination of the extended H I emission background. The latter was done to avoid the systematic errors that appear in any technique for eliminating an extended background. The declinations (1950.0) of the drift curves are indicated at the top. To save space, the drift curves are given only in a limited range of radial velocities, from about $+9$ to $−100$ km s^{-1}, and no zero lines of each curve are given. The radial velocity with respect to the Local Standard of Rest is indicated near each curve on the right. The spectrometer channel number and the antenna temperature scale are indicated on the left. Compact H I features in the SNR region were identified using linear interpolation, which, of course, can cause their antenna temperature to be slightly overestimated. In each figure, the upper curve represents the drift curve in the continuum at a frequency of 1420 MHz. The SNR is highlighted by dark shading; in the remaining part of the figure, heavy vertical lines corresponding to the SNR boundaries at minimum brightness in right ascension were drawn. The same dark shading in some of the figures highlights the absorption line from bright thermal continuum sources (W3, W4).

The thin closed lines represent the results of our attempts to combine the H I features at adjacent radial velocities into single structures surrounding the SNR. In our opinion, two such structures are observed, outer and inner. The features of the drift curves pertaining to different structures are highlighted by shading of different densities. The presence of such ringlike ($\alpha−V$) structures generally reflects the expanding shells of gas around the object. The table lists the measured and calculated parameters of neutral hydrogen in these structures under the assumption that they are indeed expanding shells around the SNR. We adopt the distance to the H I shells, 3 ± 1 kpc, in accordance with their mean radial velocities and the IAU 1986 model of Galactic rotation (Kerr et al. 1986). The mean radial velocities that we obtained are very close to the optical velocities of the HB3 features inferred from the Hα observations by Lozinskaya and Sitnik (1980), and the adopted distance corresponds to this work. We see that the H I distribution in the inner shell is highly asymmetric; therefore, the table gives two values of the H I density, for the eastern and western parts of the shell. The knife-edge beam of the radio telescope makes
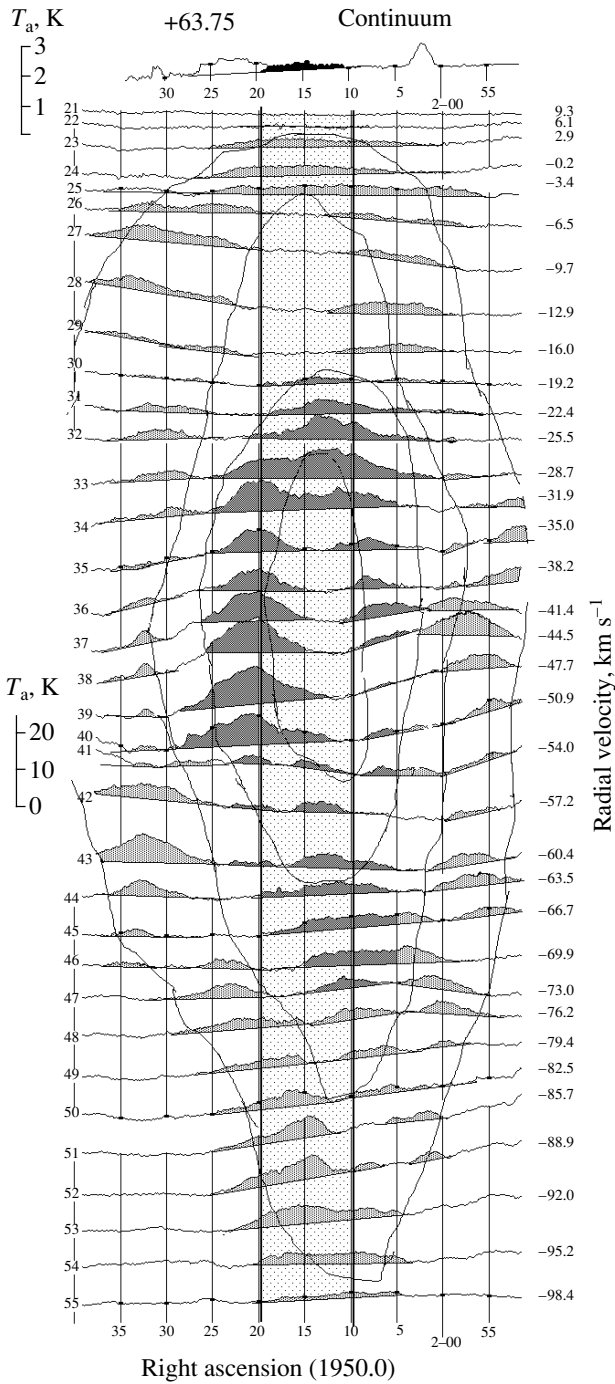
Fig. 1. Drift curves of the H I line emission in the region of the SNR HB3 at a declination of +63°.75. The SNR continuum is highlighted by the black color on the upper curve. The radial velocities of each curve are given on the right. The gray shading marks the H I features that may be associated with the SNR.
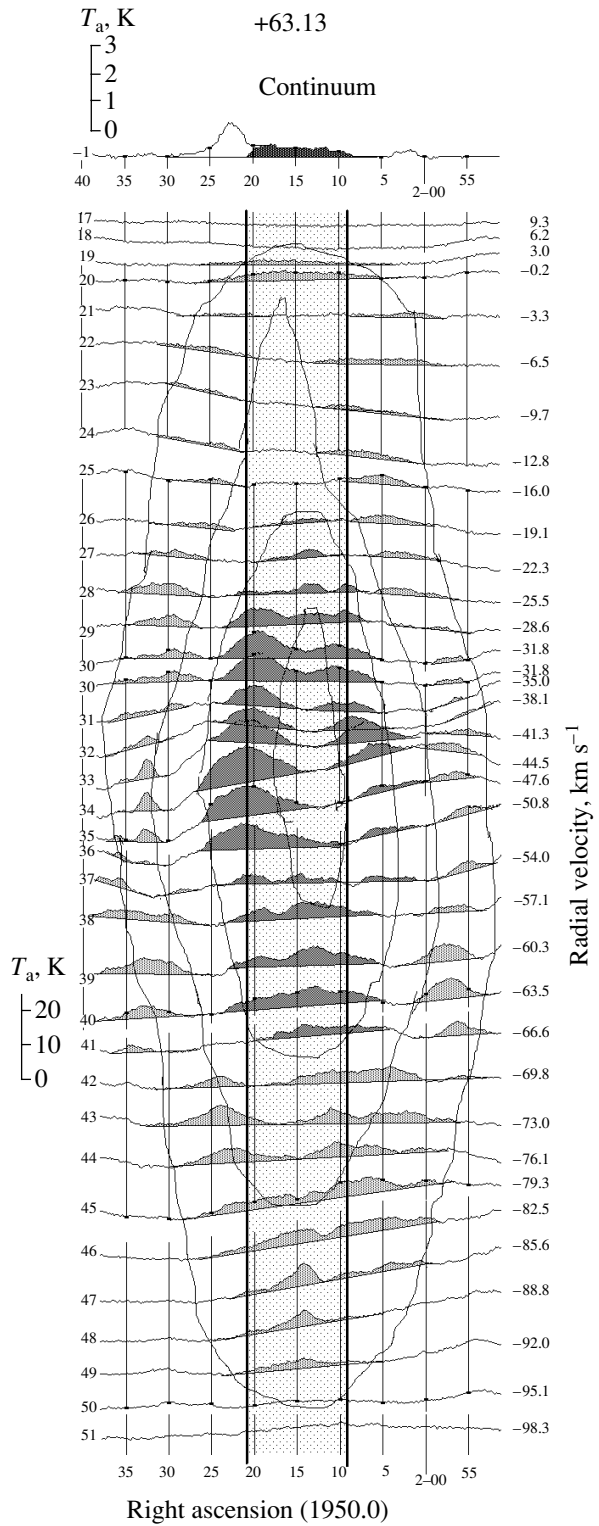
Fig. 2. Same as Fig. 1 at a declination of +63°.13.

it impossible to estimate the size of the H I features in declination with acceptable accuracy. However, it can be seen that the observed radial-velocity range of

the H I ringlike structures decreases northward and southward of the central drift curve.

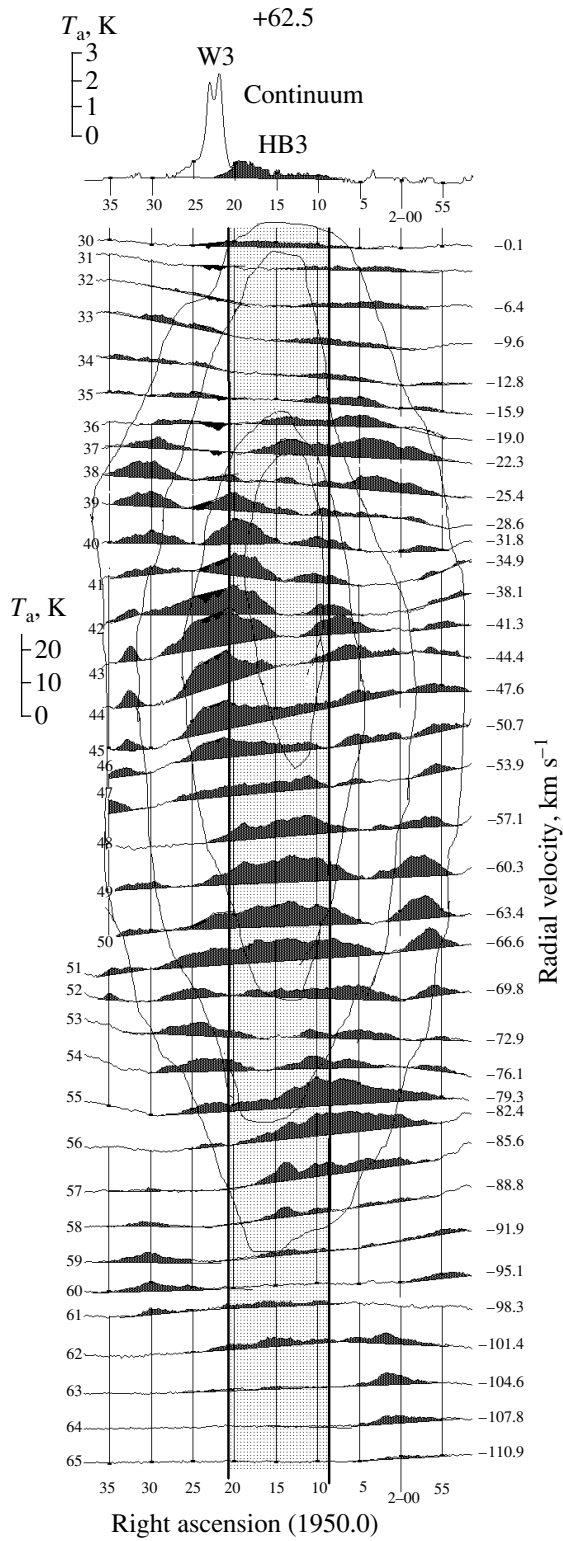The simplest physical interpretation of the H I shell structure that we detected around HB3 is that

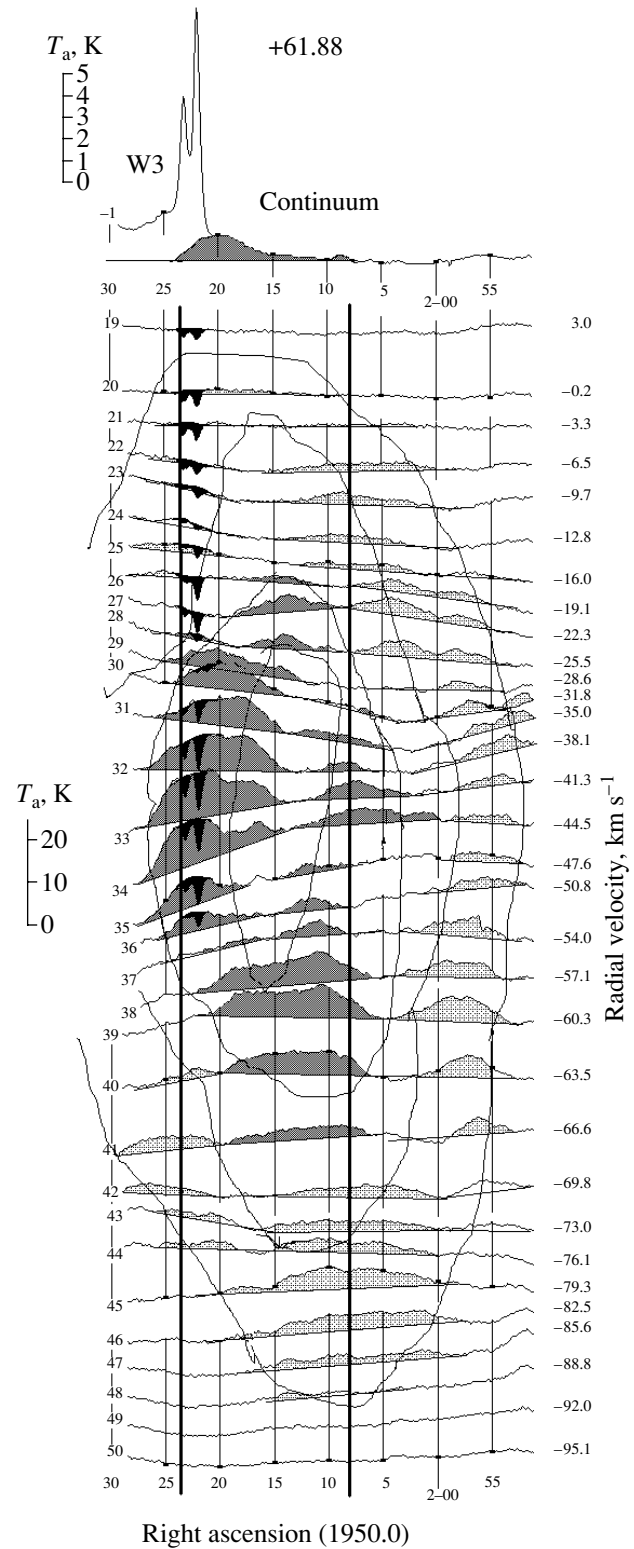**Fig. 3.** Same as Fig. 1 at a declination of +62°5.

Fig. 4. Same as Fig. 1 at a declination of +61°88.

the outer shell may be assumed to be produced by the wind from the star over its lifetime on the main sequence, while the inner shell arose from the

SNR shock. In this case, we can estimate several physical properties of the star and its explosion using theoretical estimates of the effect of shock waves from
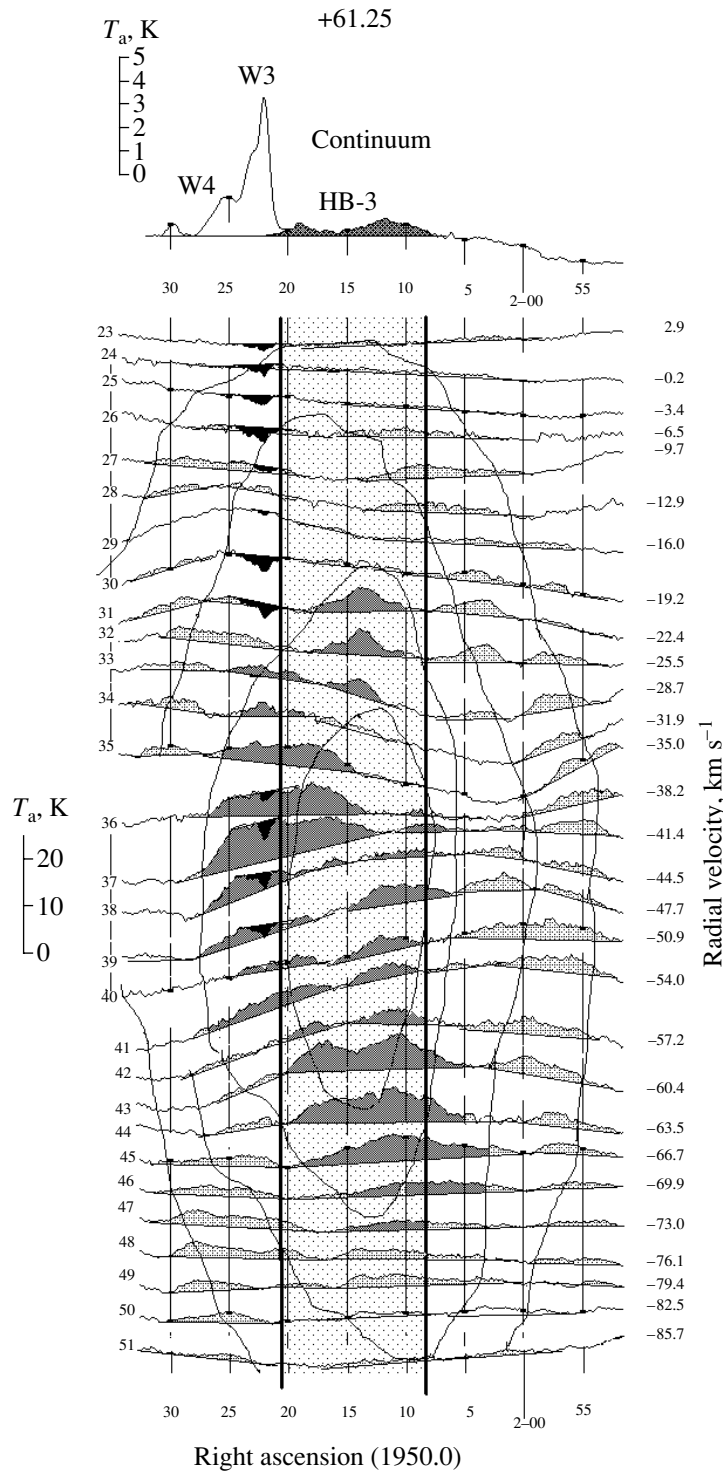
Fig. 5. Same as Fig. 1 at a declination of +61°25.

the stellar wind and SNRs on the interstellar medium. The initial density of the medium needed for such estimates can be obtained by "spreading" the total amount of gas in the shells over the entire current volume.

To estimate the mechanical luminosity of the stellar wind from the observed parameters of the outer shell, we can use the model by Weaver *et al.* (1977). Our estimations yield an age of the shell 1.7×10^6 yr and a mechanical luminosity of the stellar wind $1.5 \times$

Measured parameters of the H I shells

Parameters	Outer shell	Inner shell
Coordinates	$\alpha = 2^{\mathrm{h}}15^{\mathrm{m}}$	$\alpha = 2^{\mathrm{h}}15^{\mathrm{m}}$
(1950.0)	$\delta = +62°5$	$\delta = +62°5$
Radial velocity, km s^{-1}	-44.4	-42.3
Expansion velocity, km s^{-1}	47.5	24.3
Distance, kpc	3	3
Outer radius, pc	133	78
Thickness, pc	24	36
HI mass, $M_{\odot}$	2.3×10^5	0.9×10^5
HI density, cm^{-3}	2.1	(E) 2.46 (W) 1.23
Initial H I density, cm^{-3}	0.95	1.85

10^{38} erg s^{-1}. At a typical stellar wind particle velocity of ≈ 2000 km s^{-1}, this mechanical luminosity requires a relatively high mass loss rate, 6.8×10^{-5} $M_{\odot}$ yr^{-1}, which is characteristic of massive O-type stars or M-type supergiants (de Jager 1984). It is well known that for Wolf−Rayet stars, the wind velocity can be several times higher; however, this stage is not too long and is unlikely to play a significant role in forming the outer shell.

The kinematic parameters of the inner shell make it possible to estimate the parameters of the supernova explosion by using the model from Wheeler et al. (1980). The SNR age is $\approx 10^6$ yr, and the total explosion energy is $\approx 10^{52}$ erg.

CONCLUSIONS

Thus, in the simplest model, our observations of the neutral hydrogen distribution around the SNR HB3 have shown that, first, we see here the results of the explosion of one of the most massive stars in the Galaxy, and, second, the shells from the stellar wind and from the SNR shock are observed simultaneously. This case seems extremely rare, although traces of the action of the wind and the SNR were found in several other objects, and the model used may not be unique. Note that the presence of a dense inner HI shell around HB3 was also pointed out by Routledge et al. (1991), although these authors could not determine the full range of its velocities. In addition, they also pointed out traces of the outer H I shell, but failed to get its complete picture due to the limited field of view.

In general, our data may be considered to be consistent with the model of Routledge et al. (1991).

However, the quantitative estimates of the shell age and energetics should be considered approximate, bearing in mind the semiquantitative nature of the gas-dynamical models of Weaver et al. (1977) and Wheeler et al. (1980) used for this purpose. However, energetics of this scale is now not considered as something extraordinary. It is rather difficult to estimate the effect of the wind from the stars of the Cas OB6 association on the kinematics of the outer H I shell surrounding the SNR HB3, since the SNR, although being close to the association, is nevertheless outside it (see Fig. 1 from Routledge et al. 1991). As regards the inner shell, it probably has a much more complex structure than that assumed above in the simple model of two expanding H I shells. Nevertheless, all the above authors consider it to be the result of the action of the SNR shock on the interstellar gas. In our drift curves, especially at the southernmost declinations $+61°25$ and $+61°88$, we clearly see that the bulk of the H I is observed in the eastern and close (to the Sun) parts of the shell (the radial velocities are negative with respect to the mean velocity of the object). Incidentally, according to Huang and Thaddeus (1986), it is here that a CO cloud is located. In the western and far (from us) parts of the HB3 shell, where the SNR radio brightness is very low, Routledge et al. (1991) placed the high-velocity features of gas detected, in particular, by Lozinskaya and Sitnik (1980) in their model. According to Lozinskaya and Sitnik (1980), the high-velocity features have a very low mass. As regards the bulk of the Hα-emitting filaments, their expansion velocity obtained by Lozinskaya and Sitnik (1980), $\leq 35 \pm 25$ km s^{-1}, agrees well with our data on the expansion of the H I shell (see the table). The significant decrease in the density of the cold neutral

gas in the shell around the SNR in its northwestern and farther (from the Sun) part is also confirmed by our data shown in Figs. 3–5. The regions of very hot gas responsible for the X-ray emission could very well be located there as well (Seward 1990). A strong nonuniformity of the distribution of interstellar gas in the region of the SNR HB3 and a significant amount of the molecular component dissociated by the shock could explain the rather strange mean initial densities that we obtained for the outer and inner shells.

Finally, note that we presented here the results of RATAN-600 H I observations around one of the 104 SNRs selected from the catalog of Green (1998). The H I survey around them has now been completed, and the data are now processed. Later, drift curves in the $\alpha-V$ plane similar to those presented here will be available on the web page of the Special Astrophysical Observatory, Russian Academy of Sciences, for free use.

ACKNOWLEDGMENTS

REFERENCES

1. Z. A. Alferova, I. V. Gosachinskiĭ, S. R. Zhelenkov, *et al.*, Izv. Spets. Astron. Obs. **23**, 89 (1986).
2. N. A. Esepkina, N. S. Bakhvalov, B. A. Vasil'ev, *et al.*, Izv. Spets. Astron. Obs. **11**, 182 (1979).
3. D. A. Green, Astrophys. Space Sci. **148**, 3 (1998).
4. Y.-L. Huang and P. Thaddeus, Astrophys. J. **309,** 804 (1986).
5. G. N. Il'in, V. A. Prozorov, and A. M. Pilipenko, *Problems of Modern Radio Astronomy, XXVII Radioastronomical Conference, Theses of Papers* (Inst. PA Ross. Akad. Nauk, St. Petersburg, 1997), p. 128.
6. F. G. Kerr and D. Linden Bell, Mon. Not. R. Astron. Soc. **221**, 1023 (1986).
7. T. L. Landecker, J. F. Vaneldik, P. E. Dewdney, and D. Routledge, Astron. J. **94**, 111 (1987).
8. T. A. Lozinskaya and T. G. Sitnik, Astron. Zh. **57**, 997 (1980) [Sov. Astron. **24**, 572 (1980)].
9. M. Normandeau, A. R. Taylor, and P. E. Dewdney, Astrophys. J., Suppl. Ser. **108**, 279 (1997).
10. D. Routledge, P. E. Dewdney, T. L. Landecker, and J. F. Vaneldik, Astron. Astrophys. **247**, 529 (1991).
11. F. D. Seward, Astrophys. J., Suppl. Ser. **73**, 781 (1990).
12. A. P. Venger, I. V. Gosachinskiĭ, V. G. Grachev, and N. F. Ryzhkov, Izv. Spets. Astron. Obs. **14**, 118 (1981).
13. A. P. Venger, V. G. Grachev, T. M. Egorova, *et al.*, Soobshch. Spets. Astron. Obs. **35**, 5 (1982).
14. R. Weaver, R. McCray, J. Castor, *et al.*, Astrophys. J. **218**, 377 (1977).
15. J. C. Wheeler, T. J. Masurek, and A. Sivaramakrishnan, Astrophys. J. **237**, 781 (1980).
16. K. de Yager, *The Most Luminous Stars* (Mir, Moscow, 1984) [in Russian].

Translated by V. Astakhov

Strong Cosmic-Ray Scattering
in an Anisotropic Random Magnetic Field

Yu. P. Mel'nikov[*] and I. N. Toptygin[**]

St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia
Received October 7, 2004

Abstract—We calculate the kinetic coefficients and the transport mean free paths of high-energy particles parallel to the regular magnetic field in the approximation of a large-scale anisotropic random magnetic field by using a nonlinear collision integral, i.e., by taking into account the processes of strong random scattering. We consider the diffusion of solar and Galactic cosmic rays by two-dimensional turbulence. Strong random scattering by two-dimensional turbulence is shown to reduce the parallel transport mean free path severalfold. The momentum dependence of the parallel mean free path does not change, $\Lambda_{||} \propto p^{2-\nu}$. In the case of strong random scattering by turbulence formed by several modes, the parallel transport mean free path is $\Lambda_{||} \propto p$. We show that two-dimensional turbulence can make a major contribution to the parallel transport mean free paths of cosmic rays in the heliosphere and the interstellar medium. © *2005 Pleiades Publishing, Inc.*

Key words: *cosmic rays, nonthermal radiation, diffusion, interplanetary medium, interstellar medium.*

INTRODUCTION

It follows from the analysis of experimental data performed by Matthaeus *et al.* (1990) and Bieber *et al.* (1996) that the distribution of interplanetary magnetic field fluctuations is anisotropic. In the weakly disturbed inner heliosphere, the preferential direction of the magnetic field fluctuations is perpendicular to the the regular magnetic field. The wave vectors of the fluctuations are also mainly perpendicular to the regular magnetic field, which gives rise to two-dimensional fluctuations. In the interplanetary medium, the energy of the two-dimensional fluctuations can reach 85% of the energy of the random magnetic field. These magnetic fluctuations are associated with Alfvén and magnetosonic waves and with stationary structures in the solar wind.

Using satellite data, Bieber *et al.* (1996) showed that about a sixth of the magnetic field fluctuation energy is accounted for by fluctuations in which the wave vectors are parallel to the regular magnetic field (slab fluctuations). These fluctuations are attributable to Alfvén turbulence. Carbone *et al.* (1995) showed that collisionless solar cosmic-ray events are often associated with high-velocity solar wind flows. The random magnetic field is highly anisotropic in these. The wave vectors of the fluctuations in high-velocity

flows in the inner heliosphere are mainly parallel to the regular magnetic field (slab fluctuations). The correlation lengths of the random magnetic field perpendicular to the regular field are several dozen times larger than those parallel to the regular magnetic field.

The parallel transport mean free paths of high-energy particles in the interplanetary magnetic field, including the anisotropy of random fluctuations, were calculated numerically by Bieber *et al.* (1994), Teufel and Schlickeiser (2002, 2003), Teufel *et al.* (2003), Shalchi and Schlickeiser (2004) and analytically by Dröge (2003). These authors used a quasi-linear random magnetic field approximation. They showed that particles are scattered weakly by two-dimensional fluctuations. The calculated transport mean free paths of solar cosmic-ray protons exceed their observed values by several tens or hundreds of times. To reconcile the theoretical and experimental mean free paths, one has to introduce slab fluctuations with a magnetic field strength of the order of the strength of two-dimensional fluctuations, which is a factor of 2 or 3 larger than the values measured in interplanetary space. In this paper, we show that for nonlinear broadening of two-dimensional perturbations, the random scattering frequency increases significantly, and the transport mean free path decreases.

These authors introduced the cyclotron resonance broadening using a decorrelation in the correlation tensor of the random magnetic field. This decorrelation is attributable to the thermal motion of solar-

[*]E-mail: `quark@tu.neva.ru`
[**]E-mail: `cosmos@IT10242.spb.ru`

wind plasma ions and to the random motion of magnetic field fluctuations. The decorrelation time scale is $\sim L_c v_a^{-1}$, where L_c is the correlation length of the random field, and v_a is the Alfvén velocity. Mel'nikov (1996, 2000) showed that the decorrelation time scale of the correlation tensor for high-energy particles at $v \gg v_a$, where v is the particle velocity, is much larger than the nonlinear random resonance damping time, which is attributable to strong random scattering, is described by a nonlinear collision integral and is $\sim \omega^{-1}$, where ω is the gyrofrequency in the random magnetic field. In addition, at particle energies ~ 200 MeV, the gyroradius in the random field R_l is on the order of the irregularity size L_c; therefore, nonlinear scattering processes must be taken into account (Mel'nikov 1996). Thus, a nonlinear collision integral that includes strong random scattering must be used to properly incorporate the resonance damping in the case of large-scale resonance particle scattering in the cosmic medium for $R_1 \delta L$ and $R_1 \ll L$.

In the interstellar medium, the anisotropy of the random magnetic field and the processes of strong random scattering should also be taken into account. Minter and Spangler (1996) and Spangler (2001) showed that the knee in the spectrum of high-energy cosmic rays at energies $1-10$ PeV could be described by the change in the anisotropy of the random magnetic field, i.e., by the transition from three-dimensional to two-dimensional turbulence. Chandran (2001) showed that $k_{||} \ll k_{\perp}$ in the interstellar small-scale random magnetic field, where $k_{||}$ and $k_{\perp}$ are the wave vector components parallel and perpendicular to the regular magnetic field. Cosmic rays are also scattered weakly by such fluctuations. Therefore, the anisotropy of the random magnetic field in the quasilinear approximation strongly affects the cosmic-ray scattering frequency.

In the interstellar medium, $R_1 \ll L_c$ over the entire energy range of the high-energy cosmic rays that diffuse in the Galaxy, i.e., from 1 to 10^5 TeV (Berezisnkii $et\ al.$ 1990; Chuvilgin and Ptuskin 1993; Ptuskin 2001). Therefore, a nonlinear collision integral must be used to calculate the kinetic coefficients and transport mean free paths of these particles.

In this paper, we determine the kinetic coefficients and transport mean free paths for two-dimensional turbulence. In contrast to Bieber $et\ al.$ (1994), Teufel and Schlickeiser (2002, 2003), Teufel $et\ al.$ (2003), Dröge (2003), and Shalchi and Schlickeiser (2004), we take into account the resonance damping by using a nonlinear collision integral, i.e., by including strong random scattering on the correlation length of the random magnetic field. The form of the tensor part of the correlation tensor for the random magnetic field corresponds to two-dimensional fluctuations. Since a strong regular magnetic field produces

anisotropy of the MHD turbulence in plasma, we introduce different parallel and perpendicular correlation lengths relative to the direction of the regular field. In the nonlinear collision integral, the additional Green function that gives the resonance broadening is taken in the approximation of large-scale random scattering. This Green function includes the small (in magnitude) random magnetic field fluctuations parallel to the regular field. This allows the resonance broadening produced by non-Alfvén modes, including the mirroring processes, to be taken into account. These processes can make a large contribution to the resonance broadening and the transport mean free paths (Achterberg 1981; Dröge 2003; Shalchi and Schlickeiser 2004). This approach makes a fairly accurate allowance for the interaction conditions between cosmic rays with energies above 1 MeV and a weakly disturbed interplanetary magnetic field in the uniform case. As a result, the scattering frequencies calculated here increase, while the transport mean free paths decrease, which is in better agreement with the observational data without assuming a significant fraction of the slab turbulence.

Berezinskii $et\ al.$ (1990) and Chuvilgin and Ptuskin (1993) showed that, when the diffusion of Galactic cosmic rays in the interstellar medium is studied, the random magnetic field may be treated as a set of Alfvén waves with isotropically distributed wave vectors. In this case, the correlation tensor is anisotropic, and its anisotropy corresponds to two-dimensional turbulence. The random magnetic field is perpendicular to the regular magnetic field. Therefore, our results can be used to determine the parallel transport mean free paths in the interstellar medium.pt

TRANSFORMING THE COLLISION INTEGRAL

We will consider the kinetic coefficients and particle transport mean free paths over a wide energy range from 1 MeV to several GeV in the inner heliosphere and at energies above 10 GeV in the outer heliosphere, including those at the energies at which $R_1 \simeq L_{||}, L_{\perp}$, where R_1 is the gyroradius in the random magnetic field, and $L_{\perp}$ and $L_{||}$ are the perpendicular and parallel (relative to the regular magnetic field) correlation lengths, respectively. We use the following kinetic equation for the average particle distribution function $F(\mathbf{r}, \mathbf{p}, t)$ with a nonlinear collision integral (Mel'nikov 1996, 2000):

$$\left\{ \frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \mathbf{H}_0\mathbf{D} \right\} F(\mathbf{r}, \mathbf{p}, t) = \mathrm{St}F, \quad (1)$$

$$\mathrm{St}F = D_\alpha \int dx_1 B_{\alpha\beta}(\mathbf{r}, t, \mathbf{r}_1, t_1) \qquad (2)$$
$$\times\, G_1(x, x_1) D_{1\beta} F(x_1, x_0),$$

where $x \equiv \mathbf{r}, \mathbf{p}, t$, $\mathbf{r}$ is the coordinate, $\mathbf{p}$ is the momentum, $\mathbf{v}$ is the particle velocity, t is the time, $\mathbf{H}_0$ is the strength of the regular magnetic field, $\mathbf{D} = \frac{e}{c}\left[(\mathbf{v} - \mathbf{u}) \times \frac{\partial}{\partial\mathbf{p}}\right]$, e is the particle charge, c is the speed of light, $\mathbf{u}$ is the velocity of the plasma with a frozen-in magnetic field, and $G_1(x, x_1)$ is the particle Green function that is the solution of the linear kinetic equation.

Bieber *et al.* (1996), Carbone *et al.* (1995), and Matthaeus *et al.* (1990) showed that two-dimensional fluctuations are the main type of magnetic field fluctuations in the inner heliosphere and near the Earth's orbit. Therefore, we assume the contribution from other types of fluctuations to be negligible and choose the correlation tensor of the random anisotropic magnetic field $\mathbf{H}_1\mathbf{H}_0$ for a power-law spectrum in the form (Carbone *et al.* 1995; Matthaeus *et al.* 1990; Toptygin 1985; Chuvilgin and Ptuskin 1993)

$$B_{\alpha\beta}(\mathbf{k}) = P(\mathbf{k})\frac{[\mathbf{k} \times \mathbf{h}_0]_\alpha[\mathbf{k} \times \mathbf{h}_0]_\beta}{k_\perp^2}, \qquad (3)$$

$$P(\mathbf{k}) = A_\nu(q_\perp q_\parallel)^{-1-\frac{\nu}{2}}$$
$$\times \left(\frac{k_\parallel^2}{q_\parallel^2} + \frac{k_\perp^2}{q_\perp^2}\right)\left[1 + \frac{k_\parallel^2}{q_\parallel^2} + \frac{k_\perp^2}{q_\perp^2}\right]^{-2-\frac{\nu}{2}},$$

where $\mathbf{k}$ is the wave vector, ν is the spectral index of the random magnetic field, $\nu \sim 1.5{-}2$ for the inertial interval of wave numbers of the random magnetic field in the interplanetary medium, $\mathbf{h}_0 = \mathbf{H}_0 \cdot H_0^{-1}$, $\mathbf{k}_\parallel = (\mathbf{k}\mathbf{h}_0)\mathbf{h}_0$, $\mathbf{k}_\perp = \mathbf{k} - \mathbf{k}_\parallel$, $q_\parallel = L_\parallel^{-1}$, $q_\perp = L_\perp^{-1}$,

$$A_\nu = 2\Gamma\left(2 + \frac{\nu}{2}\right)q_\perp^{\frac{\nu}{2}-1}q_\parallel^{\frac{\nu}{2}}\langle H_1^2\rangle\left[3\pi^{\frac{3}{2}}\Gamma\left(\frac{\nu-1}{2}\right)\right]^{-1},$$

and $\Gamma(n)$ is the Gamma function. Passing to the drift approximation and averaging Eq. (1) over the particle gyration angle in the regular magnetic field and set $\mathbf{u} = 0$, $q_\perp^2 v_\perp^2 \Omega^{-2} \ll 1$, where $v_\perp^2 = [\mathbf{v} \times \mathbf{h}_0]^2$, Ω is the cyclotron frequency in the regular magnetic field. As a result, we obtain a kinetic equation for the average distribution function in the form (Toptygin 1983; Mel'nikov 2000)

$$\left\{\frac{\partial}{\partial t} + v\mu\frac{\partial}{\partial z} - \frac{1}{2}v\mathrm{div}\mathbf{h}_0\sin\vartheta\frac{\partial}{\partial\vartheta}\right\}\Phi = \langle\mathrm{St}F\rangle_\varphi,$$
$$(4)$$

where the coordinate z is along the vector $\mathbf{h}_0$, $\Phi = \langle F\rangle_\varphi$ is the distribution function averaged over the particle gyration in the regular field, ϑ is the angle between the vectors $\mathbf{p}$ and $\mathbf{h}_0$, $\mu = \cos\vartheta$, and φ is the azimuthal angle of the vector $\mathbf{p}$ in a plane perpendicular to $\mathbf{h}_0$. In this equation, the nonlinear average collision integral is (Toptygin 1983; Mel'nikov 2000)

$$\langle\mathrm{St}F\rangle_\varphi = \frac{\partial}{\partial\mu}(1 - \mu^2)b(\mu)\frac{\partial}{\partial\mu}\Phi(\mathbf{r}, \mathbf{p}, \mu, t), \qquad (5)$$

where the kinetic coefficient is

$$b(\mu) = \frac{e^2}{2m^2c^2}\int_0^\infty d\tau \int d\mathbf{k} P(\mathbf{k})\cos(\varphi_\mathbf{k} - \varphi) \qquad (6)$$
$$\times\, \cos(\varphi_\mathbf{k} - \varphi - \Omega\tau)\Gamma_0(\omega, \tau, \varphi_\mathbf{k} - \varphi)$$
$$\times\, \exp\{i\mathbf{k}\Delta\mathbf{r}(\tau) - i\omega_\mathrm{a}(k)\tau\},$$

m is the particle mass, $\varphi_\mathbf{k}$ is the azimuthal angle of the vector $\mathbf{k}$, Ω is the gyrofrequency in the regular magnetic field, ω is the gyrofrequency in the random magnetic field, $\omega = e\sqrt{\langle H_1^2\rangle}m^{-1}c^{-1}$, $\Delta\mathbf{r}(\tau)$ is the change in the radius vector of the particle in the regular magnetic field, $\omega_\mathrm{a} = v_\mathrm{a}k_\parallel$, v_a is the Alfvén velocity, and $\Gamma_0(\omega, \tau, \varphi_\mathbf{k} - \varphi)$ is a factor that is related to the additional Green function of the particle in the nonlinear collision integral and that yields the damping of the resonant wave–particle interaction (Toptygin 1985; Mel'nikov 2000),

$$\Gamma_0(\omega, \tau, \varphi_\mathbf{k} - \varphi) = \exp\left\{-\frac{\omega_\perp^2}{16}v_\parallel^2 k_\perp^2\tau^4\right. \qquad (7)$$
$$-\frac{\omega_\perp^2}{4\Omega^2}v_\perp^2 k_\parallel^2\tau^2 - \frac{\omega_\parallel^2}{4\Omega^2}v_\perp^2 k_\perp^2\tau^2$$
$$\left.-\frac{\omega_\parallel^2}{4\Omega^2}v_\perp^2 k_\perp^2\tau^2\cos[2(\varphi_\mathbf{k} - \varphi)]\right\},$$
$$\omega^2 = \omega_\perp^2 + \omega_\parallel^2, \quad \omega_\parallel^2 = \omega^2\langle(\mathbf{H}_1\mathbf{h}_0)^2/\mathbf{H}_1^2\rangle.$$

The factor $\Gamma_0(\omega, \tau, \varphi_\mathbf{k} - \varphi)$ arises when the solution of the linear kinetic equation for the particle Green function is substituted into the nonlinear diffusion coefficient. The four terms in the exponent of (7) correspond to different types of random magnetic field fluctuations. The first term and the second, third, and fourth terms make a large contribution to the damping at high parallel and perpendicular particle velocities, respectively. These terms make a large contribution at large pitch angles and limit the transport mean free path. The first two terms are associated with Alfvén waves; the last two must be taken into account in the presence of magnetosonic waves. The first, third, and fourth terms are related to the change $\Delta\mathbf{r}_1$ in the particle radius vector in the random field perpendicular to $\mathbf{h}_0$, while the second term is related to the change $\Delta\mathbf{r}_1$ parallel to $\mathbf{h}_0$.

KINETIC COEFFICIENTS AND TRANSPORT MEAN FREE PATHS FOR VARIOUS TYPES OF FLUCTUATIONS. COMPARISON WITH EXPERIMENTAL DATA

Let us first consider the limiting case of the absence of resonance broadening, $\omega = 0$ and $\Gamma_0(\omega, \tau) = 1$. In the kinetic coefficient (6), we expand the corresponding functions in terms of Bessel functions. We transform the series of Bessel functions and add the series using the addition formula for the Bessel functions

$$I_n(\rho) \exp\left[in\frac{\pi - \beta}{2}\right] = \sum_{k=-\infty}^{+\infty} I_{n+k}(z)I_k(z)\exp[ik\beta],$$

where $\rho = 2z\sin(\beta/2)$, $I_n(\rho)$ is the Bessel function of order n. The integrations in (6) yield a kinetic coefficient in the form

$$b_0(\mu) = c_0(\nu)\omega_\perp \left(\frac{\omega_\perp}{\Omega}\right)(q_\|\mu R)^{\nu-1}, \qquad (8)$$

where

$$c_0(\nu) = \frac{\sqrt{\pi}\Gamma\left(\frac{\nu}{2} + 2\right)}{3\nu\Gamma\left(\frac{\nu - 1}{2}\right)}, \quad R = v\Omega^{-1}.$$

Let us now turn to the diffusion approximation following Toptygin (1985). Expanding the distribution function into the series

$$\Phi(p, \mu) = \frac{Np}{4\pi} + \delta\Phi(p, \mu), \quad |\delta\Phi| \ll \frac{N}{4\pi},$$

we substitute this expansion into Eq. (4) and find the equation for $N(p)$ and $\delta\Phi$. Solving the equation for $\delta\Phi$ on long time scales and large distances from the source yields the expression for the transport mean free path

$$\Lambda_0 = \frac{3v}{4} \int_{-1}^{+1} \frac{1 - \mu^2}{b_0\mu} d\mu.$$

Integration over μ yields the parallel transport mean free path

$$\Lambda_0 = \frac{3R(q_\|R)^{1-\nu}}{4(2 - \nu)\left(2 - \frac{\nu}{2}\right)c_0(\nu)}\left(\frac{\Omega^2}{\omega_\perp^2}\right). \qquad (9)$$

The transport mean free path of protons with an energy of 100 MeV in the inner heliosphere calculated from this formula is a factor of 5 to 8 larger than the observed mean free path (Palmer 1982; Bieber et al. 1994). In contrast to Jokipii (1966), Bieber et al. (1994), Dröge (2003), Teufel and Schlickeiser (2002, 2003), and Shalchi and Schlickeiser (2004), the parallel transport mean free path Λ_0

in this paper is proportional to the correlation length component $L_\|^{\nu-1}$ rather than the absolute value of $L_c^{\nu-1}$. Since $\omega_\perp \gg \omega_\|$, only $\omega_\perp$ appears in the expression for $b(\mu)$. The divergence of the parallel transport mean free path for $\nu \to 2$ stems from the approximation of weak random scattering, i.e., the quasi-linear approximation, being inapplicable at low μ.

In the case of strong random scattering at $v_\|^2 q_\perp^2 \gg v_\perp^2 q_\|^2$, the following factor makes a major contribution to the resonance damping function $\Gamma_0(\omega, \tau)$ in the collision integral:

$$\Gamma_1(\omega, \tau) = \exp\left\{-\frac{\omega_\perp^2}{16}v_\|^2 k_\perp^2 \tau^4\right\}.$$

We substitute Γ_1 into (6) and expand the corresponding functions in terms of Bessel functions. We transform the series of Bessel functions also using the addition formulas for the Bessel functions. The integrations in (6) yield a kinetic coefficient in the form

$$b_1(\mu) = c_1(\nu)\omega_\perp \left(\frac{\omega_\perp}{\Omega}\right)^\nu (\mu q_\perp R)^{\nu-1}, \qquad (10)$$

where

$$c_1(\nu) = \left(\frac{5}{4}\right)^{\nu-\frac{1}{2}} \frac{\sqrt{\pi}\Gamma\left(\frac{1}{4}\right)\left(\frac{\nu}{2} + 1\right)\Gamma\left(\nu - \frac{1}{2}\right)}{3 \times 2^{\frac{\nu}{2}+1}\Gamma\left(\frac{\nu - 1}{2}\right)}.$$

For $v_\|^2 q_\perp^2 \ll v_\perp^2 q_\|^2$, the following integrand factor makes a major contribution to the damping function $\Gamma_0(\omega, \tau)$:

$$\Gamma_2(\omega, \tau) = \exp\left\{-\frac{\omega_\perp^2}{4\Omega^2}v_\perp^2 k_\|^2 \tau^2\right\}. \qquad (11)$$

Substituting it into (6), we obtain after transformations and integrations

$$b_2(\mu) = c_2(\nu)\omega_\perp \left(\frac{\omega_\perp}{\Omega}\right)(q_\|R)^{\nu-1} \qquad (12)$$

$$\times \left[\sqrt{\pi}\mu^{\nu-1} + \left(\frac{\omega_\perp}{\Omega}\right)^{\nu-1}\right],$$

where

$$c_2(\nu) = \frac{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{\nu}{2} + 2\right)}{3\nu\Gamma\left(\frac{\nu - 1}{2}\right)}.$$

At $\omega_\| \neq 0$, the following integrand factor makes an additional contribution to the damping function $\Gamma_0(\omega, \tau)$:

$$\Gamma_3(\omega, \tau) = \exp\left\{-\frac{\omega_\|^2}{4\Omega^2}v_\perp^2 k_\perp^2 \tau^2\right\}. \qquad (13)$$

Substituting it into (6) and performing integrations yields

$$b_3(\mu) = c_3(\nu)\omega_\perp \left(\frac{\omega_\perp}{\Omega}\right)\left(\frac{\omega_{||}}{\Omega}\right)^{\nu-1}(q_\perp R\sqrt{1-\mu^2})^{\nu-1} \tag{14}$$

$$\text{for } \left(\frac{\omega_{||}}{\Omega}q_\perp R\right)^2 \ll 1,$$

where

$$c_3(\nu) = \frac{\pi\left(\frac{\nu}{2}+1\right)\Gamma\left(\frac{\nu}{2}\right)}{3\nu 2^{\frac{3\nu}{2}}\Gamma\left(\frac{\nu-1}{2}\right)}.$$

It follows from formula (14) that the parallel components of the random magnetic field contribute to the scattering frequency at $\mu^2 \ll 1$.

Using numerical methods, we can show that the last term in the exponent of (7), which is proportional to $\cos(2\varphi_\mathbf{k} - 2\varphi)$, makes a minor contribution to the kinetic coefficient, and, hence, it can be disregarded.

It is convenient to combine (8), (10), (12), and (14) into a general interpolation formula for $b(\mu)$ that is valid at any pitch angle and that includes the random magnetic field components $\omega_\perp$ and $\omega_{||}$. For this purpose, we can add b_1, b_2, and b_3. This addition is possible since the kinetic coefficients b_1 and b_2 make a contribution at different μ, the coefficient b_3 is small, and all of the kinetic coefficients are proportional to $R^{\nu-1}$. As a result, we obtain

$$b(\mu) = \omega_\perp \left(\frac{\omega_\perp}{\Omega}\right)(q_{||}R)^{\nu-1} \tag{15}$$

$$\times\left[c_0\mu^{\nu-1} + c_1\left(\frac{\omega_\perp q_\perp}{\Omega q_{||}}\right)^{\nu-1}\mu^{\nu-1}\right.$$

$$\left. + c_2\left(\frac{\omega_\perp}{\Omega}\right)^{\nu-1} + c_3\left(\frac{\omega_{||}q_\perp}{\Omega q_{||}}\right)^{\nu-1}(\sqrt{1-\mu^2})^{\nu-1}\right].$$

In this expression for $b(\mu)$, the terms that include strong random scattering have a higher order of smallness in relative strength of the random magnetic field $\omega\Omega^{-1} < 1$. Including strong random scattering gives rise to a μ-independent term proportional to $c_2(\nu)$. The kinetic coefficient $b(\mu)$ has no sharp maximum at $\mu = 0$. This is because there are no components with the wave vectors parallel to the regular magnetic field in the tensor part of the correlation tensor for the random magnetic field.

Averaging $b(\mu)$ over the pitch angle yields the parallel transport mean free path

$$\Lambda_{||} = \frac{3}{4}\frac{R\left(q_{||}R\right)^{\nu-1}}{c_2(\nu)}\left(\frac{\Omega}{\omega_\perp}\right)^{\nu+1} \tag{16}$$

$$\times \int_0^1 \frac{(1-x^2)dx}{1 + f_1(\omega_\perp)x^{\nu-1} + f_2(\omega_\perp,\omega_{||})(\sqrt{1-x^2})^{\nu-1}},$$

where

$$f_1(\omega_\perp) = \frac{c_0(\nu)}{c_2(\nu)}\left(\frac{\Omega}{\omega_\perp}\right)^{\nu-1} + \frac{c_1(\nu)}{c_2(\nu)}\left(\frac{q_\perp}{q_{||}}\right)^{\nu-1},$$

$$f_2(\omega_\perp,\omega_{||}) = \frac{c_3(\nu)}{c_2(\nu)}\left(\frac{\omega_{||}}{\omega_\perp}\right)^{\nu-1}\left(\frac{q_\perp}{q_{||}}\right)^{\nu-1}.$$

The term proportional to $f_2(\omega_\perp,\omega_{||})$ in the integrand in (16) may be disregarded with an error of $\sim 5\%$. Using the interpolation formula for the integral in (16), we obtain the final formula for the parallel transport mean free path, including strong random scattering:

$$\Lambda_{||}(R) = \Lambda_0 \frac{\sqrt{\pi}(2-\nu)\left(2-\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\Omega}{\omega_\perp}\right)^{\nu-1} \tag{17}$$

$$\times \frac{1}{1.63 + \frac{\nu-1}{3}(f_1(\omega_\perp))^{0.7} + (2.13-\nu)f_1(\omega_\perp)}.$$

In this case, $\Lambda_{||} \propto p^{2-\nu}$. The contribution of strong random scattering is significant at any strengths of the random magnetic field. The momentum dependence of $\Lambda_{||}$ in this case is similar to that numerically calculated by Teufel and Schlickeiser (2002, 2003) and Shalchi and Schlickeiser (2004).

For protons with energy of 200 MeV scattered in a weakly disturbed interplanetary medium, substituting $R = 4 \times 10^8$ m, $L_{||} = L_\perp = 5 \times 10^9$ m, $\nu = 1.67$, and $\Omega^2/\omega^2 \approx 8$ yields $\Lambda_{||} \approx 0.25$ AU. This value of $\Lambda_{||}$ is close to the mean experimental values from Palmer (1982). The values of $\Lambda_{||}$ are close to those from Shalchi and Schlickeiser (2004), in which, however, slab turbulence produces the main scattering.

Thus, the weak momentum dependence of $\Lambda_{||}$ for solar cosmic rays over a wide energy range from several MeV to several GeV can be explained in terms of strong (moderate) random scattering by two-dimensional turbulence in the solar wind.

Here, we showed that two-dimensional turbulence is similar in scattering properties to slab turbulence. Therefore, let us estimate the transport mean free path of cosmic rays in corotating solar wind flows, including nonlinear scattering, by assuming that there is two-dimensional turbulence in them. For protons with an energy of 200 MeV that propagate in corotating flows with weak scattering, substituting $R = 4 \times 10^8$ m, $L_{||} = 5 \times 10^9$ m, $L_\perp = 5 \times 10^{11}$ m, $\nu = 1.7$, and $\Omega^2/\omega^2 \approx 30$ yields $\Lambda_{||} \approx 1.5$ AU. Based on

the conclusions reached by Kaminer *et al.* (1984), we take the low strength of the random magnetic field. The calculated value of $\Lambda_{||}$ is close to the experimental values for collisionless events from Palmer (1982) and Kolomeets and Sevost'yanov (1987).

For Galactic cosmic rays with energies above 4 GeV scattered in the interstellar medium, when the random magnetic field has a Kolmogorov spectrum with $\nu \approx 1.7$, we obtain the following order-of-magnitude estimate from formula (17):

$$\Lambda_{||} = \frac{1.6}{q_{||}}(Rq_{||})^{2-\nu}\left(\frac{\Omega}{\omega_\perp}\right)^{\nu+1}.$$

Assuming that $\omega \approx 1.8\Omega$, $L_{||} \approx 100$ pc, and $H_1 \approx 0.3$ nT, we obtain for relativistic protons

$$\Lambda_{||} \approx 0.3 \times 10^{18}\left(\frac{E}{1\,\text{GeV}}\right)^{2-\nu}\quad\text{cm}, \qquad (18)$$

where E is the particle kinetic energy in GeV. The calculated value of $\Lambda_{||}$ is slightly lower than the mean experimental values from Berezinskii *et al.* (1990), Chuvilgin and Ptuskin (1993), and Ptuskin (2001). This may be because the interstellar medium is fractal, i.e., because there are regions in the Galaxy with a greatly reduced density of the medium and a very low magnetic field strength (Lagutin and Tyumentsev 2003).

SCATTERING BY STRONG TURBULENCE PRODUCED BY ALFVÉN AND MAGNETOSONIC WAVES AND STATIONARY STRUCTURES

Let us consider the case of strong random particle scattering by the turbulence produced by Alfvén and magnetosonic waves and stationary structures in the corotating solar-wind interaction region. Note that strong two-dimensional turbulence is similar in scattering properties to slab turbulence. Therefore, let us estimate the transport mean free path of cosmic rays in the corotating solar-wind interaction region, including nonlinear scattering, by assuming that there is only two-dimensional turbulence in it. We assume that $\omega_\perp \gg \omega_{||}$ and $\omega_\perp \sim \Omega$. Particles are strongly scattered on the correlation length, and the resonance damping time scale in the isotropic case is

$$\tau_0 = \sqrt{\frac{1}{q_{||}q_\perp v_{||}v_\perp}}.$$

The estimates of the terms in the exponent of (7) are

$$\frac{\omega_\perp^2}{16}v_{||}^2 k_\perp^2 \tau^4 \sim \frac{\omega_\perp^2\tau^2}{16}\frac{L_{||}^2}{L_\perp^2}, \quad \frac{\omega_\perp^2}{4\Omega^2}v_\perp^2 k_{||}^2 \tau^2 \sim \frac{\omega_\perp^2\tau^2}{4}\frac{R_\perp^2}{L_\perp^2},$$

$$\frac{\omega_{||}^2}{4\Omega^2}v_\perp^2 k_\perp^2 \tau^2 \sim \frac{\omega_{||}^2\tau^2}{4}\frac{R_\perp^2}{L_\perp^2}.$$

Since $R_\perp^2 \ll L_\perp^2$, $L_{||}^2$, we may disregard all but the first term in the exponent of (7). Substituting $\Gamma_1(\omega,\tau)$ into (7), substituting τ_0^2 for τ^2 in $\Gamma_1(\omega,\tau)$, and performing transformations in the collision integral, we obtain the kinetic coefficient and the transport mean free path

$$b(\mu) = \frac{\pi\Gamma\left(\frac{\nu}{2}+2\right)\omega}{3\nu 2^{\frac{3\nu-1}{2}}\Gamma\left(\frac{\nu-1}{2}\right)}\left(\frac{\omega}{\sqrt{2}\Omega}\right)^\nu\left(\frac{v_{||}q_\perp}{v_\perp q_{||}}\right)^{\frac{\nu-1}{2}},$$

$$(19)$$

$$\Lambda_{||} = \frac{3\nu 2^{\frac{3\nu}{2}-2}\Gamma\left(\frac{\nu-1}{2}\right)}{\pi\Gamma\left(\frac{\nu}{2}+2\right)}\left(\sqrt{2}\frac{\Omega}{\omega}\right)^{\nu+1}\left(\frac{q_{||}}{q_\perp}\right)^{\frac{\nu}{2}-1}R.$$

$$(20)$$

Calculations using (20) for $\omega \approx 0.7\Omega$ yield $\Lambda_{||} \approx 1$ AU for Galactic cosmic rays with an energy of 10 GeV. The numerical value and energy dependence of $\Lambda_{||} \propto R$ match the experimental data for Galactic cosmic rays in the outer heliosphere obtained by Gerasimov *et al.* (1999). The absolute value and rigidity dependence of $\Lambda_{||}$ are also close to the experimental data on the anomalous cosmic-ray diffusion in the outer heliosphere obtained by Fujii and McDonald (1995).

DISCUSSION AND CONCLUSIONS

We have shown that, virtually irrespective of the random magnetic field strength, including strong random scattering by two-dimensional turbulence causes the scattering frequency to increase and the parallel transport mean free path to decrease. A characteristic feature of the parallel transport mean free paths calculated by including strong random scattering by two-dimensional turbulence is the dependence $\Lambda_{||} \propto p^{2-\nu}$ over a wide particle energy range. In this case, there is no knee in the energy dependence of $\Lambda_{||}$ at $R_1 \sim L_c^{-1}$ (Mel'nikov 1996, 2000). This could be the reason why the momentum dependence of the parallel transport mean free path for solar cosmic-ray particles in the inner heliosphere is weak over a wide energy range, from several MeV to 1 GeV (Palmer 1982; Bieber *et al.* 1994). It should be noted that, in the quasi-linear approximation used by Teufel and Schlickeiser (2002, 2003), Teufel *et al.* (2003), and Shalchi and Schlickeiser (2004), a knee appears in the energy dependence of the parallel

proton mean free path in the case of scattering by two-dimensional turbulence at very low energies.

Including nonlinear processes leads to finite values of the kinetic coefficient $b(\mu)$ and the parallel mean free path $\Lambda_{\parallel}$ for $\mu \to 0$. The resonance-broadening Green function $G_1(x, x_1)$ is taken in the approximation of large-scale random scattering; in addition, the correlation tensor of the random magnetic field satisfies the condition $h_{0\alpha} B_{\alpha\beta}(\mathbf{k}) = 0$. Therefore, there is no sharp maximum of the Cherenkov term in the kinetic coefficient at $\mu = 0$. As follows from formulas (15) and (16), the presence of a small random magnetic field component parallel to the regular field changes only slightly the kinetic coefficient $b(\mu)$ and the parallel mean free path $\Lambda_{\parallel}$.

The anisotropy in the random magnetic field attributable to the difference between the correlation length $L_{\perp}$ and $L_{\parallel}$ affects only slightly the parallel mean free paths when cosmic rays propagate in a weakly disturbed heliosphere. However, this anisotropy at $L_{\perp} \gg L_{\parallel}$ increases appreciably the mean free paths of high-energy cosmic rays in the interstellar medium.

Thus, the value and rigidity dependence of the transport mean free path $\Lambda_{\parallel}$ for solar cosmic-ray protons with energies from 5 MeV to 1 GeV in the inner heliosphere are consistent with strong (moderate) random scattering by two-dimensional turbulence. This is the principal mode in the inner heliosphere. The parallel mean free path agrees with the mean experimental values from Palmer (1982) (Palmer's consensus). To reconcile the theoretical and experimental transport mean free paths, there is no need to introduce large slab turbulence components, as was done by Shalchi and Schlickeiser (2004), which are in conflict with the experimental data on the interplanetary random magnetic field (Bieber *et al*. 1996; Matthaeus *et al*. 1990).

The value and rigidity dependence of the parallel transport mean free path $\Lambda_{\parallel}$ for Galactic comic-ray protons with energies from several GeV to 10^7 GeV are consistent with strong (moderate) random scattering by two-dimensional turbulence and the fractal structure of the medium.

The large transport mean free paths of solar cosmic-ray protons in collisionless events in near-Earth space can be explained in terms of scattering by two-dimensional turbulence with a spectral index $\nu \sim 1.7$, large $L_{\perp}$, and a low random magnetic field strength, as observed in the interplanetary medium (Kaminer *et al*. 1984; Matthaeus *et al*. 1990; Carbone *et al*. 1995; Bieber *et al*. 1996).

The value and rigidity dependence of the parallel transport mean free path for Galactic cosmic rays

with energies of several dozen GeV and anomalous cosmic rays with energies of several hundred MeV in the outer heliosphere (Gerasimova *et al*. 1999; Fujii and McDonald 1995) can be explained in terms of strong random scattering in the corotating interaction region by turbulence composed of Alfvén and non-Alfvén modes.

REFERENCES

1. A. Achterberg, Astron. Astrophys. **98**, 161 (1981).
2. V. S. Berezinskii, S. V. Bulanov, V. L. Ginzburg, *et al*., *Astrophysics of Cosmic Rays* (Nauka, Moscow, 1990).
3. J. W. Bieber, W. H. Matthaeus, C. W. Smith, *et al*., Astrophys. J. **420**, 294 (1994).
4. J. W. Bieber, W. Wanner, and W. H. Matthaeus, J. Geophys. Res. **101**, 2511 (1996).
5. A. M. Bykov and I. N. Toptygin, Zh. Éksp. Teor. Fiz. **97**, 194 (1990a).
6. A. M. Bykov and I. N. Toptygin, Zh. Éksp. Teor. Fiz. **98**, 1255 (1990b).
7. A. M. Bykov and I. N. Toptygin, Usp. Fiz. Nauk **163**, 19 (1993).
8. V. Carbone, F. Malara, and P. Veltri, J. Geophys. Res. **100,** 1763 (1995).
9. B. D. G. Chandran, Space Sci. Rev. **99**, 271 (2001).
10. L. G. Chuvilgin and V. S. Ptuskin, Astron. Astrophys. **279**, 278 (1993).
11. W. Dröge, Astrophys. J. **589**, 1027 (2003).
12. Z. Fujii and F. B. McDonald, J. Geophys. Res. **100**, 17 043 (1995).
13. S. K. Gerasimov, V. G. Grigor'ev, P. A. Krivoshapkin, *et al*., Izv. Ross. Akad. Nauk, Ser. Fiz. **63**, 1611 (1999).
14. J. R. Jokipii, Astrophys. J. **146**, 480 (1966).
15. N. S. Kaminer, A. E. Kuz'micheva, and N. V. Mymrina, Izv. Ross. Akad. Nauk, Ser. Fiz. **48**, 2149 (1984).
16. E. V. Kolomeets and V. N. Sevost'yanov, Geomagn. Aéron. **27**, 12 (1987).
17. A. A. Lagutin and A. G. Tyumentsev, Izv. Ross. Akad. Nauk, Ser. Fiz. **67**, 443 (2003).
18. W. H. Matthaeus, M. L. Goldstein, and D. A. Roberts, J. Geophys. Res. **95**, 20 673 (1990).
19. Yu. P. Mel'nikov, Zh. Éksp. Teor. Fiz. **109**, 1599 (1996) [JETP **82**, 860 (1996)].
20. Yu. P. Mel'nikov, Zh. Éksp. Teor. Fiz. **117**, 526 (2000) [JETP **90**, 460 (2000)].
21. A. H. Minter and S. R. Spangler, Astrophys. J. **458**, 194 (1996).

22. I. D. Palmer, Rev. Geophys. Space Phys. **20**, 335 (1982).
23. V. S. Ptuskin, Space Sci. Rev. **99**, 281 (2001).
24. A. Shalchi and R. Schlickeiser, Astrophys. J. **604**, 861 (2004).
25. S. R. Spangler, Space Sci. Rev. **99**, 261 (2001).
26. A. Teufel, L. Lerche, and R. Schlickeiser, Astron. Astrophys. **397**, 777 (2003).
27. A. Teufel and R. Schlickeiser, Astron. Astrophys. **393**, 703 (2002).
28. A. Teufel and R. Schlickeiser, Astron. Astrophys. **397**, 15 (2003).
29. I. N. Toptygin, *Cosmic Rays in the Interplanetary Magnetic Fields* (Nauka, Moscow, 1983; Reidel Publ. Company, Dordrecht, 1985).

Translated by V. Astakhov

The SCORPIO Universal Focal Reducer of the 6-m Telescope

V. L. Afanasiev and A. V. Moiseev[*]

Special Astrophysical Observatory, Russian Academy of Sciences, Nizhnii Arkhyz, 357147
Karachai-Cherkessian Republic, Russia

Received September 10, 2004

Abstract—We describe the SCORPIO focal reducer that has been used since the fall of 2000 for observations on the 6-m Special Astrophysical Observatory telescope. We give parameters of the instrument in various observing modes (direct imaging, long-slit and multislit spectroscopy, spectropolarimetry, Fabry–Perot panoramic spectroscopy). Observations of various astronomical objects are used as examples to demonstrate the SCORPIO capabilities. © *2005 Pleiades Publishing, Inc.*

Key words: *astronomical observing techniques, devices and instruments.*

INTRODUCTION

The idea of using a focal reducer on a telescope was suggested and implemented by Courtés (1960, 1964) back in the 1950 and 1960s. The focal reducer allows several problems to be solved simultaneously. First, the equivalent focal ratio of the system becomes faster and the field of view increases, which is important for studying faint extended objects. Second, the off-axis aberrations of the primary mirror can be corrected by using specially calculated optics. Third, it becomes possible to install dispersive elements (grisms, Fabry–Perot interferometers (FPI), etc.) in the parallel beam between the collimator and the camera, which turns the focal reducer into a universal spectrograph.

Instruments based on this scheme, which are primarily designed for the spectroscopy and photometry of faint extended objects, have gained wide acceptance in the last two decades. It will suffice to mention the EFOSC camera of the 3.6-m ESO telescope (Buzzoni *et al.* 1984); in many respects, it became the prototype of modern spectrographs for 8–10-m telescopes, such as FORS on VLT (Nicklas *et al.* 1997).

The first focal reducer for interferometric observations on the 6-m BTA telescope was created at the Special Astrophysical Observatory (SAO) of the Russian Academy of Sciences in the mid-1980s using commercially available photographic lenses. Despite such shortcomings as poor image quality at the edge of the field of view, low optical transmittance (about 30% at maximum), and the absence of any automation, the reducer had been used on the 6-m BTA telescope for more than ten years until the question of its upgrading arose. In 1999, work to create a new focal reducer for the prime focus of the 6-m BTA telescope began at the SAO.

The new SCORPIO (Spectral Camera with Optical Reducer for Photometric and Interferometrical Observations) focal reducer allows the following types of observations of extended and starlike objects to be performed at the prime focus of the 6-m telescope:

(1) direct imaging in broad-, medium-, and narrow-band filters;

(2) FPI panoramic spectroscopy;

(3) long-slit and slitless spectroscopy;

(4) multiobject spectroscopy with 16 remotely relocatable slits;

(5) spectropolarimetry.

The mechanical and optical parts of SCORPIO were produced at the SAO breadboard workshops. The first BTA observations were performed in September 2000 with the old version of the prime focus adapter. A new adapter platform came into use in May 2001. The first successful multislit spectroscopic observations were carried out in September 2003; in the summer of 2004, the spectropolarimetric mode was implemented, and the first observations were performed.

In the next section, we consider the optomechanical layout of SCORPIO and its basic characteristics. Subsequently, we consider the peculiarities of observations in various modes; these are illustrated using specific results obtained with the 6-m telescope when the instrument was tested. We consider prospects for further upgrading SCORPIO.

[*]E-mail: `moisav@sao.ru`

Table 1. Basic parameters of SCORPIO

Field of view:	
total	$6\rlap{.}'1 \times 6\rlap{.}'1$
in mutislit mode	$2\rlap{.}'9 \times 5\rlap{.}'9$
CCD image scale	$0\rlap{.}''18$/pixel
Spectral range	$3600-10\,000$ Å
Spectral resolution:	
with grisms (for 1″ slit width)	$1.5-20$ Å
with FPI	$0.8-2.5$ Å
Maximum quantum efficiency (telescope+SCORPIO+CCD):	
direct imaging	70%
spectroscopy	40−50%
FPI observations	20%

THE DESCRIPTION OF SCORPIO

In construction, SCORPIO consists of three parts, each of which can be used independently: a focal reducer, a prime focus adapter platform, and a CCD detector. Basic parameters of the instrument are given in Table 1.

The Optomechanical Layout the Focal Reducer

The optical layout of the focal reducer (Fig. 1) includes a field lens and a collimator (a four-lens apochromat ($F/2.2$) that forms the exit pupil of the system), a camera objective (a six-lens apochromat ($F/1.8$)), and replaceable optical elements (FPI, diffraction gratings, filters, a polarization analyzer, and phase plates). The equivalent focal ratio of the system at the prime focus of the 6-m telescope is $F/2.6$. The optical surfaces are coated with seven antireflecting layers[1] that work in the wavelength range $3500-10\,000$ Å. The laboratory measurements of the SCORPIO spectral transmission curve are shown in Fig. 2.

Since the collimator optics correct the coma and field curvature of the primary mirror of the telescope, we can abandon the use of the standard lens corrector without antireflecting coating. The diameter of the collimated beam is 40 mm. The working focal length of the camera is 14 mm. The equivalent focal length of the BTA reducer is 15.6 m, which corresponds to an image scale of 75 μm/arcsec. The linear size of the

nonvignetted field of view is 28×28 mm in the plane of the detector.

In construction, the reducer was made in the form of separate remotely controlled units mounted in a single case:

(1) a multislit unit placed in front of the focal plane;

(2) two rotating wheels, each with six positions;

(3) a polarization analyzer placed in front of the collimator;

(4) the collimator focusing mechanism;

(5) a mechanism for putting/withdrawing a dispersive element in/from the collimated beam.

The multislit unit, which is designed for multiobject spectroscopy, is an arrangement that consists of 16 metal strips with slits located in the focal plane and moved in a $2\rlap{.}'9 \times 5\rlap{.}'9$ field. The slit height is about 18″. The position of each slit is fixed using two (holding and catching) electromagnets. The holding magnets are fixed; the catching magnets are fastened to a chassis that is moved along one coordinate by a stepping motor. A separate electromagnet simultaneously fixes the positions of all the slits in the focal plane. The arrangement is put in the beam using a stepping motor.

The wheels installed in the spectrograph are designed to put various replaceable elements in the beam—filters, slits, masks, etc. All of the elements installed in the wheel are mounted in bayonet-type frames with clear apertures of 72 mm, which allows them to be replaced as needed. Medium- and narrow-band interference filters, as well as a slit for spectroscopic observations, are generally installed in the wheel located in the focal plane of the telescope. Broad-band glass filters, a phase plate, and a mask

---

[1]The antireflecting layers were put on the surfaces at the Nizhni Novgorod Institute of Applied Physics.

Focal reducer
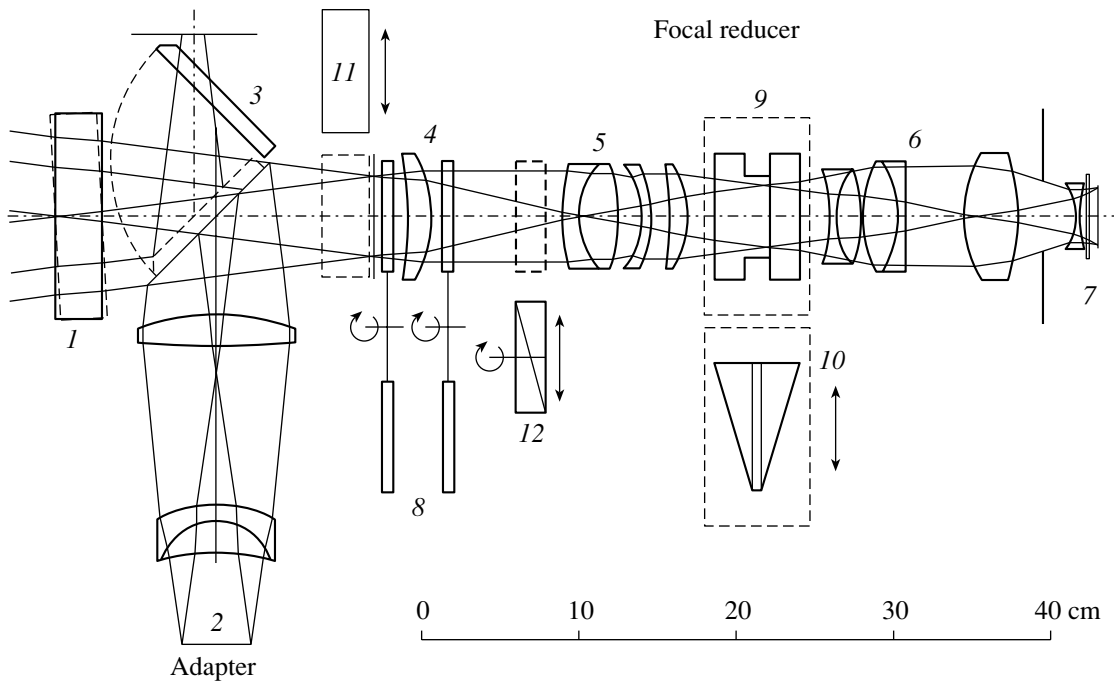
Fig. 1. Optical layout of SCORPIO: (*1*) tip-tilt plate, (*2*) calibration optics, (*3*) flat mirror, (*4*) field lens, (*5*) collimator, (*6*) camera, (*7*) CCD, (*8*) filter wheels, (*9*) FPI, (*10*) direct vision grism, (*11*) multislit unit, (*12*) polarization analyzer.
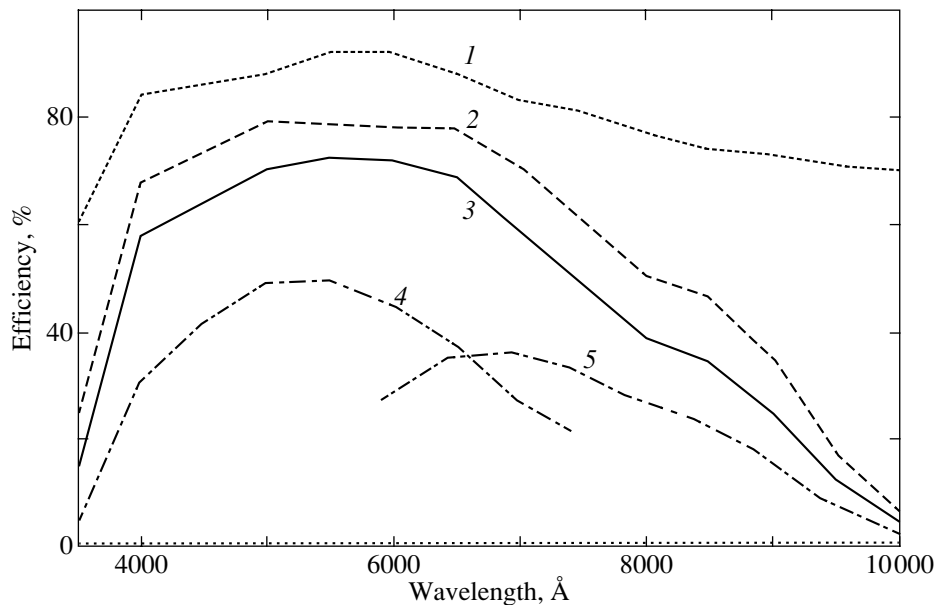
Fig. 2. Transmission curve for the SCORPIO optics (*1*); the quantum efficiency curve for the EEV 42−40 CCD, as provided by the producer (*2*); the combined quantum efficiency curve for SCORPIO+CCD (*3*); and the quantum efficiency curve in spectroscopic mode for observations with VPHG550G (*4*) and VPHG550R (*5*) low-resolution grisms.

for slitless spectroscopy are installed in the second wheel (behind the field lens and in front of the collimator). One position in each wheel always remains free so that all of the installed filters can be used in observations.

A 14-mm-thick Savart plate that separates the beams into two mutually perpendicular planes of po-

larization by 9″ in the focal plane is used as the polarization analyzer. The analyzer can be turned around the optical axis through 45°. A stepping motor is used to put the analyzer in and withdraw from the beam and to turn it.

The collimator focusing mechanism ensures a lin-

ear displacement of the collimator within 12 mm with an accuracy of 0.01 mm.

A slide with two switchable positions to put dispersive elements (FPI or direct-vision grisms) in the beam is located between the collimator and the camera. A built-in neon lamp is used to visually adjust the FPI. A central electromagnetic shutter with a shutter cycle of 0.1 s is located at the flange of the instrument closest to the primary mirror in front of the mutislit unit.

The Adapter Platform

The adapter platform is fixed on a turning table in the BTA prime-focus cage and is used for guiding based on off-axis stars and for illuminating the spectrograph by calibration lamps. Both the focal reducer and other equipment can be mounted on it. The adapter is equipped with an electromagnetic shutter that works independently of the central shutter of the focal reducer.

The adapter contains two rectangular fields to search for guiding stars[2]; the centers of the fields are offset by 12′ from the center of the field of view. A fiber bundle displaced by stepping motors in a rectangular coordinate system is located in each of the fields. The off-axis lens correction placed in front of each guiding field corrects the coma of the telescope's primary mirror. The fiber displacement range is $8.'5 \times 4.'5$, and the diameter of the field of view of each fiber is about $40''$.

The flat diagonal mirror (denoted by number 3 in Fig. 1) has two fixed positions. At one of these, the mirror throws the images from the fiber bundles to the TV view. This mode is used when exposing objects. At the other position, the mirror blocks the central beam of light from the telescope and throws the image of the field center of the instrument to the TV view, which is needed to roughly point the telescope at the required object. In addition, the light from the calibration lamps is thrown to the spectrograph at this position of the mirror.

The adapter contains the calibration illuminator optics that form a convergent beam with a focal ratio of $F/4$ at the entrance of the focal reducer, which is telecentrically equivalent to the beam formed by the primary mirror of the telescope. This scheme of the calibration unit forms the system's pupil at the same position where the image of the telescope's mirror is located. This allows us not only to properly calibrate the wavelength scale using a line-spectrum lamp, but also to calibrate the system's transmittance in various

---

[2]Since the 6-m telescope has an altazimuth mounting, both the position of the telescope in A and z and the rotation of the field of view should be controlled during the guiding. Therefore, two guiding stars are used.

operating modes (flat field). The entrance area of the calibration unit is illuminated through an integrating sphere (an Ulbricht sphere) by two calibration lamps: a He−Ne−Ar-filled lamp to calibrate the wavelength scale and a continuum halogen lamp to produce a flat field.

The Control System

SCORPIO contains a relatively large number of various electromechanical mechanisms (13 motors, two shutters, two calibration lamps, three cross-illumination LEDs, 33 electromagnets in the multislit unit) that are controlled as follows: Each of the above units (the focal reducer, the platform, and the multislit unit) includes electronic boards with a microprocessor that controls the mechanisms of the corresponding unit. Instructions to each microprocessor are issued from a remote personal computer using the standard RS232 data communication protocol. This arrangement of the instrument's remote control allows the required stability to be achieved when working with the existing long communication lines of the 6-m telescope. No continuous control of the motors and their state polling are required: the microprocessor takes over these functions. At the same time, possible changes in the observing technique and the software on the controlling computer require no reprogramming of the microprocessors.

The CCD detector

From 2000 until 2003, the detector was a TK1024 1024×1024-pixel CCD array. Since April 2003, an EEV 42−40 2048×2048-pixel CCD array has been mainly used on SCORPIO. Basic parameters of the two detectors are given in Table 2. It should be noted that a programmed choice of modes with different transformations gain and readout speeds and noise is possible for EEV 42−40. Figure 2 shows a plot of the quantum efficiency for EEV 42−40. The detector is cooled with liquid nitrogen. The complex for CCD observations (the cryostat, the electronics, and the control software) was designed and produced in the SAO Laboratory of Advanced Design (http://www.sao.ru/hq/adlab/).

As we see from Table 2, the detectors used have a high quantum efficiency (see Fig. 2), low noise, and low dark current. In addition, both detectors have an almost perfect surface: the number of bad columns and hot pixels is small. Perhaps the only shortcoming is the interference of the transmitted light (fringes) that is observed at wavelengths longer than 7500 Å for TK1024 and longer than 6600 Å for EEV 42−40. An appropriate observing technique (see section Data Reduction) is required to properly subtract the interference pattern.

PECULIARITIES OF VARIOUS OBSERVING MODES

Direct Imaging

SCORPIO is equipped with several filter sets that can be used for photometric observations. Broadband glass filters allow the Johnson−Cousins photometric $UBVR_cI_c$ system to be implemented in direct imaging mode (see Bessell 1990). Table 3 gives the limiting magnitudes for the detection of faint objects at a signal-to-noise ratio of 3 at $1''.3$ seeing. This table is based on the work of Fatkhullin (2002), who studied the capabilities of SCORPIO (with a TK1024 CCD) for the photometry of faint starlike and extended objects.

The set of medium-band interference filters with a bandwidth of 160−400 Å and central wavelengths of 3700−9700 Å was produced at the Research Institute of Applied Instrument Making (Moscow). In direct imaging mode, these filters can be used for various tasks, such as constructing the spectral energy distribution for faint objects in the field or imaging extended objects in various emission lines (Hα, [OIII], etc.) and in the continuum. Examples of such images obtained with SCORPIO are given in the paper by Lozinskaya et $al.$ (2002). I.D. Karachentsev (SAO)

Table 2. Parameters of the CCDs used

Parameters	Detector	
	TK1024	EEV 42−40
Type	Thin, back-illuminated	
Format	1024×1024	2048×2048
Pixel size	24×24	13.5×13.5
Scale	$0.32''$/pixel	$0.18''$/pixel
Field of view	$5'.4$	$6'.1$
Maximum quantum efficiency	80%	83%
Readout noise	$3\bar{e}$	$1.8{-}4\bar{e}$
Dark current	$0.1\bar{e}$/min	$0.03\bar{e}$/min

Table 3. Limiting magnitudes in broad-band filters

Filter	T_{exp}, s	Magnitude
B	2500	$27^m.0$
V	1500	26.3
R_C	1260	26.4
I_C	1800	25.1

provided a filter with a bandwidth of 75 Å centered at the wavelength of the Hα line, which was used to map the distribution of ionized hydrogen in nebulas and nearby galaxies.

The main problem of photometric observations is the interference pattern (fringes) during observations in red photometric bands (see section Data Reduction). Thus, for example, during I_c-band observations with an EEV 42−40 CCD, the fringes level reaches 8% of the sky background level.

Long-Slit and Slitless Spectroscopy

The possibility of preimaging proves to be very useful for the slit spectroscopy of both extended (since the slit position is known exactly) and starlike objects if the latter are too faint to be visible in the TV view. Thus, for example, a 1-to-2-min trial exposure in the R filter is enough for reliable pointing at $22{-}23^m$ objects at moderate seeing. In the pointing process, the object under study is set on the detector where the slit image is projected. Subsequently, the slit is set in place of the filter, and a direct-vision grism (the combination of a transparent grating and two prisms) is inserted in the collimated beam, which turns the focal reducer into a fast spectrograph; the change of the direct imaging−long slit configurations takes about one minute. Figure 3 sequentially shows the process of obtaining observational data using the spectroscopy of the radio galaxy RCJ 1154+0431 (the observations at the request of Yu.N. Pariiskii) as an example. The total V magnitude of the object is $19^m.8$; its measured redshift is $z = 1.0$ (see Afanasiev et $al.$ 2003a).
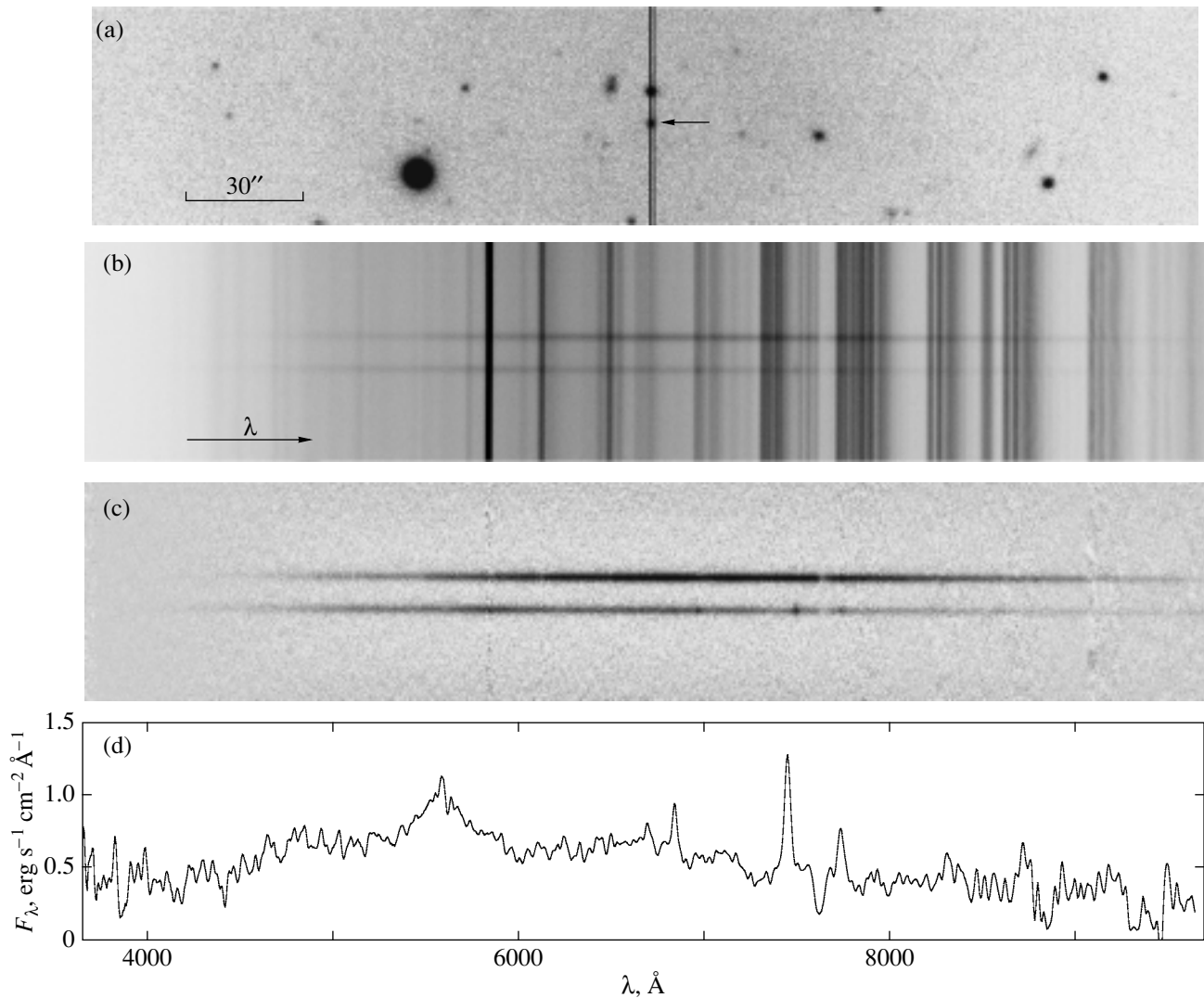
SCORPIO is equipped with a set of grisms that ensure observations with various spectral resolutions (from 1.5 to 20 Å at a slit width of $1''$) in various optical spectral ranges. Before 2003, observations were performed using transparent gratings with profiled grooves with the number of grooves from 300 to 1200 per mm. The gratings are replicas from cut gratings and were produced at the S.I. Vavilov State Optical Institute (St. Petersburg). The maximum quantum efficiency of the entire system (telescope+SCORPIO+CCD) with such gratings was 30% for low-resolution ($\delta\lambda = 15{-}20$ Å) spectra and only about 3−5% for higher-resolution ($\delta\lambda = 5{-}6$ Å) spectra. Observations with grisms using volume phase holographic gratings (VPHGs) were begun in 2003−2004; these have a high transmission and a low level of scattered light (Barden et $al.$ 2000; Habraken et $al.$ 2001). The number of lines for the available VPHGs ranges from 550 to 3000 per mm; in this case, a quantum efficiency of 20−50% is achieved with both low

Fig. 3. Spectroscopy for the radio galaxy RCJ 1154+0431 with SCORPIO: (a) a fragment of an R_C ($T_{exp} = 60$ s) image; the position of the spectrograph slit is shown; the radio galaxy is marked by the arrow; (b) a low-resolution spectrum (the sum of two 600-s exposures); (c) the same after the subtraction of the night sky's spectrum; (d) the integrated spectrum on the wavelength scale.

(see Fig. 2) and high resolutions. The main set of gratings was produced by Wasath Photonics (USA, http://wasatchphotonics.com); two gratings were kindly provided by the University of Padova (Italy).

The achieved quantum efficiency of the instrument allows one to continuously determine the redshifts and spectral classification of extragalactic radio sources, since the required low-resolution spectra of $19-21^m$ objects can be obtained even at moderate atmospheric transparency and at $3-5''$ seeing; the total exposure time is only $10-20$ min (see, e.g., Afanasiev et al. 2003b; Amirkhanyan et al. 2004). At the same time, at $1''.5$ seeing in low-resolution spectroscopy of starlike objects, a limiting magnitude of $R_c = 24^m$ is achieved over a two-hour exposure time (the signal-

to-noise ratio is 10 in the continuum of the spectra obtained). The stability of the instrumental profile of the spectrograph, which affects both the accuracy of subtracting the night sky lines and the possibility of allowance for the interference pattern in the detector material (see section Data Reduction), plays a crucial role in obtaining the spectra of such faint objects. The technique of displacing an object along the slit between exposures helps greatly in such observations. During the subsequent reduction, a pure spectrum, i.e., the spectrum of the sky taken from the same location, but on the displaced frame, is subtracted from the spectrum of the object.

A good test for the capabilities of the spectrograph is to study the kinematics of galactic stellar disks, since absorption spectra with a relatively high signal-
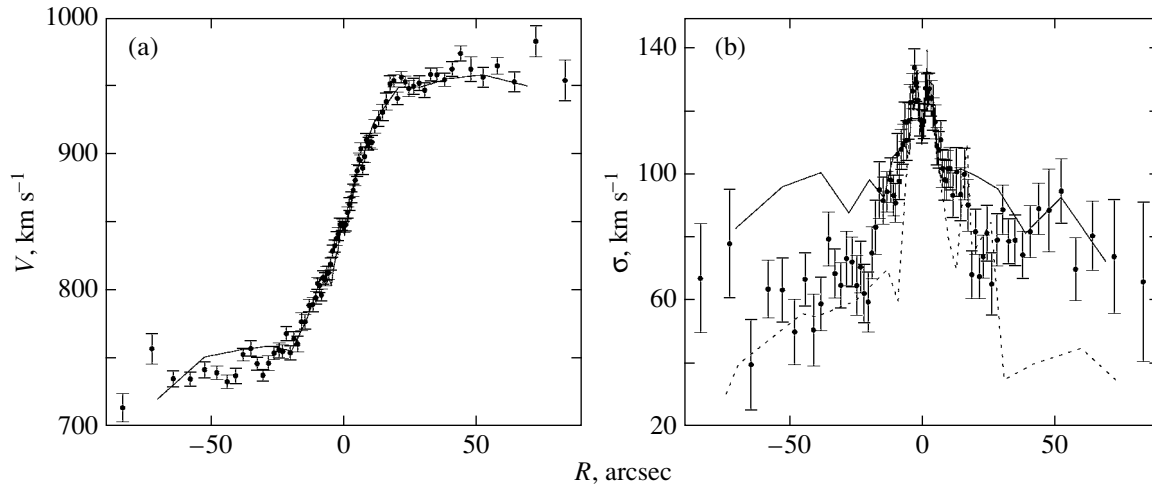
Fig. 4. Kinematics of the stellar component in the galaxy NGC 3412: the distribution of radial velocities (a) and radial velocity dispersion (b) along the major axis. The solid and dashed lines represent the published measurements by Aguerri *et al.* (2003) and Neistein *et al.* (1999), respectively.
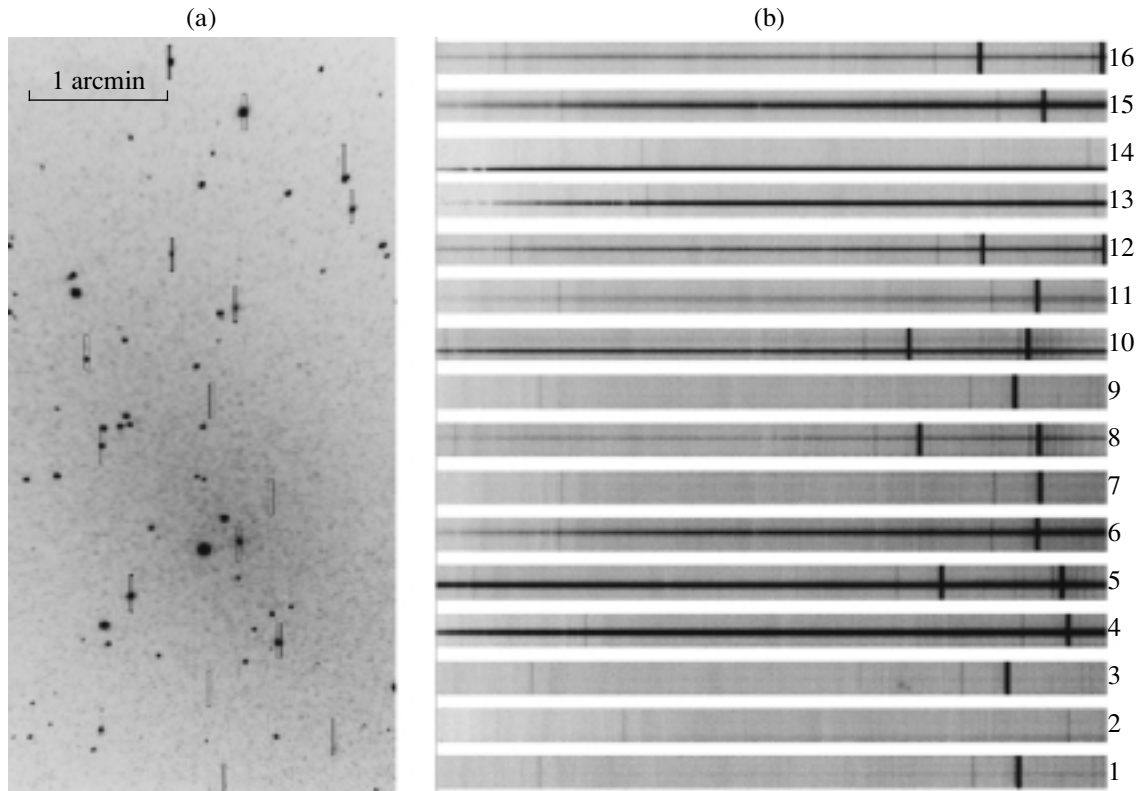

Fig. 5. Spectroscopy for globular clusters in the nearby galaxy NGC 147 (the observations at the request of M.E. Sharina): (a) a V-band image of the galaxy with marked slit positions; (b) multislit spectra of the objects.

to-noise ratio and a spectral resolution of at least $\delta\lambda = 2-4$ Å should be obtained here for regions with a surface brightness of $21-23^m/\square''$. Figure 4 shows an example of measuring the parameters of the stellar kinematics along the major axis of a barred lenticular galaxy, NGC 3412. Over a 1.5-h total exposure time on SCORPIO (using VPHG2310), we can measure the radial velocities and the radial velocity dispersions of stars for regions with a V-band surface brightness of $23^m/\square''$.

Slitless spectroscopy, where the slit is replaced
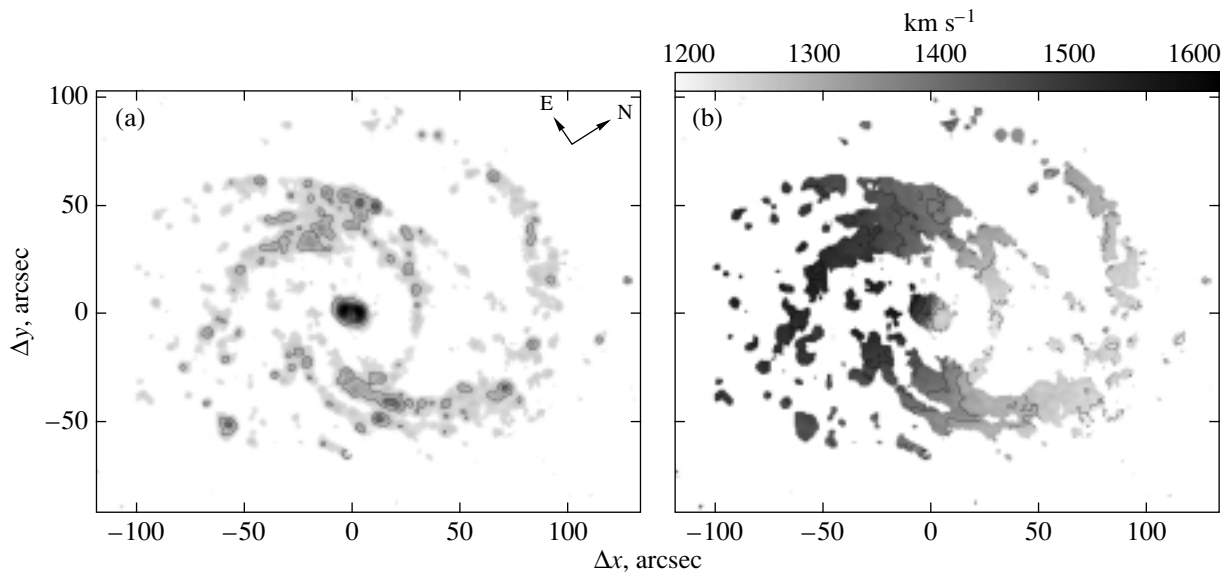
Fig. 6. FPI Hα observations of the galaxy NGC 6951 with SCORPIO: (a) an image in the emission line, (b) the radial velocity field.

with a circular mask about $30''$ in diameter, is used during the observations of spectrophotometric standard stars. This technique allows one to completely eliminate the problems of light losses on the slit and distortion of the spectral energy distribution due to the effect of differential atmospheric refraction.

Multislit Spectroscopy

The available set of grisms can also be used in multislit observations, although the total spectral range decreases in comparison with the long-slit case due to the slit displacement in the field. In such observations, a direct image of the area under study is obtained, a multislit unit is introduced in the focal plane of the telescope, and the slits are set according to the coordinates of the chosen objects measured on the CCD array. The time of the complete slit arrangement (at the required accuracy of $0.2''-0.3''$) is bout 10 min. The spectra obtained are illustrated in Fig. 5.

Panoramic Spectroscopy

A scanning FPI is a highly efficient instrument for studying the kinematics of extended objects. The observations consist in sequentially obtaining several tens of images of the fringes from the object under study (or a calibration lamp) for various optical paths between the plane-parallel plates. The radius of the rings is a function of the wavelength and the FPI plate spacing. After special reduction, these interferograms can be represented as a data cube in which two coordinate axes are located in the plane of the sky and the

wavelengths (or the Doppler velocities measured from the redshifts of spectral lines) are the third coordinate. In other words, an individual spectrum is related to each image element.

The Queensgate ET-50 scanning piezoelectric interferometer is placed between the collimator and the camera where the exit pupil of the optical system is located. There are two scanning FPIs at the SAO; these were provided by the Marseilles (France) and Burakan (Armenia) Observatories. These FPIs work in 235 and 501 orders of interference (near the Hα line) and provide a spectral resolution in this line of 2.5 and 0.7 Å for spectral ranges free from order overlapping 28 and 13 Å respectively. These are used to study both Galactic (nebulae, star clusters) and extragalactic objects. For more details on the FPI observations with SCORPIO, see Moiseev (2002).

Narrow-band filters with a bandwidth of $10-20$ Å are used to separate out the required portion of the spectrum. A set of filters centered at the wavelength of the Doppler-shifted emission line under study is needed for the observations of various galaxies. With the filters available at the SAO, objects with radial velocities from -200 to $+10\,000$ km s^{-1} and from $+3500$ to $+11\,000$ km s^{-1} are currently observable in the Hα and [O III] $\lambda5007$ lines, respectively. Most of the narrow-band filters were produced at the Research Institute of Applied Instrument Making (Moscow); several filters were provided by our colleagues from Padova University (G. Barbieri) and the Burakan Observatory (T. Movsesyan).

Figure 6 shows the FPI observations of the nearby spiral galaxy NGC 6951. The constructed velocity

Table 4. Polarization observations with SCORPIO

Object	SCORPIO measurements		Schmidt *et al.* (1992)	
	P, %	θ	P, %	θ
BD+59d389	6.61 ± 0.11	$97° \pm 1°$	6.701 ± 0.015	98.09
VICyg#12	8.80 ± 0.16	117 ± 2	8.947 ± 0.088	115.03
BD+64d106	5.35 ± 0.41	86 ± 5	5.627 ± 0.037	96.63
BD+28d4211	0.33 ± 0.3	−	0.054 ± 0.030	−

field is in good agreement with similar observations by Rozas *et al.* (2002). Here, the radial velocities were measured with an accuracy of about 5 km s^{-1}. For a detailed discussion of the data obtained, see Moiseev *et al.* (2004).

Polarization Observations

In spectropolarimetric observations, a mask that forms a dotted slit is placed in front of the SCORPIO slit. The height of each slit is $7''.5$, and the separation between the slit centers is $9''.8$. After the introduction of a polarization analyzer in the beam, a series of pairs of spectra in mutually perpendicular polarization planes is obtained at the exit of the spectrograph. Comparing the spectra taken at different orientations of the analyzer, 0° and 45°, we unambiguously determine the degree of linear polarization and the position angle of the polarization plane for the object as a function of the wavelength. Apart from the count statistics determined by the exposure, the actual accuracy of measuring the degree of polarization depends on the level of scattered light in the spectrograph, the accuracy of correcting the spectra for the flat field, the presence of ghosts in the analyzer (their level is about 0.2%), and the accuracy of extracting the spectra from the image.

As our measurements show, the measurement threshold for the degree of linear polarization in SCORPIO is 0.2−0.3%, and the measurement accuracy is 0.1−0.3% (depending on the exposure). Table 4 gives the measurements of polarization standards with VPHG550G (the dispersion is 2 Å/pixel) in the V band. The degrees of polarization that are compared with their published values were obtained by integrating the spectra in the V band.

Figure 7 illustrates the integrated spectrum of the central region in the Seyfert 2 galaxy Mrk 573 obtained when the SCORPIO polarization mode was tested in August 2004. This figure shows the wavelength dependence of the degree of linear polarization. The spectra were obtained at $2''$ seeing with VPHG550G. We took a series of 10-min exposures at successive rotation angles of the analyzer (0° and 45°). The total exposure time was 2 h, and the total V-band brightness of the galactic nucleus in the slit was about 16^m. A broad component of the hydrogen lines is distinguished in Fig. 7c. The result obtained is in satisfactory agreement with the observations of other authors (Nagao *et al.* 2004)

Data Reduction

The observational data are written in the FITS format. Various standard astronomical image reduction systems, such as MIDAS or IRAF, can be used for their reduction. The primary data reduction can be easily automated, since all of the necessary information about the spectrograph configuration is written in the FITS headers. In the IDL environment, we wrote software packages for SCORPIO data reduction and analysis. The programs for the reduction of FPI, long-slit, and multiobject spectroscopy have a user-friendly interface and can be used by users that are not familiar with the IDL language itself. The reduction sequence of FPI observations with SCORPIO was described by Moiseev (2002).

In the reduction of observations, it is important to take into account the fringe pattern in the sensitive layer of the CCD array in the red spectral range (see above). In the case of spectroscopic observations, to correct this effect, the frames with the object's spectra are divided by the frames with the accumulated spectra of the built-in lamp, a spectral flat field. In this way, the fringes effect can be reduced by more than an order of magnitude, which is enough in most cases. We can also get rid of the fringe pattern by the flat-field division in the case of direct imaging, although using the frames illuminated by the built-in lamp or the twilight sky as a flat field is often not enough, particularly in the case of broad-band filters, since the spectral energy distribution of the night sky background emission differs markedly from that of the calibration frames. In this case, an optimal imaging technique is to form an image of the averaged interference pattern using a series of all accumulations in the corresponding filter during the observing night.
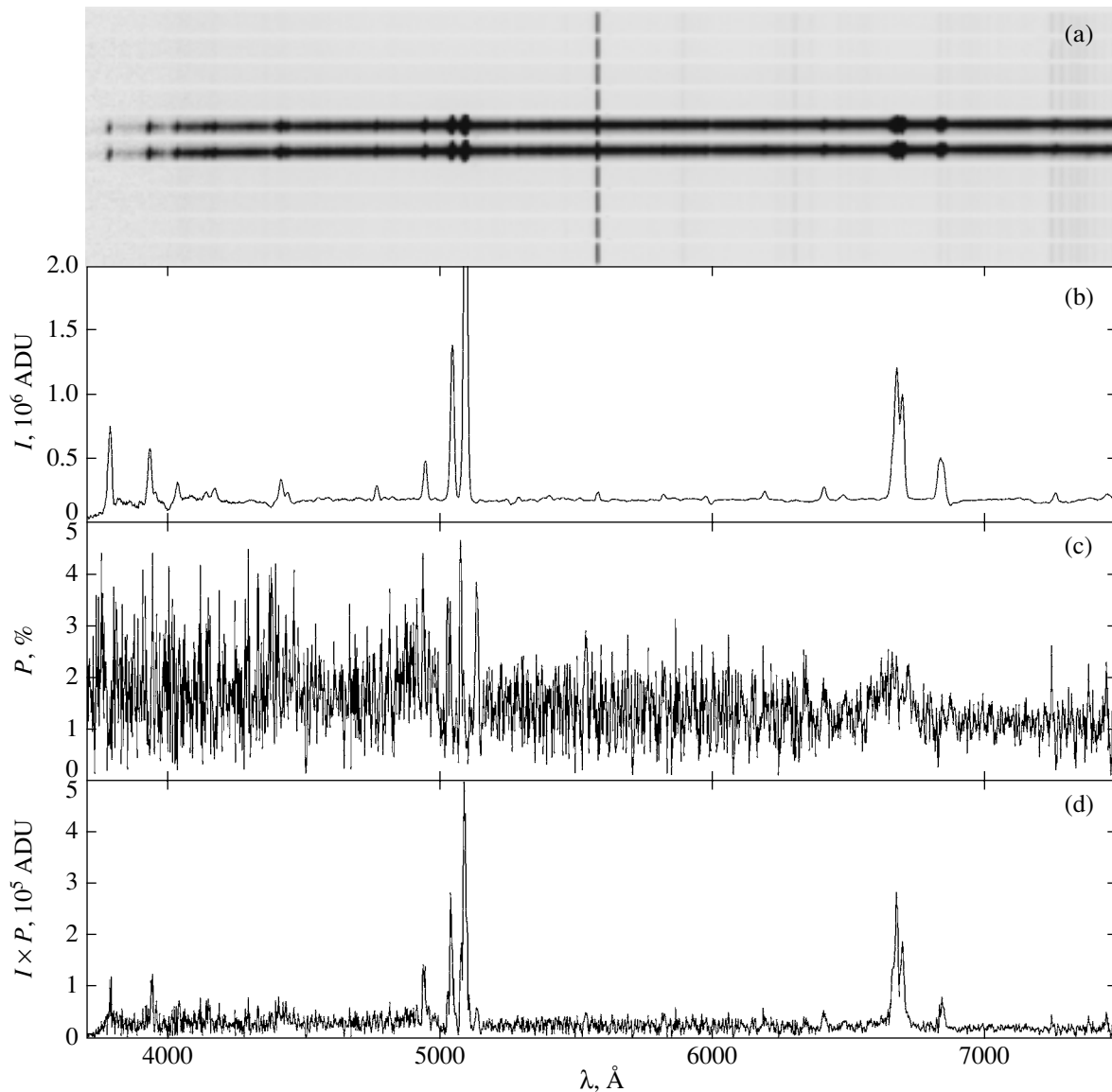
Fig. 7. Spectropolarimetry of the Seyfert 2 nucleus in the galaxy Mrk 573. The grism is VPHG550G, the exposure time is 120 min, and the slit is $1 \times 7''\!.5$: (a) the initial spectrum with the analyzer at position $0°$; (b) the integrated spectrum of the nucleus ($V = 16^m$) minus the spectrum of the surrounding galaxy (I); (c) the degree of polarization of the nucleus (P); (d) the spectrum of polarized emission ($I \times P$). ADU = $0.5\bar{e}$.

FURTHER UPGRADING OF THE INSTRUMENT

A standard autoguiding system using images on the TV view (Shergin and Maksimova 2001) is used during observations with the 6-m telescope. It allows the displacement of stars in the field to be compensated for by the corresponding motions of the entire telescope. This is a slow guiding, since the oscillations of the centers of stellar images at a frequency of 0.1 Hz are suppressed. In 2005, we plan to put into operation a fast guiding system based on a tip-tilt fused quartz plate. This system will allow oscillations with frequencies up to 10 Hz to be compensated for.

This will make it possible to significantly reduce the effect of the telescope's natural oscillations, thereby improving the quality of stellar images. For example, using a local tip-tilt corrector at the Nasmyth-2 focus of the 6-m BTA telescope would allow the limiting magnitude of the spectrographs located there to be increased by $0.5-1^m$ (Ivanov *et al.* 2001).

Equipping the spectrograph with a "semi-thin" CCD detector seems promising. This detector combines the advantages of both "thick" (directly illuminated) devices, the absence of a interference pattern, and "thin" devices (back illuminated) with a high quantum efficiency. Such a detector has a much

smaller fringes amplitude and a higher (up to 80% at 9000 Å) sensitivity in the red spectral range, which will allow the SCORPIO efficiency to be increased during near-infrared observations.

CONCLUSIONS

The high efficiency of the new instrument has been confirmed during its continuous practical use. Over the period from September 2000 through September 2004, observations were performed at the SAO on more than 200 nights; the results obtained were used in 26 papers and five doctoral dissertations. A more detailed description of the spectrograph and the observing technique can be found in the paper by Afanasiev *et al.* (2004), while the current description of the spectrograph is accessible on the Internet at http://www.sao.ru/hq/moisav/scorpio/scorpio. html; a continuously updated gallery of observed objects is also given here. A similar instrument (but without the adapter platform and the multislit unit) produced at the SAO is now being used in observations with the 2.6-m Burakan Astrophysical Observatory telescope (Armenia).

ACKNOWLEDGMENTS

REFERENCES

1. V. L. Afanasiev, S. N. Dodonov, A. V. Moiseev, *et al.*, Astron. Zh. **80**, 409 (2003a) [Astron. Rep. **47**, 377 (2003a)].

2. V. L. Afanasiev, S. N. Dodonov, A. V. Moiseev, *et al.*, Pis'ma Astron. Zh. **29**, 656 (2003b) [Astron. Lett. **29**, 579 (2003b)].

3. V. L. Afanasiev, É. B. Gazhur, S. R. Zhelenkov, *et al.*, Bull. SAO RAS (2004).

4. J. A. L. Aguerri, V. P. Debattista, and E. M. Corsini, Mon. Not. R. Astron. Soc. **338**, 465 (2003).

5. V. R. Amirkharyan, V. L. Afanasiev, S. N. Dodonov, *et al.*, Pis'ma Astron. Zh. **30**, 915 (2004) [Astron. Lett. **30**, 834 (2004)].

6. S. C. Barden, J. A. Arns, W. S. Colburn, *et al.*, Publ. Astron. Soc. Pac. **112**, 809 (2000).

7. M. S. Bessell, Publ. Astron. Soc. Pac. **102**, 1181 (1990).

8. B. Buzzoni, B. Delabre, H. Dekker, *et al.*, ESO Messenger No. 38 (1984).

9. G. Courtés, Ann. Astrophys. **23**, 115 (1960).

10. G. Courtes, Astron. J. **69**, 325 (1964).

11. T. A. Fatkhullin, Bull. Spec. Astrophys. Obs. **53**, 5 (2002).

12. S. Habraken, P.-A. Blanche, P. Lemaire, *et al.*, ESO Messenger No. 106, 6 (2001).

13. A. A. Ivanov, V. E. Panchuk, and V. S. Shergin, Preprint No. 155, SAO RAN (Spec. Astrophys. Obs., Russ. Acad. of Sci., 2001).

14. T. A. Lozinskaya, V. P. Arkhipova, A. V. Moiseev, *et al.*, Astron. Zh. **79**, 19 (2002) [Astron. Rep. **46**, 16 (2002)].

15. A. V. Moiseev, Bull. Spec. Astrophys. Obs. **54**, 74 (2002); astro-ph/0211104.

16. A. V. Moiseev, J. R. Valdés, and V. O. Chavushyan, Astron. Astrophys. **421**, 433 (2004).

17. T. Nagao, K. S. Kawabata, T. Murayama, *et al.*, Astron. J. **128**, 109 (2004).

18. E. Neistein, D. Maoz, H.-W. Rix, *et al.*, Astron. J. **117**, 2666 (1999).

19. H. Nicklas, W. Seifert, H. Boehnhardt, *et al.*, *Optical Telescopes of Today and Tomorrow, Proc. SPIE 2871*, Ed. by A. L. Ardeberg (SPIE, Sweden, 1997), p. 1222.

20. M. Rozas, M. Relano, A. Zurita, and J. E. Beckman, Astron. Astrophys. **386**, 42 (2002).

21. G. D. Schmidt, R. Elston, and O. L. Lupie, Astron. J. **104**, 1563 (1992).

22. V. S. Shergin and V. M. Maksimova, *Auto-Guiding Program TV Guide. User's Guide* (Spec. Astrophys. Obs., 2001); http://www.sao.ru/hq/vsher/vsher_ru.html.

Translated by V. Astakhov

Moments of Inertia and the Chandlerian Period for Two- and Three-Layer Models of the Galilean Satellite Io

V. N. Zharkov[*] and A. L. Sobisevich

Schmidt Joint Institute of Physics of the Earth, Russian Academy of Sciences, ul. Bol'shaya Gruzinskaya 10, Moscow, 123810 Russia

Received October 6, 2004

Abstract—We consider two-layer (Fe-FeS core+silicate mantle) and three-layer (Fe-FeS core+silicate mantle+crust) models of the Galilean satellite Io. Two parameters are known from observations for the equilibrium figure of the satellite, the mean density ρ_0 and the Love number k_2. Previously, the Radau–Darwin formula was used to determine the mean moment of inertia. Using formulas of the Figure Theory, we calculated the principal moments of inertia A, B, and C and the mean moment of inertia I for the two- and three-layer models of Io using ρ_0 and k_2 as the boundary conditions. We concluded that when modeling the internal structure of Io, it is better to use the observed value of k_2 than the moment of inertia I derived from k_2 using the Radau–Darwin formula. For the models under consideration, we calculated the Chandlerian wobble periods of Io. For the three-layer model, this period is approximately 460 days. © 2005 Pleiades Publishing, Inc.

Key words: *Galilean satellite Io, models, Chandlerian periods.*

INTRODUCTION

Previously (Zharkov *et al.* 1985; Zharkov 2004), we constructed a general theory of the equilibrium figure and gravitational field of Jupiter's Galilean satellites in the first and second approximations by expanding the tidal and centrifugal potentials in terms of the small parameter α. For Io, this parameter is

$$\alpha = \frac{3\pi}{G\rho_0\tau^2} = \frac{\omega^2 s_1^3}{Gm_0} = \left(\frac{M}{m_0}\right)\left(\frac{s_1}{R}\right)^3 \quad (1)$$
$$= 171.37 \times 10^{-5},$$

where G is the gravitational constant; M is the mass of Jupiter; R is the radius of the satellite orbit, which is assumed to be circular; and ω is the angular velocity of rotation of the satellite about its axis, which is equal to the angular velocity of revolution of the satellite about the planet for synchronous satellites. In formula (1), $\rho_0, \tau, m_0,$ and s_1 are the mean density, rotation period, mass, and mean radius of Io, respectively. The numerical value in (1) was obtained by using data from Table 1.

According to Kepler's third law, $\omega^2 = GM/R^2$ for synchronous satellites. Therefore, the same parameter α in (1) defines both the tidal, W_2 (A.1), and centrifugal, Q (A.2), potentials.

In the first approximation, formulas for the principal moments of inertia of Io are given in our previous paper (Zharkov 2004). These are

$$A = \frac{8\pi}{15} \int\limits_0^{s_1} \rho(s)d\left[s^5\left(1 + \frac{1}{2}s_2(s) - 2s_{22}(s)\right)\right] \quad (2)$$

(relative to the x axis directed from Io's center of mass to Jupiter),

$$B = \frac{8\pi}{15} \int\limits_0^{s_1} \rho(s)d\left[s^5\left(1 + \frac{1}{2}s_2(s) + 2s_{22}(s)\right)\right] \quad (3)$$

(relative to the y axis directed along the orbital motion of Io),

$$C = \frac{8\pi}{15} \int\limits_0^{s_1} \rho(s)d\left[s^5\left(1 - s_2(s)\right)\right] \quad (4)$$

(relative to the z axis directed along the rotation axis of Io).

The moments of inertia are determined by the density distribution $\rho(s)$, which is assumed to be fixed, and the figure functions $s_2(s)$ and $s_{22}(s)$, which are calculated for two- and three-layer, piecewise constant model $\rho(s)$ in the Appendix.

In the first approximation, the figure functions $s_2(s)$ and $s_{22}(s)$ define the equipotential surfaces of the satellite (Zharkov 2004)

$$r(s, \theta, \varphi) = s[1 + s_0(s) + s_2(s)P_2(t)] \quad (5)$$

---

[*]E-mail: `zharkov@ifz.ru`

Table 1. Observational data and model parameters for Io

Parameters	Values for Io
Orbital radius R, 10^3 km	421.6
Period τ, days	1.769
s_1, km	1821.6 ± 0.5
m_0, 10^{23} g	893.2
ρ_0, g cm^{-3}	3.5278 ± 0.0029
g_0, cm s^{-2}	179
$\alpha = \dfrac{3\pi}{G\rho_0\tau^2}$, 10^{-5}	171.37
$C/(m_0 s_1^2)$	0.37685 ± 0.00035
J_2, 10^{-6}	1845.9 ± 4.2
C_{22}, 10^{-6}	553.7 ± 1.2
Jupiter's mass M, 10^{30} g	1.897
$k_2 = 4(C_{22}/\alpha)$	1.2924 ± 0.0027

$$+ s_{22}(s)P_2^2(t)\cos 2\varphi],$$

where θ and φ are the polar distance and the longitude, $t = \cos\theta$, $P_2(t)$ and $P_2^2(t)$ are the Legendre polynomials. Since the function $s_0(s)$ (Zharkov 2004),

$$-s_0(s) = \frac{1}{5}s_2^2(s) + \frac{12}{5}s_{22}^2(s), \qquad (6)$$

is of the second order of smallness, it is not considered here.

Previously (Zharkov 2004), we showed that the figure functions $s_2(s)$ and $s_{22}(s)$ could be expressed in terms of the Love function $h_2(s)$ using the formulas

$$s_2(s) = -\frac{5}{6}\frac{H_2(s)}{s}\alpha = -\frac{5}{6}h_2(s)\alpha, \qquad (7)$$

$$s_{22}(s) = \frac{1}{4}\frac{H_2(s)}{s}\alpha = \frac{1}{4}h_2(s)\alpha. \qquad (8)$$

For a homogeneous model, $h_2(s) = 5/2 = \text{const}$ (A.13), and, accordingly, the expressions for the moments of inertia are

$$(A, B, C) = \frac{8\pi}{15}\rho_0 s_1^5 \left(1 + e_i\alpha\right), \qquad (9)$$

where $i = (A, B, C)$, $e_A = -55/24$, $e_B = 5/24$, and $e_C = 25/12$.

One might expect the effect of giant thermal plumes on Io to excite a Chandlerian wobble with different periods for an entirely solid model and an elastic model with a liquid core. Thus, a liquid core in the satellite could be revealed in principle by determining the Chandlerian wobble period for Io from observations. We will return to the discussion of Io's Chandlerian wobble after considering the three-layer model.

THE TWO-LAYER MODEL

Substituting the Love function $h_2(s)$ (A.14) into integrals (2)–(4) allows us to write a general expression for the principal moments of inertia in the two-layer model:

$$(A, B, C) = \frac{8\pi}{15}\rho_2 s_1^5 \left\{ \left[x_c^5\left(\frac{\rho_1}{\rho_2} - 1\right) + 1 \right] \qquad (10)$$
$$+ e_i\alpha \left[h_2(s_1) + h_2(s_c)x_c^5\left(\frac{\rho_1}{\rho_2} - 1\right) \right] \right\},$$

where

$$i = (A, B, C), \quad e_A = -11/12, \quad e_B = 1/12, \qquad (11)$$
$$e_C = 5/6, \quad x = s/s_1, \quad s_c = s_1 x_c$$

is the core radius.

Anderson *et al.* (2001) constructed a large set of two-layer (silicate mantle+Fe-FeS core) and three-layer (crust+silicate mantle+Fe-FeS core) models that satisfied the known mean density and moment of inertia of Io. We also considered this issue in detail previously (Zharkov 2003).

Anderson *et al.* (2001) took $\rho_1 = 5.15$ g cm^{-3} for the density of Io's core, which is assumed to be composed of the Fe-FeS eutectic and to be in a molten state. The mean density for the silicate mantle, $\rho_2 = 3.25$ g cm^{-3}, can then be obtained from the conservation condition for the satellite mass and the Love number k_2 ($k_2 = h_2 - 1$). Thus, for our trial two-layer model ($\rho_1 = 5.15$ g cm^{-3}, $\rho_2 = 3.25$ g cm^{-3}, $x_c = 0.5268$, $s_1 = 1821.6$ km), $h_2 = 2.2944$ (Zharkov 2004), which is almost equal to the observed k_2 in Table 1. The value of the Love function $h_2(s_c)$, which is $h_2(s_c) = 1.9956$ for our model, appears in formula (10).

The principal moments of inertia for Io $A < B < C$ and their normalized differences

$$\bar{\alpha} = \frac{(C - B)}{A}, \quad \beta = \frac{(C - A)}{B}, \quad \gamma = \frac{(B - A)}{C} \qquad (12)$$

for the homogeneous, two-layer, and three-layer models are listed in Table 2.

The Eulerian circular frequency of Io's free nutation, which is treated as a rigid, weakly triaxial ellipsoid of revolution, is given by the formula (Zharkov and Molodensky 1996)

$$\sigma_E = \omega\frac{(C - A)(C - B)}{(AB)^{1/2}} = \omega(\bar{\alpha}\beta)^{1/2}, \qquad (13)$$

where ω is the angular velocity of rotation of the satellite, $\bar{\alpha}$ and β are the normalized differences between the principal moments of inertia (12). For our two-layer model, the Eulerian free nutation period is $T_E = 393.26$ days.

Table 2. Normalized principal moments of inertia for Io, their normalized differences, the mean moment of inertia, and the model Love numbers h_2 and h_3

Io's parameters	Homogeneous model	Two-layer model	Three-layer model
$A/(m_0 s_1^2)$	0.3984	0.375886	0.3711902
$B/(m_0 s_1^2)$	0.40014	0.3773653	0.3726431
$C/(m_0 s_1^2)$	0.40143	0.3784744	0.3736409
$I = (A+B+C)/(m_0 s_1^2)$	0.39999	0.3772419	0.3724914
$\bar{\alpha}, 10^3$	3.2257	2.94982	2.6881
$\beta, 10^3$	7.4923	6.8578	6.57653
$\gamma, 10^3$	4.26676	3.90858	3.88849
h_2	2.5	2.2944	2.29704
h_3	1.75	1.6582	1.6330

In contrast to a biaxial body, the path of the pole with respect to a rotating triaxial body is elliptical (Zharkov and Molodensky 1996):

$$\omega_x = C_1 \cos(\sigma_E t), \quad \omega_y = C_2 \cos(\sigma_E t), \qquad (14)$$

where ω_x and ω_y are the components of the angular velocity vector along the principal coordinate axes located in the equatorial plane, C_1 and C_2 are constants, and their ratio is

$$C_1/C_2 = (\bar{\alpha}/\beta)^{1/2}. \qquad (15)$$

Substituting $\bar{\alpha}$ and β for our two-layer model from Table 2 into (15) yields $C_1/C_2 = 0.65546$. Thus, we conclude that the path of the pole during Io's free nutation must be highly elliptical.

THE THREE-LAYER MODEL

The choice of densities for the three-layer model is restricted by the mean density ρ_0 and the Love number $h_2 = 1 + k_2$, which are given in Table 1. In addition, the densities must be within reasonable uncertainty limits. The structure of Io's crust is most complex. We will discuss this issue following Leone and Wilson (2001) (see also Zharkov 2003).

The porosity of the material on Io's surface is assumed to be $\sim$30%. The main fluids are SO_2 and sulfur. The solid-phase SO_2 density is 1.434 g cm^{-3}; the density of the ultrabasic consolidated material is assumed to be 3.36 g cm^{-3}. The volume fractions (in percent) for SO_2 and silicates are 14.5 and 55.6%, respectively, so the surface density is 2.07 g cm^{-3}. SO_2 and sulfur melt at temperatures of 198 and 393 K, respectively; these values depend weakly on the pressure. The probable geothermal gradient on Io is $\sim$39 K km^{-1} (Keszthelyi and MacEwen 1997). The surface temperature of Io is assumed to be 130 K;

SO_2 and sulfur melt at depths of 1.75 and 6.75 km at pressures of 63 and 270 bar, respectively.

The pore volume decreases exponentially with depth due to the pressure of the overlying layers. The critical pressures for SO_2 and sulfur are 78 and 116 bar, respectively. Thus, both SO_2 and sulfur are nearly critical and supercritical liquids at all depths in Io's crust below the melting levels. The densities estimated from the van der Waals equation of state are $\sim$1 g cm^{-3}; i.e., the fluid more likely resembles a liquid than a low-temperature vapor. In our model, the crust is 50 km thick. At this depth, the pressure is $\sim$2.7 kbar and $\rho \approx 3.3$ g cm^{-3}, i.e., it is almost identical to the density of the silicate mantle in the two-layer model. In the model with a porous crust, the mean density of the crust is $\sim$2.685 g cm^{-3}.

In the three-layer density model of Io, the crust thickness and mean density are assumed to be 50 km and 2.7 g cm^{-3}, respectively. The core radius is taken to be the same as that in the two-layer model, $x_c = s_c/s_1 = 0.5268$. The mantle−crust boundary radius is $x_m = 0.97255$. When a crust is introduced in the two-layer model considered above, the density deficit is divided evenly between the Fe−FeS core and the silicate mantle. As a result, we obtain parameters of the three-layer model whose mean density $\rho_{03} = 3.5276926$ g cm^{-3} and $h_2 = 2.2970424$ satisfy the observational data given in Table 1. The trial three-layer density model of Io has the following densities and boundary radii:

$$\rho_c = 5.3 \text{ g cm}^{-3}, \quad \rho_m = 3.2785 \text{ g cm}^{-3}, \qquad (16)$$

$$\rho_{cr} = 2.7 \text{ g cm}^{-3}, \quad x_c = 0.5268, \quad x_m = 0.97255.$$

Substituting the Love function $h_2(s)$ (A.17) into integrals (2)−(4) allows us to write a general expression for the principal moments of inertia in the three-layer

model:

$$(A, B, C) = \frac{8\pi}{15} \rho_3 s_1^5 \left\{ \left[x_c^5 \left(\frac{\rho_1 - \rho_2}{\rho_3} \right) \right. \right. \tag{17}$$

$$\left. + x_m^5 \left(\frac{\rho_2}{\rho_3} - 1 \right) + 1 \right]$$

$$+ e_i \alpha \left[x_c^5 \left(\frac{\rho_1 - \rho_2}{\rho_3} \right) h_2(x_c) \right.$$

$$\left. \left. + x_m^5 \left(\frac{\rho_2}{\rho_3} - 1 \right) h_2(x_m) + h_2(s_1) \right] \right\},$$

where $e_i (i = A, B, C)$ is defined by formula (11). The specific moments of inertia are determined by the model values of the Love function $h_2(x)$ at the three levels $x = (x_c, x_m, 1)$. For model (16), the formulas in the Appendix yield

$$h_2(x_c) = 1.92315, \quad h_2(x_m) = 2.2196786, \tag{18}$$
$$h_2(1) = 2.29704.$$

The normalized principal moments of inertia for Io and their normalized differences (12) are listed in Table 2. We see that introducing a crust into the two-layer model, i.e., passing to the three-layer model, reduces appreciably the moments of inertia. Note that both the two-layer and three-layer models satisfy the observational data, but the three-layer model yields a value for the mean moment of inertia that is smaller than that for the two-layer model by five in the third decimal place. This simple calculation clearly shows that the Radau−Darwin formula

$$\frac{C}{m_0 a_1^2} = \frac{2}{3} \left[1 - \frac{2}{5} \sqrt{\frac{5\alpha}{3J_2 + \alpha} - 1} \right], \tag{19}$$

which is valid for the equilibrium figure of a rotating planet or satellite, yields noticeable errors when applied to triaxial equilibrium figures. This confirms our previous conclusion (Zharkov 2004) that, when constructing a density model for any Galilean satellite, one should use the Love number k_2 or h_2 as a constraint rather than the moment of inertia derived from formula (19) (a_1 in (19) denotes Io's equatorial semiaxis directed toward Jupiter).

The Eulerian free nutation period for the three-layer model, $T_{E3} = 420.73$ days, is almost ~ 30 days longer than that for the two-layer model.

THE CHANDLERIAN PERIOD

Real planets and satellites are not absolutely rigid bodies. In addition, a liquid core also introduces a noticeable correction to the Eulerian frequency σ_E (13). For a planet or a satellite with an elastic mantle and a liquid core, the Chandlerian frequency for a weakly triaxial body can be written with a satisfactory accuracy as (Molodensky 1961; Zharkov and Molodensky 1996)

$$\sigma_W = \frac{[(C - A)(C - B)]^{1/2}}{(AB)^{1/2} - A_c} \left(1 - \frac{k}{k_0} \right) \omega \tag{20}$$

$$= \frac{(\bar{\alpha}\beta)^{1/2}}{1 - \frac{A_c}{(AB)^{1/2}}} \left(1 - \frac{k}{k_0} \right) \omega,$$

where A_c is the moment of inertia for the liquid core, k is the Love number corrected for elasticity and the presence of a liquid core,

$$k_0 = 3G(C - \bar{A})/(\omega^2 s_1^5) \tag{21}$$

is called the secular Love number, G is the gravitational constant,

$$\bar{A} = \bar{B} = \frac{A + B}{2} \tag{22}$$

is the mean equatorial moment of inertia for the satellite.

According to (13) and (20), the expression for the Chandlerian period is

$$T_W = T_E \frac{1 - A_c/(AB)^{1/2}}{1 - k/k_0}, \tag{20a}$$

where T_E is the Eulerian period.

We used the following parameters to estimate the Love number k.

The two-layer model: $\rho_c = 5.15$ g cm^{-3}, $\rho_m = 3.25$ g cm^{-3}; the seismic velocities in the mantle $\nu_P = 7.86$ km s^{-1} and $\nu_S = 4.5$ km s^{-1}; the velocity in the core in the model with a liquid core $\nu_P(l) = 3.7$ km s^{-1}, the velocities in the model with a solid core $\nu_P(s) = 4.7$ km s^{-1} and $\nu_S(s) = 2.5$ km s^{-1};

The three-layer model: $\rho_c = 5.3$ g cm^{-3}, $\rho_m = 3.2785$ g cm^{-3}, $\rho_{cr} = 2.7$ g cm^{-3}; the seismic velocities in the mantle $\nu_P = 8$ km s^{-1} and $\nu_S = 4.6$ km s^{-1}; the velocity in the core in the model with a liquid core $\nu_P(l) = 3.7$ km s^{-1}, the velocities in the model with a solid core $\nu_P(s) = 4.95$ km s^{-1} and $\nu_S(s) = 2.65$ km s^{-1}; the velocities in the crust $\nu_P = 6.15$ km s^{-1} and $\nu_S = 3.45$ km s^{-1} (Anderson 1995; Zharkov 2003; Gudkova and Zharkov 2004).

Estimates of the quantities in formula (20a), which defines the Chandlerian period T_W, and the model periods themselves are collected in Table 3.

An interesting feature of the model estimates for Io's Chandlerian period is that the model with a liquid core and the entirely elastic model have very close values. The period for the three-layer model is about 30 days longer than that for the two-layer model, primarily because the moments of inertia for the former are appreciably lower than those for the latter. As

Table 3. Love numbers k, normalized moments of inertia for a liquid core $A_c/(m_0 s_1^2)$, Eulerian periods T_E, and Chandlerian wobble periods T_W for entirely elastic models and models with a liquid core*

Parameters	Two-layer model		Three-layer model	
	liquid core	solid core	liquid core	solid core
k	0.05274	0.03189	0.05085	0.03090
k_0	0.3623	0.3623	0.339	0.339
$A_c/m_0 s_1^2$	0.02366	–	0.02435	–
T_E, days	–	393.3	–	420.7
T_W, days	431.4	431.26	462.5	463.0

* The numbers k were calculated by S.M. Molodensky.

follows from (20a), the presence of a liquid core shortens the period, while elasticity lengthens it. The Love numbers k differ markedly for the model with a liquid core and the entirely elastic model. This ultimately leads to close Chandlerian periods for the model with a liquid core and the entirely elastic model.

Zharkov and Molodensky (1996) showed that the Chandlerian period of Mars lengthens by $\sim 2-5$ days due to its inelasticity. We still know too little about Io's inelasticity to make such estimates, although the deviation of its interiors from an ideal elastic state must be significant. It may well be that Io's interior inelasticity causes T_W to lengthen by ~ 5 days. Since the three-layer model of Io seems more realistic than its two-layer model, the Chandlerian period can be roughly estimated as $T_W \sim 460 \pm 15$ days. Clearly, the model for the internal structure of Io could be improved by determining its Chandlerian period from observations. Io's interiors are strongly heated; therefore, the model with a liquid core seems more realistic than the entirely elastic model. On the other hand, much still remains unknown about Io's interiors. Since the interiors of this satellite are strongly heated, effectively liquefied zones can in principle exist in the silicate mantle. As a result, the Chandlerian period of the satellite can deviate appreciably from our model estimate. In principle, this result would be of no less significance than our observational confirmation of the T_W estimate.

CONCLUSIONS

We have considered two-layer (Fe-FeS core+ silicate mantle) and three-layer (Fe-FeS core+silicate mantle+crust) models of the Galilean satellite Io. Two independent conditions are known from observations, the mean density of Io ρ_0 and the Love number k_2; the latter is related to the quadrupole gravitational moment of the satellite J_2 by a simple formula if Io is assumed to have an equilibrium figure. Clearly, the three principal moments of inertia for the

satellite A, B, C and, hence, the moment of inertia $I = \dfrac{1}{3m_0 s_1^2}(A + B + C)$ cannot be determined from these two conditions without additional assumptions. In their original paper, based on Galileo spacecraft data, Anderson *et al.* (2001) determined the gravitational moments J_2 and C_{22} from observations and found that the two moments are related by

$$J_2 = \frac{10}{3}C_{22} = \frac{5}{6}\alpha k_2. \tag{23}$$

This relation is valid for bodies in a state close to a hydrostatic equilibrium. The Radau−Darwin formula (19) was used to determine the mean moment of inertia I. Subsequently, we chose the density distributions in the two- and three-layer models of Io that satisfied ρ_0 and I.

The general theory of the equilibrium figure and the gravitational field constructed previously (Zharkov 2004) allowed us to derive formulas (2)−(3) for the principal moments of inertia in the form of definite integrals of the density and the figure functions $s_2(s)$ and $s_{22}(s)$. We used formulas (2)−(3) to calculate the principal moments of inertia for Io and, thus, the mean moment of inertia I for typical trial two- and three-layer models of this satellite. Both models satisfy the observational data, i.e., ρ_0 and k_2. However, the mean moment of inertia for the three-layer model is lower than that for the two-layer model by five in the third decimal place. Thus, we concluded that one must restrict the choice of trial density distributions for Io by ρ_0 and k_2 rather than go from k_2 to the moment of inertia using the Radau−Darwin formula (19). The latter is valid for a rotating planet or satellite that are not in the field of a tidal potential.

We obtained a reconnaissance estimate of Io's Chandlerian period T_W (20a). The powerful thermal plumes ejected from Io's interiors give hope that a Chandlerian wobble could be excited on this satellite. The Chandlerian period for the trial three-layer model

of the satellite was found to be $\sim$460 days; its pole must move along a highly elliptical path. Determining the Chandlerian period from observations is of great scientific interest irrespective of whether T_W will be close to our calculated value or differ markedly from our predicted value. In the latter case, this would imply that we made a gross mistake in choosing the physical model of Io.

APPENDIX

Let us introduce a spherical (r, θ, φ) coordinate system, where r is the radius, θ is the colatitude (polar distance), and φ is the longitude. In the first (linear) approximation, the tidal potential from Jupiter at the point of a satellite with (r, θ) coordinates has the standard form

$$W_2 = \frac{GMr^2}{R^3} P_2(\cos Z) = \frac{Gm_0 s^2}{s_1^3} \alpha P_2(\cos Z). \tag{A.1}$$

Here, α is defined by formula (1), and the notation is given in the Introduction; in the linear approximation, the radius r may be substituted with the radius of a sphere of equivalent volume s (Zharkov 2004); Z is the angle between the radius vector of the interior point of the satellite and the direction of Jupiter's center of mass; and P_2 is the second-degree Legendre polynomial. The equilibrium figure of the satellite is an equipotential surface of the sum of three potentials: the tidal potential W_2 (A.1), the centrifugal potential

$$Q = \frac{\omega^2 r^2}{3}[1 - P_2(\cos\theta)] = \frac{Gm_0 s^2}{3s_1^3}\alpha[1 - P_2(\cos\theta)], \tag{A.2}$$

and the potential from the mass distribution inside the satellite $V(s, \theta, \varphi)$,

$$U = W_2 + Q + V = \text{const.}$$

If the model of the satellite, i.e., its density distribution, is known and if it is close to a hydrostatic equilibrium, then the sought figure functions $s_2(s)$ and $s_{22}(s)$ that appear in integrals (2), (3), (4) and that define the principal moments of inertia for the satellite A, B, C can be expressed in terms of the Love function (see formulas (7) and (8)).

The Love numbers can be calculated by a method described by Gavrilov *et al.* (1975), Gavrilov and Zharkov (1977), and Zharkov *et al.* (1985). Let us introduce an auxiliary function $T(s)$, which is equal to the total perturbation of the satellite potential due to the action of W_2 (A.1) or Q (A.2):

$$V_i + W_k = \frac{1}{s_1 g_0} T\left(\frac{s}{s_1}\right)^n W_k, \tag{A.3}$$

$$k = 2 \text{ or } r, \quad W_r \equiv Q,$$

where n is the degree of the spherical function to which both W_2 (A.1) and Q (A.2) are directly proportional (in our case, $n = 2$), V_i is is the perturbation of the satellite potential due to the action of W_2 (A.1) $(k = 2)$ or Q (A.2) $(k = r)$, and g_0 is the gravitational acceleration on the surface of the satellite modeled by a sphere of equivalent volume with radius s_1. The function $T(s)$ satisfies the ordinary differential equation of the second order

$$T'' + \frac{2}{s}T' + \left(4\pi G\frac{\rho'}{V'} - \frac{n(n+1)}{s^2}\right)T = 0, \tag{A.4}$$

where the prime denotes differentiation with respect to s. The function $T(s)$ must be finite at zero and satisfy the following boundary condition on the $s = s_1$ surface:

$$(T')_{s_1-0} = -\frac{n+1}{s_1}(T)_{s_1-0} \tag{A.5}$$
$$+ 4\pi G\rho_{s_1}H_{s_1} + (2n+1)g_0.$$

The density distribution $\rho(s)$ in the satellite is assumed to be spherically symmetric, and, accordingly, the gravitational potential $V(s)$ depends only on the radius s.

The other function $H(s)$ that we encounter in the problem specifies the deformation and characterizes the radial displacement of the point under consideration

$$u_k = \frac{H(s)}{s_1 g_0}\left(\frac{s_1}{s}\right)^n W_k, \tag{A.6}$$
$$k = 2 \text{ or } r, \quad W_r \equiv Q.$$

This function can be expressed in terms of T and V' using the formula

$$T + HV' = 0. \tag{A.7}$$

We consider two- and three-layer piecewise constant density distributions. At density jumps, the potential $T(s)$ must satisfy the conditions

$$(T)_{s_i-0} = (T)_{s_i+0}, \tag{A.8}$$

$$\left(\frac{\partial T}{\partial s}\right)_{s_i-0} - \left(\frac{\partial T}{\partial s}\right)_{s_i+0} \tag{A.9}$$
$$= 4\pi G(\rho_{s_i-0} - \rho_{s_i+0})H(s_i),$$

where s_i is the radius of the boundary at which a density jump occurs. Thus, s_c is the core radius in both the two-layer and three-layer models of the satellite, and s_m is the radius of the boundary at which the silicate mantle of Io borders the crust in the three-layer model.

The Love numbers of the nth order are determined by the values of $T(s)$ and $H(s)$ on the $s = s_1$ surface

$$h_n = \frac{H(s_1)}{s_1}, \quad k_n = \frac{T(s_1)}{s_1 g_0} - 1 \qquad (A.10)$$

(the subscripts n of the functions $T(s)$ and $H(s)$ in Eqs. (A.3)–(A.10) was discarded). There is a simple relation between h_n and k_n,

$$h_n = 1 + k_n. \qquad (A.11)$$

The Love function

$$h_2(s) = \frac{H_2(s)}{s} \qquad (A.12)$$

allows us to determine the sought figure functions $s_2(s)$ and $s_{22}(s)$ using formulas (7) and (8) and, thus, to calculate the moments of inertia using integrals (2)–(4). For the homogeneous model, $\rho(s) = \rho_0 = \text{const}$,

$$h_2(s) = \frac{5}{2} = \text{const}, \qquad (A.13)$$

and the moments of inertia are given by formula (9).

When considering the two-layer model, we denote the core density by ρ_1 and the envelope density by ρ_2; s_c is the core radius. The solution of the problem, the Love function $h_n(s)$, is

$$h_n(s) = \begin{cases} \dfrac{3C}{4\pi G \rho_1} s^{(n-2)}, & 0 \le s \le s_c \\ \dfrac{C_1 s^{n+1} + C_2 s^{-n}}{\dfrac{4\pi G}{3}[\rho_1 s_c^3 + \rho_2(s^3 - s_c^3)]}, & s_c \le s \le s_1. \end{cases}$$
$$(A.14)$$

The coefficients C, C_1, and C_2 are

$$C = C_1 + C_2 s_c^{-(2n+1)}, \qquad (A.15)$$

$$C_1 = \frac{(2n+1)g_0 s_1^{-(n-1)} d_1 d_2}{d_1[(2n+1)d_2 - 3] - d_3},$$

$$C_2 = \frac{(2n+1)}{3} \frac{g_0 s_1^{(n+2)} d_2 d_3}{d_1[(2n+1)d_2 - 3] - d_3},$$

where the following notation is used:

$$d_1 = (2n+1) - 3(1 - \rho_2/\rho_1), \qquad (A.16)$$

$$d_2 = x_c^3(\rho_1/\rho_2 - 1) + 1,$$

$$d_3 = 9(1 - \rho_2/\rho_1)x_c^{(2n+1)}, \quad x_c = s_c/s_1.$$

Here, we are interested in the case of $n = 2$. Assuming that $s = s_1$ in (A.14), we obtain the Love number h_n for the two-layer model. The corresponding formula is given in our previous paper (Zharkov 2004).

In the three-layer model, ρ_1, ρ_2, and ρ_3 denote the core, mantle, and crust densities, respectively, s_c

and s_m are the radii of the core and the mantle–crust boundary. In this case, the sought Love function $h_n(s)$ is

$$h_n(s) = \begin{cases} \dfrac{3C s^{n-2}}{4\pi G \rho_1}, & 0 \le s \le s_c \\ \dfrac{C_1 s^{n+1} + C_2 s^{-n}}{\dfrac{4\pi}{3}G[\rho_1 s_c^3 + \rho_2(s^3 - s_c^3)]}, & s_c \le s \le s_m \\ \dfrac{C_3 s^{n+1} + C_4 s^{-n}}{\dfrac{4\pi}{3}G[\rho_1 s_c^3 + \rho_2(s_m^3 - s_c^3) + \rho_3(s^3 - s_m^3)]}, \\ s_m \le s \le s_1. \end{cases}$$
$$(A.17)$$

The coefficients C, C_1, C_2, C_3, and C_4 in (A.17) are

$$C = C_1 + C_2 s_c^{-(2n+1)}, \quad C_1 = b_3 C_2 s_c^{-(2n+1)},$$

$$C_2 = \frac{s_c^{2n+1}}{b_4}[C_3 + C_4 s_m^{-(2n+1)}], \qquad (A.18)$$

$$C_3 = -\frac{a_4}{a_3}C_4 s_m^{-(2n+1)}, \quad C_4 = -\frac{(2n+1)g_0}{a_5},$$

where the following notation is used:

$$a_1 = (x_c/x_m)^3(\rho_1/\rho_2 - 1) + 1, \qquad (A.19)$$

$$a_2 = x_c^3 \frac{(\rho_1 - \rho_2)}{\rho_3} + x_m^3\left(\frac{\rho_2}{\rho_3} - 1\right) + 1,$$

$$a_3 = \frac{b_3}{b_4}\left(n - \frac{b_2}{a_1}\right) - \frac{b_5}{b_4}\left(\frac{x_m}{x_c}\right)^{-(2n+1)} - n,$$

$$a_4 = a_3 + (2n+1),$$

$$a_5 = \frac{3}{a_2}s_1^{-(n+2)}\left\{1 + \left[\frac{a_2}{3}(2n+1) - 1\right]\right.$$

$$\left. \times \frac{a_4}{a_3}x_m^{-(2n+1)}\right\}, \quad x_m = s_m/s_1,$$

$$b_1 = 3(1 - \rho_2/\rho_1), \quad b_2 = 3(1 - \rho_3/\rho_2), \quad (A.20)$$

$$b_3 = (2n+1)/b_1 - 1, \quad b_4 = b_3 + (x_c/x_m)^{2n+1},$$

$$b_5 = (n+1) + b_2/a_1.$$

The formula for the Love number $h_n(s_1)$ for the three-layer model with a piecewise constant density distribution is

$$h_n(s_1) = \frac{2n+1}{3}a_2 x_m^{-(2n+1)} \qquad (A.21)$$

$$\times \frac{\dfrac{a_4}{a_3} - x_m^{2n+1}}{1 + \left[\dfrac{a_2}{3}(2n+1) - 1\right]\dfrac{a_4}{a_3}x_m^{-(2n+1)}}.$$

REFERENCES

1. O. L. Anderson, *Equation of State of Solids for Geophysics and Ceramic Science* (Oxford Univ. Press, New York, Oxford, 1995).
2. J. D. Anderson, R. A. Jacobson, E. L. Lau, *et al.*, J. Geophys. Res. **106** (E12), 32 963 (2001).
3. S. V. Gavrilov and V. N. Zharkov, Icarus **32**, 443 (1977).
4. S. V. Gavrilov, V. N. Zharkov, and V. V. Leont'ev, Astron. Zh. **52**, 1021 (1975) [Sov. Astron. **19**, 618 (1975)].
5. T. V. Gudkova and V. N. Zharkov, Phys. Earth Planet. Inter. **142**, 1 (2004).
6. I. Keszthelyi and A. S. MacEwen, Geophys. Res. Lett. **24**, 2463 (1997).
7. G. Leone and L. Wilson, J. Geophys. Res. **106** (E12), 32 983 (2001).
8. S. M. Molodensky, The Theory of Nutation and Diurnal Earth Tides. Communs. Observ. R. Belg., 188, S. Geophys. **58**, 25 (1961).
9. V. N. Zharkov, *Geophysical Investigations of Planets and Moons. First Reading of Shmidt, Moscow, 30 September, 2002* (Joint Institute of Earth Physics, Russ. Acad. Sci., Moscow, 2003), p. 102.
10. V. N. Zharkov, Pis'ma Astron. Zh. **30**, 548 (2004) [Astron. Lett. **30**, 496 (2004)].
11. V. N. Zharkov, V. V. Leontjev, and A. V. Kozenko, Icarus **61**, 92 (1985).
12. V. N. Zharkov and S. M. Molodensky, Planet. Space Sci. **44**, 1457 (1996).

Translated by V. Astakhov

Analysis of the Neighborhood of the 2 : 1 Resonance in the Equal-Mass Three-Body Problem

A. I. Martynova[1], V. V. Orlov[2*], and L. L. Sokolov[2]

[1]*St. Petersburg Forestry Engineering Academy, Institutskiĭ per. 5, St. Petersburg, 194022 Russia*
[2]*Sobolev Astronomical Institute, St. Petersburg State University, Universitetskiĭ pr. 28, St. Petersburg, Peterhof, 198504 Russia*
Received June 15, 2004

Abstract—We consider the trajectories in the neighborhood of a 2 : 1 resonance (in periods of osculating motions of the outer and inner binaries) in the plane equal-mass three-body problem. We identified the zones of motions that are stable on limited time intervals. All of them correspond to the retrograde motions of the outer and inner subsystems. The prograde motions are unstable: the triple system breaks up into a final binary and an escaping component. In the barycentric nonrotating coordinate system, the trajectories occasionally form symmetric structures composed of several leaves. These structures persist for a long time, and, subsequently, the trajectories of the bodies fill compact regions in coordinate space. © *2005 Pleiades Publishing, Inc.*

Key words: *celestial mechanics.*

INTRODUCTION

Resonance motions, i.e., motions in the neighborhood of commensurable osculating periods, are characterized by complex properties of their trajectories. Typical features of these motions include the emergence of new stable and unstable periodic orbits and characteristic structures of dynamic chaos in the vicinity of the perturbed separatrices of unstable periodic orbits. These structures are related to the well-known Arnold diffusion.

Resonance trajectories play an important part in the evolution of stellar and planetary systems. A whole series of orbital and spin–orbital resonances are observed in the Solar System. These include the well-known resonances in the system of Galilean satellites (Ferraz Mello 1983), spin–orbital resonances of the Moon and Mercury, etc. Spin–orbital resonances can be explained by natural causes (Beletskiĭ 1972). Gaps and condensations in the asteroid belt are directly related to stable and unstable resonance motions (Brower 1963, Bryuno 1970). The structure of the rings surrounding giant planets is also related to resonance behavior (Gor'kavyĭ and Fridman 1994).

The fact that resonance planetary systems have already been found among the few extrasolar planetary systems also corroborates the importance of

resonance effects and of their analysis for astronomy. These systems include GJ 876 (Marcy *et al.* 2001) and HD 82943 (Gozdziewsky and Maciejewski 2001; Santos *et al.* 2000)—each has two planets and one of the periods is almost exactly twice as long than the other. A 3 : 1 resonance has been found in the 55 Cnc system (Marcy *et al.* 2002). This is unlikely to be a coincidence.

Celestial mechanics usually analyzes commensurabilities in the motions of bodies with small masses compared to the mass of the central body (planets, their natural and artificial satellites, asteroids). Well developed analytical methods are known for constructing resonance solutions in these cases (see, e.g., Grebenikov and Ryabov 1971). However, fundamentally new results appear even in this thoroughly studied field. Recently, a region of stable chaos (Brasser *et al.* 2004) and other structures (Pit'ev and Sokolov 2003) have been found in the restricted spatial circular three-body problem. Resonance motions of comparable or equal masses are much more poorly studied. Of interest are nontrivial stable structures (in particular, stable periodic motions), which develop in triple and multiple stars (with more than three components) with commensurable periods.

Resonance motions of equal-mass bodies appear to be closely related to the "choreographies" discovered not too long ago (see, e.g., Moore (1993); Simo (2002), and Vanderbei (2003)). Choreographies are
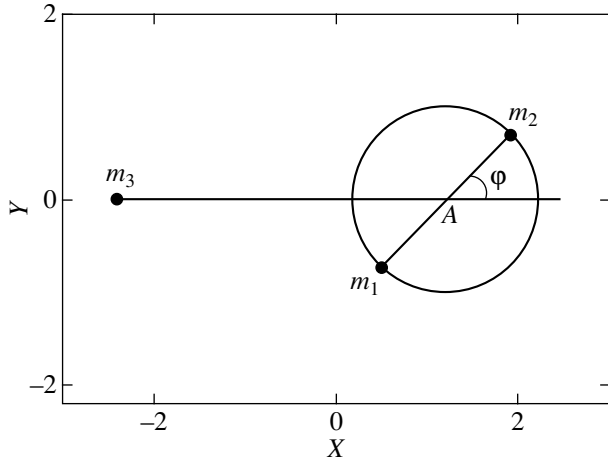

Fig. 1. Initial configuration.

specific periodic solutions, where several bodies move in the same orbit with the same period. Like other periodic motions, choreographies may be stable or unstable.

In this paper we analyze rotating triple systems with equal-mass components in the neighborhood of the $2:1$ commensurability between the osculating periods of the outer and inner binaries. The outer binary is made up of the distant body and the center of mass of the inner binary.

THE METHOD OF ANALYSIS

Consider the dynamical evolution of triple systems with equal-mass components in the vicinity of the $2:1$ resonance of the initial periods of unperturbed motion of the outer and inner binaries. The outer binary is made up of the distant body and the center of mass of the two close bodies.

In this paper we restrict our analysis to the plane problem, where all motions are in the XY plane. We choose the system of units so that gravitational constant and masses of the bodies are equal to unity ($G = 1$, $m_1 = m_2 = m_3 = m = 1$) and so is the radius of the initial circular orbit of the inner binary is equal to 1. In other words, both bodies that make up the inner binary are at a unit distance from its center of mass; the distance between these bodies (or the semimajor axis a_{in} of the relative motion of this pair) is equal to 2.

We analyze motions that are close to the $1:2$ commensurability of the initial osculating periods of the inner and outer binaries. The period of the inner binary, P_{in}, can be determined from the following relation:

$$2\pi/P_{in} = \sqrt{2Gm/a_{in}^3}. \qquad (1)$$

The period of the outer binary, P_{ex}, made up of the third body and the barycenter of the inner binary (point A in Fig. 1) can be determined from a similar relation,

$$2\pi/P_{ex} = \sqrt{3Gm/a_{ex}^3}, \qquad (2)$$

where a_{ex} is the semimajor axis of the outer binary or the distance between points m_3 and A at the initial time instant. We assume that the initial orbits of both the inner and outer binaries are circular.

The $P_{ex} : P_{in} = 2 : 1$ commensurability condition implies, in accordance with relations (1) and (2), that

$$a_{ex} = 6^{1/3} a_{in}. \qquad (3)$$

In the case of prograde motions (the outer and inner binaries rotate in the same direction), triple systems in the vicinity of the $2:1$ resonance are found to be unstable for all initial conditions considered. Such systems break into a binary and a single body that escapes from it in rather short time (several dozen periods of the inner binary). Motions in the vicinity of the $2:1$ resonance in the case of retrograde motions appear to be stable for certain initial conditions, at least over limited time intervals. Below we analyze retrograde motions exclusively. Let the inner binary made up of bodies m_1 and m_2 rotate clockwise and body m_3 move counterclockwise. At the initial instant $t = 0$ body m_3 and the center of mass A of the inner pair are located on the X axis (see Fig. 1).

We use phase φ of the orbital motion of the inner pair as a parameter. Masses of the bodies are equal to $m_1 = m_2 = 1$ (bodies are indistinguishable) and we therefore consider retrograde motions exclusively and can select angle φ in the interval $[0, \pi]$.

We numerically integrate the equations of motions in the fixed coordinate system with the origin at the center of mass of the triple system. We compute the initial coordinates and velocities of bodies in this coordinate system (Fig. 1) by the following formula:

$$x_1 = x_A - \cos\varphi, \quad y_1 = -\sin\varphi, \qquad (4)$$
$$x_2 = x_A + \cos\varphi, \quad y_2 = \sin\varphi,$$
$$x_3 = x_A - 2 \times 6^{1/3}, \quad y_3 = 0,$$

where $x_A = (16/9)^{1/3}$ is the initial abscissa of the center of mass of the inner binary;

$$\dot{x}_1 = -\frac{1}{2}\sin\varphi, \quad \dot{y}_1 = \frac{1}{2}\cos\varphi + 6^{-2/3}, \qquad (5)$$
$$\dot{x}_2 = \frac{1}{2}\sin\varphi, \quad \dot{y}_2 = -\frac{1}{2}\cos\varphi + 6^{-2/3},$$
$$\dot{x}_3 = 0, \quad \dot{y}_3 = -2 \times 6^{-2/3}.$$

The circular velocity of the motion of third body relative to the center of mass of the inner binary is equal to

$$V_{\text{ex}} = \sqrt{\frac{G(m_1 + m_2 + m_3)}{a_{\text{ex}}}} = \frac{6^{1/3}}{2}. \qquad (6)$$

We performed scanning in angle φ with different steps and identified trajectories with long lifetimes $t > t_{\text{crit}}$ during which the triple system remained within the bounded region such that the maximum distance between bodies

$$r_{\max} \leq 5d, \qquad (7)$$

where d is the average size of the triple system used in TRIPLE code (see below),

$$d = \frac{G(m_1 m_2 + m_1 m_3 + m_2 m_3)}{|E|}, \qquad (8)$$

and E is the total energy of the triple system. We set the critical time t_{crit} equal to 500τ or 1000τ, where τ is the mean time during which a component crosses of the triple system, which is also used in TRIPLE code,

$$\tau = \frac{G(m_1 + m_2 + m_3)^{5/2}}{(2|E|)^{3/2}}. \qquad (9)$$

In addition, we varied the positions of the bodies m_1 and m_2 that make up the inner binary within certain intervals of angle φ. We varied the ordinates of these bodies so as to leave the initial position of the center of mass of the inner pair unchanged. As a result, the initial orbit of the inner binary became elliptical ($e_{\text{in}} \neq 0$). We also analyzed the case where the center of mass of the inner binary shifted along the axis connecting the center of mass of this binary and the third distant component, and also the case where it shifted in the orthogonal direction (parallel to the ordinate axis).

We numerically integrated the equations of motion of the three-body problem using TRIPLE code developed by Sverre Aarseth (University of Cambridge, Great Britain). This code uses Bulirsch–Stoer integrator (Bulirsch and Stoer 1966) and numerically solves a set of regularized equations (Aarseth and Zare (1974); see also Aarseth (2003)).

RESULTS

We first scan over angle φ with a step of $\Delta\varphi = 0°.1$. As a result, we obtain the dependence of lifetime t on φ, when $\varphi \in [0, \pi]$ (see Fig. 2). At $\varphi < 55°$ and $\varphi > 124°$ triple systems usually disrupt during time $t < 1000\tau$. In the interval $\varphi \in [55°, 124°]$, the motions of bodies in systems remain bounded during this time and condition (7) is satisfied.
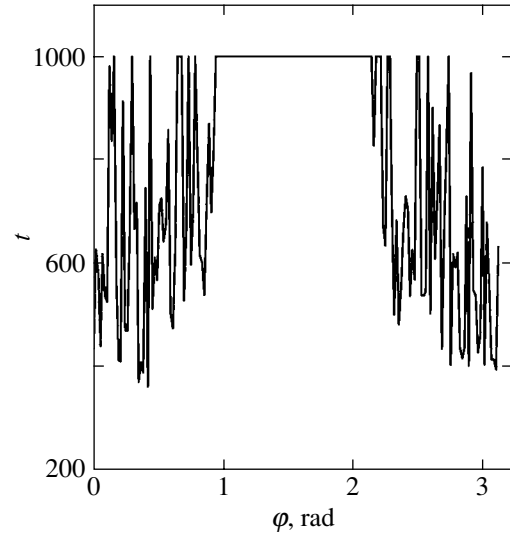
Fig. 2. Dependence of the time of disruption of the resonance structure on angle φ.

Thus, there is a region of initial conditions where the motions of three bodies remain bounded over a long time. These systems may be stable in the Lagrange sense.

Let us further analyze the vicinity of stable orbits. Trajectories that are stable in the Lagrange sense may be associated with a stable periodic orbit. To find this orbit, we introduce the following functional,

$$\Phi(\varphi) = \min_{t} \left[\frac{|\mathbf{r} - \mathbf{r}_0|}{|\mathbf{r}_0|} + \frac{|\dot{\mathbf{r}} - \dot{\mathbf{r}}_0|}{|\dot{\mathbf{r}}_0|} \right], \qquad (10)$$

where $t \in [0, t_{\text{crit}} = 1000\tau]$, $\mathbf{r}$ is the radius vector of the position of the triple system in the six-dimensional coordinate space; $\dot{\mathbf{r}}$ is the radius vector of the system in the six-dimensional velocity space, and $\mathbf{r}_0$ and $\dot{\mathbf{r}}_0$ are the corresponding radius vectors at the initial time instant.

The value of functional Φ for a periodic orbit in a fixed coordinate system must be equal to zero at $t = T$ (where T is the period). At the same time, computations show that functional does not reach close-to-zero values and that there is a lower boundary $\min_{\varphi} \Phi \approx 1.55$. It thus follows that close-to-periodic orbits are not realized in the fixed barycentric coordinate system.

Consider now the cases where the position of the center of mass of the inner binary shifts along the ordinate axis, or the orbit of the inner pair deforms (stretches or contracts):

$$y_1 = -\sin\varphi - h\sin\varphi, \qquad (11)$$
$$y_2 = \sin\varphi \mp h\sin\varphi.$$
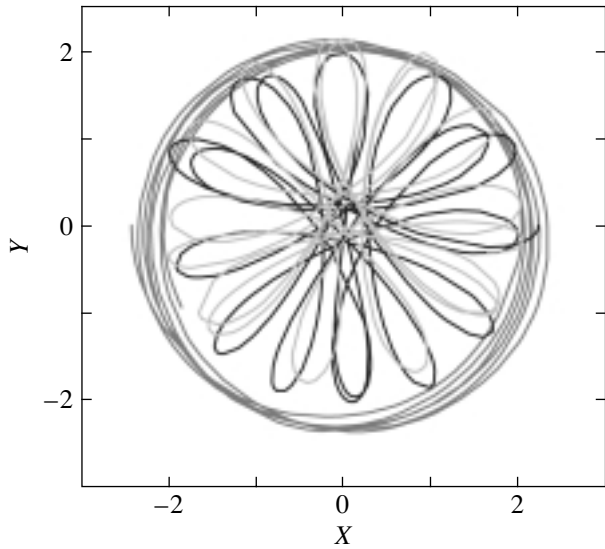
Fig. 3. Example of a trajectory with the inner pair compressed along the ordinate axis.
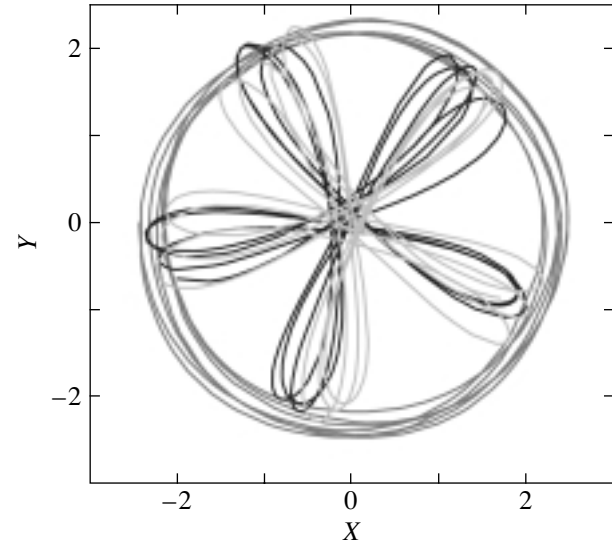
Fig. 4. Example of a five-leafed structure in the case of $\varphi = 120°$, $h = 0$.

The minus sign in the second of formulas (11) corresponds to a shift of the center of mass and plus sign, i.e., to an extension or contraction of the orbit. We vary h in the interval $[-0.2, 0.2]$. Note that introducing corrections of the form (11) puts the triple system out of the resonance. However, for small h the system remains in the neighborhood of the commensurability of the mean motions of the outer and inner binaries.

In the case of a small offset of center of mass the binary's center of mass the motion pattern remains more or less the same: the triple system remains in the neighborhood of the 2 : 1 resonance. In the case of strong stretching of the inner binary, the triple system leaves the resonance—bodies are ejected and the system disrupts in the long run. In the case of the compression of the binary, the structure of the triple system acquires hierarchical properties. Trajectories of motion of the components of the inner pair remain within the ring (Fig. 3). In the case of strong compression a gap can be seen between the upper boundary of this ring and the ring filled by the loops of the trajectory of the third body. There are "gaps" in the ring made up of the loops of the orbit of the inner binary. These gaps may be associated with higher-order resonances.

Consider now the offset of the center of mass of the inner pair along the horizontal axis. Other parameters of the inner binary are preserved. A shift of the center of mass of the pair toward the third body results in the system leaving the 2 : 1 resonance. As a result, the triple system becomes unstable. The structure of the triple system remains hierarchical if the center of mass of the inner binary is shifted, however, the 2 : 1

resonance no longer appears. Strong offsets result in the appearance of 3 : 1, 4 : 1 resonances, etc.

In some intervals of angle φ and quantities h, the trajectories of the components of the inner binary form structures made up of a system of leaves. Figures 4 and 5 show examples of such structures. In Fig. 4, the leaves of the trajectories of the inner-pair components fill the same five-leafed structure and the period during which each of the components of the inner binary covers the five-leafed structure is equal to two orbital periods of the outer binary. The process has a time lag equal to the period of the outer binary. With time, the loops of the trajectories of the inner pair begin to oscillate and their orbits gradually fill the entire region of the ring during $\sim$100 rotations of the third body. The central lines of adjacent leaves are inclined to one another at an angle of $\alpha = 360° : 5 = 72°$. The components of the inner pair pass in the vicinity of the center of mass of the triple system. As a consequence, the loops of the trajectories of these bodies densely fill the neighborhood of the center of mass.

Figure 5 shows an example of a six-leafed structure, which is realized in the case of the compression of the orbit of the inner pair. Unlike the previous case (Fig. 4), here the components of the inner pair move in two isolated three-leafed curves forming a central regular hexagon that is not filled by trajectory loops. Trajectory loops form kinks at the vortices of the three-leafed structures. The orbital period of the third body is equal to the period of motion of each component of the inner pair along the corresponding three-leafed curve. Trajectory loops gradually fill the region of the ring. This "diffusion" occurs during
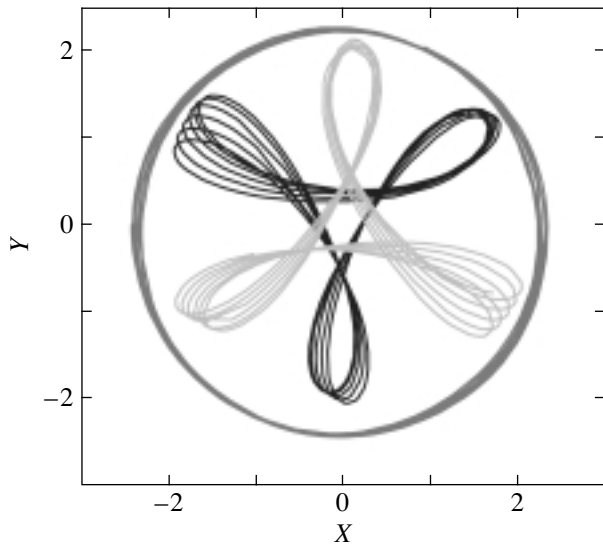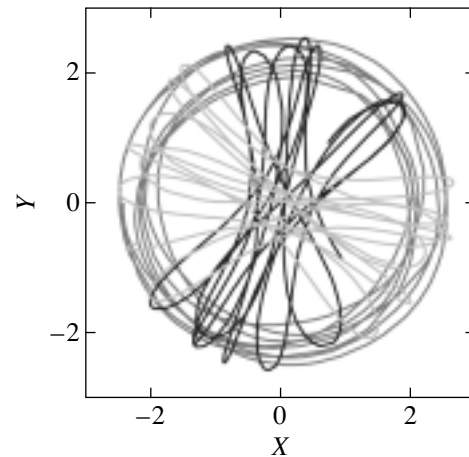
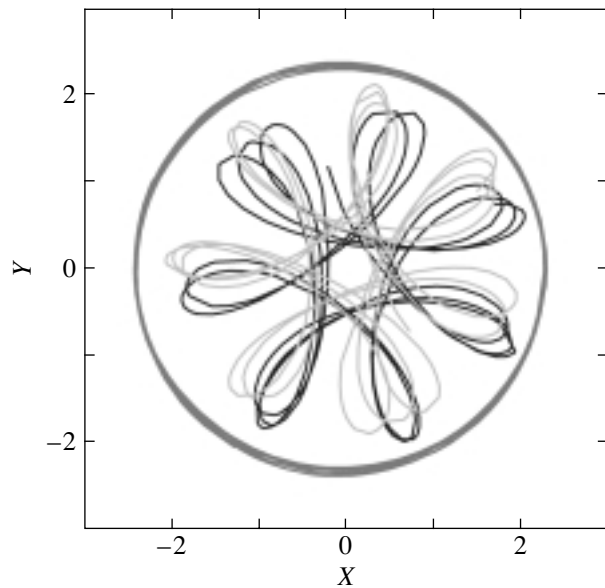**Fig. 5.** Example of a six-leafed structure in the case of $\varphi = 120°91$, $h = -0.09$.
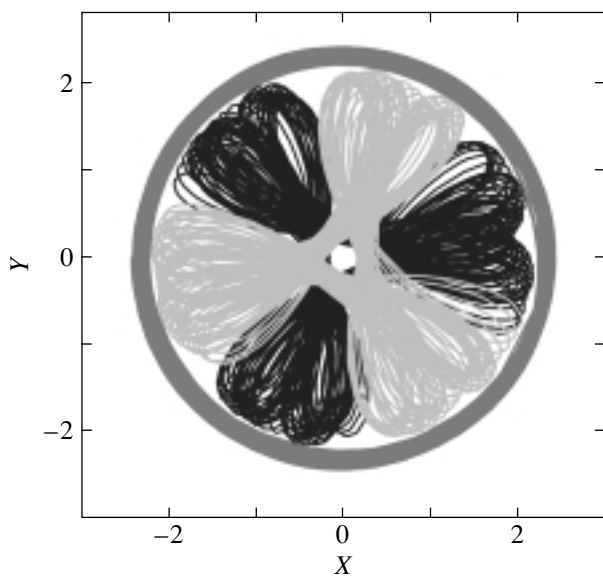
Fig. 6. Blurring of the six-leafed structure.

Fig. 7. Example of a cross-shaped structure at $\varphi = 120°91$, $h = 0.08$.

Fig. 8. Example of a seven-leafed structure at $\varphi = 60°91$, $h = -0.16$.

$\sim$100 rotations of the outer binary, like in the previous case (Fig. 6).

Figure 7 shows an example of a triple system where the loops of the trajectories of the bodies that make up the inner binary are initially (at the start of the evolution) located inside two mutually orthogonal curvilinear rectangles forming a cross-shaped structure. Subsequent evolution results in the "diffusion" of the trajectories of the components of the inner pair. As a result, the loops of the trajectories of these bodies fill the entire circle. The behavior of the system is similar to that of five- and six-leafed trajectories.

Figures 8 and 9 show examples of seven- and eight-leafed trajectories.

Note that in the case of an odd number n of leaves, the trajectory of each body of the inner pair passes through each leaf. In these cases, two three-leafed structures turned by an angle of $2\pi/n$ relative to one another are superimposed at the vertex of one of the leaves. In addition, the joining trajectories of the components of the inner pair are symmetrical with respect to the straight line passing through the center of mass of the triple system symmetrically with respect to the joining loops. Figure 10 shows an example of such a symmetrical matching of trajectories at $n = 19$.

In the case of even n each of the components of
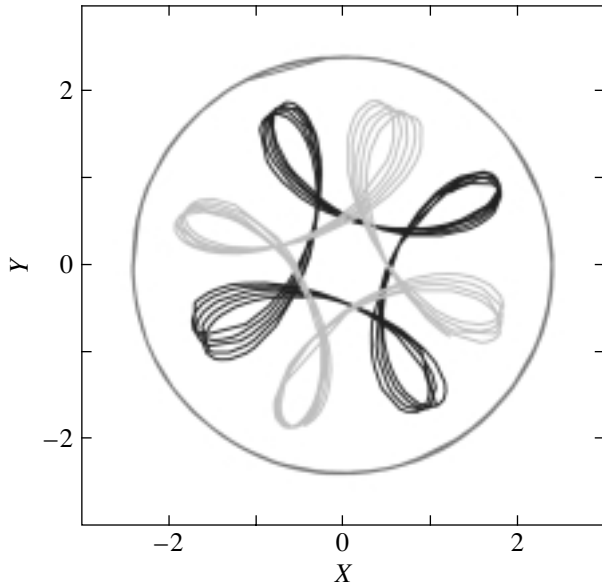
Fig. 9. Example of an eight-leafed structure at $\varphi = 90°$, $h = -0.17$.
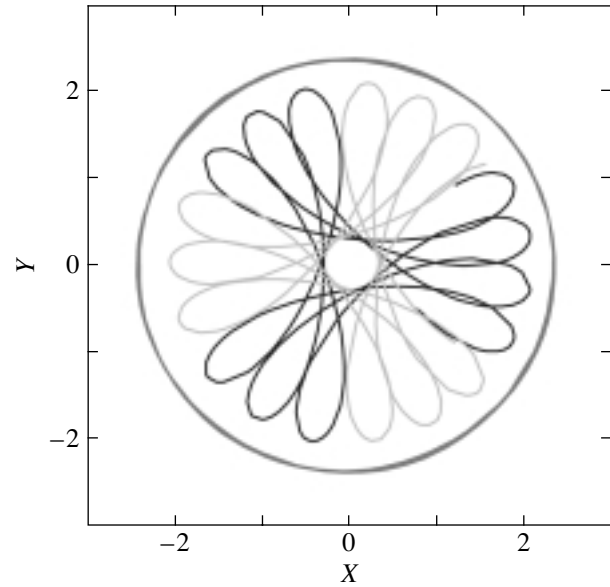
Fig. 10. Example of matching symmetrical structures at $n = 19$ (initial conditions: $\varphi = 90°$, $h = -0.09$).

the inner pair moves along its own system of leaves (the number of leaves in the system is equal to $n/2$). The angle between these systems of leaves is equal to $2\pi/n$.

DISCUSSION

A numerical analysis of the 2 : 1 resonance in the general equal-mass three-body problem in the case of retrograde motions of the inner and outer binaries showed that stable trajectories in the Lagrange sense may exist in the neighborhood of this resonance.

On the other hand, we found a number of trajectories having an n-leafed structure at the start of their evolution for a number of initial phases φ and relative compression or extension h of the inner pair. We found trajectories with $n = 3, 4, 5, 6, 7, 8$ inner-pair leaves. In this case, the distant body moves within the ring comprising the region of motions of the inner pair.

Strong compression of the inner pair or offset of its center of mass along the horizontal axis away from the distant body makes the triple system more hierarchical. In these cases, main resonances of higher orders appear: 3 : 1, 4 : 1, etc.

Note that, in the case of odd number n of loops of the trajectories of the components of the inner binary, the orbits of these bodies pass in the vicinity of the center of mass of the triple system (the center of the orbit of the external body). In the case of even n, the trajectories of the components of the inner binary do not pass in the vicinity of the center of mass of the triple system and there is a zone of "avoidance" in its vicinity in the form of a regular n-gon.

The trajectories of the inner-pair components fill the ring. The gap between the outer radius of this ring and the inner radius of the ring filled by the loops of the trajectory of the distant component increases with n.

Note that the n-leafed structures that we found are symmetrical with respect to the rotation by an angle of $2\pi/n$. If n is odd, the rotation by this angle results in an identical structure, because both components of the inner pair run through all n leaves. If n is even, a rotation by an angle of $2\pi/n$ swaps the components of this pair while leaving the shape of the trajectory unchanged.

We plan to search in the future for a reference periodic orbit, analyze its geometry and stability, and search for possible periodic orbits that generate trajectories with different numbers of leaves.

REFERENCES

1. S. J. Aarseth, *Gravitational N-Body Simulations* (Cambridge Univ. Press, Cambridge, 2003).

2. S. J. Aarseth and K. Zare, Celest. Mech. **10**, 185 (1974).

3. V. V. Beletskiǐ, *Sketchs on the Motion of Cosmic Bodies* (Nauka, Moscow, 1972) [in Russian].

4. R. Brasser, D. C. Heggie, and S. Mikkola, Celest. Mech. Dyn. Astron. **88**, 123 (2004).

5. D. Brower, Astron. J. **68**, 152 (1963).

6. A. D. Bryuno, Mat. Sbornik **83** (125), 273 (1970).

7. R. Bulirsch and J. Stoer, Num. Math. **8**, 1 (1966).

8. S. Ferraz Mellu, *Dynamics of Jupiter's Galilean Satellites* (Mir, Moscow, 1983) [in Russian].

9. K. Gozdziewsky and A. J. Maciejewski, Astrophys. J. **563**, L81 (2001).

10. N. N. Gor'kavyǐ and A. M. Fridman, *Physics of Planetary Rings* (Nauka, Moscow, 1994) [in Russian].

11. E. A. Grebenikov and Yu. A. Ryabov, *New Qualitative Methods in Celestial Mechanics* (Nauka, Moscow, 1971) [in Russian].

12. G. W. Marcy, R. P. Butler, D. A. Fisher, *et al.*, Astrophys. J. **556**, 296 (2001).

13. G. W. Marcy, R. P. Butler, D. A. Fisher, *et al.*, Astrophys. J. **581**, 1375 (2002).

14. C. Moore, Phys. Rev. Lett. **70**, 3675 (1993).

15. N. P. Pitjev and L. L. Sokolov, *Abstracts of the Conf. "Order and Chaos in Stellar and Planetary Systems," St. Petersburg, 2003*, p. 46.

16. N. C. Santos, G. Israelian, and M. Mayor, Astron. Astrophys. **363**, 228 (2000).

17. K. Simo, *Up-to-date Chaos and Nonlinearity Problems*, Ed. by A. V. Borisov and A. A. Kilin (Inst. Comp. Res., Izhevsk, 2002), p. 252 [in Russian].

18. R. J. Vanderbei, astro-ph/0303153 (2003).

Translated by A. Dambis