Astronomy Letters, Vol. 31, No. 8, 2005, pp. 501–514. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 563–576.

H I Observations of Flat Galaxies

S. N. Mitronova[1*], W. K. Huchtmeier[2], I. D. Karachentsev[1], V. E. Karachentseva[3], and Yu. N. Kudrya[3]

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

[2] *Max Planck Institut für Radioastronomie, Auf dem Hügel 69, D-53121 Bonn, Germany*

[3] *Astronomical Observatory, Kiev National University, Observatorna ul., Kiev, 304053 Ukraine*

Received March 3, 2005

Abstract—We present the H I observations of 94 flat spiral galaxies from RFGC (*Revised Flat Galaxy Catalog*) and 14 galaxies from 2MFGC (the 2MASS selected *Flat Galaxy Catalog*) performed with the 100-m radio telescope in Effelsberg (Germany). H I fluxes, heliocentric radial velocities, and H I line widths are given for 65 detected galaxies. We present a mosaic of H I profiles. We calculated some of the global parameters of the galaxies and analyzed the linear correlations between them. The ratios of the total (indicative) masses of the galaxies to their luminosities lie within the range 0.4–8.2 with a mean of 3.8 ($M_\odot/L_\odot$), and the mean mass fraction of neutral hydrogen is 13%. Upper limits are given for the radio fluxes from 43 undetected galaxies. © *2005 Pleiades Publishing, Inc.*

Key words: *galaxies, radio sources.*

INTRODUCTION

One of the central problems in extragalactic astronomy is to study the collective motions of galaxies by analyzing their peculiar velocities $V_{\rm pec} = V_{3K} - Hr$. Here, V_{3K} is the measured radial velocity of the galaxy reduced to the frame of the 3K microwave background radiation (Kogut *et al.* 1993), and Hr is the distance to the galaxy (in km s^{-1}) determined independently of its radial velocity. The main method of determining Hr for spiral galaxies is based on the Tully–Fisher (1977) relation between the absolute parameters of the galaxy (its luminosity and linear diameter) and the width of the 21-cm H I line. Karachentsev (1989) showed that the late-type edge-on spiral galaxies are appropriate objects for investigating the large-scale streams for several reasons: (1) the flat, disklike galaxies have a simple structure; (2) applying the simple selection criterion based on the apparent axial ratio $a/b > 7$ to them yields a morphologically homogeneous sample; (3) the detection probability of such galaxies in the 21-cm line is very high; and (4) since the flat galaxies are located mostly outside groups and clusters, their structure remains undistorted and they are not affected by large virial motions.

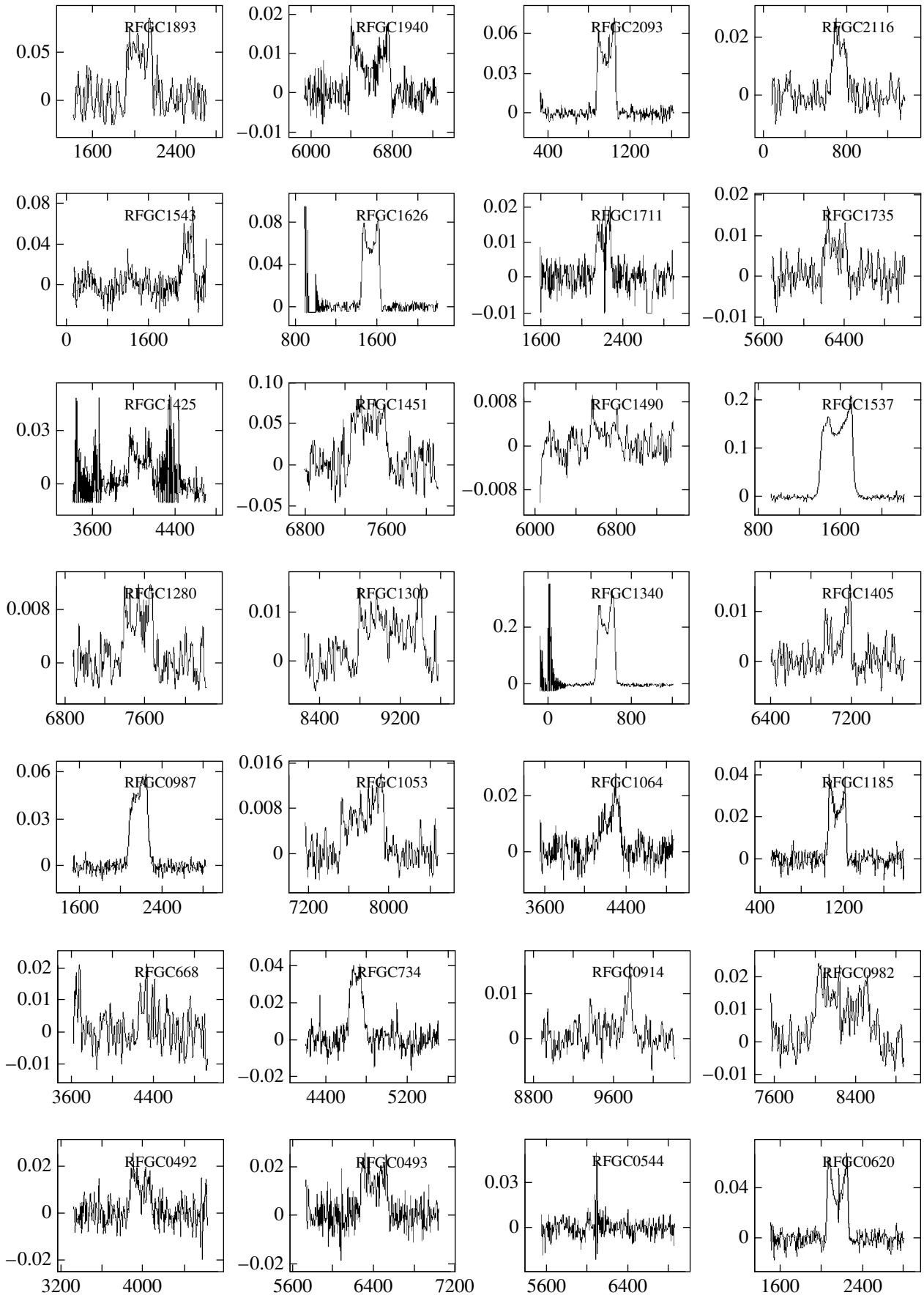
To compile our catalog of flat edge-on galaxies, we conducted an all-sky survey using blue and red POSS-I and ESO/SERC maps. FGC (*Flat Galaxy Catalog*; Karachentsev *et al.* 1993) includes galaxies with apparent axial ratios $a/b > 7$ and angular diameters $a > 0'.6$. The updated and supplemented version of this catalog, RFGC (*Revised Flat Galaxy Catalog*; Karachentsev *et al.* 1999a), contains 4236 galaxies and is one of the best samples for studying the large-scale cosmic streams of galaxies (Karachentsev *et al.* 2000a; Kudrya *et al.* 2003), since it is highly homogeneous and has the required completeness (Kudrya *et al.* 1996, 1997; Fel'dman *et al.* 2003).

Extensive high-accuracy measurements of the radial velocities ($V_{\rm h}$) and H I-line widths (W) are needed to use the Tully–Fisher (TF) relation. For this purpose, flat galaxies were observed with large radio telescopes: the 305-m radio telescope in Arecibo (Giovanelli *et al.* 1997) and the 100-m radio telescope in Effelsberg (Huchtmeier *et al.* 2005). Makarov *et al.* (2001) measured the rotation curves of 300 flat galaxies with the 6-m telescope at the Special Astrophysical Observatory of the Russian Academy of Sciences. Based on original observations and published data, Karachentsev *et al.* (2000b) compiled a list of peculiar velocities for 1327 RFGC galaxies.

The published homogeneous and complete near-infrared all-sky survey, 2MASS (the Two Micron All-Sky Survey; Skrutskie *et al.* 1997), opened up

*E-mail: mit@sao.ru

1063-7737/05/3108-0501$26.00


RFGC1893
RFGC1940
RFGC2093
RFGC2116
RFGC1543
RFGC1626
RFGC1711
RFGC1735
RFGC1425
RFGC1451
RFGC1490
RFGC1537
RFGC1280
RFGC1300
RFGC1340
RFGC1405
RFGC0987
RFGC1053
RFGC1064
RFGC1185
RFGC668
RFGC734
RFGC0914
RFGC0982
RFGC0492
RFGC0493
RFGC0544
RFGC0620

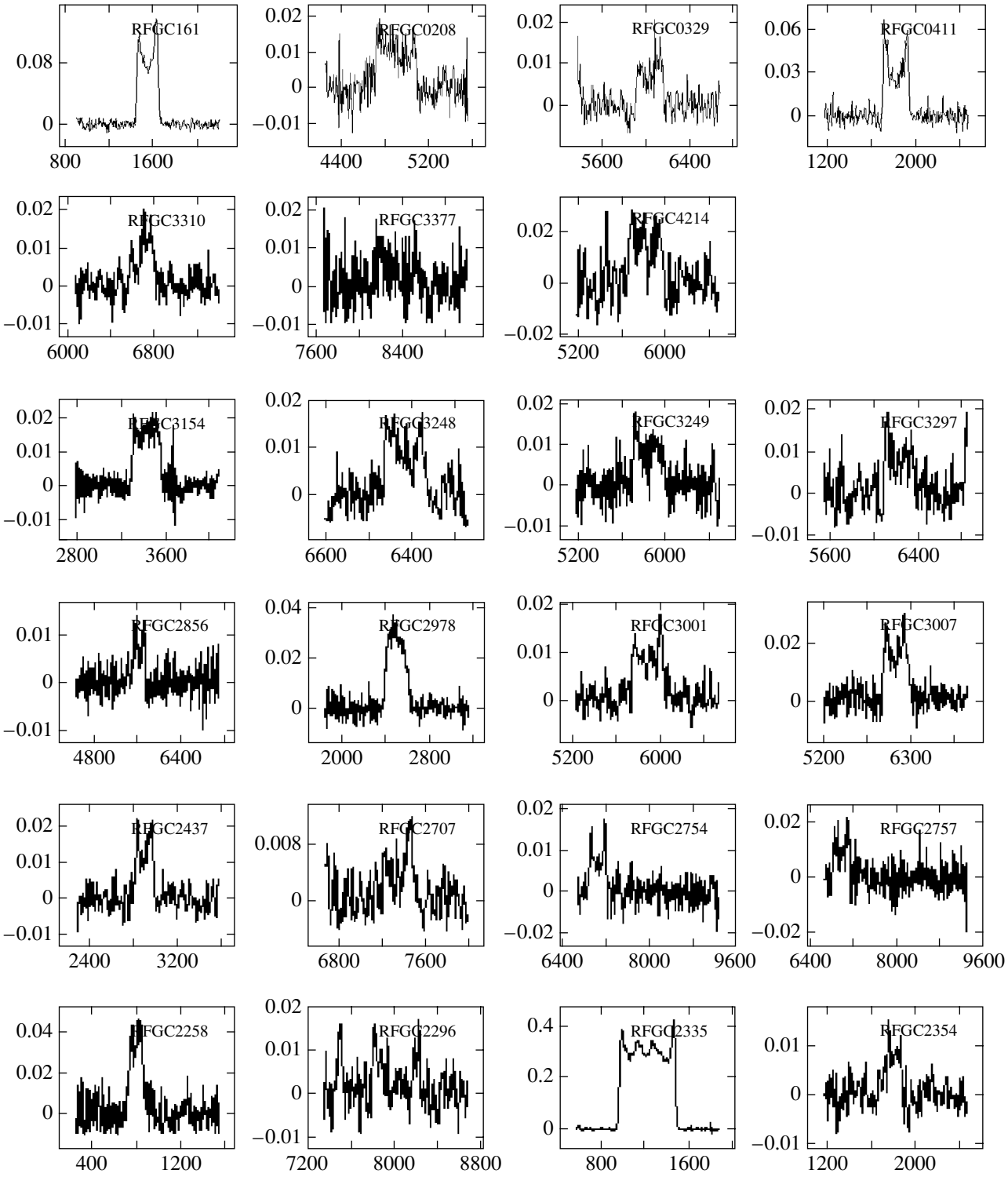


Fig. 1. HI profiles for 51 RFGC galaxies from Table 1 (see pp. 502–503).

new opportunities for studying the collective motions of galaxies based on the near-infrared TF relation (Karachentsev *et al.* 2002; Kudrya *et al.* 2003). 2MASS J-, H-, and K-band photometry is available for 71% of the RFGC galaxies, but the radial velocities and HI-line widths were reliably measured only for 25% of them. A program of HI observations of RFGC galaxies with the 100-m radio tele-

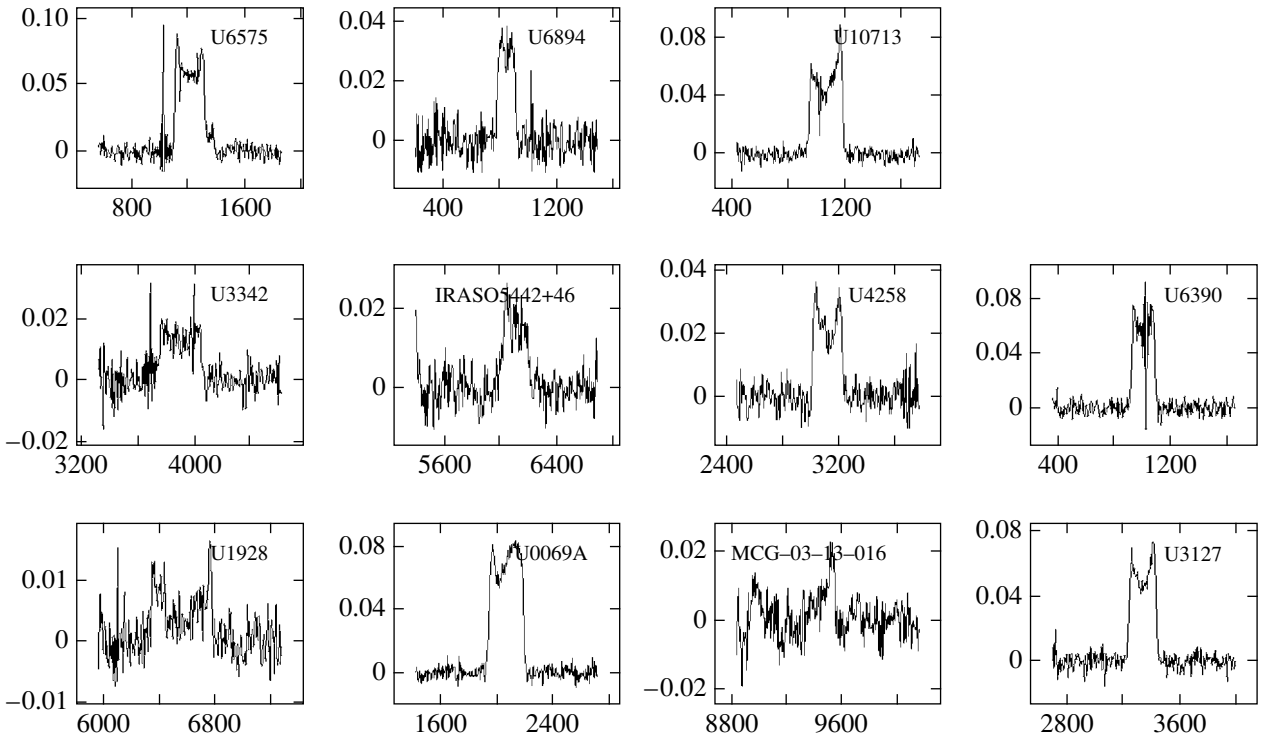


Fig. 2. H I profiles for 11 2MFGC galaxies from Table 2.

scope in Effelsberg was initiated in 2001. Huchtmeier *et al.* (2005) published the first results of observations of 268 galaxies.

Analyzing the properties of RFGC galaxies in the near infrared (2MASS) and comparing them with optical data allowed the selection criteria to be developed for compiling a new catalog of flat galaxies based on 2MASS. The 2MFGC all-sky catalog of disklike galaxies (Mitronova *et al.* 2004) contains IR photometry and LEDA and NED identifications for 18 020 objects with XSC 2MASS axial ratios $a/b \geq 3$.

In this paper, we present the observations of 94 RFGC galaxies and the test observations of 14 2MFGC galaxies performed with a 100-m radio telescope in the fall of 2004.

H I OBSERVATIONS

The H I observations of RFGC galaxies have been performed with the 100-m radio telescope in Effelsberg (Germany) since October 2001. Using the optical and infrared parameters of flat galaxies, we estimated their expected radial velocities and selected objects in the working range of the telescope $V_h \leq 9500$ km s^{-1} for our observations. We also included in our observational application galaxies for which the radial velocities were known from optical observations with an accuracy lower than 30 km s^{-1} and no H I line width measurements were available. We also included in our list galaxies for which it was necessary to improve V_h and (or) W, since they deviated by more than 3σ from the regression line in the TF relation or the magnitudes of their peculiar velocities were larger than 3000 km s^{-1}.

The spectroscopic observations with the 100-m radio telescope were performed in the full energy mode (ON-OFF), combining the measurements in the source's field with those in the comparison field whose location was earlier than the galaxy's location by 5 min in right ascension. The beam FWHM of the telescope at 21 cm is $9'.3$. The temperature of the system consisting of two receivers is $30K$. The 1024-channel autocorrelator was divided into four 256-channel bands shifted in frequency by 11 MHz. To cover the entire velocity range from 250 to 9500 km s^{-1}, we used 12.5-MHz-wide bands. In this case, the resulting resolution per channel is 10.4 km s^{-1}. For galaxies with known radial velocities, we used 6.25-MHz-wide bands, which provided a resolution of 5.2 km s^{-1} per channel. We used weighted (or equivalent) smoothing for most

galaxies to improve the signal-to-noise ratio. From one to five scans were made for each galaxy; the total accumulation time was about 2 h per galaxy.

RESULTS OF THE OBSERVATIONS

Figures 1 and 2 show the HI profiles for the detected RFGC and 2MFGC galaxies. No HI line profiles are shown for two RFGC galaxies (2553 and 4160) with complex spectra.

Table 1 contains the results of our radio observations of 94 RFGC galaxies as well as cataloged (optical) data. The columns of this table give the following: (1) the RFGC galaxy number; (2) the J2000.0 equatorial coordinates; (3) the major and minor diameters ($a \times b$) in arcmin corresponding to the B-band 25 mag arcsec^{-2} isophote; (4) the total B-band magnitudes; (5) the morphological type in the Hubble system, where Sb = 3, Sc = 5, and Sd = 7; (6) the HI-line flux in Jy km s^{-1} corrected for the ratio of the galaxy's angular diameter to the telescope's aperture; (7) the maximum flux and the rms noise error in mJy; (8) the mean heliocentric radial velocity and its error in km s^{-1}; and (9–11) the HI line widths at 50, 25, and 20% of the peak flux level in km s^{-1}. The colons (:) in the table mark unreliable measurements with low signal-to-noise ratios and a complex structure of the HI profiles.

Table 2 gives data for 2MFGC galaxies. The content of its columns corresponds to Table 1 with the following differences: (1) the galaxy names in known catalogs taken from the NED electronic database, (3) the major and minor diameters ($a \times b$) in arcmin taken from the LEDA electronic database, and (4) the total B-band magnitudes taken from the LEDA database.

For 40 RFGC and three 2MFGC galaxies, we failed to find the HI line in the range of radial velocities under study. Tables 1 and 2 give only the rms noise errors for these galaxies. The undetected galaxies probably lie outside the working spectral range, although it may well be that the HI fluxes from some of the nearby early-type galaxies could be below the instrumental detection threshold.

GLOBAL OPTICAL AND HI PARAMETERS OF GALAXIES

We calculated the global characteristics of 54 detected RFGC galaxies using the radio observations and optical parameters from Table 1. As the distance indicator, we used the radial velocity of the galaxy reduced to the centroid of the Local group,

$$V_{\rm LG} = V_{\rm h} + V_{\rm a}[\cos b \cos(b_{\rm a}) \cos(l - l_{\rm a}) + \sin b \sin(b_{\rm a})], \quad (1)$$

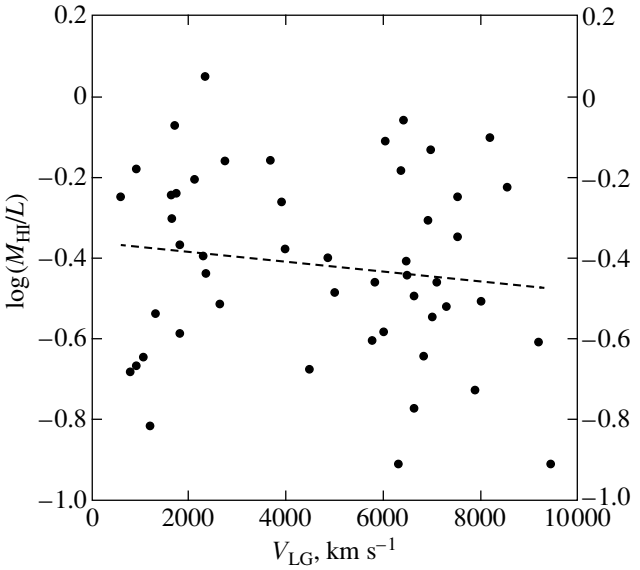


Fig. 3. HI mass-to-luminosity ratio (in solar units) versus radial velocity (in km s^{-1}).

where l and b are the Galactic coordinates and $V_{\rm h}$ is the measured heliocentric radial velocity; the apex parameters, $V_{\rm a} = 316$ km s^{-1}, $l_{\rm a} = 93°$, and $b_{\rm a} = -4°$, were taken from the paper by Karachentsev and Makarov (1996). We calculated the HI mass as

$$\log(M_{\rm HI}/M_\odot) = \log F + 2\log(V_{\rm LG}/H_0) + 5.37, \quad (2)$$

where the Hubble constant was assumed to be $H_0 = 75$ km s^{-1}Mpc^{-1}, and F is the detected 21-cm flux in Jy km s^{-1} corrected for the ratio of the galaxy's angular diameter to the telescope's aperture. We calculated the indicative mass of the galaxy within the standard optical diameter from the relation

$$\log(M_{25}/M_\odot) = 2\log(W_c) + \log(a_c) + \log(V_{\rm LG}/H_0) + 3.92, \quad (3)$$

where $W_{\rm c}$ is the hydrogen line width corrected for relativistic broadening, $W_c = W_{50}/(1 + V_{\rm h}/c)$. The angular diameter a_c was reduced to the 25 mag arcsec^{-2} isophote (Kudrya *et al.* 1997) and corrected for the Galactic absorption and the galaxy's tilt ($\log a_c = \log a + 0.09A_B - 0.2\log(a/b)$). To calculate the luminosity, we use the relation

$$\log(L/L_\odot) = 2\log(V_{\rm LG}/H_0) - 0.4(B_t - \Delta B) + 12.16, \quad (4)$$

where the total magnitude B_t was corrected for the absorption A_B (Schlegel *et al.* 1998) and the galaxy's tilt $\Delta B = A_B + 1.2\log(a/b)$.

Table 3 gives the calculated global parameters. Its columns contain the following: (1) the RFGC galaxy

Table 1. Results of the H I observations of RFGC galaxies with the 100-m radio telescope

RFGC	RA(2000)DEC	$a \times b$	B_t	T	F	S_{max}	V_h	W_{50}	W_{25}	W_{20}
1	2	3	4	5	6	7	8	9	10	11
161	004215.1−180940	3.36 × 0.30	14.5	7	19.1	139 ± 4	1553 ± 2	192	205	208
208	005332.6+025527	1.71 × 0.21	15.4	6	4.2	± 4	4843 ± 3	250		
280	011406.7+380723	1.16 × 0.16	16.0	5		± 2				
329	012737.1+510851	0.87 × 0.11	18.1	6	2.1	23 ± 2	6096 ± 11	129		
411	015616.1−225404	2.23 × 0.27	14.9	3	8.5	68 ± 9	1828 ± 1	229	240	
458	020715.4+461520	1.12 × 0.13	16.4	7		± 5				
492	021732.5−113108	1.22 × 0.17	15.8	8	3.6	32 ± 4	3957 ± 4	210		
493	021735.9−214554	0.95 × 0.09	16.6	6	4.1	26 ± 4	6423 ± 4	264		
544	023052.8+432100	1.27 × 0.11	16.2	4	0.5 :	± 3	6100:			
620	025426.6+423900	2.43 × 0.15	15.4	7	8.8	68 ± 6	2167 ± 2	203	213	
668	030854.8+703349	1.68 × 0.24	17.0	3	1.3	37 ± 9	4330 ± 5			
734	032846.7+363323	1.79 × 0.20	16.5	4	5.4	41 ± 6	4701 ± 4	129	157	
860	043018.5+884616	1.10 × 0.13	16.3	6		± 4				
914	045213.2−182335	1.53 × 0.16	15.7	6	1.4	18 ± 4	9573 ± 5	415		
974	052320.9+854023	0.83 × 0.09	16.9	5		± 6				
982	052644.4−191236	1.39 × 0.19	15.8	5	7.6	30 ± 6	8336 ± 3	512		
987	053012.0+555216	1.57 × 0.21	18.0	9	9.5	60 ± 3	2182 ± 4	187	210	214
1053	061021.8+504706	1.38 × 0.17	15.8	4	3.3	27 ± 4	7775 ± 5	420		
1063	061318.7+530643	1.19 × 0.16	16.2	7		± 4				
1064	061342.3+810425	2.04 × 0.24	15.6	4	3.2	28 ± 4	4266 ± 7	176		
1119	063900.2+572258	1.00 × 0.12	16.4	5		± 5				
1185	071554.5+675902	2.02 × 0.22	15.3	5	4.7	40 ± 3	1153 ± 3	172	183	
1207	072225.0+491642	0.84 × 0.11	16.4	9		± 4				
1280	075507.0+425728	0.78 × 0.09	16.8	6	2.2	15 ± 4	7490 ± 5	280		
1282	075524.7+561002	0.96 × 0.11	16.7	5		± 4				
1300	080009.6+562154	2.33 × 0.22	15.1	5	5.0	20 ± 3	9095 ± 5	629		
1340	081359.5+454434	5.17 × 0.58	13.6	6	43.9	322 ± 3	559 ± 1	170	189	194
1351	081709.6+560249	0.68 × 0.09	16.9	6		± 4				
1405	083812.0+455255	0.87 × 0.10	16.5	6	1.7	16 ± 3	7070 ± 7	264		
1425	084421.6+093215	1.23 × 0.10	16.4	6	4.0	26 ± 6	4064 ± 5	226		
1451	085419.0+542728	0.90 × 0.11	16.5	6	2.1	10 ± 3	7449 ± 5	350		
1490	090738.9+283808	1.10 × 0.10	16.4	7	1.0	9 ± 2	6682 ± 14	265		
1524	091831.0+493244	1.18 × 0.15	16.0	6		± 11				
1537	092145.1+641528	4.31 × 0.53	13.9	5	49.8	209 ± 4	1574 ± 2	308	336	341
1543	092311.8−265631	1.56 × 0.17	15.8	5	8 :	70 ± 11	2410 ± 3	193		
1566	093140.8−160231	1.84 × 0.21	15.3	7		± 1				
1583	093555.0+480848	1.57 × 0.22	15.6	4		± 13				

Table 1. (Contd.)

1	2	3	4	5	6	7	8	9	10	11
1612	094402.4+682212	0.74 × 0.09	16.8	6		± 5				
1626	094650.4+794839	2.49 × 0.22	15.1	6	11.3	86 ± 3	1540 ± 2	170	183	186
1711	100439.1+602759	0.88 × 0.10	16.6	6	1.7	16 ± 2	2216 ± 10	144		
1735	101001.0+715221	0.88 × 0.10	16.6	6	1.8	19 ± 3	6327 ± 5	189		
1893	105124.0−195324	4.26 × 0.48	14.0	6	13.4	86 ± 13	2067 ± 1	229		
1897	105236.0+394814	0.68 × 0.08	17.0	5		± 5				
1940	110157.1+470540	1.88 × 0.15	15.8	5	3.4	19 ± 3	6583 ± 5	378		
1988	111327.8+072553	0.73 × 0.09	16.7	6		± 8				
2056	113057.6−040507	0.92 × 0.10	16.7	7		± 5				
2093	114227.4+513551	3.70 × 0.45	14.1	6	9.3	73 ± 3	975 ± 2	178	190	193
2116	114836.0+434316	1.87 × 0.20	15.4	6	2.8	27 ± 4	735 ± 5	115		
2258	122016.6+480811	1.20 × 0.17	15.8	9	1.7	± 3	839 ± 4	122		
2296	122859.3+285143	1.12 × 0.15	16.1	5	2.0	17 ± 3	8024 ± 9	435		
2298	122908.6+575454	1.34 × 0.16	15.9	7	5.2	46 ± 7	799 ± 4	126	139	143
2335	123621.1+255906	15.90 × 1.85	10.6	4	161	420 ± 5	1226 ± 1	506	521	524
2354	124128.8−031514	0.67 × 0.09	17.1	8	1.3	15 ± 3	1803 ± 6	125		
2390	125105.0−082623	1.12 × 0.16	16.1	6		± 6				
2437	130106.7−032235	0.83 × 0.11	16.8	7	2.4	23 ± 5	2900 ± 6	156	175	
2443	130207.9+584159	3.92 × 0.32	14.5	6		± 3				
2460	130514.9−002230	1.09 × 0.13	16.3	7		± 3				
2463	130548.0+462743	1.37 × 0.19	15.7	4		± 3				
2553	132409.6−175407	1.25 × 0.10	16.4	7	1.9	26 ± 6	6994 ± 15	200		
2559	132525.7+044600	1.01 × 0.10	16.7	5		± 8				
2574	132820.2−114703	1.85 × 0.24	15.4	4		± 3				
2592	133341.5−111418	0.99 × 0.07	17.0	7		± 4				
2626	134043.3−173314	0.65 × 0.07	17.2	8		± 7				
2707	140425.4−002725	0.90 × 0.10	16.6	7	1.5	13 ± 2	7380 ± 6	301		
2713	140621.6−054313	1.28 × 0.17	15.8	4		± 4				
2737	141312.0−072648	0.67 × 0.08	17.3	9		± 6				
2754	141741.0−052748	1.29 × 0.17	15.9	5	6.1	36 ± 9	7057 ± 5	308		
2757	141833.6−050913	1.15 × 0.10	16.4	7	3.2	22 ± 5	7010 ± 11	229		
2786	143007.2−062543	1.46 × 0.17	15.9	4		± 4				
2856	144457.6+405234	1.12 × 0.16	15.9	6	1.8	14 ± 3	5640 ± 12	216		
2902	150032.6+491027	0.76 × 0.10	16.7	4		± 4				
2915	150438.9−181337	0.84 × 0.09	16.9	7		± 7				
2943	151543.2−002524	0.67 × 0.09	17.0	7		± 5				
2944	151543.2−102750	0.83 × 0.09	16.8	7		± 6				

Table 1. (Contd.)

1	2	3	4	5	6	7	8	9	10	11
2955	151836.7−011101	1.04 × 0.11	16.8	6		± 6				
2978	152628.8+411731	2.26 × 0.24	15.1	5	5.6	37 ± 5	2481 ± 1	189		
3001	153443.0+082003	1.48 × 0.12	16.3	6	2.6	20 ± 3	5803 ± 4	278		
3007	153654.0+511720	1.10 × 0.11	16.4	5	4.2	30 ± 4	5847 ± 5	209		
3154	163231.2+674449	1.12 × 0.12	16.4	4	4.0	21 ± 3	3423 ± 4	255		
3202	165140.3+532422	0.68 × 0.09	16.8	8		± 4				
3210	165450.4+702617	1.37 × 0.19	15.7	5		± 3				
3228	170247.8+444731	0.65 × 0.08	17.3	8		± 4				
3230	170338.9+454832	0.83 × 0.10	16.9	7		± 5				
3248	171145.3−254451	0.98 × 0.10	16.7	5	3.4	17 ± 4	6330 ± 6	353		
3249	171149.2+473937	0.85 × 0.10	16.7	6	2.5	15 ± 2	8320 ± 15	248		
3297	173721.6+602537	1.12 × 0.11	16.4	5	2.5	22 ± 12	6220 ± 6	254		
3310	174810.8+530908	0.95 × 0.10	16.5	6	1.5	20 ± 3	6744 ± 9	241		
3333	180645.1+874835	1.77 × 0.21	15.6	4		± 8				
3377	183312.7+525655	1.34 × 0.11	16.3	4	4.7	25 ± 6	8418:			
3680	210236.0−134753	1.68 × 0.20	15.6	3		± 2				
3870	220604.3−261107	1.27 × 0.13	16.1	6		± 2				
4160	233900.0+493531	0.83 × 0.10	16.7	6	1.2	14 ± 3	9319 ± 10			
4181	234438.4−273936	1.14 × 0.16	16.1	5		± 4				
4214	235321.6+860141	0.96 × 0.09	16.7	6	2.2	12 ± 3	5739 ± 10	361		

Table 2. Results of the H I observations of 2MFGC galaxies with the 100-m radio telescope

Name	RA(2000)DEC	$a \times b$	B_t	T	F	S_{max}	V_h	W_{50}	W_{25}	W_{20}
1	2	3	4	5	6	7	8	9	10	11
UGC256	002656.6+500150.7	1.55 × 0.29	15.28	4	1.5	11 ± 3				
UGC1928	022811.2+435109.7	1.36 × 0.30	16.47	5	2.6	16 ± 3	6561 ± 7	432		
UGC69A	032702.2+725038.3	2.10 × 0.49		3	17.7	83 ± 4	2074 ± 3	243	257	262
UGC2736	032627.6+403028.5	1.57 × 0.31	14.64	2		± 5				
MCG-03-13-016	045013.7−171557.8	1.83 × 0.53	14.25	4	5.2	22 ± 3	9248 ± 8	608		
UGC3127	044025.9−020112.6	2.10 × 0.42	14.94	7	11.0	74 ± 5	3345 ± 4	192	205	209
UGC3342	054429.7+691756.3	1.74 × 0.39	14.99	5	4.4	31 ± 4	3890 ± 4	294		
IRAS05442+46	054809.6+461531.0	0.77 × 0.60	16.23	5	3.2	27 ± 4	6083 ± 7	190		
UGC4258	081047.6+465445.7	1.43 × 0.38	18.16	6	5.3	37 ± 5	3124 ± 4	208	222	
UGC6390	112242.0+640358.7	2.07 × 0.30	19.22	6	9.3	91 ± 4	982 ± 3	163	177	180
UGC6575	113626.5+581122.0	1.88 × 0.44	14.57	5	14.6	89 ± 3	1216 ± 5	206	222	
UGC6894	115524.4+543926.3	1.45 × 0.25	15.28	5	4.0	39 ± 5	850 ± 4	128		
NGC6244	164803.9+621201.6	1.54 × 0.32	14.45	1		± 4				
UGC10713	170433.9+722647.5	1.83 × 0.32	13.97	3	12.7	89 ± 4	1072 ± 2	230	242	244

Table 3. Optical and H I parameters of galaxies

RFGC	V_{LG}	W_c	$\log(M_{HI}/M_\odot)$	$\log(M_{25}/M_\odot)$	$\log(L/L_\odot)$	M_{HI}/L	M_{25}/L	M_{HI}/M_{25}
1	2	3	4	5	6	7	8	9
161	1623	191	9.3	10.1	9.6	0.57	3.76	0.15
208	4998	246	9.6	10.6	10.1	0.33	2.89	0.11
329	6352	126	9.5	9.9	10.3	0.20	0.48	0.42
411	1824	228	9.1	10.2	9.4	0.43	5.68	0.08
492	3985	207	9.4	10.2	9.8	0.42	2.83	0.15
493	6407	258	9.8	10.5	9.9	0.87	3.55	0.25
620	2357	202	9.3	10.2	9.7	0.36	2.89	0.13
734	4850	127	9.7	10.1	10.5	0.16	0.39	0.41
914	9438	402	9.7	11.2	10.6	0.12	4.06	0.03
982	8170	498	10.3	11.3	10.4	0.79	8.23	0.10
987	2321	186	9.3	10.2	10.3	0.11	0.79	0.14
1053	7874	409	9.9	11.2	10.7	0.19	3.48	0.05
1064	4476	174	9.4	10.3	10.2	0.16	1.31	0.12
1185	1305	171	8.5	9.8	9.1	0.29	5.00	0.06
1280	7509	273	9.7	10.5	10.0	0.56	3.56	0.16
1300	9182	610	10.2	11.8	10.9	0.25	8.02	0.03
1340	588	170	8.8	9.8	9.0	0.57	5.85	0.10
1405	7095	258	9.6	10.5	10.0	0.35	2.96	0.12
1425	3898	223	9.4	10.2	9.7	0.55	3.65	0.15
1451	7517	342	9.7	10.8	10.0	0.45	5.35	0.08
1490	6608	259	9.3	10.5	10.0	0.17	3.17	0.05
1537	1692	306	9.8	10.7	9.8	0.85	7.54	0.11
1543	2114	191	9.2	10.0	9.4	0.62	3.95	0.16
1626	1732	169	9.2	9.9	9.4	0.58	3.50	0.16
1711	2314	143	8.6	9.5	9.0	0.40	3.22	0.13
1735	6483	185	9.5	10.2	9.9	0.36	1.69	0.21
1893	1796	227	9.3	10.5	9.8	0.26	4.23	0.06
1940	6617	370	9.8	11.1	10.3	0.32	5.97	0.05
2093	1043	177	8.6	10.0	9.3	0.23	4.81	0.05
2116	765	115	7.8	9.1	8.5	0.21	4.14	0.05
2258	903	122	7.8	9.1	8.4	0.22	4.53	0.05
2296	7997	424	9.7	11.1	10.2	0.31	7.09	0.04
2298	912	126	8.3	9.2	8.4	0.67	5.21	0.13
2335	1189	504	10.0	11.5	10.8	0.15	5.65	0.03
2354	1639	124	8.2	9.1	8.5	0.50	4.42	0.11
2437	2749	155	8.9	9.6	9.0	0.69	3.79	0.18
2553	6807	195	9.6	10.4	10.2	0.23	1.47	0.15
2707	7290	294	9.5	10.6	10.0	0.30	3.85	0.08
2754	6959	301	10.1	10.8	10.2	0.74	3.73	0.20
2757	6914	224	9.8	10.4	10.1	0.49	2.18	0.23
2856	5753	212	9.4	10.3	10.0	0.25	2.21	0.11
2978	2625	187	9.2	10.2	9.7	0.31	2.88	0.11
3001	5831	273	9.6	10.7	10.0	0.35	4.24	0.08
3007	6027	205	9.8	10.3	9.9	0.77	2.41	0.32
3154	3657	252	9.3	10.3	9.5	0.70	5.96	0.12
3248	6302	346	9.7	10.9	10.7	0.11	1.77	0.06
3249	8548	241	9.9	10.5	10.1	0.59	2.47	0.24
3297	6474	249	9.6	10.5	10.0	0.39	2.94	0.13
3310	6998	236	9.5	10.4	10.0	0.29	2.52	0.11
4214	5985	354	9.5	10.8	10.1	0.26	4.70	0.06

Table 4. Characteristics of the distributions of flat galaxies in integrated parameters

Parameter	Mean	Standard deviation	Skewness	Kurtosis
$\log(W_{50})$	2.36	0.17	0.33	−0.40
$\log(A_{25})$	1.17	0.29	−0.79	0.50
$\log(L_B)$	9.79	0.59	−0.65	0.01
$\log(M_{\rm HI})$	9.38	0.56	−1.10	0.94
$\log(M_{25})$	10.35	0.59	−0.11	0.10
$\log(M_{\rm HI}/L_B)$	−0.42	0.23	−0.16	−0.77
$\log(M_{\rm HI}/A_{25}^2)$	7.03	0.23	−0.20	−0.82
$\log(M_{25}/L_B)$	0.56	0.21	−0.43	0.07
$\log(M_{\rm HI}/M_{25})$	−0.98	0.28	−0.07	−0.42
$\log(A_{25}W_{50})$	3.53	0.43	−0.37	0.31

number; (2) the radial velocity $V_{\rm LG}$ reduced to the centroid of the Local group; (3) the H I line width W_c at 50% of the peak flux level corrected for relativistic broadening; (4) the logarithm of the H I mass, $\log(M_{\rm HI}/M_\odot)$; (5) the logarithm of the galaxy's indicative mass, $\log(M_{25}/M_\odot)$; (6) the logarithm of the galaxy's total luminosity, $\log(L/L_\odot)$; (7) the H I mass-to-luminosity ratio, $M_{\rm HI}/L$; (8) the galaxy's total mass-to-luminosity ratio, M_{25}/L (the values in columns 4–8 are in solar units); and (9) the mass fraction of neutral hydrogen, $M_{\rm HI}/M_{25}$.

According to Table 3, the mean depth of the sample is 4823 km s^{-1}, which corresponds to 64 Mpc.

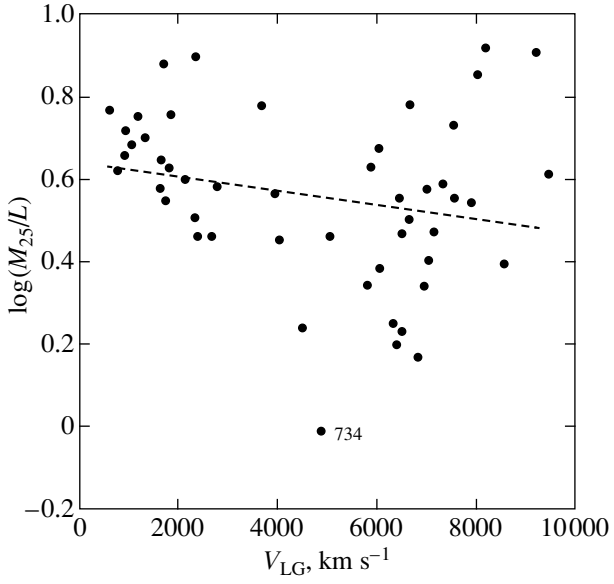


Fig. 4. Total mass-to-luminosity ratio (in solar units) versus radial velocity (in km s^{-1}).

The total mass-to-luminosity ratios of the galaxies lie within the range 0.4–8.2 with a mean of 3.8, and the mass fraction of neutral hydrogen is 13%.

Karachentsev *et al.* (1999b) performed a statistical analysis of the global parameters for 587 flat spiral galaxies using the observations performed by Giovanelli *et al.* (1997) with the Arecibo radio telescope. Huchtmeier et al. (2005) presented the statistics of global parameters for 121 detected RFGC galaxies.

Table 4 gives the statistical characteristics of the distribution of 50 flat galaxies in integrated parameters. Comparison of the mean values of all parameters with those for the sample of 121 flat galaxies shows that they are almost equal.

In Figs 3 and 4, the logarithm of the H I mass-to-luminosity ratio and the logarithm of the indicative mass-to-luminosity ratio are plotted against radial velocity $V_{\rm LG}$ for flat galaxies. Since the radial velocity

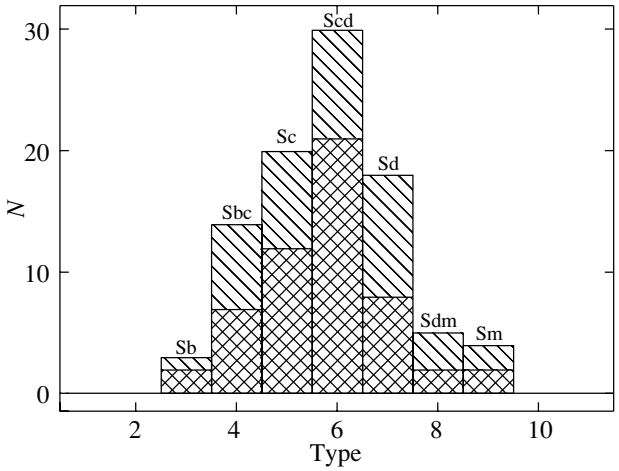


Fig. 5. Distribution of galaxies in morphological type.

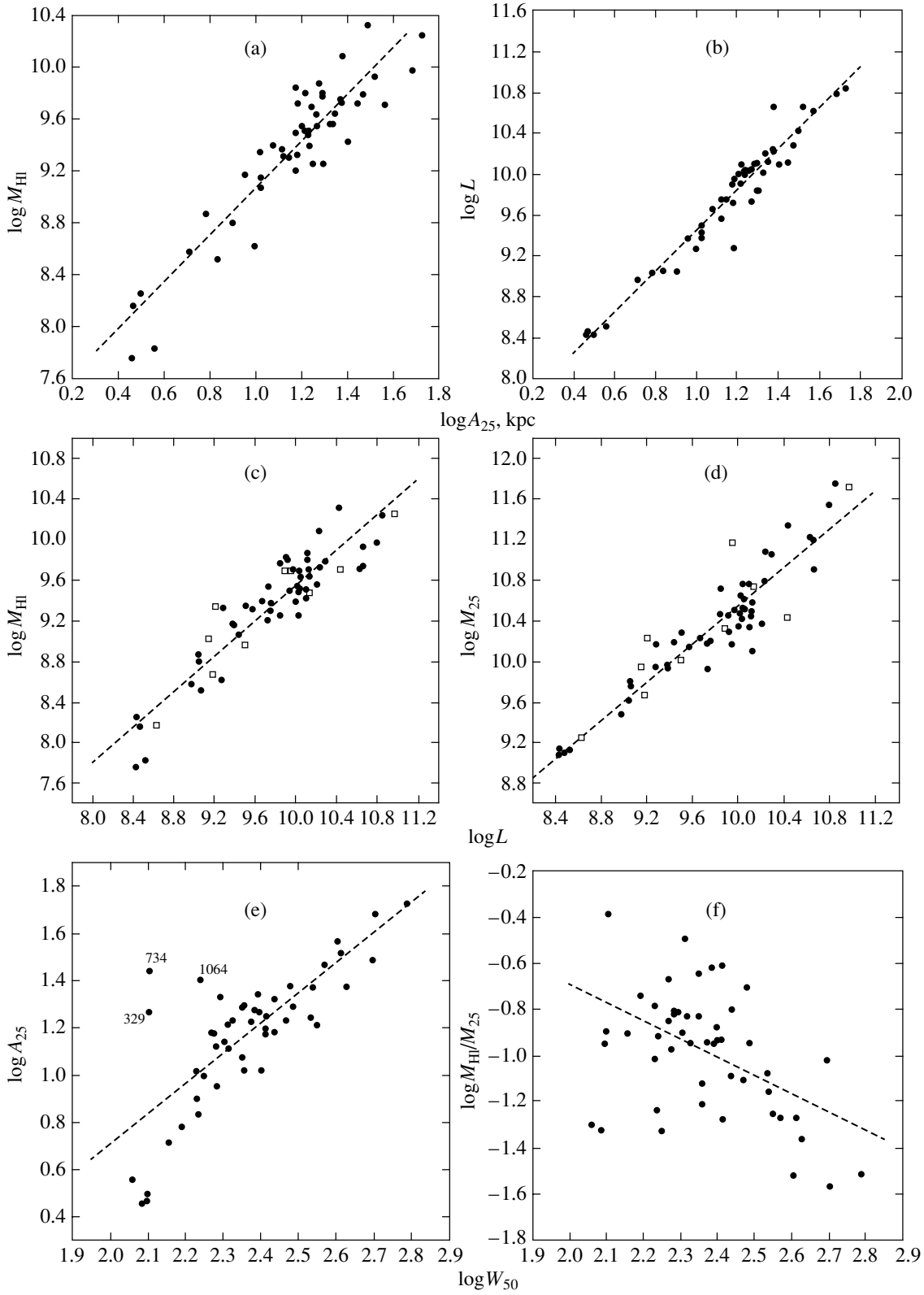


Fig. 6. Bivariate relations between the global parameters reduced to solar units: (a) logarithm of the H I mass versus linear diameter (in kpc); (b) luminosity versus linear diameter; (c) H I mass versus luminosity; (d) indicative mass versus luminosity, the squares mark the 2MFGC galaxies; (e) linear diameter of the galaxy versus H I line width (in km s^{-1}); and (f) H I mass fraction in the galaxy versus H I line width. The dashes represent the regression lines.

Table 5. Distribution of the mean parameters in morphological type for RFGC galaxies

N	Type	$\log(M_{\rm HI}/A_{25}^2)$	$\log(M_{25}/L_B)$	$\log(M_{\rm HI}/M_{25})$	$\log(M_{\rm HI}/L_B)$
18	Sb, bc, c	7.01 ± 0.06	0.60 ± 0.06	-1.03 ± 0.08	-0.43 ± 0.06
20	Scd	7.03 ± 0.06	0.53 ± 0.03	-0.96 ± 0.06	-0.43 ± 0.05
12	Sd, dm, m	7.07 ± 0.05	0.55 ± 0.05	-0.93 ± 0.06	-0.38 ± 0.07

Table 6. Parameters of the linear regression $y = kx + c$ for the global characteristics of RFGC galaxies

N	y	x	$\rho(xy)$	$\sigma(y)$	$k \pm \sigma_k$	$c \pm \sigma_c$
1	$\log(M_{\rm HI})$	$\log(A_{25})$	0.92	0.22	1.80 ± 0.11	7.26 ± 0.13
2	$\log(M_{25})$	$\log(A_{25})$	0.93	0.22	1.92 ± 0.11	8.10 ± 0.13
3	$\log(L_B)$	$\log(A_{25})$	0.97	0.15	2.00 ± 0.08	7.45 ± 0.09
4	$\log(M_{\rm HI})$	$\log(L_B)$	0.92	0.22	0.87 ± 0.05	0.81 ± 0.52
5	$\log(M_{25})$	$\log(L_B)$	0.94	0.20	0.94 ± 0.05	1.14 ± 0.48
6	$\log(A_{25})$	$\log(W_{50})$	0.77	0.19	1.28 ± 0.15	-1.86 ± 0.37
7	$\log(M_{25}/L_B)$	$\log(W_{50})$	0.39	0.19	0.47 ± 0.16	-0.54 ± 0.38
8	$\log(M_{\rm HI}/L_B)$	$\log(W_{50})$	-0.24	0.23	-0.32 ± 0.19	0.34 ± 0.45
9	$\log(M_{25}/M_{\rm HI})$	$\log(W_{50})$	-0.49	0.24	-0.79 ± 0.20	0.88 ± 0.48
10	$\log(M_{\rm HI}/A_{25}^2)$	$\log(W_{50})$	-0.05	0.23	-0.07 ± 0.19	7.19 ± 0.45
11	$\log(M_{\rm HI})$	$\log(A_{25}W_{50})$	0.91	0.23	1.19 ± 0.08	5.19 ± 0.27
12	$\log(M_{\rm HI})$	$\log(W_{50})$	0.76	0.37	2.49 ± 0.31	3.48 ± 0.73
13	$\log(M_{25}/L_B)$	$\log(V_{\rm LG})$	-0.32	0.20	-0.20 ± 0.08	1.25 ± 0.30

is a distance indicator, it can be noted that $M_{\rm HI}/L$ and M_{25}/L decrease with distance only slightly; i.e., there is virtually no distance selection effect. Arbitrarily oriented spiral galaxies are known to show no variation of M/L with distance either (Roberts and Haynes 1994).

Figure 5 shows the distribution of galaxies in morphological type. All of the 94 RFGC galaxies under study are shown on the histogram, and the types of 54 detected galaxies of them are marked by double hatching. Most of the objects belong to the morphological types $Sbc - Sd$, where the detection level is higher.

The data in Table 5 agree with the conclusions of our previous paper (Huchtmeier *et al.* 2005) that the mean values of the global parameters under consideration do not change within the error limits for galaxies of various morphological types. This conclusion is consistent with the H I observations of edge-on spiral galaxies (Karachentsev *et al.* 1999b) and arbitrarily oriented spirals of various types (Haynes and Giovanelli 1984) from the catalog of isolated galaxies by Karachentseva (1973).

BIVARIATE DISTRIBUTIONS OF GLOBAL PARAMETERS

The correlations between global parameters lead us to important conclusions about the differences between the properties of the disk structures in giant and dwarf galaxies. The observed differences may characterize the different formation and equilibrium conditions for the gaseous disks and the inequality of the star formation rates and intensities in them. It is also assumed that the contribution of dark matter depends on the linear sizes of galaxies.

Figures 6 and 7 show the bivariate distributions of global parameters for 50 detected RFGC galaxies, and Table 6 lists the parameters of the linear regression $y = kx + c$ for various relations. The logarithms of the various global parameters of flat galaxies are the variables x and y. Columns 4 and 5 give the correlation coefficients $\rho(x, y)$ and the standard deviations $\sigma(y)$; columns 6 and 7 contain the regression parameters k and c and their standard errors, respectively.

Figure 6a and the first row of Table 6 show the relation between the logarithm of the H I mass (in

$M_\odot$) and the logarithm of the galaxy's linear diameter A_{25} (in kpc) calculated from the angular diameter a_c using the formula

$$A_{25} = 0.29 a_c V_{\rm LG}/H_0. \quad (5)$$

The linear relation suggests that the H I density in the disks of spiral galaxies is virtually constant and does not depend on the linear size. The conditions for star formation in them appear to have been approximately identical in them. The statistically insignificant decrease in $\log(M_{\rm H\,I}/A_{25}^2)$ toward the massive galaxies (row 10 in Table 6) may be due to the internal absorption of the H I flux in the disks of the largest galaxies. Other parameters, the indicative mass and the total luminosity (see rows 2 and 3 in Table 6 and Fig. 6b), also correlate well with the linear diameter.

The dependences of the H I mass $\log M_{\rm H\,I}$ and the indicative mass $\log M_{25}$ on the luminosity $\log L$ show a systematic decrease in M/L from dwarf galaxies to giant spirals (see Figs. 6c and 6d and rows 4 and 5 in Table 6). This is to be expected if the relative abundance of the dark matter in giant galaxies is lower than that in dwarf galaxies.

Since the width of the 21-cm H I line does not depend on the distance to the galaxy, taken as the argument for the bivariate distributions of global parameters, it allows the morphology and luminosity selection effects to be reduced. Figure 6e (row 6 in Table 6) shows the linear relation between the galaxy's size $\log A_{25}$ (in kpc) and the amplitude of its internal rotation determined as the width of the 21-cm line. The slope of the linear regression in this Tully–Fisher (1977) relation is 1.28 with a standard deviation of 0.15. The correlations between $\log(M_{25}/L)$ and $\log(M_{\rm H\,I}/L)$ and the 21-cm line width are also well defined; their parameters are listed in rows 7 and 8 of Table 6. We see from Fig. 6f (row 9 in Table 6) that the relative hydrogen abundance increases from giant galaxies to dwarf galaxies in the same way as for arbitrarily oriented spiral galaxies whose luminosities are not affected strongly by internal absorption (Huchtmeier and Richter 1988; Stavely-Smith and Davies 1988).

The relation between the logarithm of the H I mass $\log M_{\rm H\,I}$ and the specific angular momentum $\log(A_{25}W_{50})$ (see Fig. 7 and row 11 in Table 6) has a high correlation coefficient; the slope of the linear regression is 1.19. Zasov (1974) pointed out that such a relation must hold in the gas-rich disks of spiral galaxies with active star formation. Karachentsev *et al.* (1999b) corroborated the linear relation between $\log M_{\rm H\,I}$ and $\log(A_{25}W_{50})$ for flat galaxies and Zasov's conclusion that the gaseous disks of galaxies must be near the gravitational stability boundary irrespective of the linear sizes of the galaxies.

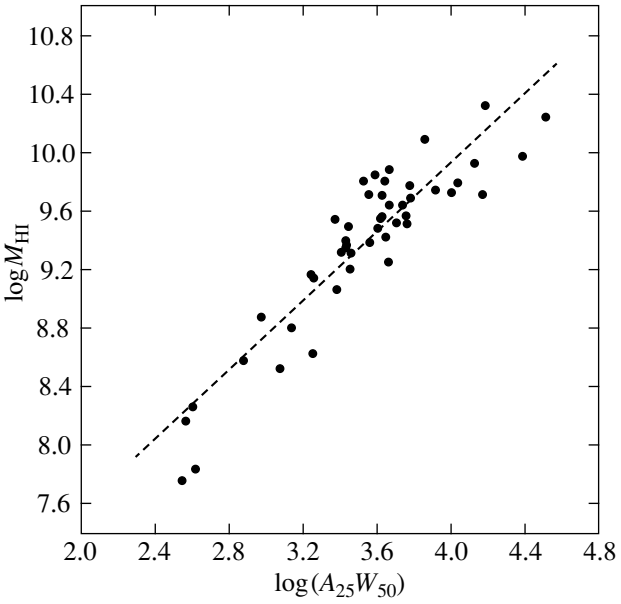


Fig. 7. H I mass versus specific angular momentum.

ACKNOWLEDGMENTS

We are grateful to the Max Plank Institut für Radioastronomie (Effelsberg, Germany) for the observational data obtained with the 100-m radio telescope. This work was supported by the Russian Foundation for Basic Research and Deutsche Forschungsgemeinschaft (DFG), grant no. 02-02-04012.

REFERENCES

1. H. Feldman, R. Yuszhkievicz, M. Davis, *et al.*, Astrophys. J. **596**, L131 (2003).
2. R. Giovanelli, E. Avera, and I. D. Karachentsev, Astron. J. **114**, 122 (1997).
3. M. P. Haynes and R. Giovanelli, Astron. J. **89**, 758 (1984).
4. W. K. Huchtmeier, I. D. Karachentsev, V. E. Karachentseva, *et al.*, Astron. Astrophys. **435**, 2, 459 (2005).
5. W. K. Huchtmeier and O. G. Richter, Astron. Astrophys. **203**, 237 (1988).
6. I. D. Karachentsev, Astron. J. **97**, 1566 (1989).
7. I. D. Karachentsev, V. E. Karachentseva, Yu. N. Kudrya, *et al.*, Bull. Spec. Astron. Obs. Russ. Acad. Sci. **47**, 5 (1999a).
8. I. D. Karachentsev, V. E. Karachentseva, and Yu. N. Kudrya, Pis'ma Astron. Zh. **25**, 3 (1999b) [Astron. Lett. **25**, 1 (1999b)].
9. I. D. Karachentsev, V. E. Karachentseva, Yu. N. Kudrya, *et al.*, Astron. Zh. **77**, 175 (2000a) [Astron. Rep. **44**, 150 (2000a)].
10. I. D. Karachentsev, V. E. Karachentseva, Yu. N. Kudrya, *et al.*, Bull. Spec. Astron. Obs. Russ. Acad. Sci. **50**, 5 (2000b).
11. I. D. Karachentsev, V. E. Karachentseva, and S. L. Parnovskii, Astron. Nachr. **314**, 97 (1993).

12. I. D. Karachentsev and D. I. Makarov, Astron. J. **111**, 794 (1996).
13. I. D. Karachentsev, S. N. Mitronova, and V. E. Karachentseva, Astron. Astrophys. **396**, 431 (2002).
14. V. E. Karachentseva, Soobshch. Spets. Astron. Obs. Akad. Nauk SSSR **8**, 3 (1973).
15. A. Kogut, C. Lineweaver, G. F. Smoot, *et al.*, Astrophys. J. **419**, 1 (1993).
16. Yu. N. Kudrya, I. D. Karachentsev, V. E. Karachentseva, and S. L. Parnovskii, Pis'ma Astron. Zh. **22**, 330 (1996) [Astron. Lett. **22**, 295 (1996)].
17. Yu. N. Kudrya, I. D. Karachentsev, V. E. Karachentseva, and S. L. Parnovskii, Pis'ma Astron. Zh. **23**, 730 (1997) [Astron. Lett. **23**, 633 (1997)].
18. Yu. N. Kudrya, V. E. Karachentseva, I. D. Karachentsev, *et al.*, Astron. Astrophys. **407**, 889 (2003).
19. D. I. Makarov, A. N. Burenkov, and N. V. Tuyrina, Astron. Lett. **27**, 213 (2001).
20. S. N. Mitronova, I. D. Karachentsev, V. E. Karachentseva, *et al.*, Bull. Spec. Astron. Obs. Russ. Acad. Sci. **57**, 5 (2004).
21. M. S. Roberts and M. P. Haynes, Annu. Rev. Astron. Astrophys. **32**, 115 (1994).
22. D. J. Schlegel, D. P. Finkbeiner, and M. Davis, Astrophys. J. **500**, 525 (1998).
23. M. F. Skrutskie, S. E. Schneider, R. Stiening, *et al.*, in *The Impact of Large Scale Near-IR Sky Surveys*, Ed. by F. Garzon *et al.* (Kluwer, Dordrecht, 1997), Vol. 210, p. 25.
24. L. Stavely-Smith and R. D. Davies, Mon. Not. R. Astron. Soc. **231**, 833 (1988).
25. R. B. Tully and J. R. Fisher, Astron. Astrophys. **54**, 661 (1977).
26. A. V. Zasov, Astron. Zh. **51**, 1225 (1974) [Sov. Astron. **18**, 730 (1974)].

Translated by A. Dambis

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 515–527. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 577–591.

Formation of Galactic Subsystems in Light of the Magnesium Abundance in Field Stars: The Thick Disk

V. A. Marsakov* and T. V. Borkova**

Rostov State University, Rostov-on-Don, Russia

Received December 14, 2004

Abstract—The space velocities and Galactic orbital elements of stars calculated from the currently available high-accuracy observations in our compiled catalog of spectroscopic magnesium abundances in dwarfs and subgiants in the solar neighborhood are used to identify thick-disk objects. We analyze the relations between chemical, spatial, and kinematic parameters of F–G stars in the identified subsystem. The relative magnesium abundances in thick-disk stars are shown to lie within the range $0.0 < [\mathrm{Mg/Fe}] < 0.5$ and to decrease with increasing metallicity starting from $[\mathrm{Fe/H}] \approx -1.0$. This is interpreted as evidence for a longer duration of the star formation process in the thick disk. We have found vertical gradients in metallicity ($\mathrm{grad}_Z[\mathrm{Fe/H}] = -0.13 \pm 0.04$ kpc^{-1}) and relative magnesium abundance ($\mathrm{grad}_Z[\mathrm{Mg/Fe}] = 0.06 \pm 0.02$ kpc^{-1}), which can be present in the subsystem only in the case of its formation in a slowly collapsing protogalaxy. However, the gradients in the thick disk disappear if the stars whose orbits lie in the Galactic plane, but have high eccentricities and low azimuthal space velocities atypical of the thin-disk stars are excluded from the sample. The large spread in relative magnesium abundance ($-0.3 < [\mathrm{Mg/Fe}] < 0.5$) in the stars of the metal-poor "tail" of the thick disk, which constitute ≈8% of the subsystem, can be explained in terms of their formation inside isolated interstellar clouds that interacted weakly with the matter of a single protogalactic cloud. We have found a statistically significant negative radial gradient in relative magnesium abundance in the thick disk ($\mathrm{grad}_R[\mathrm{Mg/Fe}] = -0.03 \pm 0.01$ kpc^{-1}) instead of the expected positive gradient. The smaller perigalactic orbital radii and the higher eccentricities for magnesium-richer stars, which, among other stars, are currently located in a small volume of the Galactic space near the Sun, are assumed to be responsible for the gradient inversion. A similar, but statistically less significant inversion is also observed in the subsystem for the radial metallicity gradient. © *2005 Pleiades Publishing, Inc.*

Key words: *Galaxy (Milky Way), stellar chemical composition, thick disk, Galactic evolution.*

INTRODUCTION

An insufficient accuracy of determining the ages of not only single stars, but also huge stellar ensembles, globular and open clusters, which are believed to consist of coeval stars, is a serious obstacle to reconstructing the formation history of the structure in the early Galaxy. Here, a "chemical" age indicator could be of help. According to present views, all of the chemical elements heavier than boron have been produced in nucleosynthesis reactions in stars of various masses. It thus follows that the number of atoms of heavy elements will inevitably increase as the Galaxy evolves, and the ratio of the number of atoms of heavy elements to the number of hydrogen atoms in the atmospheres of unevolved stars (i.e., their total metallicity) must serve as an indicator of their age. However, numerous studies show that no age–metallicity relation can be traced for old Galactic populations (see, e.g., Norris *et al.* 1985). It proved to be more convenient to use the relative abundances of various chemical elements instead of the total metallicity as a chemical age indicator of stars, since the formation of most elements can be attributed rather reliably to a particular nucleosynthesis process in stars of certain masses that evolve in a theoretically determined time. In particular, high-mass type II supernovae (SNe II) are the main suppliers of α-capture and r-process elements and a small amount of iron-group elements to the interstellar medium. In contrast, the bulk of the iron-group elements are synthesized in lower-mass stars that are members of close binaries and that explode as type Ia supernovae (SNe Ia). The α-elements are produced in a shorter time than iron, because the evolution times of SNe II (≈30 Myr) and SNe Ia (≈1–2 Gyr) differ (see, e.g., Matteucci and Greggio 1986; Thielemann *et al.* 1990; Tsujimoto *et al.* 1995; Matteucci 2001) (although these authors believe that isolated, the very first SNe Ia could appear considerably earlier than this time, only

*E-mail: marsakov@ip.rsu.ru
**E-mail: borkova@ip.rsu.ru

≈0.5 Gyr later). Since the contribution of SNe Ia to the synthesis of iron-group elements is significantly larger than their contribution to the synthesis of α-elements, the [α/Fe] ratio will decrease in the Galaxy as the interstellar medium is enriched with the remnants of these supernovae. Thus, at least ≈1 Gyr after the initial burst of star formation will pass by the time [α/Fe] will begin to decrease.

It is well known that the boundary between the halo and the thick disk can be drawn through the point [Fe/H] ≈ −1.0. Indeed, the distributions of subdwarfs, RR Lyrae variable stars, and globular clusters in heavy-element abundance exhibits a clear deficit of objects in the vicinity of precisely this point (see, e.g., Marsakov and Suchkov 1977; Zinn 1985; Borkova and Marsakov 2000, 2002). At the same time, there is evidence that the knee in the [α/Fe]–[Fe/H] relation also occurs approximately at this location in our Galaxy. Nevertheless, there is no consensus about whether the relative abundance of α-elements decreases with increasing total metallicity inside the thick disk. In particular, having analyzed their rather small samples of stars, Gratton *et al.* (2000) and Fuhrmann (2000) argue that the abundance of α-elements in thick-disk stars is approximately the same and decreases abruptly as one goes to thin-disk stars. At the same time, Prochaska *et al.* (2000) and Feldzing *et al.* (2003) derived the [α/Fe]–[Fe/H] relation inside the thick disk from their small samples, with the knee occurring precisely at the point [Fe/H] ≈ −1.0. The difference between these two results is of fundamental importance in correctly reconstructing the formation history of the thick disk. Indeed, if the thick disk were formed during the collapse of a protogalactic cloud, then a trend in the relative abundance of α-elements with increasing heavy-element abundance would inevitably arise in it (Prochaska *et al.* 2000). The thin disk could subsequently be formed in the Galaxy as a result of the ongoing collapse (see Burkert *et al.* 1992). The difficulty of this model lies in the fact that the total collapse time scale of the protogalactic cloud is much shorter than the evolution time scale of SNe Ia progenitors. This difficulty can be easily circumvented in a different class of models where the thick disk is formed from the very first thin-disk stars through the interaction of the Galaxy with a close satellite galaxy (see Kroupa 2002, Quinn *et al.* 1993). If, furthermore, the stars of the newly forming thin disk are assumed to be formed from interstellar matter with additions of accreted intergalactic metal-poor gas, then the existence of metal-poor stars with thin-disk kinematics and the metallicity overlap between thick- and thin-disk stars in the range $-0.8 < [Fe/H] < -0.3$ can be explained (Fuhrmann *et al.* 1995; Fuhrmann 1998). The difficulty of this model lies in explaining the existence of metal-rich globular clusters in the thick disk (Marsakov and Suchkov 1977). They cannot be heated, as stars, by a satellite galaxy and can appear only as a result of the star formation that accompanies this interaction (Gratton *et al.* 2000).

Magnesium is the best-studied α-element, since it has visible lines of various intensities and degrees of excitation in unevolved F–G stars. According to present views, almost all of the magnesium is synthesized in the envelopes of high-mass ($M > 10M_\odot$) presupernovae with hydrostatic carbon burning and is subsequently ejected into the interstellar medium by SNe-II explosions. To perform a comprehensive statistical analysis of the chemical, physical, and spatial–kinematic properties of stellar populations in an effort to reconstruct the formation and evolution history of the Galactic subsystems, we compiled a summary catalog of spectroscopic magnesium abundances in stars with accurate parallaxes (Borkova and Marsakov 2005). In this catalog, we collected almost all of the magnesium abundances in dwarfs and subgiants in the solar neighborhood published in the past fifteen years and calculated the space velocity components and Galactic orbital elements for them. The size of our catalog is severalfold larger than the size of any sample that has been used so far to analyze the chemical evolution of the Galaxy; this allows us to largely get rid of the observational selection that is often inherent in smaller samples. In this paper, we consider the stellar population of the Galactic thick disk.

INPUT DATA

In our summary catalog, the effective temperatures, surface gravities, and [Fe/H] values for each star were obtained by averaging some 2000 determinations from 80 publications. The accuracies of the mean values of these quantities estimated from the scatters of the deviations of individual determinations about the calculated means for each star of the sample are $\epsilon T_{\rm eff} = \pm 56^\circ$ and $\pm 82^\circ$, $\epsilon \log g = \pm 0.12$ and ± 0.24, and ϵ[Fe/H] = ±0.07dex and ±0.13dex for stars with metallicities [Fe/H] higher and lower than −1.0, respectively. To avoid the uncertainties related to non-LTE effects for iron lines, we used everywhere the iron abundances determined solely from Fe II lines. Since the non-LTE corrections for magnesium in F–G dwarfs and subgiants are small and do not exceed 0.1 dex (Shimanskaya *et al.* 2000), they were not considered separately in our catalog. The relative magnesium abundances were derived from 1412 determinations in 31 publications for 867 dwarfs and subgiants by using a three-pass iterative averaging procedure with the assignment of

a weight to each primary source and to each individual determination. The systematic shifts of all scales relative to the reduced mean scale were taken into account. Assigning the lowest weight to the least reliable determinations, this procedure yields final values close to those given by most sources without excluding any determination. The internal accuracies of the [Mg/Fe] values for metal-rich and metal-poor stars are ϵ[Mg/Fe] = ±0.05dex and ±0.07dex, respectively.

The distances to the stars and their proper motions were calculated using data from currently available high-accuracy catalogs, both orbital (Hipparcos–Tycho) and ground-based (PPM-N, PPM-S, PPM-add) ones. We used the trigonometric parallaxes with errors of less than 25% and, in the absence of the latter, the published photometric distances determined from *uvbyb* photometry. The radial velocities were taken mostly from the catalogs by Nidever *et al.* (2002), Barbier-Brossat and Figon (2000), and Nordstrem *et al.* (2004). In the absence of necessary data in these catalogs, they were taken from other sources (see the references to Table 2 in our previous paper (Borkova and Marsakov 2005)).

The components of the total space velocity (U, V, W) relative to the Sun were calculated for 844 stars with measured distances, proper motions, and radial velocities, where U is directed toward the Galactic anticenter, V is directed in the sense of Galactic rotation, and W is directed toward the North Galactic Pole. At mean errors in the distances of 15% and a mean heliocentric distance of the stars from our sample of ≈60 pc, the mean error in the space velocity components was ≈±2 km s^{-1}.

Based on the multicomponent model of the Galaxy containing a disk, a bulge, and an extended massive halo (Allen and Santillan 1991), we computed the Galactic orbital elements by simulating 30 revolutions of a star around the Galactic center. The Galactocentric distance of the Sun was assumed to be 8.5 kpc, the rotational velocity of the Galaxy at the solar Galactocentric distance was 220 km s^{-1}, and the velocity of the Sun with respect to the local standard of rest is $(U_\odot, V_\odot, W_\odot) = (-11, 14, 7.5)$ km s^{-1} (Ratnatunga *et al.* 1989).

For our studies, we retained only 847 dwarfs and subgiants in the solar neighborhood with $T_{\rm eff} >$ 5000 K; in cooler stars, the elemental abundances are determined less accurately. For 545 stars of our sample, we used the ages determined from theoretical isochrones in a paper by Nordstrem *et al.* (2004). The absolute magnitudes of the stars were found in this paper, as in our paper, mainly from the trigonometric parallaxes (from the photometric distances for some of the stars with large errors in the parallaxes), while the metallicities were derived from *uvby* photometry. A check shows that their photometric metallicities and temperatures are in good agreement with our spectroscopic data (the mean deviation is $\delta([{\rm Fe/H}]_{\rm ph} - [{\rm Fe/H}]_{\rm sp}) = 0.07$dex for the iron abundance at the dispersion of the deviations $\sigma_{\delta[{\rm Fe/H}]} = 0.09$dex and $\delta T_{\rm eff} = 45°$ for the temperature at $\sigma_{\delta T_{\rm eff}} = 70°$, i.e., of the order of the above errors in the corresponding quantities in our catalog), which allows us to use the ages from this paper. The absolute values of the age errors given by the authors are rather large: the error is less than 1 Gyr only for ≈30% of the stars; ranges from 1 to 2 Gyr for another ≈25% of the stars; and is even larger, up to ≈7 Gyr, for the remaining stars.

All stars of our sample fill the entire sky rather uniformly. Nevertheless, it exhibits observational selection: the apparent magnitude depends on metallicity (see Fig. 7 in our previous paper (Borkova and Marsakov (2005)). On the one hand, it stemmed from the fact that mostly metal-richer stars lie near the Galactic plane (where the Sun is located). On the other hand, observers knowingly use more sensitive instrumentation to increase the number of studied metal-poorer stars. As a result, almost all of the stars with [Fe/H] ≥ −1.0 were brighter than $V \approx 9^m$, while the bulk of the metal-poorer stars were fainter. However, our results are not affected by this selection.

IDENTIFICATION OF STARS OF THE SUBSYSTEMS

There is no single, necessary and sufficient criterion that would allow each star to be individually identified with absolute confidence by its belonging to a particular Galactic subsystem. Only the mean values and dispersions of such key parameters of the stellar subsystems as the ages, velocities, spatial locations, Galactic orbital elements, metallicities, and relative abundances of certain chemical elements can be determined more or less accurately from reliably identified members of the subsystem. Any of the above parameters can act as a criterion for identifying the stars of various subsystems. For globular star clusters, the morphological structure of their horizontal branches and the mean periods of their RR Lyrae stars (cf. Oosterhof's types) can also be attributed to these determining parameters. Choosing any of these parameters, we get an opportunity to study the distributions of the subsystem's objects in other parameters. Such studies have revealed that no sharp boundaries exist between the subsystems and that stars with identical characteristics can belong to different subsystems. It would be appropriate to use complex criteria that include several parameters to identify the subsystem's objects more reliably.

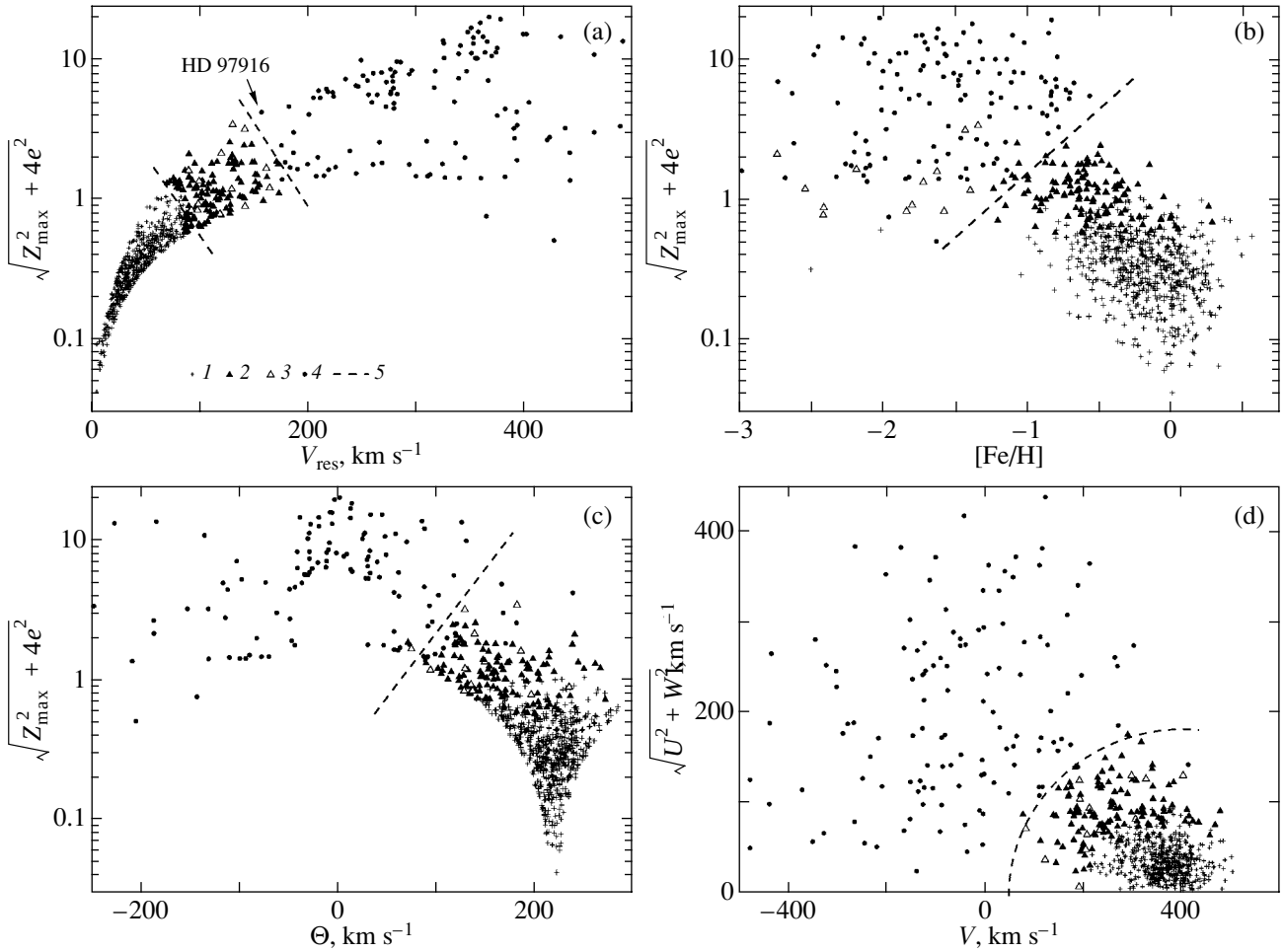


Fig. 1. Relations between $\sqrt{Z_{max}^2 + 4e^2}$ and the residual stellar velocities with respect to the local standard of rest (a), metallicities (b), azimuthal velocities (c) and between the residual stellar velocity components (d): *1*, thin-disk stars; *2*, thick-disk stars; *3*, stars of the metal-poor thick-disk tail; *4*, halo stars; and *5*, demarcation lines drawn by eye. The dashed line in panel (d) was drawn with a radius of 175 km s^{-1} separating the halo and thick-disk stars (the highest velocity and metal-poorest halo stars are outside the frames of the panels).

The metallicity of stars is commonly used to separate the stars of the spherical Galactic subsystems from the disk ones, since it remains, though rough, an age indicator. However, in this paper, we chose exclusively kinematic parameters as the initial criteria, since they shape the spatial distribution of objects that manifests itself in the morphological identification of several subsystems in the Galaxy. In this approach, the possibility of observational selection that distorts the abundance distributions of various chemical elements in the stars of the subsystems, particularly in the near-boundary regions, is ruled out. We separated the thick-disk stars from the halo using the following condition: $V_{res} = \sqrt{(U^2 + V^2 + W^2)} <$ 175 km s^{-1}, where U, V, W, and V_{res} are the space velocity components and the total residual stellar velocity relative to the local standard of rest, respectively. We optimized the specific value of this criterion by minimizing the numbers of metal-rich ([Fe/H] > -1.0) sample stars in the identified halo and metal-poor stars in the thick disk. The derived critical value of the residual velocity for the selection of thick-disk stars from our sample is almost equal to that in a paper by Fuhrmann (2000): $V_{res} = 180$ km s^{-1}. Chiappini *et al.* (1997) suggested a kinematic index independent of the stellar positions in the orbits that is based on the same three space velocity components or, more precisely, on the Galactic orbital elements calculated from them. This is $\sqrt{Z_{max}^2 + 4e^2}$, where Z_{max} is the maximum distance of the orbital points from the Galactic plane and e is the orbital eccentricity of the star. Figure 1a shows the relation between these two kinematic indices. Even a distinct gap between the halo and the thick disk, which is indicated by the dotted line, is observed in the diagram (however, it could also be the result of selection in the

original lists of stars used to compile the summary catalog). The identified subsystems are also naturally separated in the [Fe/H] – $\sqrt{Z_{\max}^2 + 4e^2}$ plane, i.e., composed of the two independent criteria that we did not use, and form two clusters that can be arbitrarily separated (see the dotted line in Fig. 1b). The stars lying in the immediate vicinity of this demarcation line occupy controversial positions here. However, according to our kinematic criterion, almost all of them fall into the thick disk, as can be seen from the diagram. As a check shows, they fill rather uniformly the same regions in all kinematic diagrams of Fig. 1 as the remaining thick-disk stars. However, excluding these stars from the thick disk would introduce artificial chemical selection into both separated systems, and a certain number of stars with an intermediate metallicity and with kinematics typical of the thick-disk stars would be attributed to the halo. Very metal-poor, but rapidly revolving (around the Galactic center), stars are commonly called the metal-poor tail of the thick disk (Prochaska *et al.* 2000). We singled out 13 stars with [Fe/H] < -1.25 from the kinematically selected thick-disk objects in order to subsequently trace the correspondence of their properties to those of the subsystem's remaining stars. The metal-poor tail stars are denoted in Fig. 1 by open triangles. Figure 1c shows the $\Theta - \sqrt{Z_{\max}^2 + 4e^2}$ diagram, where Θ is the azimuthal velocity of the star. We see from this figure that the halo and thick-disk stars (except several stars) in this diagram are separated by a clear gap, while the metal-poor tail stars fill uniformly the region occupied by all of the remaining thick-disk stars. The location of the gap is highlighted in the diagram by the inclined dotted line. The distribution of stars of the identified subsystems in Toomre's diagram is shown in Fig. 1d. We emphasize that any criterion is purely statistical and the membership of each specific star remains doubtful. For example, according to the residual velocity criterion, the star HD 97916 (marked in Fig. 1a) falls into the thick disk, but its high value of $Z_{\max} > 4$ kpc, which is atypical of the stars of this subsystem (see Fig. 6 below), suggests that it can also belong to the halo with equal success. Based on the 3σ criterion applied to the parameter $Z_{\max}$, we excluded this star from the subsequent analysis among the thick-disk stars.

The thin-disk stars are known to have low residual velocities with respect to the local standard of rest and nearly circular orbits, all the points of which do not rise high above the Galactic plane. Therefore, for reliability, we identified this subsystem using simultaneously two kinematic criteria: $V_{\rm res} < 85$ km s^{-1} (see also Furmann 1998) and $\sqrt{Z_{\max}^2 + 4e^2} < 1.05$. Higher-velocity stars were included in the thick disk. In this case, the values of the criteria were chosen to minimize the number of stars with a high relative abundance of magnesium in the thin disk and with its low abundance in the thick disk. It turned out that we could not get rid of the overlap between the abundance ranges of various chemical elements in the stars of the subsystems being separated by any combination of any kinematic parameters. However, the $V_{\rm res} - \sqrt{Z_{\max}^2 + 4e^2}$ diagram (Fig. 1a) reveals even sparseness of stars between the subsystems. Figures 1a and 1d show that we cannot reliably determine which subsystem a star belongs to from any single velocity component. Indeed, according to the adopted criteria, a star in the thick disk can have any arbitrarily low velocity component, but the other velocity components (or only one component) will necessarily be appreciably higher than those for the disk stars.

If we follow the kinematic criterion, then the metallicity is represented by not quite a proper criterion. Indeed, a significant overlap of the metallicity ranges between the thick and thin disks is seen from Fig. 1b. Consequently, the criterion based on the position of the dip found in the metallicity function of field stars near [Fe/H] $= -0.5$ (see, e.g., Marsakov and Suchkov 1977), which is commonly used to separate the stars of the subsystems, can be used only when there are no data on the total space velocities of the stars. The situation with the age is also difficult. Based on the accurately determined ages of only seven subgiants, Bernkopf *et al.* (2001) concluded that the thin disk began to form ~9 Gyr ago, while the thick-disk stars are definitely older than 12 Gyr. Nevertheless, the ages, unfortunately, cannot yet be used as a reliable criterion for separating the two disk subsystems, because the individual ages of most stars are determined with a low accuracy. Therefore, in this paper, we use the ages only to reveal statistical trends.

RELATION BETWEEN THE IRON AND MAGNESIUM ABUNDANCES

Let us consider the properties of the stars of the thick disk that we identified. Elucidating the question of whether there is a trend in the relative magnesium abundance with metallicity inside the thick disk is of greatest importance here. Figure 2a shows the [Mg/Fe]–[Fe/H] diagram for all stars of our catalog where the identified thick-disk stars are highlighted by triangles. To check the results using more reliable data, exactly the same diagram, but only for the stars in which the magnesium abundances were averaged over several sources and with the total weight of the mean value larger than unity, is shown in Fig. 2b. Figure 2c shows the stars of only the thick disk and plots the median curve that roughly fits the pattern of [Mg/Fe] variation with increasing

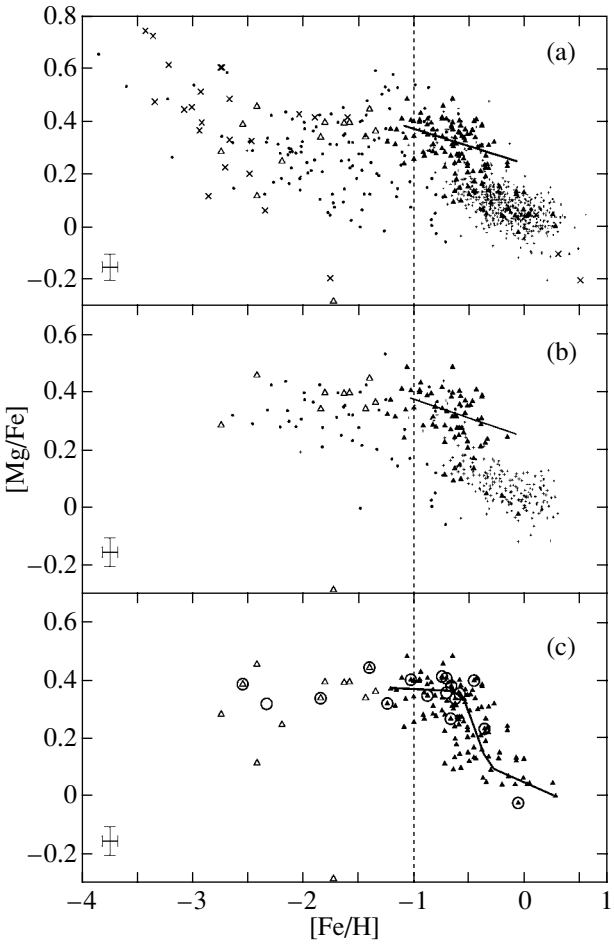


Fig. 2. Relations between metallicity and relative magnesium abundance (a) for all stars of our catalog, (b) for the stars in which the magnesium abundances were averaged over several sources, and (c) only for the thick-disk stars. The notation is the same as that in Fig. 1; the crosses denote the stars with unmeasured radial velocities. The vertical dotted lines indicate the point at which the knee of the ([Mg/Fe]–[Fe/H]) relation begins, and the inclined straight lines represent the regression lines for metal-rich and, simultaneously, magnesium-rich thick-disk stars; the broken curve represents the median line drawn by eye. The open circles in the panel denote the stars of the Arcturus moving group. The error bars are shown.

metallicity. Unfortunately, this dependence cannot be constructed mathematically rigorously; therefore, we drew it by eye halfway between the upper and lower envelopes of the diagram. We see from the figure that an appreciable number of stars with thick-disk kinematics have a chemical composition typical of the thin-disk stars, with the magnesium abundance trend with increasing metallicity for these stars being very similar to that in the thin disk. Most stars (9 of the 13) of the metal-poor thick-disk tail (open triangles) show approximately the same magnesium abundance, $\langle[\mathrm{Mg/Fe}]\rangle = 0.40 \pm 0.01$, but some of them (four stars) at metallicity $[\mathrm{Fe/H}] < -1.7$ exhibit a very low abundance ratio, $[\mathrm{Mg/Fe}] = (-0.3 \div 0.3)$dex, which is atypical of stars formed from matter with the same chemical evolution. Starting from $[\mathrm{Fe/H}] \approx -1.0$ (the vertical dotted line in the figure), the tendency for the relative magnesium abundance to decrease with increasing metallicity begins to show up increasingly clearly in the thick disk. Even if we restrict our analysis to the magnesium abundance range $[\mathrm{Mg/Fe}] > 0.25$, the correlation coefficient turns out to be larger than zero beyond the error limits. However, even when the boundary is lowered to $[\mathrm{Mg/Fe}] = 0.20$, i.e., to the upper limit for the overwhelming majority of thin-disk stars, the correlation coefficient becomes equal to $r = 0.4 \pm 0.1$ (see the inclined straight lines in Figs. 2a and 2b). If, however, we include metal-poorer thick-disk star up to $[\mathrm{Fe/H}] = -1.25$, then the slopes of the regression lines in both diagrams will decrease slightly. However, the figure shows a considerable number of thick-disk stars with even lower magnesium abundances, and the correlation coefficient increases to $r = 0.65 \pm 0.05$ for all of the thick-disk stars with metallicities $[\mathrm{Fe/H}] > -1.0$. This increase in correlation is attributable mainly to a sudden decrease in the mean magnesium abundance starting from $[\mathrm{Fe/H}] \approx -0.6$, which showed up as a sharp bend of the median sequence in Fig. 2c. The sequence again becomes flatter starting from $[\mathrm{Fe/H}] \approx -0.35$. Thus, the data from our catalog suggest that the [Mg/Fe] ratio decreases with increasing [Fe/H] in the thick disk and that the explosions of the first SNe Ia began when the mean metallicity of the interstellar medium in the Galaxy reached $[\mathrm{Fe/H}] \approx -1.0$. However, the explosions of SNe Ia began later than the mass star formation in the thick disk, which occurred (as can be seen from Fig. 2) even at $[\mathrm{Fe/H}] \approx -1.25$. These results confirm the conclusion about a fairly long duration of the star formation in the subsystem that was drawn by Prochaska *et al.* (2000) from the trends in three other α-elements with metallicity detected by the authors. Prochaska *et al.* (2000) argue that such a long time scale is possible only in the thick-disk formation scenario during dissipational collapse. The formation of metal-poor tail stars in this scenario can be explained by the fact that a number of stars with kinematics characteristic of the thick-disk stars were formed from isolated clouds of interstellar matter less enriched with heavy elements. The existence of stars with a low magnesium abundance ($[\mathrm{Mg/Fe}] <$ 0.2dex) in the metal-poor tail at low heavy-element abundances most likely implies that these clouds were in weak contact with the bulk of the protogalactic cloud in which the matter was mixed fairly well.

A further argument for the latter suggestion is the conclusion by Navarro *et al.* (2004). They provide evidence that the stars of the long known Arcturus moving group (see Eggen 1998) may well originate from a massive satellite disrupted by the Galactic tidal forces at an early formation phase of the Galaxy. We selected stars with close angular momenta from the original catalog according to the criteria in the paper by Navarro *et al.* (2004): $(85 < \Theta < 130)$, $|W| < 50$ km s^{-1} and $|\Pi| < 150$ km s^{-1}, where Π, Θ, and W are the space velocity components of the stars with respect to the local standard of rest in the cylindrical coordinate system. It turned out that 15 of the 17 stars selected in this way belong to our thick disk, and 3 of them belong to its metal-poor tail (two stars fell into the protodisk halo). Figure 2 shows the [Mg/Fe]–[Fe/H] diagram for the thick-disk stars (triangles); the circles denote the stars of the Arcturus group. The considerably smaller spread in [Mg/Fe] ratios than that for the thick-disk stars at given metallicity for Arcturus group candidates argues that these stars may well have a common origin. Thus, we see that at least some of the stars with thick-disk kinematics can be extragalactic in origin, i.e., be formed from matter with a different history of chemical evolution.

All of the currently existing thick-disk formation scenarios can be divided into two classes: (1) the collapse of a single protogalactic cloud and (2) the result of the interaction between the early Galaxy and nearby satellite galaxies. The behavior of the subsystem's stars in different scenarios must differ. Thus, for example, according to the theoretically modeled first scenario, a vertical gradient in chemical composition must be formed in the thick disk through the collapse of the protogalactic cloud (Burkert *et al.* 1992). However, free collapse takes a very short time, ≈400 Myr, while the relative abundances of chemical elements in the interstellar medium can change appreciably only on much longer time scales. The above evidence for the existence of SNe Ia traces in the chemical composition of the subsystem's stars suggests that the overwhelming majority of stars in it were formed after the mass explosions of supernovae of this type began, i.e., no earlier than 1–2 Gyr (see, e.g., Matteucci 2001) after the initial star formation in the protogalactic cloud. By this time, according to the hypothesis of rapid collapse, all of the interstellar matter in the protogalaxy had already concentrated near the plane. This contradiction is naturally circumvented in the theoretical model where as the result of a close encounter with a fairly massive satellite galaxy, the already formed stars of the primordial thin disk acquire a significant acceleration mainly perpendicular to the Galactic plane, producing the thick disk (Kroupa 2002). In addition, at all evolutionary phases, the tidal forces of the early Galaxy could disrupt dwarf satellite galaxies, capturing their stars, and, as predicted by the corresponding theoretical model (Quinn *et al.* 1993), it is then unlikely that a vertical gradient in chemical composition will arise in its thick disk. The latter assumption is supported by the possible extragalactic origin of the Arcturus moving group, almost all the stars of which belong to the thick disk.

To concretize the model, let us analyze the relation between the chemical composition and other parameters of the subsystem's stars.

RELATION OF THE IRON AND MAGNESIUM ABUNDANCES TO ORBITAL ELEMENTS AND AGE

Having reconstructed the Galactic orbits of the stars, we can trace the variations in metallicity and relative magnesium abundance in the subsystem with maximum distance of the stars from the Galactic center and plane. Figure 3a shows the $Z_{\max}$–[Fe/H] diagram for all thick-disk stars. The thirteen metal-poor tail stars with [Fe/H] < -1.25 are highlighted by the open triangles and are not used to determine the gradients. The regression line constructed for the remaining stars exhibits a fairly large negative vertical metallicity gradient: $\mathrm{grad}_Z[\mathrm{Fe/H}] = -0.13 \pm 0.04$ kpc^{-1} at $r = -0.27 \pm 0.08$. The result is stable, since even eliminating the four most distant stars lying farther than 2 kpc does not change the slope of the regression line at all. The correlation is highly significant, since the probability of random occurrence of the correlation is $P_N \ll 1\%$. The vertical gradient in the thick disk derived from the nearest stars is equal, within the error limits, to the value of -0.2 ± 0.1 kpc^{-1} obtained from RR Lyrae variable field stars (see Borkova and Marsakov 2002). The thick-disk stars also exhibit a vertical gradient in relative magnesium abundance (see Fig. 3b): $\mathrm{grad}_Z[\mathrm{Mg/Fe}] = (0.06 \pm 0.02)$ kpc^{-1} at $r = 0.33 \pm 0.08$ and at a probability of random occurrence of the correlation $P_N \ll 1\%$; i.e., the correlation can be recognized to be highly significant. Here, removing the four most distant stars changes neither the correlation coefficient nor the gradient. Note that the correlations in Figs. 3a and 3b stem exclusively from the existence of two groups of points that are separated by a distinct deficit of stars near $Z_{\max} \approx 0.5$ kpc; it is not quite clear why it arose (see also the $Z_{\max}$ distribution of stars in Fig. 6a). There are no correlations inside each group, but they differ in both mean magnesium abundance and metallicity. Therefore, if we had used a more stringent criterion and did not include low-orbit stars in the subsystem (in Fig. 1, this roughly corresponds to the satisfaction of the condition $\sqrt{Z_{\max}^2 + 4e^2} > 0.8$ for the thick disk),

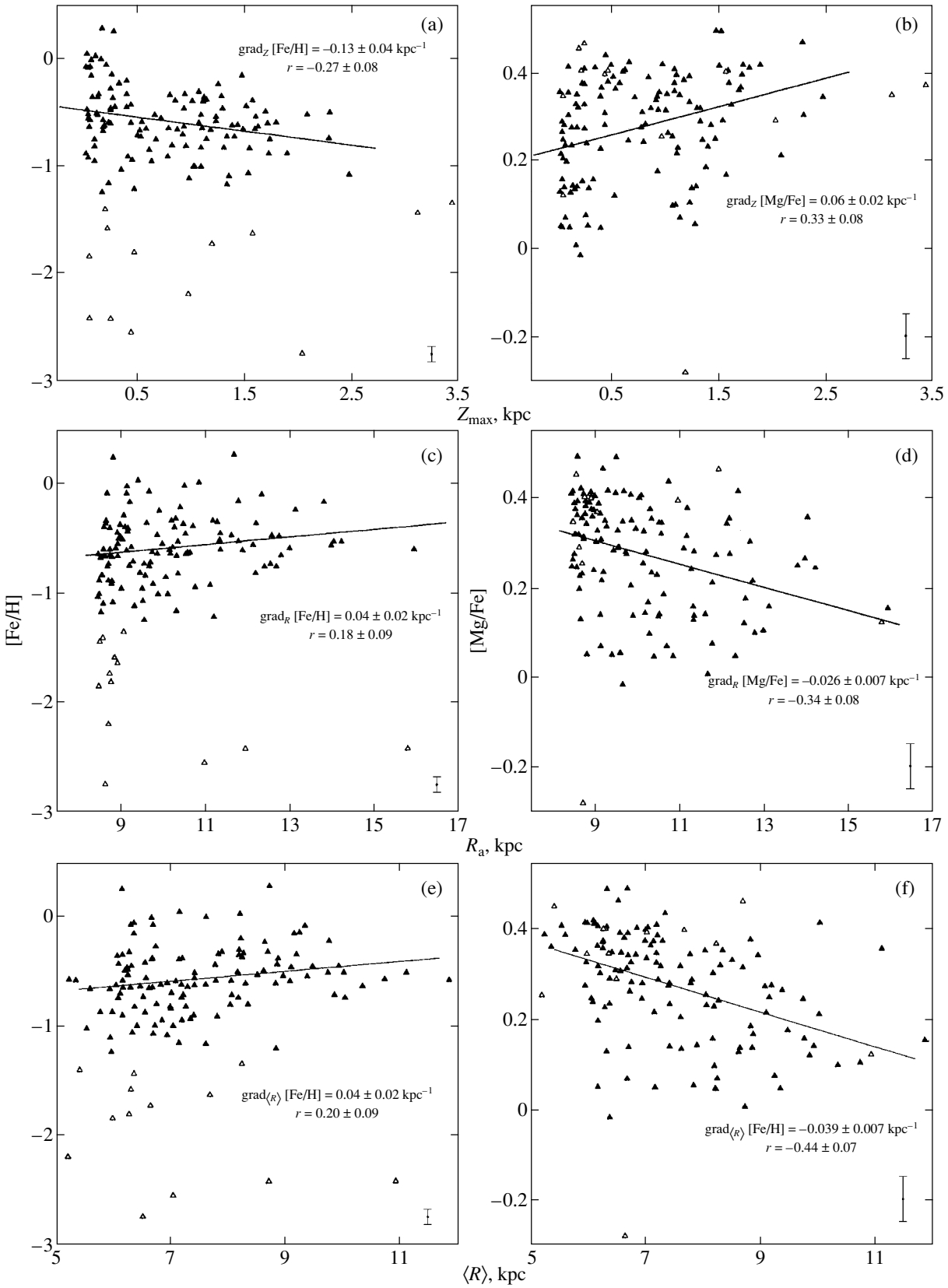


Fig. 3. Relations of the metallicity and relative magnesium abundance in the atmospheres of thick-disk stars to their maximum distance from the Galactic plane (a), (b); to the apogalactic distance (c), (d); and to the mean orbital radius (e), (f). The notation is the same as that in Fig. 1. The solid lines represent the regression lines for the thick-disk stars. The gradients and correlation coefficients are shown. The error bars are given.

then the two vertical gradients in it would disappear. However, the high orbital eccentricities ($\langle e \rangle = 0.33 \pm 0.01$), residual velocities ($\langle V_{\rm res} \rangle = 97 \pm 2$ km s^{-1}), and ages ($\langle t \rangle = 9.7 \pm 0.9$ Gyr) and the low azimuthal space velocity ($\langle \Theta \rangle = 190 \pm 6$ km s^{-1}) even at low magnesium abundance ($[{\rm Mg/Fe}] = 0.21 \pm 0.02$) in the excluded stars do not allow them to be attributed to the thin disk. Therefore, some of the authors (e.g., Furmann 1998; Bensby *et al.* 2003) did not include them in any of the disk subsystems, calling them transitional stars. Thus, the question as to whether a vertical gradient in chemical composition exists in the thick disk will be solved only after it will be established which subsystem this group of stars belongs to.

The radial metallicity gradient in the thick disk (Fig. 3c) was found to be positive rather than negative, but almost indistinguishable from zero outside the error limits ($r = 0.18 \pm 0.09$). Since the probability of random occurrence of this correlation is $P_N = 5\%$, it cannot be recognized as statistically significant. The RR Lyrae variable stars in the thick disk reveal no radial metallicity gradient at all (Borkova and Marsakov 2002). In contrast, the radial gradient in relative magnesium abundance is not only nonzero outside the 3σ limits, but also is negative, instead of the expected increase in the relative magnesium abundance with increasing apogalactic orbital radius of the stars: ${\rm grad}_{R_{\rm a}}[{\rm Mg/Fe}] = -0.026 \pm 0.007$ kpc^{-1} at $r = -0.34 \pm 0.08$ and at a probability of random occurrence of the same correlation $P_N \ll 1\%$ (see Fig. 3d). Note that the corresponding radial gradients and correlation coefficients in Figs. 3e and 3f were found to be almost the same when using the mean orbital radii of the stars. The latter are believed to be less subject to variations during the stellar lifetime and, therefore, are commonly used to find the initial radial metallicity gradients. Remarkably, eliminating the stars with $\sqrt{Z_{\rm max}^2 + 4e^2} < 0.8$ from the subsystem has virtually no effect on the results, since stars with any orbital radii are eliminated more or less equiprobably. We clearly see from the diagrams that the density of stars with a high magnesium abundance increases continuously with decreasing galactocentric distance, and the overwhelming majority of them are found near the solar orbital radius, while the density of magnesium-poor stars in the diagrams is almost independent of both $R_{\rm a}$ and $\langle R \rangle$. Therefore, even if we reject the seven most distant stars in the diagrams (with $R_{\rm a} > 13$ kpc in Fig. 3d and with $\langle R \rangle > 10$ kpc in Fig. 3f), both the slopes of the regression lines and the correlation coefficients will only increase. The relative numbers of stars with different metallicities in the [Fe/H]$-R_{\rm a}$ (Fig. 3c) and [Fe/H] $-\langle R \rangle$ (Fig. 3e), where positive correlations are observed instead of the expected negative correlation, behave in a similar way. This gradient inversion is most likely caused by the selection effect related to the predominantly smaller perigalactic orbital radii and larger eccentricities for the metal-poorer and, simultaneously, magnesium-richer stars in the solar neighborhood (see the correlations in the diagrams of Fig. 4). Indeed, the probability of finding a star near its apogalactic orbital radius is highest. Therefore, among the thick-disk stars with fairly high orbital eccentricities ($\langle e \rangle = 0.34$ at $\sigma_e = 0.13$), we will see a larger fraction of slow stars near the apogalactic radius equal to the solar orbital radius than stars with large $R_{\rm a}$ and $\langle R \rangle$ passing rapidly by the Sun. At the same time, the relative velocities for the stars with the perigalactic radii equal to the solar orbital radius are low; therefore, among the stars with large $R_{\rm a}$ and $\langle R \rangle$ in the solar neighborhood, the fraction of such stars is larger. Since both the magnesium abundance ($r = -0.44 \pm 0.07$ at $P_N \ll 1\%$) and, to a small degree, the metallicity ($r = 0.18 \pm 0.09$ at $P_N = 5\%$) in the thick disk correlate with $R_{\rm p}$ (see Figs. 4a and 4c, respectively), magnesium-poorer and, simultaneously, metal-poorer stars will prevail at large $R_{\rm a}$. The correlation between magnesium abundance and eccentricity ($r = 0.34 \pm 0.08$ at $P_N \ll 1\%$) in Fig. 4b also contributes to the enhancement of this selection. In Fig. 4d, metallicity does not correlate with eccentricity; nevertheless, its sign is negative: $r = -0.11 \pm 0.09$ at $P_N = 28\%$. We emphasize that both correlations with the magnesium abundance should be recognized to be highly significant, and eliminating the low-orbit stars from the subsystem does not change the result. Thus, the usually declared absence of a radial metallicity gradient and the existence of the negative (inverse) radial magnesium abundance gradient in the thick disk found above can be largely attributed to the selection related to the observations of stars in a limited region of Galactic space.

The distributions of the stars that are currently in the local volume of the Galaxy in orbital eccentricity and maximum distance from the Galactic plane are less dependent on the observer's location, i.e., the Sun, than their distributions in maximum and minimum Galactocentric orbital radii. Therefore, these orbital elements may be expected to also correlate with stellar age. We see from Fig. 5 that the two parameters indeed exhibit a correlation with age. However, the $e(t)$ correlation is actually significant (the correlation coefficient is $r = 0.35 \pm 0.10$ at $P_N = 2\%$), while the $(t-Z_{\rm max})$ diagram exhibits no correlation, but it clearly shows the absence of large distances from the Galactic plane among the young stars (see the empty upper left corner in Fig. 5b separated by the dotted line). The magnesium abundance also correlates with age ($r = 0.27 \pm 0.11$ at $P_N = 2\%$, i.e.,

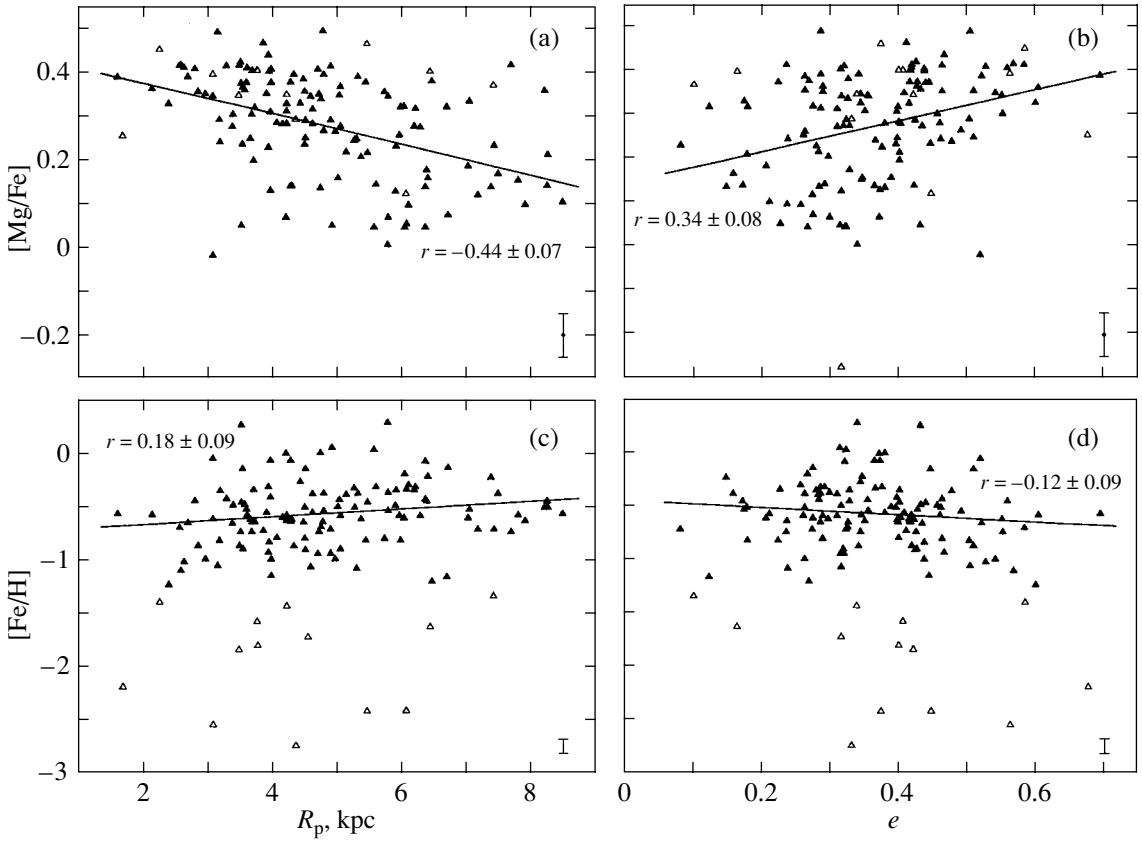


Fig. 4. Relations of the metallicity and relative magnesium abundances in the thick-disk stars to their perigalactic distances (a), (c) and orbital eccentricities (b), (d). The correlation coefficients are shown. The error bars are given.

the correlation should still be recognized to be significant), while the metallicity does not correlate with age at all (see Figs. 5c and 5d, respectively). Eliminating the low-orbit stars changes nothing. Let us consider the t–[Mg/Fe] diagram in more detail. According to Fuhrmann (2000), there can be no stars younger than ≈10 Gyr in the thick disk, but they are present in our diagram. The larger triangles in the diagram denote the stars with age errors of less than 2 Gyr. We see that the number of stars decreased by a factor of 2, but the overall structure of the diagram did not change and the correlation was retained. The stars in the diagram are naturally separated into two groups by the dotted line drawn almost perpendicularly to the regression line. The groups of different ages turned out to differ in both mean space velocity and galactic orbital elements. In particular, their mean residual velocities $\langle V_{\rm res}\rangle = 95 \pm 3$ km s^{-1} and 116 ± 3 km s^{-1}, dispersions $\langle \sigma_{V_{\rm res}}\rangle = 13 \pm 2$ km s^{-1} and 25 ± 2 km s^{-1}, azimuthal velocities $\langle \Theta\rangle = 200 \pm 7$ and 156 ± 4 km s^{-1}, orbital eccentricities $\langle e\rangle = 0.31 \pm 0.01$ and 0.36 ± 0.01, and the maximum distances of orbital points from the Galactic plane $\langle Z_{\rm max}\rangle = 0.55 \pm 0.10$ and 0.87 ± 0.07 kpc, respectively, for the young and old groups. A subgroup of six stars with a high magnesium abundance (the metal-poor tail stars were not included) stands somewhat apart from the stars with young ages. The mean values of its kinematic parameters listed above are almost equal to those for the main subgroup of young stars. The subgroups differ only in magnesium abundance and metallicity: for the subgroup of six stars, [Fe/H] $= -0.74 \pm 0.10$dex is almost equal to the metallicity of the old group -0.63 ± 0.03dex, while for the main young subgroup, it is -0.35 ± 0.05dex. The fact that 8 of the 24 stars in the young, magnesium-poor subgroup have an azimuthal velocity $\Theta > 220$ km s^{-1} and $R_a > 12$ kpc was unexpected. There are virtually no such stars rapidly revolving around the Galactic center in the old group. Judging by their characteristics, these stars could well be attributed to the thin disk if it were not for the larger $Z_{\rm max} > 1$ kpc for half of them and $e > 0.3$, which is atypical of the thin-disk stars, for the remaining stars. Thus, it seems that the thick disk consists of two discrete populations that differ in ages, elemental abundances, angular momenta, space velocity dispersions, and orbital elements. However, the

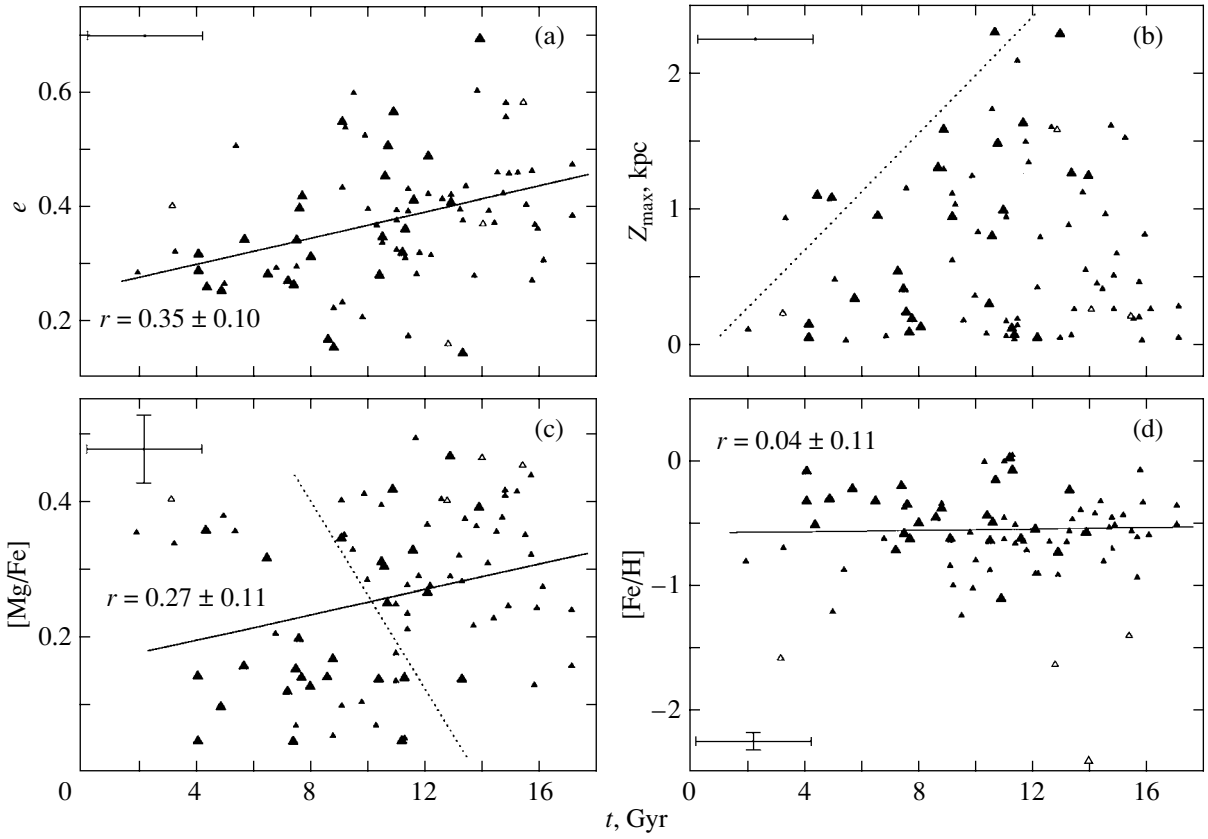


Fig. 5. Relations of the orbital eccentricities (a), the maximum distances of orbital points from the Galactic plane (b), and the relative magnesium abundances (c) and metallicities (d) in thick-disk stars to the age. The larger triangles denote the stars with age errors of less than 2 Gyr. The solid lines represent the regression lines. The correlation coefficients are shown. See the text for an explanation of the dotted lines in panels (b) and (c). The error bars are given.

belonging of the young group to the thick disk requires additional proofs.

The scale height is one of the most important parameters of the subsystem. It characterizes the subsystem's thickness and is equal to the height above the Galactic plane at which the density of its objects decreases by a factor of e. This parameter can also be determined from nearby stars by fitting their Z_{max} distribution by an exponential law. The following should be taken into account: first, the kinematic selection, since the probability of detecting a star at a given time near the Sun decreases with its increasing relative velocity; and, second, we must not just include each star in the histogram with its Z_{max}, but, as it were, "spread" it over the orbit from $-Z_{max}$ to $+Z_{max}$ proportional to the probability density of finding it at different Z. This is because all stars of the subsystem cannot be simultaneously at the maximum distances from the Galactic plane. To allow for the kinematic selection, we assigned a weight to each star proportional to its residual velocity with respect to the local standard of rest, V_{res}. The initial Z_{max} histogram and the reconstructed actual Z distribution of thick-disk stars are shown in Fig. 6. The solid curve in Fig. 6b represents a function of the form $n(Z) = C\exp(-Z/Z_0)$, where Z_0 is the scale height (for more details on the procedure for reconstructing the height distribution of stars, see Marsakov and Shevelev (1995)). The scale height for the thick disk was found from the stars of our sample to be $Z_0 = 0.6 \pm 0.1$ kpc, where the uncertainty was estimated from the maximum and minimum values of Z_0 obtained by eliminating the extreme columns of the histogram. The derived scale height is slightly smaller than its value that we determined also from the RR Lyare variable field stars: $Z_0 = 0.74 \pm 0.05$ kpc (see Borkova and Marsakov 2002), while the latter matches the estimates of other authors (in particular, Robin *et al.* (1996) gave a value averaged over several source of $Z_0 = 0.76 \pm 0.05$ kpc). The size of the sample used must be increased significantly to reliably estimate the scale height for the subsystem. Note also that the scale height slightly increases if the low-orbit stars are eliminated from the subsystem.

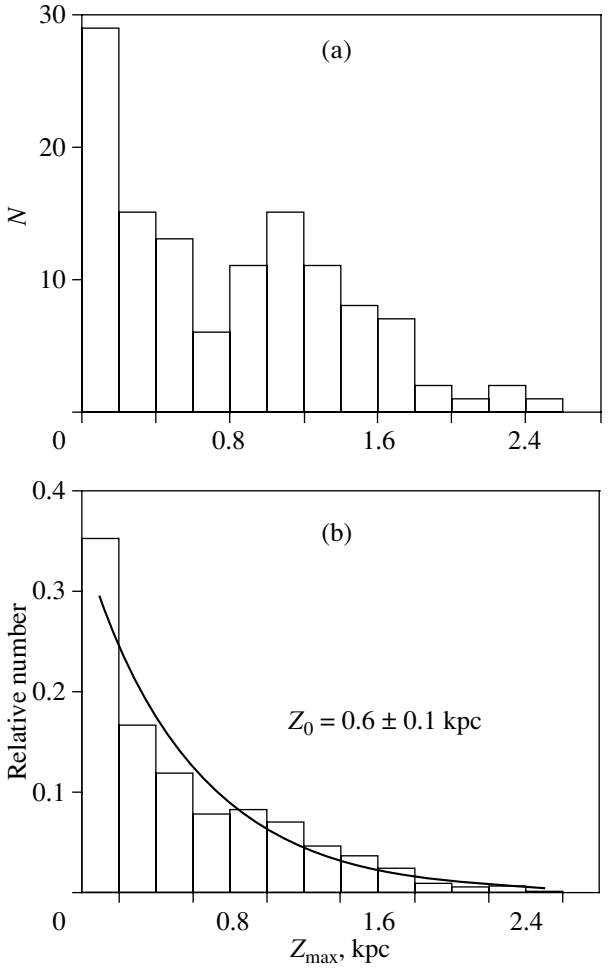


Fig. 6. (a) Distribution in maximum distance from the Galactic plane and (b) reconstructed Z distribution for thick-disk stars. The solid curve indicates the exponential fit to the distribution, and the scale height is indicated together with its uncertainty.

The mean velocity components and their dispersions calculated from all 133 thick-disk stars of our sample are: $\langle U\rangle = 4 \pm 7$ km s^{-1}, $\sigma_U = 80 \pm 5$ km s^{-1}; $\langle V\rangle = -49 \pm 4$ km s^{-1}, $\sigma_V = 46 \pm 3$ km s^{-1}; $\langle W\rangle = 2 \pm 4$ km s^{-1}, and $\sigma_W = 45 \pm 3$ km s^{-1}. They are in satisfactory agreement with the determinations of other authors even in the case of a radical difference between the initial prerequisites in forming the samples of thick-disk stars. For example, Soubiran *et al.* (2003) obtained $(\sigma_U, \sigma_V, \sigma_W) = (63 \pm 6, 39 \pm 4, 39 \pm 4)$ km s^{-1} for a sample of ≈400 thick-disk red giants lying toward the North Galactic Pole. On the other hand, Gilmore and Wyse (1987) give the following velocity dispersions for the thick-disk field stars: (70, 50, 45) km s^{-1}.

THE FORMATION TIME SCALE OF THE THICK DISK

Thus, the trends that we found here strongly suggest that the formation time scale of the thick disk was rather long and in no way fits into the scenario of a rapidly collapsing protogalaxy (Eggen *et al.* 1962). The formation of the first stars in it began long before the onset of mass type-Ia supernovae explosions when the metallicity had not yet reached a turning value of [Fe/H] ≈ −1.0. Since the evolution time scale of close binary stars that end their lives with SNe Ia explosions is ≈(1−2) Gyr, the formation of the subsystem began before the Galaxy reached this age. By this time, the young Galaxy already had an almost formed spherical subsystem. By studying the variations in the abundances of r- and s-elements in thick-disk stars with magnesium abundance [Mg/Fe] > 0.25, Mashonkina *et al.* (2003) showed that the overwhelming majority of thick-disk stars were formed within 1.1−1.6 Gyr after the protogalactic cloud began to collapse. However, the last stars in the thick disk turned out to have been born after the formation of a younger subsystem, the thin disk, began. This is evidenced by our largely selection-free sample that contains an appreciable number of stars with thick-disk kinematics, but lying in the ranges of metallicities and relative magnesium abundances extending to their solar values. Indeed, according to the isochrone ages, these stars are no older than the Sun. The angular momentum difference between the halo and the thick disk suggests the absence of a smooth transition between these two subsystems (Wyse and Gilmore 1992). A long delay of star formation after an active phase in the evolution of the Galaxy associated with mass supernova explosions (see Berman and Suchkov 1991) could be responsible for this difference. This explanation simultaneously resolves the contradiction with the short time scale of free collapse of the protogalactic cloud (≈0.4 Gyr), and the age at which the mass formation of thick-disk stars begins could prove to be even older than 1 Gyr. In other words, the delay of star formation lengthens the time frames of the natural processes of collapse and chemical enrichment of the interstellar medium.

The vertical gradients in metallicity and magnesium abundance that we found as well as the correlations between the magnesium abundances (and, possibly, also metallicities) and the perigalactic distances and the magnesium abundances and the orbital eccentricities naturally fit into this picture. However, the question arises as to whether the stars with thick-disk kinematics, but lying in low orbits, belong to the subsystem, since, if they do not belong to the thick disk, there are no vertical gradients in it. In this case, the model of the subsystem's formation through

the interaction of the Early Galaxy with its satellites becomes of current interest. The stars of the so-called metal-poor thick-disk tail do not follow the patterns related to [Fe/H] and exhibited by all the remaining stars of the subsystem. In addition, they show an anomalously large spread in [Mg/Fe]. Therefore, we believe that they were most likely formed inside isolated interstellar clouds that were enriched with chemical elements according to scenarios different from the evolution of the actively mixing main body of the interstellar matter of a single protogalactic cloud. Further studies and data on the abundances of other chemical elements in the subsystem's stars are required to refine the thick-disk formation time scale and model.

In our next papers, we will consider the properties of stars in other Galactic subsystems and further supplement the existing database of stars in the solar neighborhood with published determinations of the abundances of various chemical elements.

ACKNOWLEDGMENTS

We are grateful to an anonymous referee for valuable remarks that forced us to present our results in a more reasoned way.

REFERENCES

1. C. Allen and A. Santillan, Rev. Mex. Astron. Astrofis. **22**, 255 (1991).
2. M. Barbier-Brossat and P. Figon, Astron. Astrophys., Suppl. Ser. **142**, 217 (2000).
3. T. Bensby, S. Feldsing, and I. Lundstrem, Astron. Astrophys. **410**, 527 (2003).
4. B. G. Berman and A. A. Suchkov, Astrophys. Space Sci. **184**, 169 (1991).
5. J. Bernkopf, A. Fiedler, and K. Fuhrmann, in *Astrophysical Ages and Times Scales*, Ed. by T. von Hippel, Ch. Simpson, and N. Manset (Astron. Soc. Pac., San Francisco, 2001); Astron. Soc. Pac. Conf. Ser. **245**, 207 (2001).
6. T. V. Borkova and V. A. Marsakov, Astron. Zh. **77**, 750 (2000) [Astron. Rep. **44**, 665 (2000)].
7. T. V. Borkova and V. A. Marsakov, Astron. Zh. **79**, 510 (2002) [Astron. Rep. **46**, 460 (2002)].
8. T. V. Borkova and V. A. Marsakov, Astron. Zh. **82**, 453 (2005) [Astron. Rep. **49**, 405 (2005)].
9. A. Burkert, J. W. Truran, and G. Hensler, Astrophys. J. **391**, 651 (1992).
10. C. Chiappini, F. Matteycci, and R. G. Gratton, Astrophys. J. **477**, 765 (1997).
11. O. J. Eggen, Astron. J. **115**, 2397 (1998).
12. O. J. Eggen, D. Linden-Bell, and A. Sandadge, Astrophys. J. **136**, 765 (1962).
13. S. Feldzing, T. Bensby, and I. Lundstrom, Astron. Astrophys. **397**, L1 (2003).
14. K. Furmann, Astron. Astrophys. **338**, 183 (1998).
15. K. Furmann, in *The First Stars. Proceedings of the MPA/ESO Workshop, Garching, Germany, 1999,* Ed. by A. Weiss, T. G. Abel, and V. Hill (Springer-Verlag, Berlin, 2000), p. 68.
16. K. Furmann, M. Axer, and T. Gehren, Astron. Astrophys. **301**, 492 (1995).
17. G. Gilmore and R. F. G. Wyse, in *The Galaxy. Proceedings of the NATO Advanced Study Institute, Cambridge, England, 1986,* Ed. by G. Gilmore and B. Carswell (Reidel, Dordrecht, 1987), pp. 247–279.
18. R. Gratton, E. Carretta, F. Matteucci, and C. Sneden, Astron. Astrophys. **358**, 671 (2000).
19. P. Kroupa, Mon. Not. R. Astron. Soc. **330**, 707 (2002).
20. V. A. Marsakov and Yu. G. Shevelev, Astron. Zh. **72**, 321 (1995) [Astron. Rep. **39**, 284 (1995)].
21. V. A. Marsakov and A. A. Suchkov, Astron. Zh. **54**, 1232 (1977) [Sov. Astron. **21**, 700 (1977)].
22. L. Mashonkina, T. Gehren, C. Travaglio, and T. Borkova, Astron. Astrophys. **397**, 275 (2003).
23. F. Matteucci, *The Chemical Evolution of the Galaxy* (Kluwer, Dordrecht, 2001), Astrophys. and Spase Sci. Library, Vol. 253.
24. F. Matteucci and I. Greggio, Astron. Astrophys. **154**, 279 (1986).
25. J. F. Navarro, A. Helmi, and K. C. Freeman, Astrophys. J. **601**, L43 (2004).
26. D. L. Nidever, G. W. Marcy, R. P. Butler, *et al.*, Astrophys. J., Suppl. Ser. **141**, 503 (2002).
27. B. Nordstrem, M. Mayor, J. Andersen, *et al.*, Astron. Astrophys. **418**, 989 (2004).
28. J. Norris, M. Bessel, and A. J. Piscles, Astrophys. J., Suppl. Ser. **58**, 463 (1985).
29. J. X. Prochaska, S. O. Naumov, B. W. Carney, *et al.*, Astron. J. **120**, 2513 (2000).
30. P. J. Quinn, L. Hernquist, and D. P. Fullagar, Astrophys. J. **403**, 74 (1993).
31. K. U. Ratnatunga, J. N. Bahcall, and S. Casertano, Astrophys. J. **339**, 106 (1989).
32. A. C. Robin, M. Hauwood, M. Creze, *et al.*, Astron. Astrophys. **305**, 125 (1996).
33. N. N. Shimanskaya, L. I. Mashonkina, and N. A. Sakhibullin, Astron. Zh. **77**, 599 (2000) [Astron. Rep. **44**, 530 (2000)].
34. C. Soubiran, O. Bienaume, and A. Siebert, Astron. Astrophys. **398**, 141 (2003).
35. F.-K. Thielemann, M. Hashimoto, and K. Nomotto, Astrophys. J. **349**, 222 (1990).
36. T. Tsujimoto, K. Nomoto, Y. Yoshii, *et al.*, Mon. Not. R. Astron. Soc. **277**, 945 (1995).
37. R. F. G. Wyse and G. Gilmore, Astron. J. **104**, 144 (1992).
38. R. Zinn, Astrophys. J. **293**, 424 (1985).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 528–536. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 592–600.

Universal Mechanism of Thermonuclear Explosion for Single (Type I + 1/2) and Binary (Type Ia) Presupernovae: The History and Prospects

V. S. Imshennik and N. V. Dunina-Barkovskaya*

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

Received March 18, 2005

Abstract—We discuss the history and prospects for the one-dimensional models of thermonuclear explosions in carbon–oxygen stellar cores. In connection with the recently studied SN 2002ic, which combines the properties of classical type Ia and IIn supernovae, we hope that our delayed detonation mechanism is applicable not only to binary, but also to single presupernovae. Since a large amount of ^{56}Ni is synthesized, it can also describe adequately the light curves of supernovae similar to SN 2002ic.

Key words: *supernovae, supernova remnants.*

INTRODUCTION

In connection with the recently studied SN 2002ic (Hamuy *et al.* 2003), which combines the signatures of type II and I supernovae (SNe), i.e., the presence of an observable hydrogen envelope and a theoretically possible thermonuclear explosion mechanism, the possibility of a thermonuclear explosion for single asymptotic-giant-branch (AGB) stars is again of current interest. Such an explosion is possible within the framework of the universal thermonuclear explosion mechanism for carbon–oxygen (CO) stellar cores suggested in our recent papers (Dunina-Barkovskaya and Imshennik 2000, 2003; Dunina-Barkovskaya *et al.* 2001). In this paper, we first briefly review the corresponding models of the 1970–1980s and critically analyze their drawbacks (despite the well-known advantages in developing the deflagration regime of thermonuclear burning). Subsequently, in the spirit of partly rectifying these drawbacks, we consider our works of the early 2000s in which results similar to the previous ones (pulsational deflagration with delayed detonation) were obtained by taking into account the evolution of the presupernova's CO core and convection in the mixing-length approximation (at both the evolutionary and explosion phases). In conclusion, we attempt to use the most recent results to theoretically interpret the thermonuclear explosions of AGB stars.

*E-mail: dunina@itep.ru

CRITICAL REVIEW

Remarks on a Paper by Ivanova et al. (1974)

(1) Under the influence (at that time) of a paper by Bruenn (1972), who considered the detonation self-consistency effects by Arnett (1969) that led to a weakening of detonation to the point of the formation of a gravitationally bound remnant at initial central densities $\rho_{c0} > (1.5–2) \times 10^{10}$ g cm^{-3}, the conclusions in our paper shifted in favor of such a weakening due to pulsations (almost the entire discussion of our results).

(2) Nowhere was it emphasized that, due to the self-consistency of the burning and hydrodynamic processes, subsonic burning, i.e., deflagration (this term is absent in the paper), was substantiated for the first time: after all, the pulsations themselves were just the result of deflagration. In addition, it was shown that a transition to detonation after a number of pulsations (delayed detonation) is possible at a fairly high initial temperature "seed," which was chosen quite arbitrarily.

(3) Nothing was said about the role of a finite boundary pressure $P_b \neq 0$ that could actually imitate not only the outer envelope with radius R_{00}, but also the accretion of matter ($P_b = P_0(R_{00}/R)^4$, where $P_0 \sim \Delta M_{ac}$).

(4) Deflagration and the role of the initial temperature distribution in its initiation were clearly discussed in a review by Imshennik and Nadezhin (1982) ($M = 1.4\ M_\odot$, $10^9 < \rho_{c0} < 3 \times 10^{10}$ g cm^{-3}). These authors also presented the velocity profiles for the

1063-7737/05/3108-0528$26.00

erroneous model of delayed detonation with $\rho_{c0} = 5.03 \times 10^9$ g cm^{-3} in graphical form (according to the theory of evolution, a rise in the central density is possible at a low accretion rate), while Chechetkin *et al.* (1977) in their review paper presented a collapsing model and established its boundary, $\rho_{c0} \simeq 4 \times 10^9$ g cm^{-3}. In their review paper, Imshennik and Nadezhin (1982) attached fundamental importance to a nonzero boundary pressure, in particular, as a means of constructing initial models with a density ρ_{c0} in the above range of interest.

Remarks on a Series of Papers by Ivanova et al. (1977a, 1977b)

(1) The physical formulation of the problem of the thermonuclear explosion of a carbon core with a given mass $M \lesssim M_{Ch}$ and an arbitrary central density ρ_{c0} (with the corresponding external boundary pressure P_b) to include the secondary nuclear reactions (in the approximation of nuclear statistical equilibrium (NSE)), the kinetics of β-interactions, and the neutrino energy losses was set out in great detail in a paper by Ivanova *et al.* (1977a). The kinetics of the thermonuclear reaction $C^{12} \rightarrow Ni^{56}$ was brought into coincidence with the high-temperature region $T > T_{cr} \simeq 3 \times 10^9$ K, where the NSE conditions are assumed to be satisfied and all of the possible β-interactions are included. All these changes are an improvement of a paper by Ivanova *et al.* (1974), while the key points were highlighted more clearly in the Introduction and the deflagration regime of burning was declared.

(2) The considerable significance of an oxygen admixture in the initial state and the convection and heat conduction processes at the phase of a thermal flash was mentioned, but these effects had not yet been included in the formulation of the problem! A nonzero external pressure was again attributed to the influence of outer envelopes (rather than accretion, since the scenarios of thermonuclear supernova explosions through accretion in binary systems became popular only in the 1990s); it was again (Ivanova *et al.* 1977a) declared that "It is quite clear that the tenuous outer envelope plays no significant role in the thermonuclear explosion of a star," which could be even erroneous. This "small" boundary pressure (see the plots of P_c and ρ_c) could be important due to the proximity to the outer boundary of the fuel ignition zone.

(3) In their thermonuclear explosion calculations for $\rho_{c0} = 2.33 \times 10^9$ g cm^{-3} and $M = 1.40 M_\odot$, Ivanova *et al.* (1977b) reached the pessimistic conclusion that no total expansion of a carbon stellar core with SN-Ia-scale energy release is possible for a realistic (from the standpoint of previous rough estimates of the role of convection) initial temperature (its seed) profile: for a seed of $\Delta m/M \simeq 3 \times 10^{-2}$, the expansion after five pulsations had a total energy of 1.3×10^{50} erg (and a mean velocity of 3×10^8 cm s^{-1}). Instead of criticizing this convection estimate, the authors discussed a different path, a severalfold increase in the central density, which subsequently showed up in a review by Imshennik and Nadezhin (1982), where the main model with a "good" explosion had $\rho_{c0} = 5.03 \times 10^9$ g cm^{-3}. However, there is a hint at noncentral ignition of the thermonuclear reaction through the initial temperature inversion in a paper by Ivanova *et al.* (1977b). The external pressure was estimated to be $P_b \sim 2 \times 10^{-2}$, while $P_c \sim 10^2$ ($[P] = 1.36 \times 10^{25}$ dyn cm^{-2}).

Remarks on a Paper by Ivanova et al. (1982)

(1) In a paper by Ivanova *et al.* (1982) on the structure of the detonation front, it would be better to use the results of the model with a low central density $\rho_{c0} \simeq 2 \times 10^9$ g cm^{-3} instead of $\rho_{c0} \simeq 5 \times 10^9$ g cm^{-3} when computing the structure of the deflagration burning front to better correspond to the evolutionary initial models by Dunina-Barkovskaya and Imshennik (2000, 2003) and Dunina-Barkovskaya *et al.* (2001) and to eliminate the effects of weak interactions. Note that in the latter mentioned papers, the growth rate of the CO core was assumed to be $\dot{M} = 5 \times 10^{-7} M_\odot$ yr^{-1}, which can correspond both to accreting white dwarfs in close binary systems (the substantiated case of SNe Ia) and to the growing CO core of a single AGB star (e.g., for SN 2002ic).

(2) We see that the calculation of the jump in pressure at the burning front Δp, to which the sought-for change in entropy ΔS is very sensitive, was too inaccurate:

$$\Delta S \propto \frac{(\Delta p)^3}{T_1 \rho_1^{8/3}}. \quad (1)$$

Unfortunately, the thickness of the burning front is still poorly resolved spatially; $\Delta r \simeq 10^6$ cm is grossly overestimated, but even this resolution is still unattainable, since the spatial step is, on average, $\Delta r_j \simeq 10^6$ cm, $R/j \simeq (2 \times 10^8)/(3 \times 10^2)$ cm $\simeq 10^6$ cm.

(3) The resolute opinion that the introduction of artificial convection in the equation is "unnecessary or even unnatural" (see the last note in the paper by Ivanova *et al.* (1982)) must be "abolished" precisely for the deflagration burning front. The latter, first, is actually blurred by turbulent convection (Lisewski *et al.* 2000) and, second, has no jumps in matter velocity when artificial viscosity acts to blur the shock front. (In fact, this was a barely noticeable footnote on p. 24.)

Critical Remarks on a Scenario of Supernova Explosions Through a Single Neutrino Ignition Mechanism Suggested in 1978–1983

The pretentious (with regard to the interpretation of the two types of supernovae) paper entitled "Types I and II Supernovae and the Neutrino Mechanism of Thermonuclear Explosion of Degenerate Carbon–Oxygen Stellar Cores" was published initially as a preprint (Chechetkin *et al.* 1978) and subsequently as a paper (Chechetkin *et al.* 1980). These works gave the full physical formulation of the problem, including all the significant neutrino processes, that sums up a large series of previous papers by the same authors, evidently from an extensive list of references. It can be said that allowance for the neutrino processes in the hydrodynamic deflagration–pulsational regime of thermonuclear CO-fuel burning begun in the paper by three authors (Ivanova *et al.* 1977a) was brought to perfection, though without including any neutrino self-absorption effects. Nevertheless, a strong explosion (with a total energy of 1.8×10^{51} erg, the supernova explosion scale) resulted from the treacherous artificial effect of an increase in the matter entropy in the case of developing large-amplitude CO-core pulsations. Unfortunately, this result for the model with an initial central density of $\rho_{c0} = 5.03 \times 10^9$ g cm^{-3} (note that the central density ρ_{c0} of the initial hydrostatic equilibrium CO-core configuration was the main specified parameter of the physical formulation of the problem) inspired the authors to consider CO cores with ρ_{c0} in a wide range of values ($10^9 < \rho_{c0} < 9 \times 10^9$ g cm^{-3}) as the progenitors for all SNe II. In the hydrodynamic calculations with this ρ_{c0} range, the total energy reached its maximum at $\rho_{c0} = 5 \times 10^9$ g cm^{-3}, and an explosion with total expansion of the CO core, i.e., without any gravitationally bound remnant, resulted, which definitely cannot be acceptable for all SNe II from the standpoint of observations of their remnants.

The methodical error made in the difference scheme was soon corrected (Ivanova *et al.* 1983). Several cases of the formulated problem were carefully recalculated; in the model with a density $\rho_{c0} \simeq 5 \times 10^9$ g cm^{-3}, the total explosion energy was 0.4×10^{50} erg, i.e., it was negligible compared to the observed supernova explosion energy. The result of the previous paper by Chechetkin *et al.* (1980) was erroneous not only quantitatively, but also qualitatively: in this paper, the detonation regime of thermonuclear burning arose after eight (!) pulsations and did not arise at all in the series of check calculations in 1983 at any parameter ρ_{c0} (the fraction of the burned CO fuel barely exceeded 0.6 even in the model with $\rho_{c0} = 3.5 \times 10^9$ g cm^{-3}. Figures 1–8 from Ivanova *et al.* (1983) were compared with Figs. 1–5 from Chechetkin *et al.* (1980). For the sake of justice, it should be noted that the suspicious rise in temperature at the contraction phase ahead of the burning front during the completion of the last (eighth) expansion to $T \simeq 10^9$ K (the CO-fuel detonation temperature) was pointed out specifically by Chechetkin *et al.* (1980), but it was erroneously attributed to the previously disregarded neutrino ignition effect. In fact, it was the result of an error in the calculations of the local entropy of the matter.

According to the above critical paper, which, unfortunately, was not published subsequently, the deflagration–pulsational regime that did not eventually transform to detonation proved to be essentially inconsistent with SNe I (to be more precise, SNe Ia, given the collapsing subtypes of type I supernovae, SNe Ib and SNe Ic, introduced in the past twenty years). It should be emphasized that the model with the lowest value of $\rho_{c0} = 2.34 \times 10^9$ g cm^{-3}, where the role of neutrino ignition was insignificant, compared to the previous hydrodynamic calculation (Ivanova *et al.* 1977b) remained unshakable. As was pointed out above, the total explosion energy in this paper was modest, 1.3×10^{50} erg. Note that in the 1978–1983 papers for the previous models with higher central densities, the decay energy of radioactive nuclides and the free neutron capture energy could have been added to the total explosion energy. However, even these roughly estimated additions gave 2×10^{50} erg, which does not change the above conclusions.

The calculations of collapsing CO cores with central densities for an initially hydrostatic equilibrium configuration $\rho_{c0} \gtrsim 9 \times 10^9$ g cm^{-3} must be interpreted quite differently. Chechetkin *et al.* (1980) with $\rho_{c0} = 9.22 \times 10^9$ g cm^{-3} showed that, given the new effect of neutrino ignition, the CO-core collapse, which was presented in the cited paper in the same detail (in Figs. 6–11) as the erroneous case of a strong explosion, occurs after short weak pulsations. Interestingly, neutrino ignition in the case of collapse ensures detonation (!) burning of the outermost layers of the CO core and their ejection with significant velocities $v \gtrsim 10^9$ cm s^{-1} that is subsequently capable of ejecting an envelope with a mass of $\sim 0.1 M_\odot$, if it is present, simultaneously with the collapse. This paper corrected significantly the results obtained previously by Ivanova *et al.* (1977c) for the models with $\rho_{c0} \gtrsim 4 \times 10^9$ g cm^{-3} corresponding to collapse. Nevertheless, the significant decrease in entropy attributable mainly to matter neutronization that was investigated in detail in the last mentioned paper received confirmation; the critical density (the boundary of collapsing and exploding CO cores) established by Chechetkin *et al.* (1980) differs markedly from

$\rho_{c0} \gtrsim 2 \times 10^{10}$ g cm^{-3} (Bruenn 1972). In fact, such a change draws attention to this exotic (in its low dimensionless entropy $S \leq 1.0$ (in units of k_B/m_0)) interval of collapsing stellar CO cores. Indeed, the Chandrasekhar upper limit for the central density is known to be $\rho_{c\,ch} = 3 \times 10^{10}$ g cm^{-3}; i.e., in the case of the upper limit by Bruenn (1972), it shrinks to an insignificant interval of central densities.

Neutrino Processes in the Thermonuclear Explosion of a Stellar CO Core: Remarks Mostly for Collapsing Stellar CO Cores

(1) Allowance for the bulk neutrino energy losses ε_ν, first using the formulas by Beaudet *et al.* (1967) in the paper by Ivanova *et al.* (1974) and then additionally including the URCA losses in the paper by Ivanova *et al.* (1977a) calculated by considering more fully the neutrino processes of the general kinetics of β processes, was formulated in detail in the last mentioned paper.

(2) As was shown by Ivanova *et al.* (1977b), allowance for the total neutrino energy losses at an initial central density of $\rho_{c0} = 2.34 \times 10^9$ g cm^{-3} plays no significant role in a thermonuclear explosion. However, in a subsequent paper, Chechetkin *et al.* (1980) analyzed in detail the neutrino processes (the general kinetics of β processes) and showed that their contribution to the hydrodynamic explosion process increases greatly with central density due to the neutronization of thermonuclear burning products, the iron-group elements. A sharp enhancement of the CO-fuel neutrino ignition ahead of the deflagration burning front with increasing density was also established in the cited paper, roughly speaking, for energy release (per unit mass) $\varepsilon_{ig} \propto \rho^{11/3}$. The interesting effect in the scattering of neutrinos emitted from the region behind the burning front by strongly degenerate electrons ahead of the front (the specific elementary ignition process), which consists in an increase in the scattering cross section by an order of magnitude, should also be emphasized.

(3) In previous papers (Arnett 1969; Bruenn 1972; and others) where the detonation regime of CO-fuel burning was assumed, the collapsing stellar CO cores occupied a narrow interval of central densities, 2×10^{10} g cm$^{-3} < \rho_{c0} < 3 \times 10^{10}$ g cm^{-3}. However, with the universally accepted transition to the deflagration regime of burning, this interval initially widened significantly, 4×10^9 g cm$^{-3} < \rho_{c0} < 3 \times 10^{10}$ g cm^{-3} (Chechetkin *et al.* 1977; Ivanova *et al.* 1977c) and subsequently, given the neutrino ignition effect, again slightly narrowed, 9×10^9 g cm$^{-3} < \rho_{c0} < 3 \times 10^{10}$ g cm^{-3} (Chechetkin *et al.* 1980), acquiring a prominent and intriguing role in the theory of thermonuclear explosions of stellar CO cores. It should be added to this important result that, as was already noted above, the collapse of a CO core with an enhanced central density was accompanied by a noticeable surface explosion and the ejection of outer layers with a mass up to $0.1M_\odot$ and an expansion energy of $\sim 10^{49}$ erg, while the gravitationally bound remnant itself had a very low entropy. Could these collapsars later become strange stars due to this entropy in contrast to those born during the collapses of iron stellar cores?

(4) For the relatively wide interval of initial central densities 10^9 g cm$^{-3} < \rho_{c0} < 9 \times 10^9$ g cm^{-3}, the deflagration regime of burning was ultimately confirmed in a preprint by Ivanova *et al.* (1983), though with an increasing role of neutrino processes as ρ_{c0} rises in this interval. However, the model with $\rho_{c0} \lesssim 2.0 \times 10^9$ g cm^{-3} remains unshakable in comparison with the paper by Ivanova *et al.* (1977b) and even the first paper on deflagration by Ivanova *et al.* (1974).

DELAYED DETONATION AND OBSERVATIONS OF THERMONUCLEAR SUPERNOVAE

Major Differences Between the New Series of Papers and the Previous Series of the 1970–1980s

Thus, a large series of works on the thermonuclear explosions of stellar CO cores was completed some 20 years ago. At present, we have managed to supplement and reconsider the model with a low central density, $\rho_{c0} \simeq 2 \times 10^9$ g cm^{-3}, by eliminating two major drawbacks of the previous series (Dunina-Barkovskaya and Imshennik 2000, 2003; Dunina-Barkovskaya *et al.* 2001). First, we included the evolutionary phase of a CO core in our calculations and began our calculations with reliable evolutionary models of low-mass CO cores. Therefore, the initial parameter ρ_{c0} was not specified arbitrarily (though in a fairly narrow interval, 2×10^9 g cm$^{-3} < \rho_{c0} < 3 \times 10^{10}$ g cm^{-3}), but was unambiguously obtained in the long phase of hydrostatic equilibrium evolution simulated using a hydrodynamic code, which allowed a smooth transition to an explosion to be made. We emphasize that our calculations were performed by including convection and the growth rate of the CO-core mass, which could correspond both to the accretion of matter in a close binary system of stars and to the growth of the CO core of a single AGB star through helium burning in its shell source. It is in this sense that the thermonuclear explosion mechanism considered in our recent series of papers is called universal. Second, the explosion per se was no longer provoked by a seemingly insignificant temperature "seed" near the center of the CO core, but was caused

by a gradual (evolutionary) rise in the density and temperature with allowance made for the ever increasing role of convection. Thus, in fact, the significant arbitrariness in choosing the initial conditions for the previous calculations of the thermonuclear explosion of a CO core has been removed both with regard to the choice of an initial central density ρ_{c0} and, in particular, with regard to specifying the temperature seed near the center of the stellar CO core.

Unfortunately, we had to quantitatively describe the convection, which, strictly speaking, requires using non-one-dimensional hydrodynamic calculations, in terms of a one-dimensional theory in the standard mixing-length approximation. This approximation, which is well known in the theory of stellar structure and evolution (the interpretation of red giant stars with convective outer envelopes), was used in the most general nonstationary representation, which, of course, is obligatory against the background of an intense hydrodynamic explosion process. This series of works (Dunina-Barkovskaya and Imshennik 2000, 2003; Dunina-Barkovskaya *et al.* 2001) was performed without including any neutrino processes, which is most likely admissible at the derived low matter densities, $\rho < \rho_{c0} \sim 2 \times 10^9$ g cm^{-3}.

An Attempt to Compare the Results of the New Series of Papers with the Observations of SN 2002ic and Similar Supernova

In connection with the above correction from the preprint by Ivanova *et al.* (1983), it should be particularly emphasized that the total explosion energy in the new series of works (Dunina-Barkovskaya and Imshennik 2000, 2003; Dunina-Barkovskaya *et al.* 2001) increased significantly and is $\sim 1.4 \times 10^{51}$ erg for the models in which pulsational deflagration was followed by detonation from the outer layers toward the center. It is important that the rate of increase in the mass of the CO core of a supernova's progenitor was chosen in a universal way in order that the mechanism of delayed detonation under consideration could be applied to both single and binary stars (Dunina-Barkovskaya *et al.* 2001). It is of no less importance that the external boundary conditions in the form of a radial dependence of the boundary pressure P_b closely match for both cases (Dunina-Barkovskaya *et al.* 2001).

A relatively large mass of radioactive nickel, about $1.2M_\odot$ (see Table 2 from Dunina-Barkovskaya and Imshennik (2003)), must be formed behind the detonation front in the above models. In this case, the entire mass traversed by the convergent detonation wave from its generation place to the deflagration burning front located at a mass of $0.16M_\odot$ from the center (see Table 2 in the cited paper) is taken as this mass. Over a long time of pulsational deflagration (dozens of pulsations with a total duration of several dozen seconds), this mass is certain to have turned into stable iron-group elements through β-processes, as was shown in detail previously (Ivanova *et al.* 1977a, 1977b). Next, given the possibility of incomplete thermonuclear CO-fuel burning in the outermost layers, because the thickness of the burning zone is comparable to the distance from the generation place of detonation to the surface of the model, according to the calculations of this burning zone by Imshennik and Khokhlov (1984), the mass of the synthesized ^{56}Ni can be roughly estimated to be no less than $1M_\odot$. A rough criterion that follows from the cited paper consists in the assertion that incomplete burning takes place in all of the outer layers of the CO core in which the density is less than 10^7 g cm^{-3}.

Such a density is obtained in the models for the detonation generation time at a mass of about $1.30M_\odot$ (from the center), while the detonation generation place corresponds to a mass of $1.33M_\odot$ (see Figs. 9–11 from Dunina-Barkovskaya *et al.* (2001)); i.e., detonation between these layers can take place with incomplete burning (in fact, up to the ^{28}Si-type elements (Imshennik and Khokhlov 1984)). Nevertheless, the above ^{56}Ni mass estimate of $1M_\odot$ is considerably larger than the typical (mean) mass of $0.6M_\odot$ in *bona fide* thermonuclear SNe Ia established during the calculations of light curves. It should be noted, however, that having analyzed the light curves for eight classical SNe Ia (without any hydrogen lines in the spectrum), Contardo *et al.* (2000) obtained the spread in presumed ^{56}Ni masses from $0.4M_\odot$ to $1M_\odot$, which, in the opinion of these authors, could be suggestive of a certain variety of explosion mechanisms. Our delayed detonation mechanism may be applicable to the supernovae that correspond to the upper boundary of this ^{56}Ni mass range, while the models obtained, for example, in the three-dimensional calculations by Reinecke *et al.* (2002), where the mass of the synthesized ^{56}Ni ranged from $0.526M_\odot$ to $0.616M_\odot$, are applicable to the supernovae in the middle of this range. It should be noted that a significant asymmetry was introduced in the initial explosion conditions in the model by Reinecke *et al.* (2002), which, of course, cannot be realized in our one-dimensional models, and the spread in ^{56}Ni masses in actual supernovae could also be attributable to different degrees of explosion asymmetry.

A detailed analysis of SN 2002ic (Hamuy *et al.* 2003) may suggest that single intermediate-mass ($M_{\rm ms} = 4{-}8M_\odot$) stars can still explode as thermonuclear supernovae at the AGB phase, but developing this hypothesis will require detailed calculations of the evolution of an AGB star at the presupernova phase with allowance made for the (possibly nonstationary) helium burning in its envelope and the accompanying mass loss.

Hamuy *et al.* (2003) convincingly estimated the luminosity rise in SN 2002ic due to the interaction of the CO core with the circumstellar hydrogen envelope (the SN/CSM interaction). Thus, these authors showed that light curves similar to those of ordinary SN Ia remain after subtracting the luminosity rise in the three photometric bands under consideration attributable to the SN/CSM interaction. However, for late times, the residual light curves of SN 2002ic begin to deviate seriously from the ordinary light curves. In our opinion, this deviation could be attributed to an enhanced abundance of radioactive nickel, which follows, for example, from our models. Unfortunately, there are no observational data on such late times when the light curves, roughly speaking, follow a simple law of radioactive ^{56}Co decay, while at a known distance to SN 2002ic $\approx$ 307 Mpc, according to Hamuy *et al.* (2003), if $H_0 =$ 65 km s^{-1} Mpc^{-1} in this paper is taken, they yield an estimate of the initial ^{56}Ni mass. As regards the universal thermonuclear explosion mechanism, it is also interesting to note that delayed detonation is obtained for a single presupernova in a wider range of the free model parameter, the mixing length in the convection theory. In fact, this was shown by Dunina-Barkovskaya and Imshennik (2003), who greatly reduced the boundary pressure of the CO core, which is natural only for a binary system rather than for a single AGB star in which the boundary pressure $P_{\rm b}$ imitates the presence of a hydrogen envelope. There is no doubt that a finite boundary pressure $P_{\rm b}$ plays a major role in the initiation of detonation from the surface.

Another peculiarity of the suggested universal mechanism, more specifically, the proximity of the synthesized ^{56}Ni to the outer edge of the exploded stellar CO core, can show up in supernovae observations similar to those performed by Hamuy *et al.* (2003). Above, we have already given the mass coordinates of the detonation generation place ($1.33M_\odot$), but even if we take into consideration the effect of incomplete CO-fuel burning (see above), the mass of the products of this incomplete burning is unlikely to exceed $M_{\rm CO} = 0.1M_\odot$. Clearly, this layer will relatively soon become transparent to the γ-ray emission of ^{56}Ni (and subsequently ^{56}Co) as the exploded CO core expands. Below, we give a rough estimate for the transparency time of this envelope by taking the effective opacity for the γ-ray emission under consideration to be $\kappa_\gamma = 0.03$ cm^2 g^{-1} (it may be assumed that this is the most adequate value, but we can specify κ_γ in the range $0.03 < \kappa_\gamma < 1.2$ cm^2 g^{-1}, as derived by Arnett and Fu (1989) (see also Imshennik and Popov (1992)).

Thus, let us write the column density of the partially burned matter in a spherical envelope screening the γ-ray emission as

$$L_{\rm CO}(t) = \int_{R_{\rm Ni}}^{R_{\rm CO}} \rho(t)dr \simeq \bar{\rho}(t)(R_{\rm CO} - R_{\rm Ni}) \simeq \bar{\rho}(t)R_{\rm CO}(t), \quad (2)$$

where $R_{\rm CO}$ and $R_{\rm Ni}$ denote the time-dependent radii of the outer and inner boundaries of this envelope, with $R_{\rm CO} \gg R_{\rm Ni}$ even at the very beginning of the expansion (see below), and $\bar{\rho}(t)$ is the mean density in this envelope. Clearly, the latter is

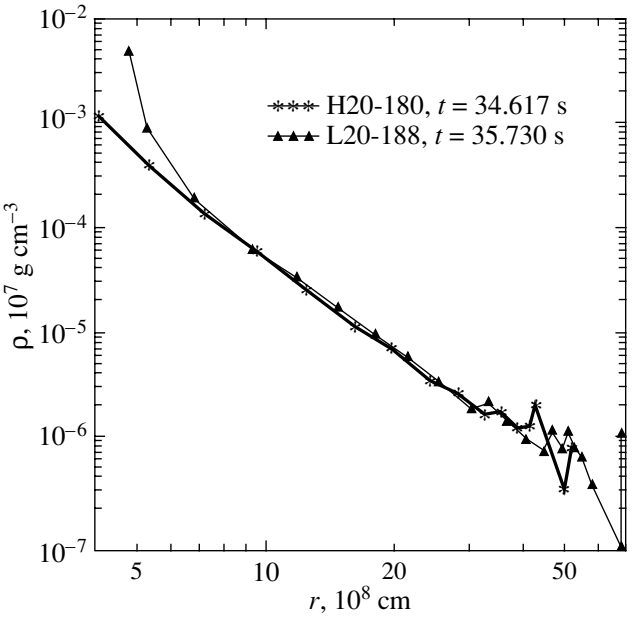
$$\bar{\rho}(t) = M_{\rm CO} \Big/ \left(\frac{4\pi}{3}R_{\rm CO}^3\right) \simeq M_{\rm CO}\left[\frac{3}{4\pi}(v_0 t)^{-3}\right], \quad (3)$$

where, in turn, $R_{\rm CO} = R_{\rm CO}^{(0)} + v_0 t \simeq v_0 t$, since another strong inequality, $v_0 t \gg R_{\rm CO}^{(0)}$, where $R_{\rm CO}^{(0)}$ is the initial value of this radius (according to our models, at the generation time of detonation from the edge), which is $R_{\rm CO}^{(0)} \simeq 4 \times 10^9$ cm (Dunina-Barkovskaya and Imshennik 2003), while $R_{\rm Ni}^{(0)} \simeq 2 \times 10^8$ cm, is soon satisfied. The following simple expression for the sought-for quantity can then be derived from (2) and (3):

$$L_{\rm CO}(t) \simeq \left(\frac{3M_{\rm CO}}{4\pi}\right)\frac{1}{(v_0 t)^2}. \quad (4)$$

The characteristic velocity of the matter of the supernova envelope, including the envelope under consideration, which contains absolutely no radioactive nickel and cobalt, the products of complete CO-fuel burning, can be expressed in terms of the total explosion energy $E_{\rm expl} = 1.4 \times 10^{51}$ erg and the total CO-core mass $M = 1.4M_\odot$: $v_0 = (2E_{\rm expl}/M)^{1/2} \simeq 10^9$ cm s^{-1}. We then immediately obtain the optical depth of the screening envelope from (4) for any time, except, of course, the initial times, when the above inequality had not yet held:

$$\tau_\gamma(t) = \kappa_\gamma L_{\rm CO}(t) \simeq \left(\frac{3\kappa_\gamma M_{\rm CO}}{4\pi v_0^2}\right)\frac{1}{t^2}. \quad (5)$$



Density distribution between the radii $R_{\rm Ni}^{(0)}$ and $R_{\rm CO}^{(0)}$ for times close to the onset of delayed detonation for the H20-180 and L20-188 models from Dunina-Barkovskaya and Imshennik (2003).

Substituting all of the quantities in parentheses into (5), we find with $\kappa_\gamma = 0.03$ cm^2 g^{-1}:

$$\tau_\gamma(t) \simeq 190/t_{\rm days}^2;$$

whence it follows that the transparency condition for γ-ray emission at $\tau_\gamma(t_{\rm tr}) \simeq 1$ occurs at $t_{\rm days} = (190)^{1/2} \simeq 14$. At an opacity of $\kappa_\gamma = 0.12$ cm^2 g^{-1}, we obtain approximately one month for the beginning of an active effect of γ-ray emission on the supernova photosphere. It remains to note that the supernova photosphere will be in the hydrogen envelope, somewhere near the shock front that also propagates with the characteristic velocity v_0 in order of magnitude, i.e., at a distance of about $v_0 t \simeq 0.864 \times 10^{14} t_{\rm days} \simeq 1.2 \times 10^{15}$ cm. The value of $R_{\rm CO}(t)$ is the same in our estimate, but, naturally, the shock radius is considerably larger than this value. It can be shown that, given the ejection rate of the hydrogen envelope $\dot{M} = 10^{-2.4}\ M_\odot$ yr^{-1} (Hamuy *et al.* 2003) and the ejection velocity $v_0' = 300$ km s^{-1}, the contribution of this envelope to $L_{\rm CO}(t)$ will be negligible: it can be estimated from the initial density distribution in the envelope, $\sim 10^{-6}$ of the value found.

The density distribution in the envelope under consideration was disregarded in the previous estimate. Including this distribution would slightly increase the column density $L_{\rm CO}(t)$, but would not change the order of magnitude of this estimate. We can generalize this estimate for a power-law density distribution, $\rho = \rho_0 (r/R_{\rm Ni})^\alpha (-3 < \alpha < -1)$:

$$L_{\rm CO}(t) \simeq \bar{\rho}(t) R_{\rm CO}(t) \left[-\frac{\alpha+3}{3(\alpha+1)}\right] \times \left[R_{\rm Ni}(t)/R_{\rm CO}(t)\right]^{\alpha+1}. \quad (6)$$

Clearly, in the special case of $\alpha = -3$, we have a different formula:

$$L_{\rm CO}(t) = \bar{\rho}(t) R_{\rm CO}(t) \left\{\frac{200}{3\ln[R_{\rm CO}(t)/R_{\rm Ni}(t)]}\right\}. \quad (7)$$

This case is most suitable for fitting the density distribution between the radii $R_{\rm Ni}^{(0)}$ and $R_{\rm CO}^{(0)}$ shown in the figure by a power law for times close to the onset of delayed detonation for the H20-180 and L20-188 models from Dunina-Barkovskaya and Imshennik (2003). Since these authors did not consider the questions of envelope transparency, no plots for the radial density profiles were given. These profiles are very similar in shape, although the boundary temperatures of the models are different (high for H20-180 and low for L20-188). If we assume an inertial law of envelope expansion with time, which asymptotically holds in hydrodynamics only several days later (Imshennik and Nadezhin 1988), and attribute the velocity v_0 estimated above precisely to the outer boundary, then the ratio under the logarithm sign will be constant with time (a close similarity of the envelope configuration):

$$R_{\rm CO}(t)/R_{\rm Ni}(t) = R_{\rm CO}^0/R_{\rm Ni}^0 \simeq 4 \times 10^9/2 \times 10^8 = 20.$$

Since the factor in the braces in (7) is found to be approximately equal to 20, the estimate for the optical depth of the partially burned CO fuel is eventually given by the relation

$$\tau_\gamma(t) = \kappa_\gamma \bar{\rho}(t) R_{\rm CO}(t) \left(\frac{200}{3\ln 20}\right) \simeq 25 \kappa_\gamma \bar{\rho}(t) R_{\rm CO}(t). \quad (8)$$

Thus, the expected envelope transparency time may be only a few months, exceeding the above estimate by a factor of $25^{1/2} = 5$. It should be emphasized that the well-known mixing of the supernova envelope matter, which was studied in detail for SN 1987A, was disregarded in this estimate. Therefore, the values of (5)–(8) must be considered overestimated for the transparency time of interest.

Finally, Hamuy *et al.* (2003) pointed out that at least two supernovae (SN 1997cy and SN 1988Z) that were previously classified as SNe IIn (for which the theory did not recognize the thermonuclear explosion mechanism) are virtually identical to SN 2002ic in their spectra at a time that roughly corresponds to day 70 after the explosion, while Wood-Vasey *et al.* (2004) suggested that the classification of other SNe IIn could also be reconsidered. Considering the possible scenarios of thermonuclear supernova explosions in a hydrogen circumstellar medium, Nomoto *et al.* (2004) surmised that the presence of this medium could be attributable to a stellar wind in a binary system. However, they also recognize the possibility of the explosion of a single AGB star, although they believe that the metallicity of such stars must be very low for this scenario to take place.

Chugai and Yungelson (2004) constructed theoretical light curves for SN 2002ic by assuming that the ^{56}Ni mass is $0.7M_\odot$ and is uniformly distributed in the inner region of the model with a mass of $0.8M_\odot$. As a result, the best agreement of the light curve with the observations was achieved for the model of a circumstellar envelope with a constant density, while the brightness of the computed light curve for the models of an envelope attributable to a constant wind ($\rho \propto r^{-2}$) was well below its observed value. However, it is not quite clear how a circumstellar envelope with a constant density could appear during the evolution of the presupernova. In the paper by Chugai *et al.* (2004), good agreement with the observations was achieved for the models in which the density of the circumstellar envelope was constant for $\log(r/10^{15}\mathrm{cm}) < 0.4$ and decreased as a power law as the radius increased further. In this case, the ^{56}Ni mass was taken to be $0.92M_\odot$, in accordance with the DET1 detonation model from Khokhlov *et al.* (1993), although Chugai *et al.* (2004) emphasized that the calcium lines observed in the spectrum of SN 2002ic are incompatible with the low calcium mass fraction synthesized in the DET1 model.

CONCLUSIONS

Theoretically, the problem of the thermonuclear explosion of a single AGB star is very attractive, since a single star is much more convenient for numerical simulations than a binary system (even when the mass loss and the nonstationary processes in its envelope are included). It is possible that with an appropriate development of the theory and observations, such thermonuclear supernovae in a hydrogen circumstellar medium could also be used as standard candles.

ACKNOWLEDGMENTS

We wish to thank S.I. Blinnikov for inspiring discussions and valuable advice regarding the comparison of our model with superbright SNe Ia or I + 1/2. We are also grateful to V.P. Utrobin for stimulating discussions, particularly in connection with the possibility of the envelopes of such supernovae becoming transparent to γ-ray emission relatively early.

REFERENCES

1. W. D. Arnett, Astrophys. Space Sci. **5**, 180 (1969).
2. W. D. Arnett and A. Fu, Astrophys. J. **340**, 396 (1989).
3. G. Beaudet, V. Petrosian, and E. E. Salpeter, Astrophys. J. **150**, 979 (1967).
4. S. W. Bruenn, Astrophys. J., Suppl. Ser. **24**, 283 (1972).
5. V. M. Chechetkin, S. M. Gershtein, V. S. Imshennik, *et al.*, Preprint No. OTF 78-158, IFVÉ (Institute for High-Energy Physics, Serpukhov, 1978).
6. V. M. Chechetkin, S. M. Gershtein, V. S. Imshennik, *et al.*, Astrophys. Space Sci. **67**, 61 (1980).
7. V. M. Chechetkin, V. S. Imshennik, L. N. Ivanova, and D. K. Nadyozhin, in *Supernovae,* Ed. by Schramm (Astrophys. and Space Sci. Library, 1977), Vol. 66, p. 159.
8. N. N. Chugai, R. A. Chevalier, and P. Lundqvist, Mon. Not. R. Astron. Soc. **355**, 627 (2004).
9. N. N. Chugai and L. P. Yungelson, Pis'ma Astron. Zh. **30**, 83 (2004) [Astron. Lett. **30**, 65 (2004)].
10. G. Contardo, B. Leibundgut, and W. D. Vacca, Astron. Astrophys. **359**, 876 (2000).
11. N. V. Dunina-Barkovskaya and V. S. Imshennik, Tr. Fiz. Inst. im. P. N. Lebedeva, Ross. Akad. Nauk **227**, 32 (2000).
12. N. V. Dunina-Barkovskaya and V. S. Imshennik, Pis'ma Astron. Zh. **29**, 13 (2003) [Astron. Lett. **29**, 10 (2003)].
13. N. V. Dunina-Barkovskaya, V. S. Imshennik, and S. I. Blinnikov, Pis'ma Astron. Zh. **27**, 412 (2001) [Astron. Lett. **27**, 353 (2001)].
14. M. Hamuy, M. M. Phillips, N. B. Suntzeff, *et al.*, Nature **424**, 651 (2003).
15. V. S. Imshennik and A. M. Khokhlov, Pis'ma Astron. Zh. **10**, 631 (1984) [Sov. Astron. Lett. **10**, 262 (1984)].
16. V. S. Imshennik and D. K. Nadezhin, Itogi Nauki Tekh., Ser.: Astron. **21**, 63 (1982).
17. V. S. Imshennik and D. K. Nadezhin, Usp. Fiz. Nauk **156**, 576 (1988) [Sov. Phys. Usp. **31**, 1 (1988)].
18. V. S. Imshennik and D. V. Popov, Astron. Zh. **69**, 497 (1992) [Sov. Astron. **36**, 251 (1992)].
19. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Astrophys. Space Sci. **31**, 497 (1974).
20. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Astron. Zh. **54**, 354 (1977a) [Sov. Astron. **21**, 197 (1977a)].
21. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Astron. Zh. **54**, 661 (1977b) [Sov. Astron. **21**, 374 (1977b)].

22. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Astron. Zh. **54**, 1009 (1977c) [Sov. Astron. **21**, 571 (1977c)].
23. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Pis'ma Astron. Zh. **8**, 17 (1982) [Sov. Astron. Lett. **8**, 8 (1982)].
24. L. N. Ivanova, V. S. Imshennik, and V. M. Chechetkin, Preprint No. 109, ITÉF (Institute for Theoretical and Experimental Physics, Moscow, 1983).
25. A. Khokhlov, E. Müller, and P. Höflich, Astron. Astrophys. **270**, 223 (1993).
26. A. M. Lisewski, W. Hillebrandt, and S. E. Woosley, Astrophys. J. **538**, 831 (2000).
27. K. Nomoto, T. Suzuki, J. Deng, *et al.*, astro-ph/0406263 (2004).
28. M. Reinecke, W. Hillebrandt, and J. C. Niemeyer, Astron. Astrophys. **391**, 1167 (2002).
29. W. M. Wood-Vasey, L. Wang, and G. Aldering, Astrophys. J. **616**, 339 (2004).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 537–545. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 601–610.

Comparison of the Fermi and Betatron Acceleration Efficiencies in Collapsing Magnetic Traps

S. A. Bogachev* **and B. V. Somov**

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

Received March 4, 2005

Abstract—The acceleration of charged particles in the solar corona during flares is investigated in terms of a model in which the electrons and ions preaccelerated in the magnetic reconnection region are injected into a collapsing magnetic trap. Here, the particle energy increases rapidly simultaneously through the Fermi and betatron mechanisms. Comparison of the efficiencies of the two mechanisms shows that the accelerated electrons in such a trap produce more intense hard X-ray (HXR) bursts than those in a trap where only the Fermi acceleration mechanism would be at work. This effect explains the *Yohkoh* and *RHESSI* satellite observations in which HXR sources more intense than the HXR emission from the chromosphere were detected in the corona. © *2005 Pleiades Publishing, Inc.*

Key words: *Sun: solar flares, magnetic reconnection, particle acceleration, X-ray emission.*

INTRODUCTION

Intense bursts in the energy range 20–100 keV observed during solar flares arise when fast electrons decelerate in the plasma of the solar atmosphere. Since the emission of a photon with energy $\mathcal{E}_X$ requires an electron of no less kinetic energy, $\mathcal{K} = \mathcal{E} - mc^2$ (here, $\mathcal{E}$ is the total particle energy), the electrons emitting in flares must be accelerated from energy $\mathcal{K}_\mathrm{T} \sim 0.1$ keV, which corresponds to the coronal temperature, to the above energies, i.e., by more than two orders of magnitude, in a time no longer than a few seconds or, occasionally, a few tenths of a second.

Two regions are considered with regard to solar flares: electrons accelerate in the first region and decelerate in the second region, generating hard X-ray (HXR) emission. The second region is called a target. The bremsstrahlung intensity and spectrum depend on the number of *emitting* electrons in the target and their energy distribution. The relationship between the bremsstrahlung spectrum and the injection spectrum of the *accelerated* electrons into the target depends on the properties of the target. In the thick-target model, the plasma density in the emission source is assumed to be enough for the injected electrons to fully decelerate in a fairly short time (Brown 1971; Syrovatsky and Shmeleva 1972; Hudson 1972). Since the deceleration of slow electrons in plasma is stronger than that of fast electrons, the spectrum of the electrons becomes harder as they penetrate into the target; consequently, X-ray emission with a harder spectrum is formed. In the thin-target model, the plasma density is assumed to be high enough to account for the intensity of the emission, but the thickness of the target is not large enough for the electrons to change their spectrum in it (Datlowe and Lin 1973). The spectrum of the emitting electrons in the thin-target model is identical to the injection spectrum.

Today, data for an enormous number of solar flares recorded with time and spectral resolutions high enough to reconstruct the number of accelerated electrons and their spectrum have been obtained onboard the *Yohkoh* and *RHESSI* satellites. The currently available HXR observations of solar flares are appreciably more accurate than the above classical models, which disregard several significant physical factors: the electric field of the backscatter (compensating) current in the plasma of the target and the dispersion of angular scattering during Coulomb collisions (see Somov 2000). Nevertheless, it is customary to compare the number of accelerated electrons and their spectrum calculated using greatly simplified models with a particular theoretical model of particle acceleration in solar flares. This is known to present great difficulty, although there is no shortage of models that purport to explain the particle acceleration processes in solar flares (see the monographs by Miroshnichenko (2001) and Aschwanden (2002)).

Below, we proceed from a model in which the electrons and ions acquire an initial energy in the magnetic reconnection region or, more precisely, in the collisionless current sheet located on the separator in the corona. Subsequently, these particles are

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

1063-7737/05/3108-0537$26.00

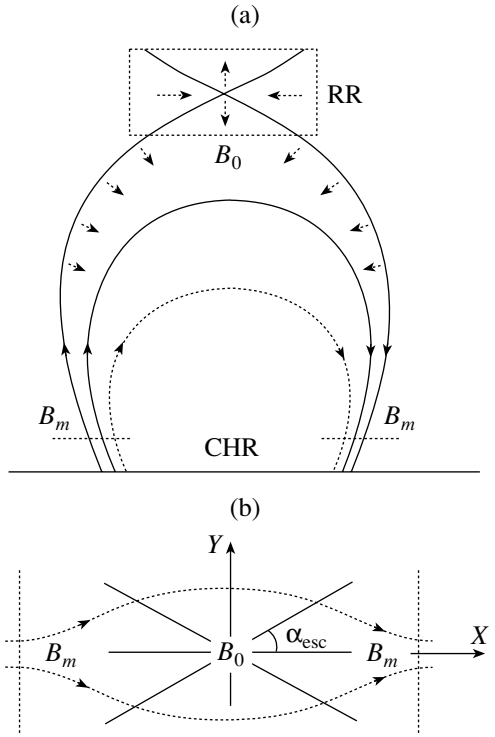


Fig. 1. Formation of contracting magnetic traps in the solar corona during a flare. (a) The magnetic field configuration: the reconnection region (RR) is located above the closed magnetic field loops that rest on the chromosphere (CHR). (b) The schematic view of a magnetic trap.

captured into a collapsing magnetic trap where they accelerate further to high energies (Somov and Kosugi 1997; Bogachev and Somov 2001). Apart from the first-order Fermi acceleration (an increase in the longitudinal particle momentum with decreasing separation between the magnetic mirrors), Somov and Bogachev (2003) suggested taking into account the betatron effect in collapsing traps, i.e., an increase in the transverse momentum as the trap contracts. The additional increase in energy caused by the betatron acceleration was shown to be exactly offset by the decrease in the time of particle confinement in the trap, because the loss cone becomes larger during the contraction. As a result, the particle energy at the time of its escape from the trap remains the same as it would be during the escape from a trap without contraction.

In this paper, we develop a trap model in which both the Fermi and betatron accelerations are at work. This model allows us to compare the efficiencies of the two mechanisms, as applied to the problem of electron acceleration in solar flares and to the observed HXR sources in the solar corona. We determine the maximum kinetic energy that a particle can acquire in a trap under the Fermi and betatron accelerations. We derive formulas for the spectrum of the accelerated particles and their pitch-angle distribution. Subsequently, we determine the number of emitting electrons and their density in a collapsing trap. Finally, we calculate the total kinetic energy of the trapped electrons. The results obtained are summarized in the last section.

THE FORMATION OF A TRAP, ITS CONTRACTION, AND PARTICLE ACCELERATION

Figure 1a schematically shows the magnetic field configuration in the solar corona during a flare. We assume that the electron energy in the magnetic reconnection region (RR) or, more precisely, in the high-temperature turbulent current sheet (HTTCS; Somov 2000) increases from a coronal thermal energy $\mathcal{K}_T \sim 0.1$ keV at least to an energy $\mathcal{K}_{eff} \sim 10$ keV, which corresponds to the effective HTTCS temperature. Each magnetic flux tube consisting of reconnected magnetic field lines is a trap for the fast particles in it, since the magnetic field strength B_m at the footpoints of the tube is larger than that at its top, B_0.

The preaccelerated electrons are injected from the HTTCS into the central part of the trap and are confined in it by converging magnetic field lines, magnetic mirrors. We assume that the particle injection is impulsive, i.e., the electrons fall into the trap at the initial (reconnection) time and subsequently either precipitate into the loss cone or become trapped and acquire additional energy. In addition, the electrons decelerate through Coulomb collisions, producing a burst of hard X-ray bremsstrahlung (the magneto-bremsstrahlung of fast electrons in a collapsing trap will be described in another paper).

Let us simplify this configuration and assume that the electrons are confined in an axially symmetric flux tube with a field B_0 that is minimal at the center of the trap and increases at its edges to B_m, as shown in Fig. 1b. B_m and B_m/B_0 are commonly called the field in the magnetic mirrors and the mirror ratio, respectively. Clearly, this simplification of the actual magnetic configuration does not affect the results of our calculations, since the Larmor radius of the electrons in the trap is much smaller than the characteristic length scales on which the magnetic field varies and since the electron acceleration time in the trap is much shorter than the characteristic drift times. In this case, the curvature of the field lines is unimportant (see Somov and Bogachev 2003).

As the length of the trap decreases as a result of its motion from the reconnection region (RR) toward the chromosphere, the energy of the trapped particles increases through the first-order Fermi mechanism. When the magnetic trap contracts transversely, the particles in it are accelerated by the betatron

mechanism. If the trap contracts simultaneously in both (longitudinal and transverse) directions, the two mechanisms tend to increase the particle energy.

We will describe the transverse contraction of the tube by the quantity

$$b(t) = B(t)/B_0, \quad (1)$$

which changes from the initial value of $b(0) = 1$ at which $B(0) = B_0$ to

$$b_m = B_m/B_0, \quad (2)$$

at which the field at the center of the trap, $B(t)$, becomes equal to the field in the magnetic mirrors, B_m, and the trap ceases to confine the particles.

To describe the change in the trap length L with time, we introduce the parameter

$$l(t) = L(t)/L_0, \quad (3)$$

which decreases from $l(0) = 1$ to $l = 0$ or to a certain very low value corresponding to the residual trap length.

If the period of the particle motion between the magnetic mirrors is much shorter than the lifetime of the trap, if the Larmor radius of the particle is much smaller than the length scales for the magnetic field variations, and if other conditions of the adiabatic approximation are satisfied (see, e.g., Somov 2000), then the adiabatic invariants are conserved as the trap contracts: the longitudinal invariant

$$p_{||}L = p_{0||}L_0 = \text{const} \quad (4)$$

and the transverse invariant

$$p_\perp^2/B = p_{0\perp}^2/B_0 = \text{const}. \quad (5)$$

Combining (1) and (3) with (4) and (5), we find how the particle momentum components change as the length of the trap decreases and as it contracts:

$$p_{||} = \frac{p_{0||}}{l} \quad (6)$$

and

$$p_\perp = p_{0\perp}\sqrt{b}. \quad (7)$$

In the case of Fermi acceleration, according to (6), the longitudinal particle momentum increases, while the transverse momentum remains constant. The betatron acceleration (7) increases the transverse momentum and does not change the longitudinal momentum. As a result, the Fermi and betatron accelerations are always accompanied, respectively, by a decrease and an increase in the particle pitch angle. When the two mechanisms act simultaneously, the pitch angle changes as follows:

$$\tan\alpha = \frac{p_\perp}{p_{||}} = l\sqrt{b}\left(\frac{p_{0\perp}}{p_{0||}}\right) = l\sqrt{b}\tan\alpha_0. \quad (8)$$

If $l\sqrt{b} < 1$, then the pitch angle of the accelerated particles decreases. Although both mechanisms act in such a trap, it is closer in particle behavior to a collapsing trap with Fermi acceleration. If, alternatively, $l\sqrt{b} > 1$, then the pitch angle of the particles during their acceleration increases. In this case, the trap is closer in properties to a trap with betatron acceleration. The condition

$$l\sqrt{b} = 1 \quad (9)$$

separates these two cases. When condition (9) is satisfied, the trapped particles accelerate without any change in their pitch angle: $\tan\alpha = \tan\alpha_0$. If the physical model with a collapsing trap as its element allows the functions $b(t)$ and $l(t)$ to be determined theoretically or observationally, then we can determine which type of acceleration dominates throughout the trap contraction process or even at its different stages by comparing these functions. The kinetic energy of the nonrelativistic particles captured into a contracting magnetic trap is

$$\mathcal{K} = \frac{1}{2m}\left(p_{||}^2 + p_\perp^2\right) = \frac{1}{2m}\left(\frac{p_{0||}^2}{l^2} + bp_{0\perp}^2\right) \quad (10)$$
$$= \mathcal{K}_0\left(\frac{\cos^2\alpha_0}{l^2} + b\sin^2\alpha_0\right);$$

it can only increase, since $l \le 1$ and $b \ge 1$. Here, $\mathcal{K}_0 = p_0^2/2m$ is the initial kinetic energy of the particle. The energy is transferred mainly to the longitudinal (along the magnetic field lines) and transverse degrees of freedom under the dominant Fermi ($l\sqrt{b} < 1$) and betatron ($l\sqrt{b} > 1$) accelerations, respectively. When $l\sqrt{b} = 1$, the energy is distributed uniformly between the degrees of freedom, and its increase does not depend on the pitch angle: $\mathcal{K} = \mathcal{K}_0/l^2$. We call this case isotropic acceleration.

In general, the energy of the particle increases until it falls into the loss cone. The particle pitch angle at this time is defined by the equation

$$\tan\alpha_{\text{esc}} = \frac{1}{\sqrt{B_m/B - 1}} = \frac{1}{\sqrt{b_m/b - 1}}. \quad (11)$$

Comparing (8) and (11), we find that the relation

$$l^2 b\tan^2\alpha_0 = \frac{1}{b_m/b - 1}$$

or

$$\frac{1}{l^2\tan^2\alpha_0} = b_m - b \quad (12)$$

must hold at the time of particle escape from the trap. Let us transform formula (10) for the particle energy:

$$\mathcal{K} = \mathcal{K}_0\sin^2\alpha_0\left(\frac{1}{l^2\tan^2\alpha_0} + b\right).$$

Substituting in condition (12) yields the particle energy at the escape time

$$\mathcal{K}_{\rm esc} = \mathcal{K}_0 b_m \sin^2\alpha_0. \quad (13)$$

If the trap contracts not only in the longitudinal direction, but also in the transverse direction, then the particle in the trap acquires an additional energy through its betatron acceleration. The acceleration rate increases, but the final energy that the particle acquires does not change. This is because the loss cone (11) depends on the mirror ratio. As the magnetic flux tube contracts, the mirror ratio decreases, while the loss cone grows. This reduces the particle confinement time exactly to the extent needed to compensate for the effect of their faster acceleration.

For any combination of the Fermi and betatron accelerations, the final particle energy remains constant and is defined by formula (13). This property simplifies the analysis of the problem in which particles with initial energy $\mathcal{K}_0$ arrive from the coronal reconnection region, while particles with energy $\mathcal{K}_{\rm esc}$ precipitate from the trap into the chromosphere. The particle acceleration mechanism inside the trap may not be concretized; each of them or both lead to the same final energy at the exit from the trap.

The second conclusion drawn from formula (13) is a constraint on the energy that a particle gains in the trap. The maximum energy depends only on the mirror ratio:

$$\mathcal{K}_{\rm max} = \mathcal{K}_0 b_m. \quad (14)$$

In solar flares, electrons with initial energy $\mathcal{K}_{\rm eff} \sim 10$ keV can be accelerated to energies $\sim$1 MeV only inside traps with a mirror ratio $b_m = B_m/B_0 \geq 100$. For this reason, no efficient particle acceleration is possible inside the trap formed above the front of an adiabatic shock, since the jump in field at its front does not exceed $B_m/B_0 = 4$. Bogachev *et al.* (1998) showed that the mirror ratio could increase by more than an order of magnitude if the adiabatic approximation is inapplicable to the shock, for example, due to rapid postshock plasma cooling (see also Somov *et al.* 1999).

THE SPECTRUM AND ANISOTROPY OF TRAPPED ELECTRONS

The distribution of the particles captured in a collapsing trap transforms with time, because their energy changes and they escape into the loss cone. Let us first consider the former effect. The initial electron distribution is determined by the energization processes that precede their injection into the trap, i.e., by their heating to anomalously high temperatures and their preacceleration in the HTTCS. We assume that the initial electron distribution $f_0(\mathcal{K}_0)$ is given and normalized in the standard way:

$$dN(\mathcal{K}_0, \alpha_0) = N_0 f_0(\mathcal{K}_0) d\mathcal{K}_0 2\pi \sin\alpha_0 d\alpha_0, \quad (15)$$

$$\int_0^\infty f_0(\mathcal{K}_0) d\mathcal{K}_0 \int_0^\pi 2\pi \sin\alpha_0 d\alpha_0 \quad (16)$$

$$= 4\pi \int_0^\infty f_0(\mathcal{K}_0) d\mathcal{K}_0 = 1,$$

where N_0 is the number of electrons injected into the trap. Here, the initial electron distribution is assumed to be independent of the pitch angle. We also assume that the trap is symmetric relative to the magnetic mirrors. Both these assumptions may be abandoned when needed.

The relationship between the initial distribution function $f_0(\mathcal{K}_0)$ and the distribution function $f(\mathcal{K}, \alpha)$ of the trapped ($\alpha > \alpha_{\rm esc}$) particles is defined by the condition for the conservation of the number of particles,

$$2\pi f(\mathcal{K}, \alpha) \sin\alpha = 2\pi \mathcal{P} f_0(\mathcal{K}_0) \sin\alpha_0, \quad (17)$$

where the transformation Jacobian $\mathcal{P}$ is

$$\mathcal{P} = \begin{pmatrix} \dfrac{\partial\alpha_0}{\partial\alpha} & \dfrac{\partial\mathcal{K}_0}{\partial\alpha} \\ \dfrac{\partial\alpha_0}{\partial\mathcal{K}} & \dfrac{\partial\mathcal{K}_0}{\partial\mathcal{K}} \end{pmatrix}.$$

Since, according to (8), the change in pitch angle does not depend on energy, the Jacobian $\mathcal{P}$ is

$$\mathcal{P} = \frac{\partial\alpha_0}{\partial\alpha}\frac{\partial\mathcal{K}_0}{\partial\mathcal{K}}. \quad (18)$$

Substituting (18) into (17) yields

$$f(\mathcal{K}, \alpha) = \frac{\sin\alpha_0}{\sin\alpha}\frac{\partial\alpha_0}{\partial\alpha}\frac{\partial\mathcal{K}_0}{\partial\mathcal{K}} f_0(\mathcal{K}_0). \quad (19)$$

Let us first find the derivative $\partial\mathcal{K}_0/\partial\mathcal{K}$. The relationship between the initial and final particle energies is defined by formula (10). Let us transform it as follows:

$$\mathcal{K} = \frac{\mathcal{K}_0}{l^2}\left(\frac{1 + bl^2\tan^2\alpha_0}{1 + \tan^2\alpha_0}\right). \quad (20)$$

According to (8),

$$bl^2\tan^2\alpha_0 = \tan^2\alpha. \quad (21)$$

We substitute (21) into (20):

$$\mathcal{K}_0 = \left(\frac{\sin^2\alpha}{b} + l^2\cos^2\alpha\right)\mathcal{K}; \quad (22)$$

whence we find the derivative

$$\frac{\partial \mathcal{K}_0}{\partial \mathcal{K}} = \frac{\sin^2\alpha}{b} + l^2\cos^2\alpha. \quad (23)$$

Let us introduce the function

$$A_\alpha = \frac{\sin^2\alpha}{b} + l^2\cos^2\alpha \quad (24)$$

and rewrite (22) and (23) as

$$\mathcal{K}_0 = A_\alpha \mathcal{K}, \quad (25)$$

$$\frac{\partial \mathcal{K}_0}{\partial \mathcal{K}} = A_\alpha. \quad (26)$$

Let us now calculate the first two factors in formula (19). First, we transform them:

$$\frac{\sin\alpha_0}{\sin\alpha}\frac{\partial\alpha_0}{\partial\alpha} = -\frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\cos\alpha_0$$
$$= -\frac{1}{\sin\alpha}\frac{\partial}{\partial\alpha}\frac{1}{\sqrt{1+\tan^2\alpha_0}}.$$

Using (21), we substitute $\tan\alpha$ for $\tan\alpha_0$ and obtain, after the transformation,

$$\frac{\sin\alpha_0}{\sin\alpha}\frac{\partial\alpha_0}{\partial\alpha} \quad (27)$$
$$= \frac{l}{b}\frac{1}{\left((\sin^2\alpha)/b + l^2\cos^2\alpha\right)^{3/2}} = \frac{l}{b}\frac{1}{A_\alpha\sqrt{A_\alpha}}.$$

Substituting (25), (26), and (27) into (19) yields the distribution of the trapped electrons in kinetic energy and pitch angle

$$f(\mathcal{K},\alpha) = \frac{l}{b\sqrt{A_\alpha}} f_0\left(\mathcal{K}A_\alpha\right). \quad (28)$$

Recall that the function A_α is defined by formula (24) and does not depend on the particle energy.

Integrating now (28) over the energy yields the particle pitch-angle distribution

$$f(\alpha) = \int_0^\infty f(\mathcal{K},\alpha)d\mathcal{K} = \frac{l}{b\sqrt{A_\alpha}}\int_0^\infty f_0(\mathcal{K}A_\alpha)d\mathcal{K}. \quad (29)$$

The function A_α was factored outside the integral sign, since it does not depend on energy. Let us make the substitution $\mathcal{K}_0 = \mathcal{K}A_\alpha$:

$$f(\alpha) = \frac{l}{bA_\alpha^{3/2}}\int_0^\infty f_0(\mathcal{K}_0)d\mathcal{K}_0.$$

According to (16), the integral on the right-hand side is equal to $1/4\pi$. Thus, the function

$$f(\alpha) = \frac{1}{4\pi}\frac{l}{bA_\alpha^{3/2}} = \frac{1}{4\pi}\frac{l\sqrt{b}}{\left(\sin^2\alpha + bl^2\cos^2\alpha\right)^{3/2}} \quad (30)$$

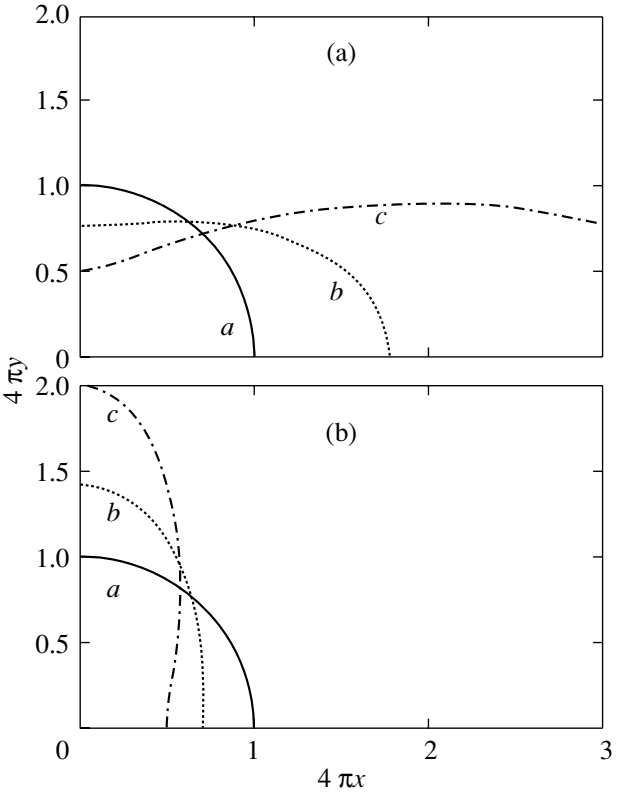


Fig. 2. Change in the pitch-angle distribution of the trapped electrons as the trap contracts. (a) A longitudinally contracting trap; the particles are accelerated by the Fermi mechanism. Curves a, b, and c correspond to the decreases in length $l = L/L_0 = 1.0$, 0.75, and 0.50, respectively. (b) A transversely contracting trap; the particles are accelerated by the betatron mechanism. Curves a, b, and c correspond to the field contractions $b = B/B_0 = 1$, 2, and 4, respectively.

is the sought-for pitch-angle distribution of the trapped electrons. It is shown in Fig. 2a for the Fermi acceleration ($b \equiv 1$) and in Fig. 2b for the betatron acceleration ($l \equiv 1$). In both cases, the initial particle distribution ($l = 1$, $b = 1$) was assumed to be isotropic. The subsequent evolution depends on which of the acceleration mechanisms dominates. For the Fermi acceleration mechanism, the electron pitch angles decrease and the pitch-angle distribution is elongated along the trap axis, the x axis. In traps with the betatron acceleration, the number of electrons with large pitch angles increases and the pitch-angle distribution grows in the direction perpendicular to the x axis. In the intermediate case ($l\sqrt{b} = 1$), the electrons are accelerated with the isotropic distribution retained.

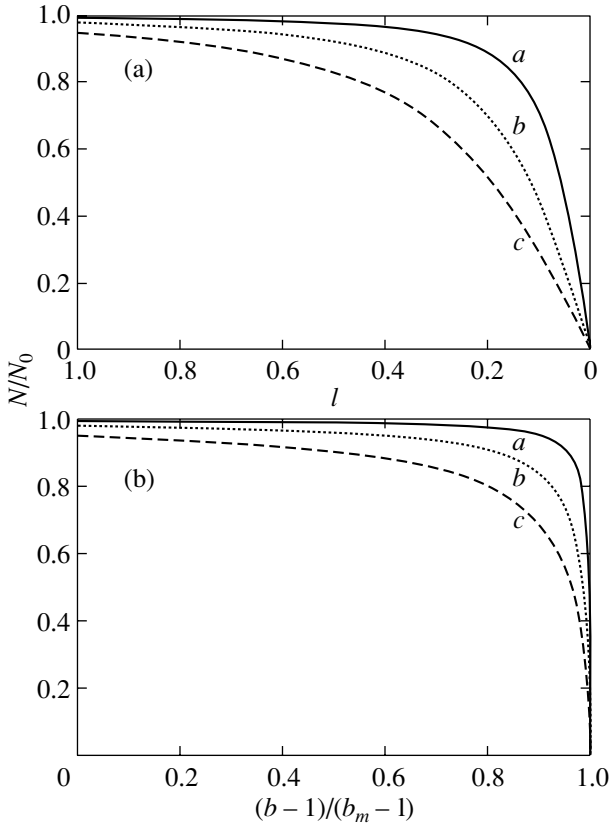


Fig. 3. Change in the number of electrons in traps of different types. Curves a, b, and c correspond to the mirror ratios $b_m = B_m/B_0 = 100$, 25, and 10, respectively; (a) Fermi acceleration and (b) betatron acceleration.

THE NUMBER OF ELECTRONS IN A TRAP AND THEIR DENSITY

The intensity of the HXR emission observed in the solar corona is directly proportional to the number of electrons N inside the trap. Let us find this number. Knowing how the particle pitch-angle distribution (30) changes, let us determine how many particles are outside the loss cone, i.e., the total number of particles confined in the trap:

$$N = N_0 \int_{\alpha_{\rm esc}}^{\pi-\alpha_{\rm esc}} f(\alpha) 2\pi \sin\alpha d\alpha \qquad (31)$$

$$= N_0 l\sqrt{b} \int_{\alpha_{\rm esc}}^{\pi/2} \frac{\sin\alpha d\alpha}{\left(\sin^2\alpha + bl^2\cos^2\alpha\right)^{3/2}}.$$

Here, $\alpha_{\rm esc}$ is the loss cone (11), which, as the trap, is assumed to be symmetric relative to the magnetic mirrors (see Fig. 1b).

Transforming the integrand

$$\frac{\sin\alpha d\alpha}{\left(\sin^2\alpha + bl^2\cos^2\alpha\right)^{3/2}} = \frac{d(\tan^2\alpha)}{2\left(\tan^2\alpha + bl^2\right)^{3/2}}$$

and making the substitution $t = \tan^2\alpha$ yields

$$N = N_0 l\sqrt{b} \int_{t_0}^{\infty} \frac{dt}{(t + bl^2)^{3/2}}, \qquad (32)$$

where

$$t_0 = \frac{b}{b_m - b}.$$

Performing the integration, we find the number of electrons in the trap,

$$N = N_0 \frac{l\sqrt{b_m - b}}{\sqrt{1 + (b_m - b)l^2}}. \qquad (33)$$

If the trap length decreases, while the parameter b does not change, then the particles in the trap are accelerated by the first-order Fermi mechanism. In this case, setting $b = 1$ in (33), we obtain

$$N = N_0 \frac{l\sqrt{b_m - 1}}{\sqrt{1 + (b_m - 1)l^2}}, \qquad (34)$$

where l decreases from 1 to 0. For a transversely contracting trap corresponding to the betatron acceleration, the parameter $l = 1$, while b increases from 1 to b_m. The change in the number of particles in such a trap is defined by the formula

$$N = N_0 \frac{\sqrt{b_m - b}}{\sqrt{b_m - b + 1}}. \qquad (35)$$

The results of the corresponding calculations are shown in Figs. 3a and 3b for traps with different mirror ratios b_m. In Fig. 3b, the parameter $(b - 1)/(b_m - 1)$, which changes from 0 to 1 at any b_m, rather than b, which changes from $b = 1$ to $b = b_m$, is plotted along the horizontal axis.

The number of electrons confined in the trap that produce a HXR burst in the coronal emission source is always smaller than the number N_0 of electrons initially injected into it, since some of the electrons (denote their number by N_{pr}) immediately precipitate into the loss cone and escape from the trap through the magnetic mirrors. According to (33), the fraction of the electrons that directly fall into the chromosphere from the magnetic reconnection region is

$$\frac{N_{pr}}{N_0} = 1 - \sqrt{1 - 1/b_m}. \qquad (36)$$

Thus, for example, a trap with the mirror ratio $b_m = 100$ captures 99.5% of the injected particles, while a trap above an adiabatic shock front with $b_m = 4$ captures about 87%.

Next, let us define the mean electron density in the trap as the number of electrons N divided by the volume V of the trap:

$$n = N/V. \qquad (37)$$

$n_0 = N_0/V_0$ is the initial electron density in the trap; hence

$$n = n_0 \frac{N}{N_0}\frac{V_0}{V}. \quad (38)$$

In general, the change in the number of particles N/N_0 in a collapsing trap is defined by formula (33). Let us find the change in trap volume V/V_0. If L is the trap length and S is its cross sectional area, then

$$\frac{V_0}{V} = \frac{L_0}{L}\frac{S_0}{S} = \frac{L_0}{L}\frac{B}{B_0} = \frac{b}{l}; \quad (39)$$

here, we use the conservation of the magnetic flux through the trap cross section: $BS = B_0S_0$. Combining (33), (38), and (39), we obtain

$$n = n_0 \frac{b\sqrt{b_m - b}}{\sqrt{1 + (b_m - b)l^2}}. \quad (40)$$

For the Fermi acceleration ($b = 1$),

$$n = n_0 \frac{\sqrt{b_m - 1}}{\sqrt{1 + (b_m - 1)l^2}}, \quad (41)$$

while for the betatron acceleration ($l = 1$),

$$n = n_0 \frac{b\sqrt{b_m - b}}{\sqrt{b_m - b + 1}}. \quad (42)$$

The results of our calculations using formulas (41) and (42) are shown in Figs. 4a and 4b, respectively.

The volume of a collapsing trap approaches zero. In principle, this could lead to an unbounded increase in the density of the trapped particles. Figures 4a and 4b show that this does not occur. In a trap with the Fermi acceleration, the density of the accelerated electrons in it increases monotonically to

$$n_0\sqrt{b_m - 1}. \quad (43)$$

In traps with the betatron acceleration, the density at $b \to b_m$ becomes zero, passing through its maximum before this.

Substituting the inequalities $l \geq 0$ and $b \leq b_m$ into (40), let us estimate the maximum density in the trap for an arbitrary combination of the Fermi and betatron accelerations:

$$n \leq 2n_0 \left(\frac{b_m}{3}\right)^{3/2}. \quad (44)$$

Based on the increase in the density of the accelerated electrons inside the coronal HXR source during a solar flare, we can estimate the mirror ratio in a collapsing trap from (44):

$$b_m \geq 3\left(\frac{1}{2}\frac{n}{n_0}\right)^{2/3}. \quad (45)$$

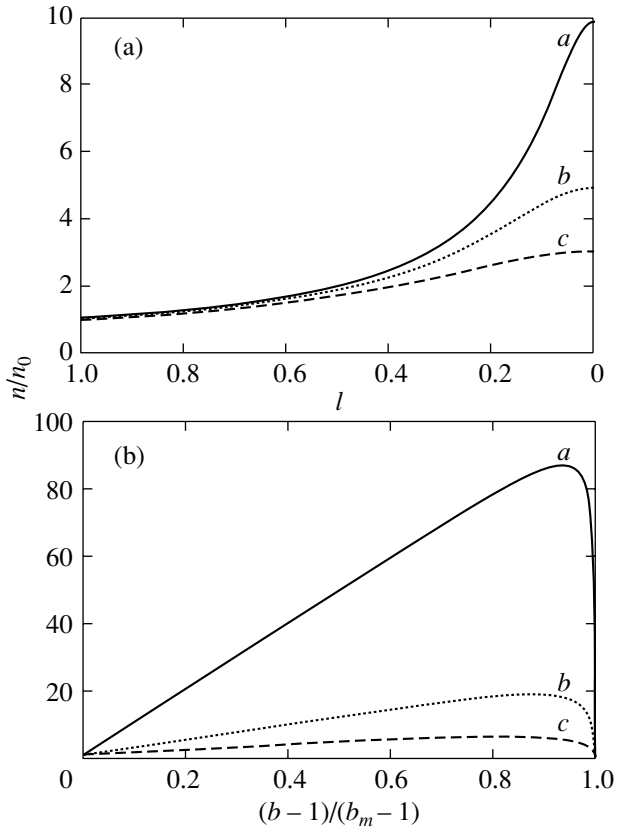


Fig. 4. Change in the density of the particles captured in traps of different types. Curves *a*, *b*, and *c* correspond to the mirror ratios $b_m = B_m/B_0 = 100$, 25, and 10, respectively; (a) Fermi acceleration, and (b) betatron acceleration.

It follows from formulas (40)–(42) and Fig. 4 that the betatron particle acceleration is accompanied by the largest increase in their density among all types of traps. Since the collision frequency between the trapped protons and electrons increases with their density, the betatron particle acceleration is accompanied by stronger bursts of HXR bremsstrahlung than the Fermi acceleration. This conclusion is confirmed by the results of the next section.

THE TOTAL KINETIC ENERGY OF TRAPPED PARTICLES

Two main factors determine the total kinetic energy of the trapped particles: the increase in the energy of each particle as it accelerates and the decrease in the number of particles. The combined effect of these processes can result in an increase in the total energy of all particles in the trap whose time profile is similar to the HXR burst intensity during a solar flare. The total kinetic energy of the accelerated electrons can be estimated from observations using the thick- and thin-target models.

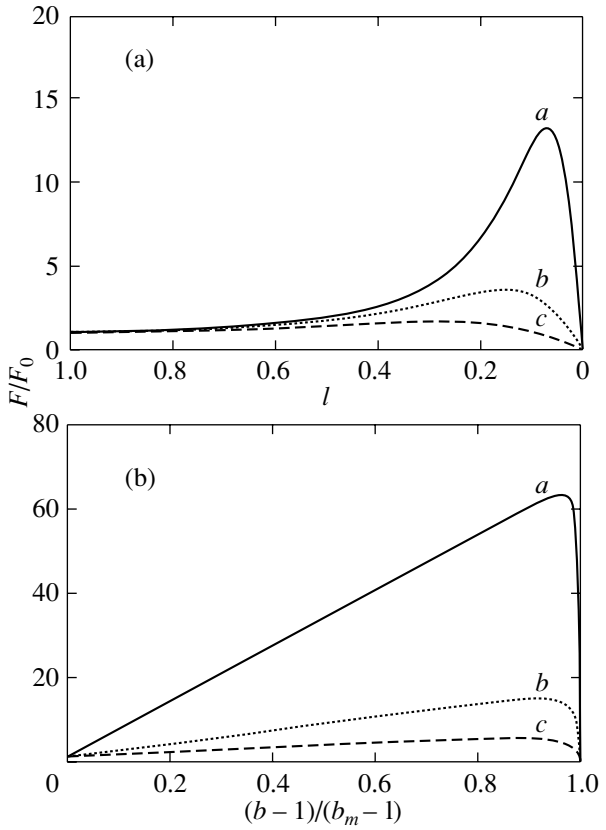


Fig. 5. Change in the total kinetic energy of the particles captured in traps of different types. Curves a, b, and c correspond to the mirror ratios $b_m = B_m/B_0 = 100$, 25, and 10, respectively; (a) Fermi acceleration, and (b) betatron acceleration.

The distribution of the trapped particles in energy and pitch angle is given by formula (28). Let us multiply it by the particle kinetic energy $\mathcal{K}$ and the total number of injected particles N_0 and integrate:

$$\mathcal{F} = N_0 \int_0^\infty \int_{\alpha_{\rm esc}}^{\pi-\alpha_{\rm esc}} f(\mathcal{K},\alpha)\mathcal{K}d\mathcal{K}2\pi\sin\alpha d\alpha; \qquad (46)$$

the integration over the pitch angle is performed outside the loss cone $\alpha_{\rm esc}$ (see (11)) that is symmetric in pitch angle, since the trap itself is assumed to be symmetric. Substituting in (28) yields

$$\mathcal{F} = 4\pi N_0 \frac{l}{b}\int_{\alpha_{\rm esc}}^{\pi/2}\frac{\sin\alpha}{\sqrt{A_\alpha}}d\alpha\int_0^\infty f_0\left(\mathcal{K}A_\alpha\right)\mathcal{K}d\mathcal{K}.$$

Next, we make the substitution $\mathcal{K}_0 = \mathcal{K}A_\alpha$:

$$\mathcal{F} = N_0\frac{l}{b}\int_{\alpha_{\rm esc}}^{\pi/2}\frac{\sin\alpha}{A_\alpha^2\sqrt{A_\alpha}}d\alpha\int_0^\infty 4\pi f_0(\mathcal{K}_0)\mathcal{K}_0 d\mathcal{K}_0. \qquad (47)$$

The second integral in (47) is equal to the mean kinetic energy of the electrons at the time of their injection:

$$\mathcal{F}_0 = \int_0^\infty f_0(\mathcal{K}_0)\mathcal{K}_0 d\mathcal{K}_0\int_0^\pi 2\pi\sin\alpha_0 d\alpha_0 \qquad (48)$$
$$= 4\pi\int_0^\infty f_0(\mathcal{K}_0)\mathcal{K}_0 d\mathcal{K}_0.$$

The mean kinetic energy of the electrons at the time of their injection can be determined from observations or by theoretically analyzing the reconnection process in the current sheet. For example, if the temperature $T_{\rm eff}$ of the superhot plasma injected from the HTTCS is known, then

$$\mathcal{F}_0 = \frac{3}{2}kT_{\rm eff}.$$

In general, combining (47) and (48), we obtain

$$\mathcal{F} = N_0\mathcal{F}_0\frac{l}{b}\int_{\alpha_{\rm esc}}^{\pi/2}\frac{\sin\alpha}{A_\alpha^2\sqrt{A_\alpha}}d\alpha, \qquad (49)$$

where the function A_α is defined by formula (24). Let us transform the integrand in (49):

$$\frac{\sin\alpha}{A_\alpha^2\sqrt{A_\alpha}}d\alpha = \frac{b^2\sqrt{b}\sin\alpha d\alpha}{\cos^5\alpha(\tan^2\alpha + bl^2)^{5/2}}.$$

Further,

$$\frac{\sin\alpha}{A_\alpha^2\sqrt{A_\alpha}}d\alpha = \frac{b^2\sqrt{b}}{2}\frac{(\tan^2\alpha+1)}{(\tan^2\alpha+bl^2)^{5/2}}d(\tan^2\alpha). \qquad (50)$$

Substituting (50) into (49) and changing the variables $t = \tan^2\alpha$ yields

$$\mathcal{F} = N_0\mathcal{F}_0\frac{lb\sqrt{b}}{2}\int_{t_0}^\infty\frac{t+1}{(t+bl^2)^{5/2}}dt, \qquad (51)$$

where

$$t_0 = \frac{b}{b_m - b}.$$

Performing the integration in (51), we get

$$\mathcal{F} = \frac{\mathcal{F}_0 N}{3}\left(2b + \frac{b_m}{l^2(b_m-b)+1}\right), \qquad (52)$$

where N is the number of particles in the trap defined by formula (33). In particular,

$$\mathcal{F} = \frac{\mathcal{F}_0 N}{3}\left(2 + \frac{b_m}{l^2(b_m-1)+1}\right) \qquad (53)$$

for traps with the Fermi acceleration ($b = 1$) and

$$\mathcal{F} = \frac{\mathcal{F}_0 N}{3}\left(2b + \frac{b_m}{b_m - b + 1}\right) \qquad (54)$$

for traps with the betatron acceleration ($l = 1$). The results of our calculations are shown in Figs. 5a and 5b. The change in the total kinetic energy of the particles has the form of a burst. The burst height is at a minimum for a trap with the Fermi acceleration ($b = 1$) and at a maximum for a trap with the betatron acceleration ($l = 1$). Any other combination of l and b leads to a result that is limited by these two extreme cases.

CONCLUSIONS

Collapsing magnetic traps arise during solar flares through fast magnetic reconnection and provide an efficient acceleration of the trapped electrons to energies of dozens of MeV. The bremsstrahlung of the trapped electrons is identified with coronal HXR sources. We investigated two main mechanisms that increase the particle energy in a collapsing trap in terms of a unified model: the first-order Fermi acceleration and the betatron acceleration. For each type of acceleration and, in general, for their combined effect, we derived formulas for the number of electrons in the trap, their density, and their total kinetic energy.

The highest energy that an electron acquires in a collapsing trap does not depend on the acceleration mechanism. It is the same for all types of traps, irrespective of which of the two mechanisms dominates. However, a trap with the dominant betatron acceleration confines the particles better. The density of the trapped particles in it and their total kinetic energy reach values that are severalfold higher than the maximum possible values of these quantities in a trap with the dominant Fermi particle acceleration. As a result, traps with the dominant betatron acceleration must produce HXR bursts in the solar corona with amplitudes higher than and durations shorter than those in the bursts generated in collapsing traps with the dominant Fermi acceleration.

REFERENCES

1. M. J. Aschwanden, *Particle Acceleration and Kinematics in Solar Flares* (Kluwer, Dordrecht, 2002).
2. S. A. Bogachev and B. V. Somov, Astron. Zh. **78**, 187 (2001) [Astron. Rep. **45**, 157 (2001)].
3. S. A. Bogachev, B. V. Somov, and S. Masuda, Pis'ma Astron. Zh. **24**, 631 (1998) [Astron. Lett. **24**, 543 (1998)].
4. J. C. Brown, Sol. Phys. **18**, 489 (1971).
5. D. H. Datlowe and R. P. Lin, Sol. Phys. **32**, 459 (1973).
6. H. S. Hudson, Sol. Phys. **24**, 414 (1972).
7. S. Masuda, T. Kosugi, H. Hara, *et al.*, Nature **371**, 495 (1994).
8. L. I. Miroshnichenko, *Solar Cosmic Rays* (Kluwer, Dordrecht, 2001).
9. B. V. Somov, *Cosmic Plasma Physics* (Kluwer, Dordrecht, 2000).
10. B. V. Somov and S. A. Bogachev, Pis'ma Astron. Zh. **29**, 701 (2003) [Astron. Lett. **29**, 621 (2003)].
11. B. V. Somov and T. Kosugi, Astrophys. J. **485**, 859 (1997).
12. B. V. Somov, Yu. E. Litvinenko, T. Kosugi, *et al.*, in *Magnetic Fields and Solar Processes. The 9th European Meeting on Solar Physics, Florence, Italy, 1999,* Ed. by A. Wilson (1999), ESA SP-448, p. 701.
13. S. I. Syrovatsky and O. P. Shmeleva, Astron. Zh. **49**, 334 (1972) [Sov. Astron. **16**, 273 (1972)].
14. S. Tsuneta and T. Natio, Astrophys. J. Lett. **495**, L67 (1998).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 546–556. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 611–621.

Structure of Solar-Wind Streams at the Maximum of Solar Cycle 23

N. A. Lotova[1], K. V. Vladimirskii[2], V. N. Obridko[1]*, B. P. Filippov[1], and O. A. Korelov[3]

[1] *Institute of Terrestrial Magnetism, Ionosphere, and Radiowave Propagation, Russian Academy of Sciences, Troitsk, Moscow oblast, 142092 Russia*

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

[3] *Research Radio-Physical Institute, Nizhni Novgorod, Russia*

Received February 25, 2005

Abstract—We study the formation of solar-wind streams in the years of maximum solar activity 2000–2002. We use observations of the scattering of radio emission by solar-wind streams at distances of ~ 4–$60R_S$ from the Sun, data on the magnetic field structure and strength in the source region ($R \sim 2.5R_S$), and observations with the LASCO coronagraph onboard the *SOHO* spacecraft. Analysis of these data allowed us to investigate the changes in the structure of circumsolar plasma streams during the solar maximum. We constructed radio maps of the solar-wind transition, transonic region in which the heliolatitudinal stream structure is compared with the structure of the white-light corona. We show that the heliolatitudinal structure of the white-light corona largely determines the structure of the solar-wind transition region. We analyze the correlation between the location of the inner boundary of the transition region R_{in} and the magnetic field strength on the source surface $|B_R|$. We discuss the peculiarities of the $R_{in} = F(|B_R|)$ correlation diagrams that distinguish them from similar diagrams at previous phases of the solar cycle. © *2005 Pleiades Publishing, Inc.*

Key words: *Sun*.

INTRODUCTION

The peculiarities of the formation of the solar-wind stream structure at the solar maximum in 2000–2002 are being studied. The stream structure and its changes over a solar cycle have been the subject of many studies for several years (Schwenn 1990; Bird and Edenhofer 1990; Bird *et al*. 1994, 1996; Tu and Marsch 1999; Janardhan *et al*. 1999; Efimov 1994; Lotova 1992; Lotova *et al*. 1985, 2000, 2002, 2003; Wohlmuth *et al*. 2001). A substantial fraction of these studies were performed in regions far from the Sun, $R \geq 100R_S$ (Veselovskiĭ *et al*. 1998; Kojima and Kakinuma 1987, 1990; Kojima *et al*. 2000), which made it difficult to investigate the specific stream formation mechanisms. Nevertheless, these works yielded extensive statistics of solar-wind stream parameters at various distances from the Sun (Rickett and Coles 1991; Muhleman and Anderson 1981; Veselovskiĭ *et al*. 1999; Fujiki *et al*. 2003; Lotova *et al*. 2004). At the same time, the formation and development of streams with their different subsequent fates as the distance from the Sun increases are important not only per se, but also in investigating the solar-wind acceleration mechanisms (Efimov *et al*. 1990; Tokumaru *et al*. 1995; Fujiki *et al*. 2003; Lotova *et al*. 1995, 1997). Experimental data on the radio sounding of interplanetary plasma regions close to the Sun, at radial distances of ~ 4–$60R_S$ from the Sun, form the basis for our studies. The observations were performed at the Pushchino Radio Astronomy Observatory of the Lebedev Physical Institute, Russian Academy of Sciences. The main results of our experimental studies are the locations of the inner (R_{in}) and outer (R_{out}) boundaries of the solar-wind transition, transonic region where the subsonic flows give way to supersonic flows. There is a mixed flow in the transition region; the subsonic and supersonic flows coexist and interact (Lotova *et al*. 1985; Lotova 1992). In our experimental studies, the transition region is revealed as a vast region of enhanced and variable scattering, a natural result of the instability of the boundaries of the mixed flow components. An important result of our analysis of these observations that confirms the very existence of the solar-wind stream structure is the correlation of R_{in} with the calculated magnetic field strength B_R at the point where the onset of solar-wind stream acceleration can be suggested: at the point of the sphere with a radius of $2.5R_S$ with the same angular coordinates as the boundary R_{in} (Lotova *et al*. 2000). The previously developed method for analyzing the R_{in}–$|B_R|$ correlation diagrams (Lotova *et al*. 2002,

*E-mail: solter@izmiran.rssi.ru

2003) reveals several types of dependence of the acceleration intensity on the magnetic field strength in the source region, $R_{\rm in} = F(|B_{\rm R}|)$. This method for analyzing the results of radio-wave scattering observations is of particular importance in analyzing the observations performed at the maximum of solar activity, since the coronal magnetic field structure is particularly complex at the solar maximum in conditions close to magnetic field polarity reversal.

THE HELIOLATITUDINAL STRUCTURE OF PLASMA STREAMS FROM RADIO-ASTRONOMICAL DATA

The spatial flow inhomogeneity, the solar-wind stream structure, is of greatest importance in the problem of the formation of a supersonic flow. The flow inhomogeneity in the solar plasma is revealed by a sharp increase in the random velocity spread in the main acceleration zone, the transonic solar-wind transition region (Lotova 1992). The radial dependence of the radio-wave scattering by plasma streams is studied in experiments: the scattering angle $2\theta(R)$ and the scintillation index $m(R)$. The velocity variations in individual solar-wind plasma streams cannot be judged from these parameters. Nevertheless, the variations in the locations of the boundaries of the transition region are direct evidence for the variations in the characteristic velocity (Lotova *et al.* 1995). The spatial locations of the inner ($R_{\rm in}$) and outer ($R_{\rm out}$) boundaries of the transition region can be determined from the radial dependence of the parameters $2\theta(R)$ and $m(R)$. The transition region is detected here as an extended region of enhanced scattering compared to the asymptotic decrease in scattering with increasing distance from the Sun in the subsonic region $R \leq 10$–$20R_{\rm S}$ (Lotova 1992; Lotova *et al.* 2004). The determination of the location of the inner boundary $R_{\rm in}$ is facilitated significantly by the use of a precursor, a narrow region of sharply reduced scattering located near the inner boundary of the region of enhanced scattering (Lotova *et al.* 2004). The large-scale stream structure is reconstructed in the form of two-dimensional radio maps of the transition region with the heliolatitude-dependent boundaries $R_{\rm in}$ and $R_{\rm out}$. The stream structure and its evolution over an 11-year solar cycle can be visualized by constructing radio maps (Lotova *et al.* 1997).

The radio maps at the solar maximum in 2000–2002 are shown in Figs. 1a–1c. The circles correspond to the observations of sources approaching the Sun in June and August: the compact source in the Crab Nebula 3C144 and the quasars 3C133, 3C152, 3C154, 3C166, 3C208, 3C215, 3C225, and 3C228. The diamonds correspond to the observations of maser sources approaching the Sun in June (GGD4, S252A, U Ori) and December (W31(2), IRC-20431, W28A(1)). The triangles are used for the quasars 3C2 and 3C5 approaching the Sun in March; the squares are used for the quasars 3C273, 3C275, and 3C279 approaching the Sun in October. The transition region in Figs. 1a–1c is hatched. The open and filled symbols correspond to the inner (closer to the Sun), $R_{\rm in}$, and outer, $R_{\rm out}$, boundaries of the transition region, respectively. Figure 1d also displays the time variations in solar activity in units of R_Z (http://www.dxlc.com/solar/solcycle.html). Figures 1a–1c show that the boundaries $R_{\rm in}$ and $R_{\rm out}$ at the solar maximum are mainly within radial distances of $\sim$15–$60R_{\rm S}$ from the Sun. For comparison, note that at the preceding epoch of increase in activity, 1997–1998, the boundaries of the transonic region were within $\sim$10–$40R_{\rm S}$ (Lotova *et al.* 2000). Thus, in the years of solar maximum, the transition region was displaced to farther regions of interplanetary space; its extent increased. Given the correlation (Lotova *et al.* 1995) between the flow characteristics, the velocity V and the location of the boundary $R_{\rm in}$, it can be concluded that a distinctive feature of the plasma flow formation at the solar maximum is the predominance of slow streams in the solar wind. This fact has long been known (Kojima and Kakinuma 1990; Kojima *et al.* 2000; Lotova *et al.* 2002); we only concretized it in a numerical expression of the parameters. Comparison of Figs. 1a–1c with Fig. 1d indicates that the slowest streams were observed in 2000 during the first, highest level of activity. The second peak of activity was observed in 2002. Here, the boundaries of the transition region $R_{\rm in}$ and $R_{\rm out}$ also moved away from the Sun compared to their locations in 2001, but not so far as in 2000. The numerical values of the solar-wind characteristics $R_{\rm in}$ and $R_{\rm out}$ as a function of the level of solar activity R_Z give a clearer insight into the influence of solar activity on the evolution of solar-wind streams. Below, this dependence is analyzed in more detail. Here, we note that the predominance of slow solar-wind streams in the period 2000–2002 leads us to conclude that the energy of the magnetic fields forming the streams in the solar corona decreases at the solar maximum.

THE MATTER DENSITY DISTRIBUTION IN THE SOLAR CORONA

When studying the formation of the solar wind at the initial stage, we compared the radio maps of the transition region (Figs. 1a–1c) with the structure of the white-light corona. Such a comparison is possible, because, apart from the rare epochs of total solar eclipses, the white-light corona is observed with coronagraphs, telescopes equipped with an artificial moon. This method is most efficient for

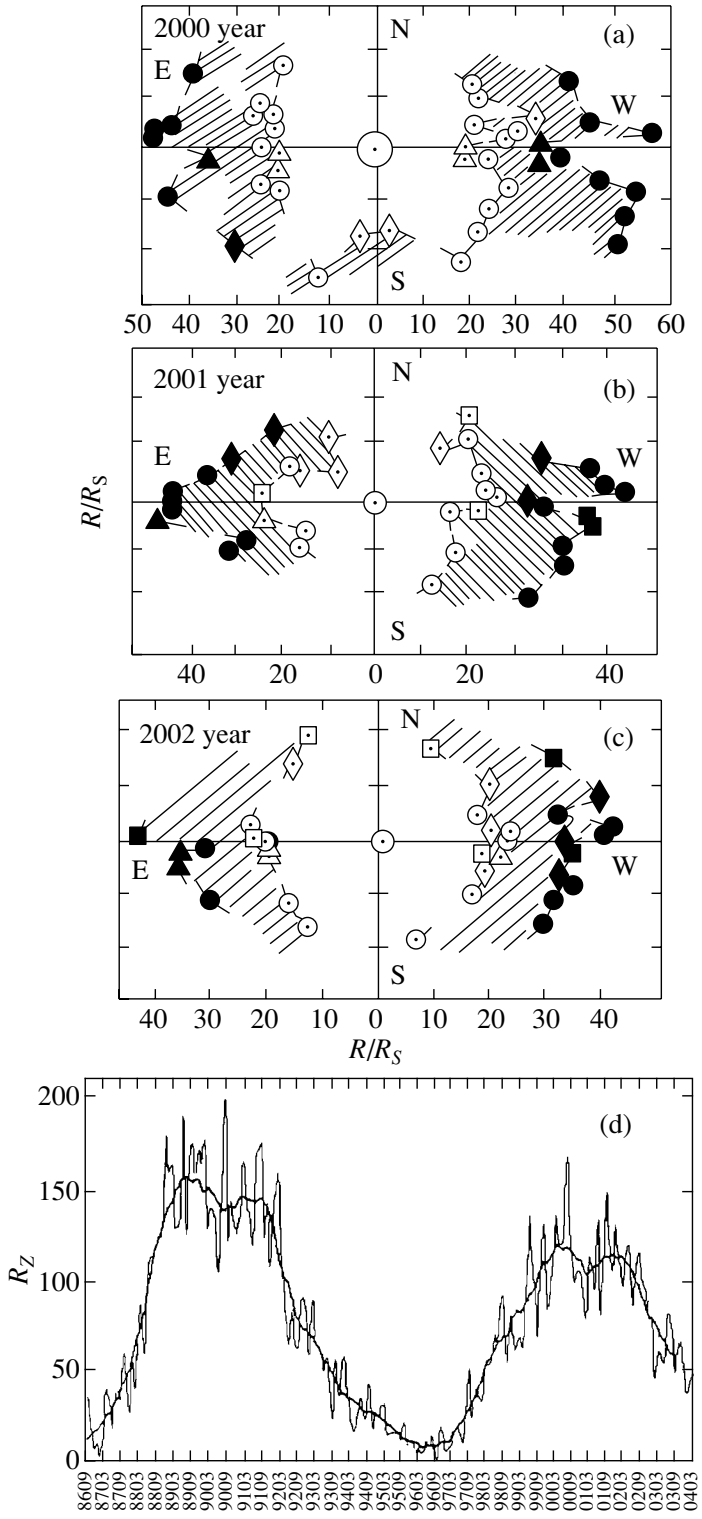


Fig. 1. Radio maps of the solar-wind transition region at the solar maximum for (a) 2000, (b) 2001, and (c) 2002; (d) the time variations in solar activity in units of R_Z.

orbiting observatories, for which the effect of the light scattered in the atmosphere is eliminated. It is this method that is used in daily *SOHO* observations of the white-light corona. The package of three LASCO coronagraphs allows the corona to be observed in the range from 1.1 to $30R_S$ (Bruckner *et al.* 1995). For our analysis, we used data from the LASCO C2 coronagraph with a field of view from 2 to $6R_S$.

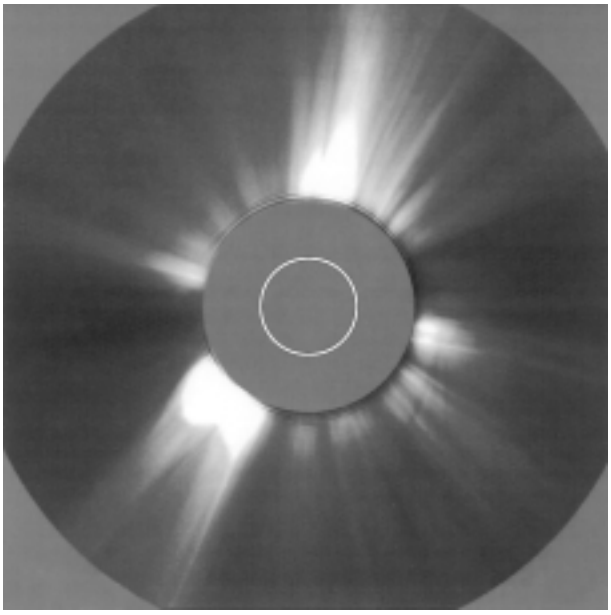


Fig. 2. White-light image of the solar corona on June 19, 2000, obtained with the C2 coronagraph onboard the SOHO space observatory.

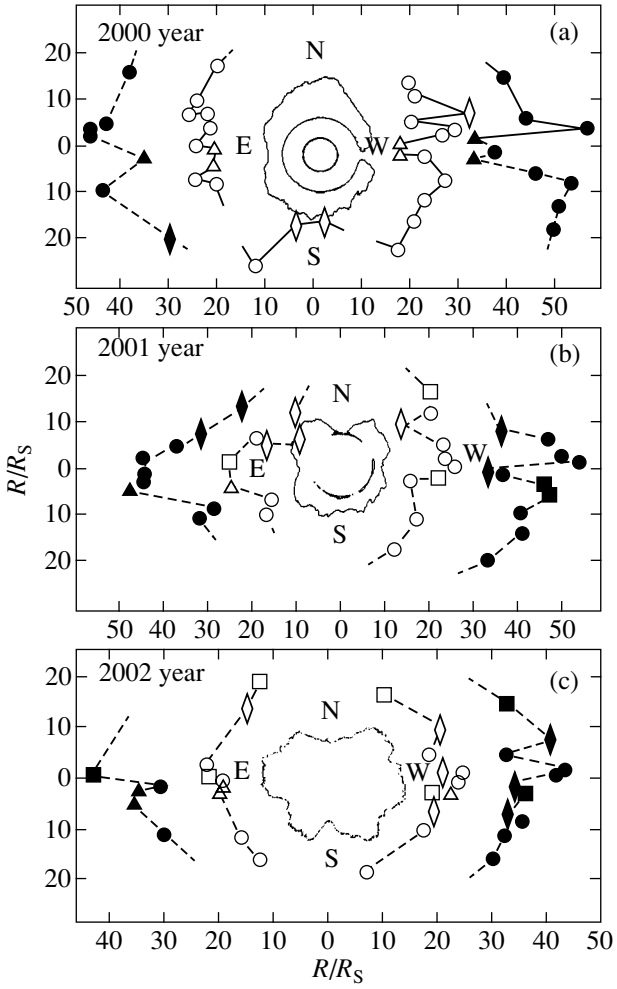


Fig. 3. Comparison of the isophotal shape of the white-light corona with the structure of the solar-wind transition region. The isophotes were obtained by averaging over the images for (a) June 2000, (b) 2001, and (c) 2002.

Although the data from the C3 coronagraph cover the region up to $30R_S$, they are less suitable for estimating the density distribution due to the large background irregularities; the influence of planets, stars; etc. At the same time, according to the studies by Woo and Habbal (1999, 2001), the density variations in the lower solar atmosphere are transferred almost radially to the outer solar corona up to $20–30R_S$. The coronal images were retrieved from the site at http://sohowww.nascom.nasa.gov. Figure 2 shows a typical image of the corona for the maximum of solar activity with many thin coronal streamers and several large streamers. Since there is no hydrostatic equilibrium in the outer corona, the individual streamers are plasma streams that continuously outflow from the Sun into interplanetary space and that give rise to the solar wind. When comparing the structure of the white-light corona with the radio maps of the transition region, we must bear in mind that the radio maps in Figs. 1a–1c are not snapshots of the transition region. They were obtained from the data accumulated over a year from several series of observations, each approximately a month in duration. Therefore, they should be compared with an averaged coronal matter density distribution. Although the overall coronal structure (except the time intervals when coronal mass ejections occur) is preserved for a fairly long time, different structures happen to be in the line of sight due to the solar rotation, and the visible picture changes noticeably from day to day. To eliminate the longitudinal density variations in the corona, which are quasi-random for different epochs of radio observations, we added up the daily coronal images over one solar rotation and constructed the density isoline, which can be roughly considered to be a plot of density against angle in polar coordinates where the brighter regions have a larger radial extent. Figures 3a–3c show a comparative picture of the heliolatitudinal structure of the solar-wind transition region and the structure of the white-light corona represented by the density isoline. Given the inevitable nonstationarity of the solar wind and the averaged image of the white-light corona, the agreement between the heliolatitudinal shape of the white-light corona and the structure of the solar-wind transition region may be considered quite satisfactory. This leads us to conclude that the heliolatitudinal structure of the white-light corona largely determines the heliolatitudinal structure of the transition region, the solar-wind stream structure.

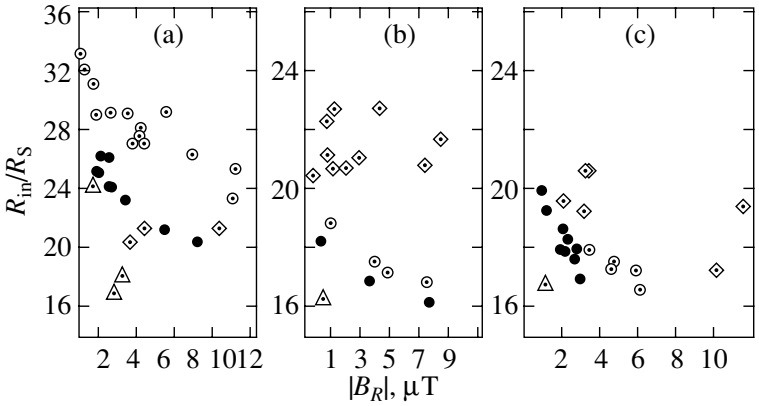


Fig. 4. Correlation between the location of the inner boundary of the transition region R_{in} and the coronal magnetic field strength $|B_R|$, $R = 2.5R_S$, as inferred from the data of table for (a) 2000, (b) 2001, and (c) 2002.

COMPONENTS OF THE SOLAR-WIND STREAM STRUCTURE

Since the deviations of solar-wind streams from the radial direction at distances of the order of a few dozen solar radii are small, the time and angular coordinates of the radio-astronomical observations were directly transferred to the solar-wind source region, to an arbitrary source surface, a sphere of radius $2.5R_S$, which was considered as the region in which the acceleration of circumsolar plasma streams begins. Thus, a unique opportunity to compare the data on the plasma stream acceleration and velocity with the magnetic field data in the source region was created. The data on the magnetic field strength and structure in the source region were obtained by solving the field equations in the region $R_S \le R \le 2.5R_S$. The calculations were performed in the potential-field approximation using the method by Hoeksema *et al.* and Obridko (Hoeksema *et al.* 1982, 1983; Obridko and Schelting 1999; Lotova *et al.* 2002). The results of our calculations were presented in two forms: we determined the absolute values of the radial field component on the source surface $|B_R|$ reconciled in time and angular coordinates with the detection time of the inner boundary of the transition region R_{in}. In addition, we obtained data on the closed or open type of structure of the magnetic field lines in the source region in order to be able to judge the conditions for the penetration of the frozen-in magnetic field into solar-wind streams (Lotova *et al.* 2002).

The table presents the results of our solar-wind studies based on the 2000–2002 observations; the stream data are in the form of a pair of parameters R_{in}/R_S and $|B_R|$. The table gives the names of the sources used, the detection date of R_{in}, the sounding geometry (the E or W hemisphere and the angle φ), and the results of our magnetic field calculations: the field strength on the source surface $|B_R|$ and the field structure: open, closed, or mixed field line configurations. The last column contains the data on the structure of the white-light corona obtained from observations with the LASCO C2 coronagraph onboard the *SOHO* spacecraft. This table does not directly reveal any correlation between the field strength and the geometry of the transition region.

As was noted above, the interrelated study of the solar-wind sources and the stream structure is based on the analysis of the correlation between the location of the inner boundary of the transition region R_{in} and the coronal magnetic field $|B_R|$ on the source surface ($R = 2.5R_S$). The emergence of several branches in the $R_{in} = F(|B_R|)$ correlation diagram reveals the existence of a stream structure of plasma flows, flows of various types that have differing magnetic field structures at their bases.

The correlation diagram (Figs. 4a–4c) illustrates the main result of our comparison of R_{in} and $|B_R|$. The groups forming almost unequivocal correlations between R_{in} and $|B_R|$ stand out in the general field. The data of the table contain two groups of $R_{in} = F(|B_R|)$ correlations. The first group is denoted by the filled circles in Figs. 4a–4c. These are the streams that have large, closed magnetic field structures as their sources; the main body of the streamers corresponds to them in the optical range. The streams attributable to the mixed, open, and closed magnetic field structures and the raylike structure of the white-light corona in the source region correspond to the second group (the open circles in Fig. 4). Analysis of the acceleration $V = F(R_{in})$ suggests that these two groups of experiments relate to slow solar-wind streams. In Fig. 5, these two groups of points R_{in} and $|B_R|$ are presented on a logarithmic scale. The notation here is the same as that in Fig. 4. The two branches exhibit a nearly linear correlation. The fitting straight lines were drawn by

Solar-wind stream structure from the radio-sounding data, the coronal magnetic field, and the morphology of the white-light corona in 2000, 2001, 2002

No.	Source	Date	E/W	φ, deg	R_{in}/R_S	$\lvert B_R\rvert$, μT	Magnetic field structure B_R	White-light corona structure
				2000				
1	3C133	June 9	E	14.0	24	0.6	Open	Streamer sidelobe
2	S255	June 22	E	−82.0	18	3.1		
3	S255	June 24	W	−79.0	17	2.65		
4	3C2	Mar. 16	E	−9.4	20	8.13	Closed	Streamer
5	IRS-20431	Dec. 17	E	16.0	26	1.81		
6	3C225	Aug. 9	E	−2.0	26	1.95		
7	3C133	June 21	W	29.0	25	2.33		
8	3C144	June 22	W	−4.0	24	2.60		
9	3C228	Aug. 24	W	12.0	23	3.25		
10	3C228	Aug. 12	E	9.0	21	5.86		
11	3C138	June 15	W	−51.0	29	6.11	Mixed	Raylike emission
12	IRC-20431	Jan. 1	W	5.0	33	0.18		
13	W31(2)	Dec. 16	E	30.0	32	0.74		
14	W28A2(1)	Dec. 30	W	−11.4	31	1.30		
15	3C225	Aug. 24	W	3.0	29	1.47		
16	3C215	Aug. 15	W	5.0	29	2.39		
17	3C138	June 7	E	−66.0	29	3.60		
18	3C208	Aug. 11	W	−26	28	4.38		
19	3C212	Aug. 12	W	−17.8	27.5	4.26		
20	3C207	Aug. 7	W	−35.0	27	4.58		
21	3C245	Aug. 25	E	40.0	27	3.87		
22	3C172	June 28	E	14.0	26	7.94		
23	3C154	June 30	W	32.0	25	10.91		
24	3C166	June 26	E	−22.0	23	10.54		
25	3C2	Mar. 25	W	−5.4	21	4.06		Complex structures near dark region
26	3C5	Mar. 18	E	−5.8	20	3.54		
27	3C5	Mar. 29	W	−1.4	21	9.66		
28	W31(2)	Jan. 2	W	14.3	35	4.53		
29	GGD4	June 27	W	9.9	37	6.65		
				2001				
1	3C144	June 19	W	−09.0	16.0	1.07	Open	Ephemeral coronal hole
2	3C154	June 19	E	22.5	19.0	0.87	Closed	Streamer
3	IRC-20431	Dec. 18	E	22.0	17.0	2.71		
4	3C144	June 11	E	−23.0	16.0	6.78		

Table. (Contd.)

No.	Source	Date	E/W	φ, deg	R_{in}/R_S	$\lvert B_R\rvert$, μT	Magnetic field structure B_R	White-light corona structure
				2000				
5	3C208	Aug. 9	W	−33.0	20.0	1.43	Mixed	Raylike emission
6	3C166	June 27	E	−26.0	18.0	2.91		
7	W31(2)	Dec. 27	W	39.5	17.5	3.51		
8	W31(2)	Dec. 21	E	52.0	17.0	6.60		
9	3C207	Aug. 5	W	−46.3	23.0	2.18	Low loops, local coronal hole periphery	Between dark region and streamer
10	3C154	June 30	W	33.0	23.5	2.94		
11	3C225	Aug. 22	W	03.0	26.0	3.91		
12	3C279	Oct. 2	E	04.0	24.5	7.14		
13	3C279	Oct. 15	W	−03.0	23.0	6.35		
14	3C228	Aug. 23	W	13.0	22.5	0.35		
15	3C273	Oct. 3	W	41.2	26.0	1.53		
16	3C215	Aug. 12	W	05.0	25.5	1.30		
17	3C225	Aug. 10	E	−02.0	23.5	1.28		
18	3C2	Mar. 16	E	−09.0	23.0	1.50		
				2002				
1	3C228	Aug. 22	W	15.0	18.0	1.08	Open	Streamer sidelobe
2	3C2	Mar. 18	E	−8.0	20.0	2.77	Closed	Streamer
3	3C2	Mar. 28	W	−5.1	21.0	2.00		
4	3C207	Aug. 02	W	−70.2	19.5	2.70		
5	3C208	Aug. 1	E	−50.0	18.5	3.00		
6	3C212	Aug. 1	E	−35.0	20.0	1.87		
7	3C215	Aug. 1	E	−2.0	20.5	1.75		
8	3C215	Aug. 13	W	5.4	23.0	0.89		
9	3C225	Aug. 10	E	−2.1	20.0	1.86		
10	W31(2)	Dec. 29	W	28.0	22.0	1.10		
11	3C5	Mar. 19	E	−5.8	20.0	3.43	Mixed	Raylike emission
12	3C212	Aug. 12	W	−28.0	19.0	4.64		
13	3C273	Oct. 1	W	63.5	19.5	4.75		
14	3C275	Oct. 10	W	−7.1	18.0	6.09		
15	IRC-20431	Dec. 28	W	6.0	19.0	5.86		
16	W31(2)	Dec. 19	E	42.0	22.0	3.23		
17	3C228	Aug. 11	E	7.0	24.0	3.23	Low loops, equatorial coronal loop periphery	Between dark region and streamer
18	3C273	Sep. 17	E	54.6	24.0	3.30		
19	3C275	Sep. 28	E	0.3	22.5	11.58		
20	W28A2(1)	Dec. 27	W	−14.0	19.0	10.17		
21	3C225	Aug. 23	W	3.4	22.5	1.94		

the least-squares method and have different slopes, which may be considered as independent evidence that the sources in the two groups are different in nature. Apart from these previously known types of slow solar-wind streams (Vladimirskii *et al.* 2003), the 2000–2002 data (the table, Fig. 4) revealed the hitherto unknown lowest-speed streams denoted in Fig. 4 by the diamonds, which show no $R_{\rm in} = F(|B_{\rm R}|)$ correlations. On the solar surface, the small-scale looplike structures pressed to the surface correspond to them in the optical range (Fig. 6a). According to the LASCO C2 coronagraph/*SOHO* data, they are located in a narrow region on the streamer periphery between the streamer and the black band formed by local coronal holes near the zero magnetic field line (Fig. 6b). This uncorrelated stream component arises from the complex interaction in the corona between the small-scale magnetic field structures located near the zero magnetic field line.

DISCUSSION

The table and the $R_{\rm in} = F(|B_{\rm R}|)$ correlation diagrams (Fig. 4), as well as the data for the previous epoch of the solar cycle (Lotova *et al.* 2000, 2002, 2003; Vladimirskiĭ *et al.* 2003), reveal a pattern in the variations of the solar-wind stream composition: the number of streams from sources with a mixed magnetic field configuration increases appreciably as the level of solar activity rises. When interpreting this pattern in the variations of the composition of the solar-wind components, we should turn to the basic factors of the solar cycle: the evolution of the magnetic field on the Sun and in the solar corona. The role of small-scale magnetic field structures is known to increase in years of maximum solar activity (see, e.g., Ivanov *et al.* 1999): the magnetic loops become lower, their number increases, and the number of local sources with an open magnetic field configuration also increases. A change, a decrease in the magnetic field scales and the interaction between open and closed field structures cause the number of streams with mixed magnetic fields to increase. Analysis of the $R_{\rm in} = F(|B_{\rm R}|)$ correlation diagrams in Fig. 4 and in a paper by Lotova *et al.* (2003) indicates that the solar-wind streams from sources with a mixed magnetic field structure (in Fig. 4, this branch is denoted by the open circles) lie higher than the branch with closed field structures at its base (in Fig. 4, this branch is denoted by the filled circles). According to the previously known results of the analysis of the correlation between the flow velocity and the location of the inner boundary of the transition region $R_{\rm in}$ (Lotova *et al.* 1995), in the streams of two types considered in Fig. 4 at the same magnetic field strength $|B_{\rm R}|$, the streams from sources with a closed field structure

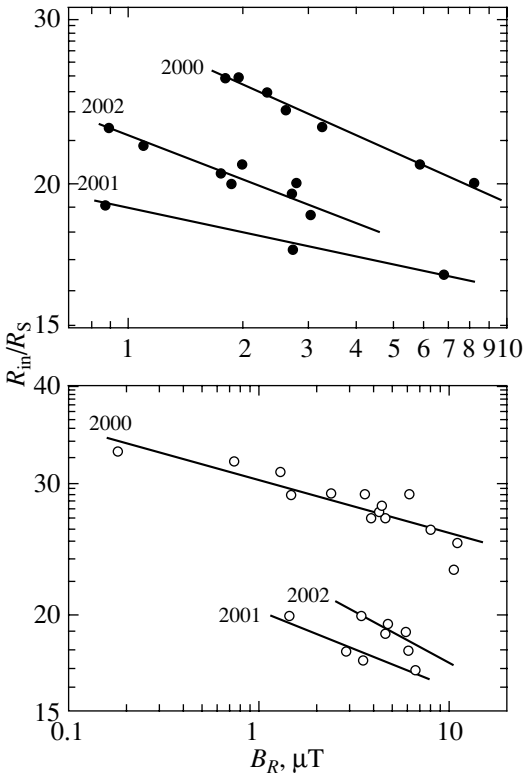


Fig. 5. Evolution of the $R_{\rm in} = F(|B_{\rm R}|)$ correlation at the solar maximum for two slow stream components on a logarithmic scale.

(streamers) have higher velocities than the streams from sources with a mixed field structure. Therefore, the magnetic field of a streamer has a high loop that opens at large heights in the corona where the stream-decelerating transverse magnetic field component B_Z is small. The streams from sources with mixed magnetic fields are associated with the periphery of the streamer or its sidelobes, i.e., with a smaller-height looplike magnetic field structure where the magnetic field component B_Z increases. This explains the relative positions of the slow streams in the $R_{\rm in} = F(|B_{\rm R}|)$ correlation diagram. Figure 5 shows that the enhancement of solar activity at the solar maximum (2000 and 2002) is accompanied in the $R_{\rm in} = F(|B_{\rm R}|)$ correlation diagram by the rise of both branches. Thus, the evolution of the streams at the solar maximum can be associated with the decrease in the scales of the coronal magnetic field structure and with the formation of small-scale loops pressed to the solar surface. Very small-scale structures are illustrated by Fig. 6a. Note that the transverse magnetic field component B_Z strengthens in small-scale structures of this type.

The emergence of an uncorrelated stream component is an unexpected feature of the solar maximum. It is denoted in Fig. 4 by diamonds. In the

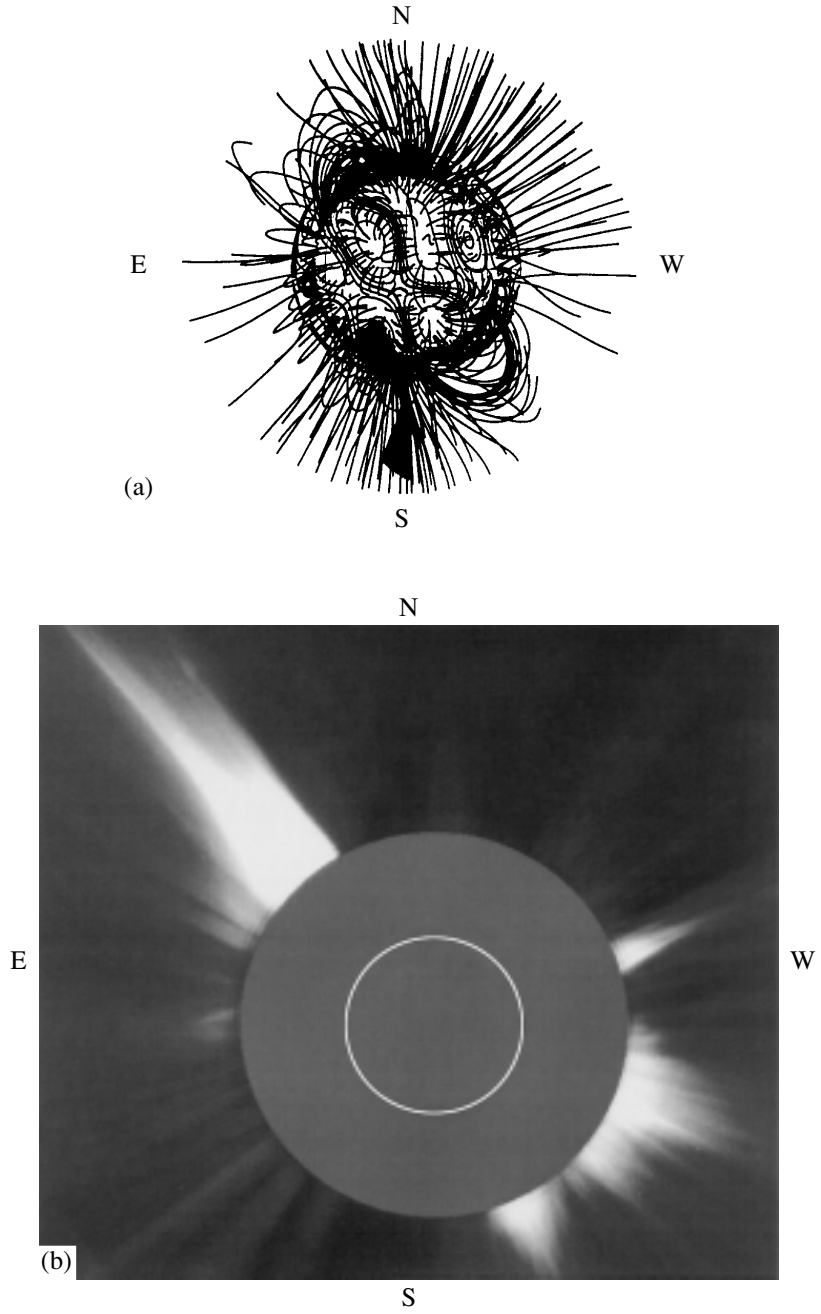


Fig. 6. (a) Small-scale coronal magnetic field structures in the source region; (b) localization of the source of the uncorrelated component on the streamer periphery.

$R_{\rm in} = F(|B_{\rm R}|)$ correlation diagram, this component lies above the previously known correlated stream components and corresponds to the slowest solar-wind streams. The uncorrelated stream component probably arises from the complex interaction between magnetic fields of different types. The photographs taken with the LASCO C2 coronagraph from the *SOHO* spacecraft show (Fig. 6a) that this interaction takes place in a narrow coronal region on the streamer periphery where closely spaced local coronal holes and small-scale magnetic loops are observed. In 2000, the zero magnetic field line of the Sun was strongly bent (Fig. 6b). The open local fields are associated here with the bends of the zero magnetic field line near which conglomerates of local coronal holes emerge. Thus, the structure of the small-scale magnetic field becomes much more complex in years of maximum solar activity. When discussing the nature of the uncorrelated stream component, one should also bear in mind that, as the sizes of the magnetic

field structures decrease, the computational model of the magnetic field B_R in the potential-field approximation becomes unsuitable due to an insufficient field resolution on the source surface ($R = 2.5R_S$).

CONCLUSIONS

Our analysis of the experimental data obtained in the period 2000–2002 in comparison with previous results has revealed that significant changes in the solar-wind stream structure arise at the epoch of maximum solar activity. In the period 2000–2002, the transonic transition region of the solar wind was displaced to outer regions of interplanetary space with the characteristic locations of the boundaries within ~ 15–$60R_S$. In the period 1997–1998, the boundaries were at distances of ~ 10–$40R_S$ from the Sun. The recession of the transition region from the Sun is attributed to the predominance of slow plasma streams in the stationary solar wind. The correspondence between the heliolatitudinal structure of the isophotes of the white-light corona and the radio maps of the solar-wind transition region suggests that the structure and energy of the coronal magnetic fields determine the entire formation of the solar-wind plasma stream structure. Our analysis of the correlation between the location of the inner boundary R_{in} and the coronal magnetic field strength $|B_R|$ on the source surface $R = 2.5R_S$, $R_{in} = F(|B_R|)$, showed the following: the higher the level of solar activity, the larger the slowdown of the previously known slow stream components with differing sources. This implies that the energy of the magnetic field forming the solar-wind streams in the corona decreases with increasing solar activity. At maximum activity, the coronal magnetic field structure and, accordingly, the stream structure change profoundly. A hitherto unknown component without any $R_{in} = F(|B_R|)$ correlation emerges. These streams are the slowest in the overall picture of the solar wind.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 05-02-16090 nd 05-02-31007) and the "Solar Wind: Generation and Interaction with the Earth and Other Planets" program (no. 18).

REFERENCES

1. M. K. Bird and P. Edenhofer, in *Physics of the Inner Heliosphere,* Ed. by R. Schwenn and E. Marsch (Springer-Verlag, Berlin, 1990), p. 13.
2. M. K. Bird, M. Patzold, P. Edenhofer, *et al.*, Sol. Phys. **316**, 441 (1996).
3. M. K. Bird, H. Volland, M. Patzold, *et al.*, Astrophys. J. **426**, 373 (1994).
4. G. E. Bruckner, R. A. Hovard, M. J. Koomen, *et al.*, Sol. Phys. **162**, 357 (1995).
5. A. I. Efimov, Space Sci. Rev. **70**, 397 (1994).
6. A. I. Efimov, I. V. Chasheĭ, V. I. Shishov, and O. I. Yakovlev, Kosm. Issled. **28**, 581 (1990).
7. K. Fujiki, M. Kojima, M. Tokumaru, *et al.*, *Solar Wind Ten. Proceedings of the Tenth International Solar Wind Conference,* Ed. by M. Velly, R. Bruno, and F. Malara; AIP Conf. Proc. **679**, 141 (2003).
8. J. T. Hoeksema, J. M. Wilcox, and P. H. Scherer, J. Geophys. Res. **87**, 10331 (1982).
9. J. T. Hoeksema, J. M. Wilcox, and P. H. Scherer, J. Geophys. Res. **88**, 9910 (1983).
10. E. V. Ivanov, V. N. Obridko, E. V. Nepomnyashchaya, and V. N. Kutilina, Sol. Phys. **184**, 369 (1999).
11. P. Janardhan, M. K. Bird, P. Edenhofer, *et al.*, Sol. Phys. **184**, 157 (1999).
12. M. Kojima, K. Fujiki, K. Hakamada, *et al.*, Adv. Space Res. **25**, 1893 (2000).
13. M. Kojima and T. Kakinuma, J. Geophys. Res. **92**, 7269 (1987).
14. M. Kojima and T. Kakinuma, Space Sci. Rev. **53**, 173 (1990).
15. N. A. Lotova, in *Solar Wind Seven. Proceedings of the 3rd COSPAR Colloquium, Goslar, Germany, 1991,* Ed. by E. Marsch and R. Schwenn (Pergamon, Oxford, 1992), Vol. 3, pp. 217–220.
16. N. A. Lotova, D. F. Blums, and K. V. Vladimirskii, Astron. Astrophys. **150**, 266 (1985).
17. N. A. Lotova, V. N. Obridko, and K. V. Vladimirskii, Astron. Astrophys. **357**, 1051 (2000).
18. N. A. Lotova, V. N. Obridko, K. V. Vladimirskii, *et al.*, Sol. Phys. **79**, 377 (2002).
19. N. A. Lotova, V. N. Obridko, and K. V. Vladimirskii, in *Solar Wind Ten. Proceedings of the Tenth International Solar Wind Conference,* Ed. by M. Velly, R. Bruno, and F. Malara; AIP Conf. Proc. **679**, 110 (2003).
20. N. A. Lotova, K. V. Vladimirskii, and O. A. Korelov, Sol. Phys. **172**, 225 (1997).
21. N. A. Lotova, K. V. Vladimirskii, V. N. Obridko, and I. A. Subaev, Pis'ma Astron. Zh. **30**, 387 (2004) [Astron. Lett. **30**, 343 (2004)].
22. N. A. Lotova, K. V. Vladimirskii, I. V. Yurovskaya, and O. A. Korelov, Astron. Zh. **72**, 757 (1995) [Astron. Rep. **39**, 675 (1995)].
23. D. O. Muhleman and J. D. Anderson, Astrophys. J. **247**, 1093 (1981).
24. V. N. Obridko and B. D. Schelting, Sol. Phys. **187**, 185 (1999).
25. B. J. Rickett and W. A. Coles, J. Geophys. Res. **96**, 1717 (1991).
26. R. Schwenn, in *Physics of the Inner Heliosphere*, Ed. by R. Schwenn and E. Marsch (Springer-Verlag, Berlin, 1990), p. 99.
27. M. Tokumaru, H. Mori, T. Tanaka, and Konolo, J. Geomagn. Geoelectr. **47**, 1113 (1995).

28. C.-Y. Tu and E. Marsch, in *Solar Wind Nine. Proceedings of the Ninth International Solar Wind Conference, Nantucket, MA, 1998,* Ed. by Sh. R. Habbal, R. Esser, J. Hollwed, and P. A. Isenberg; AIP Conf. Proc. **471**, 373 (1999).
29. I. S. Veselovskiĭ, A. V. Dmitriev, O. A. Panasenko, and A. V. Suvorova, Astron. Zh. **76**, 558 (1999) [Astron. Rep. **43**, 485 (1999)].
30. I. S. Veselovskiĭ, A. V. Dmitriev, and A. V. Suvorova, Astron. Vestn. **32**, 352 (1998) [Sol. Syst. Res. **32**, 310 (1998)].
31. K. V. Vladimirskiĭ, N. A. Lotova, and V. N. Obridko, Kratk. Soobshch. Fiz., No. 2, 23 (2003).
32. P. Wohlmuth, D. Plettemeier, P. Edenhofer, *et al.*, Space Sci. Rev. **97**, 9 (2001).
33. R. Woo and S. R. Habbal, Astrophys. J. **510**, L69 (1999).
34. R. Woo and S. R. Habbal, Astrophys. J. **549**, L253 (2001).

Translated by V. Astakhov

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 557–562. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 622–627.

To the Description of Long-Term Variations in the Solar Magnetic Flux: The Sunspot Area Index

Yu. A. Nagovitsyn*

Pulkovo Astronomical Observatory, Russian Academy of Sciences, Pulkovskoe sh. 65, St. Petersburg, 196140 Russia

Received December 28, 2004

Abstract—We show that the Wolf sunspot numbers W and the group sunspot numbers GSN are physically different indices of solar activity and that it is improper to compare them. Based on the approach of the so-called "primary" indices from the observational series of $W(t)$ and $GSN(t)$, we suggest series of yearly mean sunspot areas beginning in 1610 and monthly mean sunspot areas beginning in 1749. © *2005 Pleiades Publishing, Inc.*

Key words: *Sun, solar activity.*

INTRODUCTION

Since the publication of a paper by Hoyt and Schatten (1998), in which a new long (from 1610 until now) series of group sunspot numbers $GSN(t)$ was derived by revising the historical observations, heliophysicists have been discussing questions that in one way or another relate to the reliability of this series, its comparison with the traditional series of Wolf sunspot numbers $W(t)$, and even the possibility of replacing the "old" index by a "new" one (Hathaway *et al.* 2002; Usoskin and Mursula 2003; Usoskin and Kovaltsov 2004). At the same time, it is clear that both the Wolf sunspot numbers and the group sunspot numbers are not well-grounded indices from a physical point of view. When describing the solar magnetic field variations, it would be desirable to have the same long series of physically more obvious characteristics in place of them.

An index that largely meets this requirement exists: this is the total (over the solar disk) sunspot area $A(t)$ that is related to the total (absolute) magnetic flux of the sunspot component of solar activity. Indeed, taking the approximation for the flux (in Mx) of an individual sunspot in the form (Allen 1973)

$$\Phi_j = 0.39 B_{j0} \pi r_j^2 = 1.21 \times 10^{16} B_{j0} A_j,$$

where B_{j0} is the magnetic field strength at the sunspot center in G; r_j is the effective radius of the sunspot penumbra in cm; A_j is the sunspot area in millionths of the solar hemisphere (msh, the standard unit of area in observations); and, accordingly, the total absolute sunspot magnetic flux is

$$\Phi_\Sigma = \sum |\Phi_j| = 1.21 \times 10^{16} \sum |B_{0j}|\, A_j,$$

we notice that the spot-to-spot variations in the central magnetic field strengths are much smaller than the accompanying sunspot area variations (thus, for example, according to Allen (1973), as A increases by a factor of 400, B_0 increases, on average, by only a factor of 4). On this basis, we can factor the mean value of B_0 outside the sum sign and assume that

$$\Phi_\Sigma(t)\,[\mathrm{Mx}] \approx 2.49 \times 10^{19} \sum A_j(t) \qquad (1)$$
$$= 2.49 \times 10^{19} A(t)\,[\mathrm{msh}].$$

Here, we took $B_0 = 2050$ G obtained using the Pulkovo combined database of sunspot magnetic fields (1957–1997), which can be found at the site http://www.gao.spb.ru/database/mfbase, as the mean value. Thus, we can estimate the corresponding changes in the total sunspot magnetic flux from the observations of variations in the total sunspot areas. However, this is beyond the scope of this paper, and below we will talk mainly about the area index.

Unfortunately, the Greenwich series of $A(t)$ that is most commonly used in practice cannot compete in its duration (1874–1976) with the Wolf and Hoyt–Schatten series: it is shorter than these series by factors of 3 and 4, respectively.

*E-mail: nag@gao.spb.ru

1063-7737/05/3108-0557$26.00

WOLF SUNSPOT NUMBERS AND CHANGES IN THE SYSTEM OF THEIR CALCULATION

Despite the apparent simplicity of its calculation, the Wolf sunspot number or relative sunspot number

$$W = k(10G + f), \quad (2)$$

where G is the number of sunspot groups, f is the number of all sunspots, and k is the reduction coefficient of the system of observations to the standard system, is rather "treacherous" from the standpoint of preserving the homogeneity of the system. Thus, for example, Gnevyshev *et al.* (1985) showed that among the 15 stations of the Worldwide Solar Service considered by them in the selected 20-year-long period, only the Kislovodsk Mountain Station of the Pulkovo Astronomical Observatory, Russian Academy of Sciences, preserved its mutual homogeneity with the reference Zurich series. The observations at other stations were recognized to one degree of another to be unsatisfactory; this was probably the result of a difference in the optical parameters of the instruments and in the particular features of the calculation technique, an insufficient number of sunny days, and subjective factors related to the observations.

As regards the long-term internal homogeneity of the Zurich series itself, the following should be noted here. The Wolf sunspot numbers before 1749 cannot be recognized to be quite reliable, because they were reconstructed by Wolf (and, in part, later by Waldmeier) from very fragmentary data, Therefore, between 1700 to 1749, we have only yearly mean estimates of this index. The remark on the insufficient reliability of W also largely applies to the monthly mean data from 1749 until 1825, because Schwabe began truly regular ($\sim$200–270 days per year) sunspot observations only in 1826.

The year 1848 marks the beginning of systematic observations of solar activity by Wolf, and since that time we has been possible to speak about fairly reliable daily and, accordingly, monthly and yearly mean values of $W(t)$. In 1874, after almost 10-year-long photographic observations in Kew (De la Rue *et al.* 1870), the Royal Greenwich Observatory began its (also photographic) observations of the Sun with the determination of a number of sunspot activity parameters, including the number of groups $G(t)$ and the sunspot areas. Recognizing that the W observations since 1848 have been performed on a professional basis, we nevertheless must remember the following. Wolfer, who replaced Wolf as the main observer of sunspot activity, changed the technique of calculating the relative sunspot numbers. Wolf did not consider isolated solitary pores as groups, and his method of counting small sunspots in complex groups is not quite clear at all (Vitinsky *et al.* 1986). After 1894, Wolfer assigned weights from 1 to 5 to sunspots with different areas, while individual pores became groups in his technique (i.e., they gave the contribution $W = 11$). To correct such a change in the system, Wolfer multiplied numbers derived using formula (2) by $k = 0.6$.

Such a calculation technique was also preserved by Brunner, who replaced Wolfer. At the same time, the next main observer, Waldmeier, again changed the method of calculating W. He included the introduced Zurich class of a group (Vitinsky *et al.* 1986). In addition, Waldmeier began to perform several rather than one observation per day, which increased the contribution of one-day groups and also could not but affect the system of Wolf sunspot numbers. In 1980, the service of Wolf sunspot numbers was moved from Zurich to Brussels. The implications were quick to have an effect (Gnevyshev *et al.* 1986).

Thus, the criticism of the Zurich international series of Wolf sunspot numbers has some justification.

GROUP SUNSPOT NUMBERS AND POSSIBLE LOSSES OF INFORMATION

By 1998, Hoyt and Schatten had carried out a huge archival work to create a new time series of group sunspot numbers $GSN(t)$ that was completed by a generalizing publication (Hoyt and Schatten 1998). In particular, this series includes the unprecedentedly regular (for the 17th century) solar observations by the French school (mainly Picard and de la Hire) during the Maunder minimum, which were not used by Wolf in constructing the $W(t)$ series. The GSN index itself is not new: the number of groups G was given in the Greenwich catalogs. Hoyt and Schatten only supplied it with a factor of 12.08, reducing GSN to the Wolf scale:

$$GSN = 12.08G. \quad (3)$$

The long-term behavior of the $GSN(t)$ series differs from that of the $W(t)$ series, demonstrating a clear tendency to decrease when moving backward in time. This aroused criticism of using the $GSN(t)$ series by some researchers, the meaning of which can be formalized as follows.

The Wolf and Hoyt–Schatten works are separated by almost 100 years. The useful historical information I can be divided into published I_r (scientific journals, monographs, etc.) and purely archival I_a (unpublished logs of observations, reports, lists, etc.) types. Whereas I_r is preserved fairly well with time and we can assume that

$$\frac{\partial I_r}{\partial t} \approx 0,$$

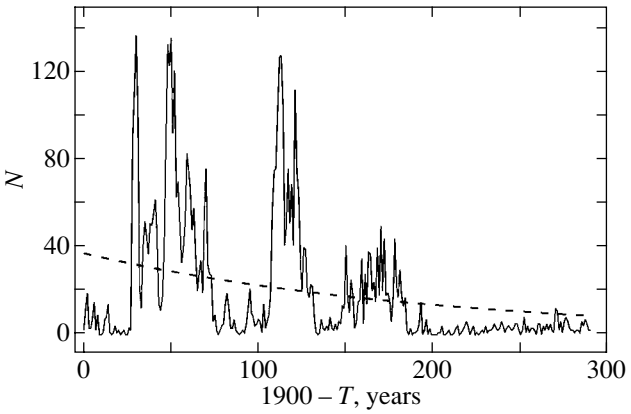


Fig. 1. Losses of information with time: number of reports on aurorae N versus time; the number of years until 1900 is along the horizontal axis.

the second type of information is subject to temporal dispersion (losses from analysis):

$$\frac{\partial I_{\rm a}}{\partial t} = -\gamma I_{\rm a},$$

where γ is the dispersion decrement (constant only in the first approximation). This effect can be estimated, e.g., from the series of aurorae by Krivsky (1984) compiled largely from $I_{\rm a}$-type information. Calculations yield $\gamma = 0.005$ year^{-1}, and the corresponding information "half-life" $T_{1/2}$ is 140 years (see Fig. 1). This time is comparable to the interval between Wolf's work on the compilation of pre-Zurich observations (the early 1890s) and the studies by Hoyt and Schatten (the mid- to late 1990s), so we may conclude that part of the information on sunspot observations used by Wolf might not be included in the $GSN(t)$ series. This can explain, in particular, the fact (Hathaway *et al.* 2002) that a number of standard solar-activity "rules" manifest themselves to a larger extent precisely for the Wolf sunspot numbers rather than for the group sunspot numbers.

PRIMARY INDICES AND SUNSPOT AREAS IN THE INTERVAL 1610–2000

Thus, in the two previous sections, we pointed out the drawbacks of both the $W(t)$ and $GSN(t)$ series. It seems to us that the problem of choosing between these series as formulated by a number researchers still does not exist. In our opinion, the $W(t)$ and $GSN(t)$ series are just physically different indices, although they may be represented by time series of slightly different quality. Indeed, let us consider the time dependence of the difference $\Delta(t) = W(t) - GSN(t)$ (see Fig. 2). We see that $\Delta(t)$ depends on the phase of the 11-year cycle (at maxima, this quantity

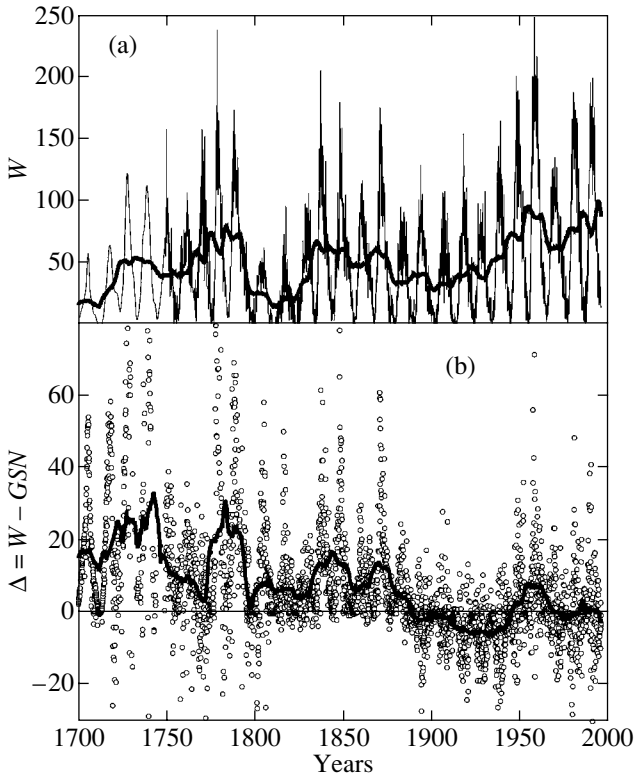


Fig. 2. Difference in the behavior of the Wolf sunspot numbers W and the Hoyt–Schatten GSN index: (a) W numbers and their values smoothed over 132 months; (b) time dependence of the difference between W and GSN and the corresponding smoothed values.

is maximal; this is especially notable for powerful cycles); closely follows the phase dependence of secular cycles; and, at the same time, has an opposite trend with a long (more than 300 years) variation of activity. Thus, we conclude that a simple global linear relationship between $W(t)$ and $GSN(t)$ is not possible.

This conclusion becomes understandable if we recall that, according to a number of papers of the 1960s–1970s (Vitinsky *et al.* 1986), the main sunspot activity indices (sunspot areas, Wolf sunspot numbers, and group sunspot numbers), to within some general assumptions, can be represented using (2) and (3) as

$$A = \frac{1}{m} f_0 T_0^2, \qquad (4)$$
$$W = k f_0 T_0 10 + M,$$
$$GSN = 12.08 k f_0 T_0,$$
$$M = \int_0^\infty T F(T) \overline{M}(T) dT,$$

where $T_0(t)$ and $f_0(t)$ are the so-called "primary" indices, which reflect such physical characteristics of

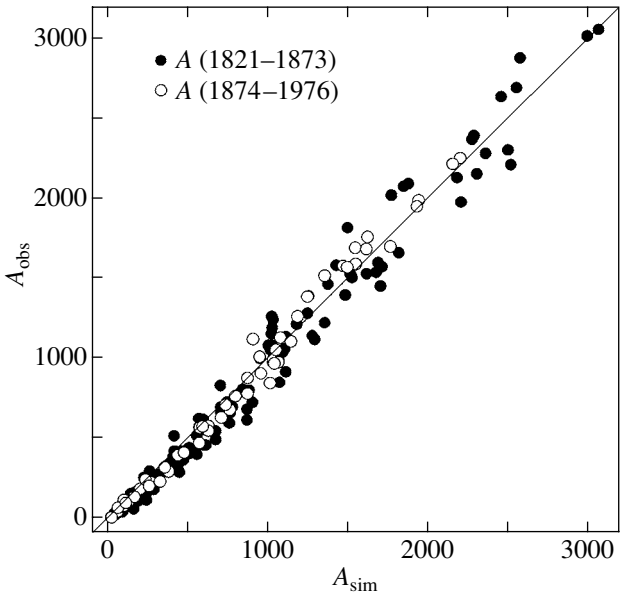


Fig. 3. Comparison of the yearly mean total sunspot areas A_{sim} calculated using formula (10) with the observed areas A_{obs} for various time intervals.

the sunspot formation process as the power and frequency, respectively. Here, we took into account the fact that the factor 12.08 in (3) corrects the coefficient k in formula (2) to unity. The index $M(t)$ is the total number of individual sunspots in groups, $\overline{M}(t)$ is the mean number of sunspots in a group as a function of its lifetime T, and $F(T)$ is the density function of the distribution of sunspot groups in their lifetime. Based on data from the Greenwich catalogs, Ringnes (1964) calculated $F(T)$ using the technique by Gnevyshev (1938). Based on these calculations, Vitinsky *et al.* (1986) take this function in the form

$$F(T) = \lambda e^{-\lambda T}, \quad T_0 = \lambda^{-1}. \tag{5}$$

As we see, from the viewpoint of the approach of primary indices (4), it is hard to imagine a logically justified relationship linking the main indices in pairs: $W = aGSN + b$ or, e.g., $A = aW + b$, $A = aGSN + b$. Therefore, in particular, Vaquero *et al.* (2004), who used the last two relationships to compile a prolonged homogeneous series of sunspot areas on the basis of Greenwich and pre-Greenwich observations, were unlikely to have achieved their main goal.

To find an expression for the Wolf sunspot numbers in (4), Kopecky (1983) and Vitinsky *et al.* (1986) assumed that

$$\overline{M}(T) \sim T. \tag{6}$$

In this case, we obtain

$$W = 10kf_0T_0 + mf_0T_0^2. \tag{7}$$

However, more recent studies (Kuklin and Kopecky 1988), which were based on the Greenwich catalogs (i.e., on reliable observational data), showed that relation (7) does not hold in practice: statistical estimations of the coefficient m yield a significantly negative value, while this is not possible, because the second term in (7) means the number of observed sunspots. To find the correlations between the indices, this forced the authors of the above paper to search for empirical statistical relationships, which, generally speaking, do not follow in any way from the theory that led to system (4).

To resolve this contradiction, we assumed that the dependence of the mean number of sunspots in a group on its lifetime is stronger than (6), more specifically,

$$\overline{M}(T) \sim T^2. \tag{8}$$

In this case, given (4), the expression for the Wolf sunspot numbers in terms of the primary indices takes the form

$$W = 10kf_0T_0 + m_1f_0T_0^3. \tag{9}$$

To substantiate assumption (6), Kopecky (1983) provided experimental data based on the Greenwich catalog (the total (over 11-year cycles) indices) and statistically found them to be consistent with a dependence of form (6). We tested our assumption (8) using the same data. It turned out that it is even in better agreement with the observations than (6), although the difference in this case is statistically insignificant. This may have forced Kopecky to dwell on a simpler, linear form of the $\overline{M}(T)$ dependence.

As regards system (4), using (5) and (8), excluding f_0 and T_0, we can obtain from it the relationship of the sunspot areas to W and GSN:

$$A^2(t) = aW(t)GSN(t) - bGSN^2(t). \tag{10}$$

Thus, we can obtain a prolonged series of physically proper sunspot areas required to meet the goal of this paper from the observational series of Wolf sunspot numbers and group sunspot numbers. For this purpose, we find the coefficients in formula (10) from the yearly mean values by the least-squares method on the entire time interval of the existence of the original 1874–1976 Greenwich series. We obtain

$$a = 539 \pm 60, \quad b = 276 \pm 61. \tag{11}$$

As we see, the coefficients differ significantly from zero; the derived correlation coefficient is $r = 0.989$.

To independently test the reliability of the derived values and the validity of formula (10), we use our reconstruction of the series of areas $A_L(t)$ on the interval 1821–1873 (Nagovitsyn 1997a, Nagovitsyn *et al.* 2004) obtained from the observations by Schwabe, Carrington, and de la Rue. Comparing $A_L(t)$ and the

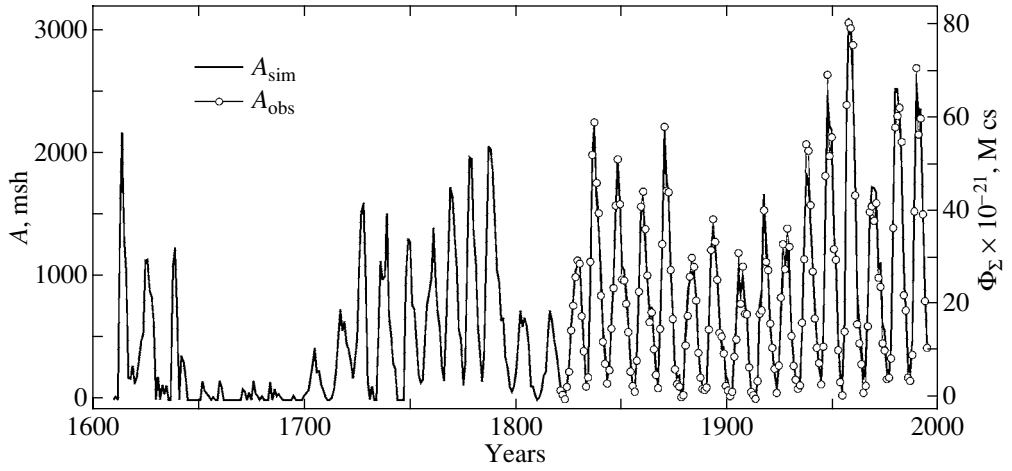


Fig. 4. Comparison of the series of yearly mean total sunspot areas $A_{\rm sim}(t)$ calculated using formula (10) with the series of observed sunspot areas $A_{\rm obs}(t)$ and the corresponding total sunspot magnetic flux Φ_Σ.

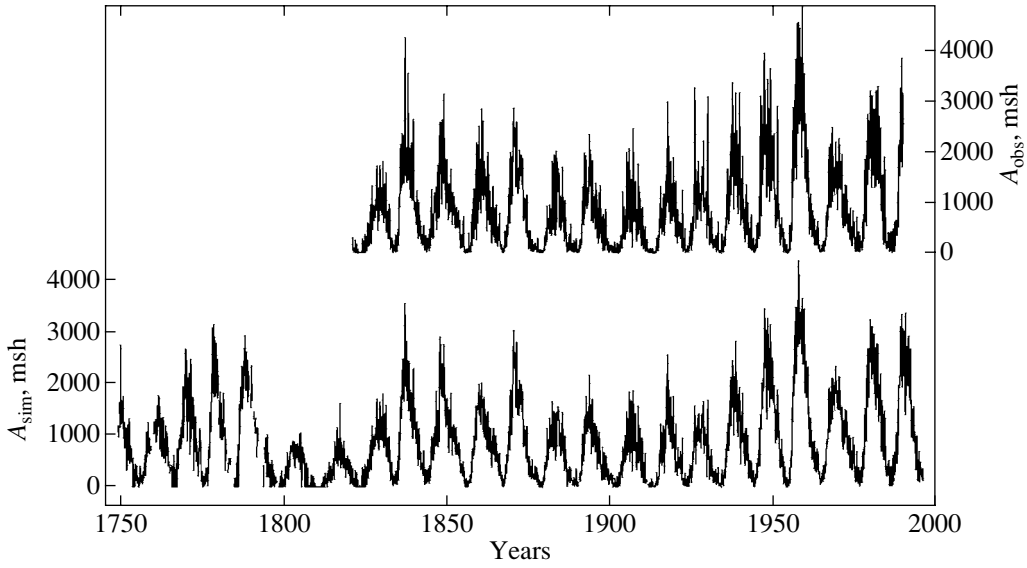


Fig. 5. Comparison of the series of monthly mean total sunspot areas $A_{\rm sim}(t)$ calculated using formula (10) with the series of observed sunspot areas $A_{\rm obs}(t)$.

areas calculated using formula (10) with the values of a and b (11), we find that the correlation coefficient $r = 0.994$ (see Fig. 3); i.e., our procedure passes the test. Therefore, we use the entire interval 1821–1994 to determine the coefficients in (10), and, having obtained

$$a = 582 \pm 34, \quad b = 319 \pm 36 \qquad (12)$$

with the correlation coefficient $r = 0.990$, we construct a series of yearly mean sunspot areas $A_{\rm sim}(t)$ on the interval 1610–1994 (see Fig. 4, where the right vertical axis scales the derived series to the total absolute sunspot magnetic flux according to (1)).

Since the Zurich series of Wolf sunspot numbers begins only from 1700, for the interval 1610–1699, we used the model values $W(t)$ from our version (Nagovitsyn 1997b) constructed using the approach of nonlinear oscillations as well as the data by Schove (1983) on the epochs of extrema of the 11-year cycles in the past. As was shown by Miletsky *et al.* (2004), the behavior of this series is in satisfactory agreement with data from other sources. In our version, the maximum values of W for the period of the Maunder minimum were also taken from the monograph by Schove (1983). They could be slightly overestimated, and, consequently, the sunspot areas in Fig. 4 for the second half of the 17th century represent an upper limit for this index. In general, the values of $A(t)$ before 1700 are approximate.

In addition to the yearly mean sunspot areas, we also calculated the monthly mean sunspot areas since 1749 using (10). The results are presented in Fig. 5. The correlation coefficient between the calculated values and the actual observations is slightly smaller than that for the yearly mean sunspot areas $A(t)$, but it is not low either, 0.951.

RESULTS

We showed that the Wolf sunspot number W and the group sunspot number GSN are physically different indices of solar activity, and that it is improper to compare them.

Based on the approach of the so-called "primary" indices, we constructed the series of total sunspot areas and total sunspot magnetic flux since 1610 for the yearly mean values and since 1749 for the monthly mean values from the observational series of $W(t)$ and $GSN(t)$.

To conclude, we note once again that the solar magnetic field indices, which have a clear physical meaning, must be used in heliophysical studies.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project nos. 03-02-17505 and 04-02-17560), INTAS (grant 2002-00550), and the "Solar Activity and Physical Processes in the Sun-Earth System" Program of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. C. W. Allen, *Astrophysical Quantities*, 3rd ed. (Athlone, London, 1973; Mir, Moscow, 1977).
2. W. De la Rue, B. Stewart, and B. Loewy, Phil. Trans. **160**, 389 (1870).
3. M. N. Gnevyshev, Tsirk. Gl. Astron. Obs. **24**, 37 (1938).
4. M. N. Gnevyshev, Yu. A. Nagovitsin, and E. Yu. Nagovitsin, Soln. Dannye **2**, 72 (1985).
5. M. N. Gnevyshev, Yu. A. Nagovitsin, and E. Yu. Nagovitsin, Soln. Dannye **3**, 57 (1986).
6. D. H. Hathaway, R. M. Wilson, and E. J. Reichmann, Sol. Phys. **211**, 357 (2002).
7. D. V. Hoyt and K. Schatten, Sol. Phys. **179**, 189 (1998).
8. M. Kopecky, Bull. Astron. Inst. Czech. **34**, 23 (1983).
9. L. Krivsky, Sol. Phys. **93**, 189 (1984).
10. G. V. Kuklin and M. Kopecky, Bull. Astron. Inst. Czech. **39**, 141 (1988).
11. E. V. Miletsky, V. G. Ivanov, Yu. A. Nagovitsyn, and H. Jungner, Sol. Phys. **224**, 77 (2004).
12. Yu. A. Nagovitsyn, Soln. Dannye, 38 (1997a).
13. Yu. A. Nagovitsyn, Pis'ma Astron. Zh. **23**, 851 (1997b) [Astron. Lett. **23**, 742 (1997b)].
14. Yu. A. Nagovitsyn, V. G. Ivanov, E. V. Miletsky, and D. M. Volobuev, Sol. Phys. **224**, 103 (2004).
15. T. S. Ringnes, Astrophys. Norv. **9**, 95 (1964).
16. D. J. Schove, *Sunspot Cycles* (Hutchinson Ross. Publ., Stroudsburg, 1983).
17. I. G. Usoskin and G. A. Kovaltsov, Sol. Phys. **224**, 37 (2004).
18. I. G. Usoskin and K. Marsula, Sol. Phys. **218**, 319 (2003).
19. J. M. Vaquero, F. Sanchez-Bajo, and M. C. Gallego, Sol. Phys. **221**, 179 (2004).
20. Yu. I. Vitinsky, M. Kopetskiĭ, and G. V. Kuklin, *Statistics of Sunspot Activity of the Sun* (Nauka, Moscow, 1986) [in Russian].

Translated by G. Rudnitskii

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 563–568. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 628–633.

Conservative Two-Body Problem with Variable Masses

L. G. Luk'yanov*

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

Received February 18, 2005

Abstract—We consider the conservative two-body problem with a constant total mass, but with variable individual masses. The problem is shown to be completely integrable for any mass variation law. The Keplerian motion known for the classical two-body problem with constant masses remains valid for the relative motion of the bodies. The absolute motions of the bodies depend on the center-of-mass motion. Hitherto unknown quadratures that depend on the mass variation law were derived for the integrals of motion of the center of mass. We consider some of the laws that are of interest in studying the motion of close binary stars with mass transfer. © *2005 Pleiades Publishing, Inc.*

Key words: *two-body problem with variable masses, Gylden–Mestschersky problem, center-of-mass motion, close binary stars.*

INTRODUCTION

In astronomy, one often has to deal with variable-mass bodies. In contrast to the two-body problem with constant masses, the problem of two point bodies with variable masses is not integrable in general form. A rigorous solution of the equations of motion can be obtained only for certain mass variation laws.

Gylden (1884) was the first to write and analyze the differential equations of relative motion in the two-body problem with variable masses. Mestschersky (1893, 1902) obtained rigorous solutions in this problem for specific mass variation laws, which are now known as Mestschersky's first, second, and combined laws. In 1902, Mestschersky published the equations of motion for variable-mass bodies, which bear his name. Mestschersky derived the equations of motion in Gylden's problem from these equations by assuming the absence of propulsive forces. Given Mestschersky's contribution, Gylden's problem came to be called the *Gylden–Mestschersky problem*. Much later, Gelfgat (1959) found other mass variation laws in this problem that allowed the equations of motion to be integrated.

Based on the evolution of main-sequence stars, Jeans (1924) formulated the differential mass variation law for stars. Mestschersky's and Gelfgat's laws proved to be special cases of Jeans' law.

Whereas the relative motions of variable-mass bodies were determined, the determination of the absolute motions of these bodies was shown by Luk'yanov (1983) to be reduced to quadratures. Therefore, the absolute motions of the bodies can be determined for Mestschersky's and Gelfgat's integrable cases.

The Gylden–Mestschersky problem found wide applications in astronomy: when allowing for the corpuscular or photon radiation from stars, when mass is lost in binary systems, when mass sticks in nebulae, etc. Subsequently, many authors analyzed this problem. Many modifications and generalizations of these equations are now known. These can be learned about in review papers (see, e.g., Polyakhova (1989)).

Studying the motions of the stars in close binary systems with mass loss or transfer is an important application of the two-body problem with variable masses. Hadjidemetriou (1967), Kruszevski (1966), and Piotrovskiĭ (1967) explored the possibilities of mass ejection from one of the stars and determined the particle trajectories during mass transfer from one star to the other.

In this paper, we consider the two-body problem by assuming that both masses are variable, but the sum of the masses is always constant. We call this problem the *conservative two-body problem with variable masses*.

This formulation of the problem has not been considered, apparently due to its triviality, since the relative motions of the bodies in this problem are known in advance. These are completely identical to the Keplerian motions. Determining the absolute motions of the bodies is reduced to quadratures for any mass variation law. The problem under consideration is of great importance in studying the motions of the stars in close binary systems with mass transfer from one star to the other.

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

1063-7737/05/3108-0563$26.00

THE EQUATIONS OF MOTION

The motions of two mutually attracting bodies M_1 and M_2 with variable masses $m_1(t)$ and $m_2(t)$ in an absolute rectangular coordinate system with the origin at point O are described by the system of differential equations

$$m_1\ddot{\boldsymbol{\rho}}_1 = \frac{fm_1m_2}{r^3}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1), \quad (1)$$
$$m_2\ddot{\boldsymbol{\rho}}_2 = -\frac{fm_1m_2}{r^3}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1),$$

where $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are the radius vectors of bodies M_1 and M_2, f is the universal gravitational constant, $\mathbf{r} = \boldsymbol{\rho}_2 - \boldsymbol{\rho}_1$, and $r = |\mathbf{r}|$ is the distance between the bodies.

The equations of relative motion for body M_2 can be easily derived from Eqs. (1) by the parallel translation of the coordinate origin to point M_1 :

$$\ddot{\mathbf{r}} = -\frac{fm}{r^3}\mathbf{r}, \quad (2)$$

where $m = m_1 + m_2$. According to the formulation of the problem, the total mass of the two bodies is constant: $m = \text{const}$. Equations (2) then closely match the equations of relative motion in the classical two-body problem with constant masses and a total mass m. Therefore, Eqs. (2) are completely integrable, and the trajectory of body M_2 relative to body M_1 is the conic section $r = p/(1 + e\cos v)$, where p is the focal parameter, e is the eccentricity, and v is the true anomaly. The Keplerian elements p, e, i, Ω, ω, and τ can be used as arbitrary constants for the general solution of Eqs. (2).

Below, we use a special system of units. We take the focal parameter of the orbit as the unit of length, $p = 1$, the total mass of the two bodies as the unit of mass, $m = 1$, and choose the unit of time in such a way that the gravitational constant is equal to unity, $f = 1$. We denote the masses of bodies M_1 and M_2 in this system of units by $1 - \mu$ and μ, respectively. The equations of motion (1) and (2) can then be rewritten as

$$(1-\mu)\ddot{\boldsymbol{\rho}}_1 = \frac{\mu(1-\mu)}{r^3}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1), \quad (3)$$
$$\mu\ddot{\boldsymbol{\rho}}_2 = -\frac{\mu(1-\mu)}{r^3}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1),$$

$$\ddot{\mathbf{r}} = -\frac{\mathbf{r}}{r^3}. \quad (4)$$

Since the general integral of Eqs. (4) is well known, finding the general integral of Eqs. (3) requires finding six more independent first integrals. In the problem with constant masses, the integrals of center-of-mass motion are such integrals. As will be seen below, in the problem under consideration, the equations of motion (3) also admit of the existence of integrals of center-of-mass motion.

THE MOTION OF THE CENTER OF MASS

To derive the integrals of center-of-mass motion, we assume that the function $\mu(t)$ and its first derivative $\dot{\mu}(t)$ are integrable.

The relation

$$(1-\mu)\ddot{\boldsymbol{\rho}}_1 + \mu\ddot{\boldsymbol{\rho}}_2 = 0 \quad (5)$$

can be easily derived from Eqs. (3). Integrating it twice yields

$$(1-\mu)\dot{\boldsymbol{\rho}}_1 + \mu\dot{\boldsymbol{\rho}}_2 - \int \dot{\mu}\dot{\mathbf{r}}dt = \mathbf{a}, \quad (6)$$

$$(1-\mu)\boldsymbol{\rho}_1 + \mu\boldsymbol{\rho}_2 - \int \dot{\mu}\mathbf{r}dt \quad (7)$$
$$- \int\int \dot{\mu}\dot{\mathbf{r}}dtdt = \mathbf{a}t + \mathbf{b},$$

where $\mathbf{a}$ and $\mathbf{b}$ are arbitrary vector constants.

If we assume that the function $\mu(t)$ is given and that the general solution of Eqs. (4) has been determined, then relations (6) and (7) are the first integrals of the equations of motion (3). Together with the general integral of Eqs. (4), they form the general integral of Eqs. (3).

To elucidate the mechanical meaning of integrals (6) and (7), let us denote the radius vector of the center of mass of bodies M_1 and M_2 by $\boldsymbol{\rho}_G$. Then, the following relations are valid for $\boldsymbol{\rho}_G$ and $\dot{\boldsymbol{\rho}}_G$:

$$\boldsymbol{\rho}_G = (1-\mu)\boldsymbol{\rho}_1 + \mu\boldsymbol{\rho}_2, \quad (8)$$
$$\dot{\boldsymbol{\rho}}_G = (1-\mu)\dot{\boldsymbol{\rho}}_1 + \mu\dot{\boldsymbol{\rho}}_2 + \dot{\mu}\mathbf{r},$$

which allow integrals (6) and (7) to be represented as

$$\dot{\boldsymbol{\rho}}_G = \mathbf{a} + \dot{\mu}\mathbf{r} + \int \dot{\mu}\dot{\mathbf{r}}dt, \quad (9)$$

$$\boldsymbol{\rho}_G = \mathbf{a}t + \mathbf{b} + \int \dot{\mu}\mathbf{r}dt + \int\int \dot{\mu}\dot{\mathbf{r}}dtdt. \quad (10)$$

The derived integrals describe the motion of the center of mass of the two bodies in the absolute coordinate system. At $\dot{\mu} = 0$, these integrals transform into the well-known integrals for constant-mass bodies.

At $\mathbf{a} = \mathbf{b} = 0$, the coordinate origin O in the constant-mass problem coincides with the center of mass G, i.e., $\boldsymbol{\rho}_G = 0$. There is no such coincidence for variable-mass bodies. To estimate the deviation of the center of mass from the coordinate origin on a short time interval, $\dot{\mu}$ can be roughly assumed to be

constant. We then derive the following approximate formulas from Eqs. (9) and (10) at $\mathbf{a} = \mathbf{b} = 0$:

$$\dot{\boldsymbol{\rho}}_G \approx 2\dot{\mu}\mathbf{r}, \qquad \boldsymbol{\rho}_G \approx 2\dot{\mu}\int \mathbf{r}dt. \tag{11}$$

Formulas (11) become exact in the case of a constant mass transfer rate (see below).

We see from Eqs. (9) and (10) that at a low mass transfer rate $\dot{\mu}$, the center of mass is located near the coordinate origin, and its velocity is also low. If $\dot{\mu} > 0$, the motion of body M_2 relative to body M_1 is directed toward the origin O of the absolute coordinate system. At $\dot{\mu} < 0$, the motion of body M_2 mentioned above is directed away from the origin of the absolute coordinate system. If, for example, we imagine that the derivative $\dot{\mu}$ changes its sign instantaneously at a certain instant of time and assume that the vectors $\mathbf{a}$ and $\mathbf{b}$ are still equal to zero, then the coordinate origin O will change instantaneously to the symmetric location O' on the opposite side of segment M_1M_2 (in this case, the relative motion of body M_2 will not change in any way).

The general solution of the equations of relative motion (4) and the integrals of center-of-mass motion (9) and (10) can be used to construct the general solution of Eqs. (3) without any additional integration in the form

$$\begin{aligned} \boldsymbol{\rho}_1 &= \boldsymbol{\rho}_G - \mu\mathbf{r}, \\ \dot{\boldsymbol{\rho}}_1 &= \dot{\boldsymbol{\rho}}_G - \mu\dot{\mathbf{r}} - \dot{\mu}\mathbf{r}, \end{aligned} \tag{12}$$

$$\begin{aligned} \boldsymbol{\rho}_2 &= \boldsymbol{\rho}_G + (1-\mu)\mathbf{r}, \\ \dot{\boldsymbol{\rho}}_2 &= \dot{\boldsymbol{\rho}}_G + (1-\mu)\dot{\mathbf{r}} - \dot{\mu}\mathbf{r}. \end{aligned}$$

The motion of the barycenter of the system for variable-mass bodies follows a law that differs from the uniform and rectilinear law. The barycentric coordinate system for these bodies ceases to be inertial. Given the acceleration in the center-of-mass motion $\ddot{\boldsymbol{\rho}}_G = 2\dot{\mu}\dot{\mathbf{r}} + \ddot{\mu}\mathbf{r}$, the equations of barycentric motion for bodies M_1 and M_2 can be written as

$$\begin{aligned} \ddot{\mathbf{s}}_1 &= \frac{\mu}{r^3}\mathbf{r} - 2\dot{\mu}\dot{\mathbf{r}} - \ddot{\mu}\mathbf{r}, \\ \ddot{\mathbf{s}}_2 &= \frac{1-\mu}{r^3}\mathbf{r} - 2\dot{\mu}\dot{\mathbf{r}} - \ddot{\mu}\mathbf{r}, \end{aligned} \tag{13}$$

where $\mathbf{s}_1 = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_G$ and $\mathbf{s}_2 = \boldsymbol{\rho}_2 - \boldsymbol{\rho}_G$ are the barycentric radius vectors of bodies M_1 and M_2, respectively.

To summarize, we may note that the general conservative two-body problem with variable masses is completely integrable (i.e., is reduced to quadratures) for any mass variation law. The relative motion closely matches the Keplerian motion, and the absolute motion is defined by Eqs. (9), (10), and (12).

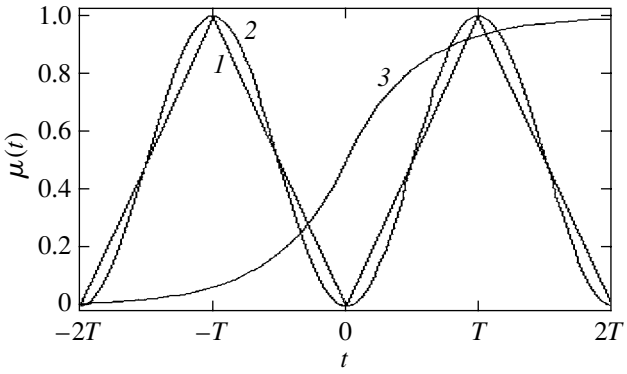


Fig. 1. Mass variation laws $\mu(t)$ considered: (*1*) linear law; (*2*) trigonometric law; and (*3*) exponential law.

The explicit formulas for the indefinite integrals appearing in Eqs. (9) and (10) depend on the specific mass variation law $\mu(t)$. In the next section, we consider some of the mass variation laws that, in our view. are of greatest interest in describing the mass transfer in close binary stars.

MASS VARIATION LAWS

Under additional assumptions, the derived quadratures can be expressed in terms of elementary functions. First, we assume that the relative motion of the bodies is circular, i.e., $e = 0$ and $r = 1$. Second, we consider the three simplest mass variation laws $\mu(t)$.

The Linear Law

This law corresponds to a zero exponent in Jeans' law. Let us define the mass variation by the formula

$$\mu(t) = \left|\frac{t}{T}\right| \tag{14}$$

on the time interval $[-T, T]$ and assume the function $\mu(t)$ to be periodic with the period $2T$, i.e., $\mu(t + 2T) = \mu(t)$, outside this interval. The physical meaning of this law corresponds to periodic multiple complete mass transfer from one body to the other at a constant rate. At $t = kT$, where $k = 0, \pm1, \pm2, \ldots$, the increase in mass changes to its decrease and vice versa; the total mass is always equal to unity. Curve *1* in Fig. 1 represents the function $\mu(t)$.

The integrals of center-of-mass motion (9) and (10) for this law can then be represented as

$$\dot{\boldsymbol{\rho}}_G - 2\dot{\mu}\mathbf{r} = \mathbf{a}, \quad \boldsymbol{\rho}_G + 2\dot{\mu}\dot{\mathbf{r}} = \mathbf{a}t + \mathbf{b}. \tag{15}$$

We thus see that the motion of the center of mass of the two bodies in stationary space is the sum of the uniform and rectilinear motion defined by the vectors

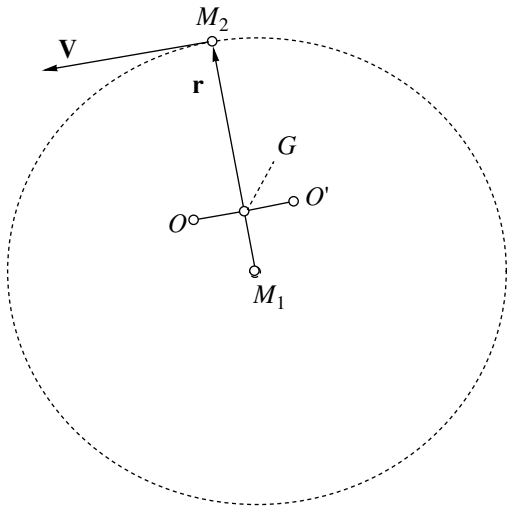


Fig. 2. Relative positions of bodies M_1 and M_2, the center of mass G, and the origin O (or O') of the absolute coordinate system .

a and **b** and the circular motion defined by the constant $\dot{\mu}$ and the relative motion of the bodies with a unit angular velocity. Each component of the center-of-mass motion has its own orientation in space. If the direction of the uniform and rectilinear motion is perpendicular to the plane of the circular relative motion, then the center of mass will move along a spatial curve, a helix. If, however, the vector **a** lies in the plane of the vectors **r** and $\dot{\mathbf{r}}$, then the center of mass will move along a plane curve, a cycloid (including the prolate and curtate cycloids). Equations (15) allow us to consider any other orientations of the components of the center-of-mass motion.

If we exclude the uniform and rectilinear motion, i.e., set $\mathbf{a} = \mathbf{b} = 0$, then the center of mass G in absolute space at $\dot{\mu} = \text{const}$ will move along a circumference of constant radius $\rho_G = |\boldsymbol{\rho}_G| = 2|\dot{\mu}|$.

Figure 2 shows the positions of bodies M_1 and M_2 and points O, O', and G in the relative coordinate system. The dotted line illustrates the trajectory of circular motion of body M_2 around body M_1. The points O and O' indicate the positions of the coordinate origin of the absolute coordinate system at $\dot{\mu} > 0$ and $\dot{\mu} < 0$, respectively.

At $t = 0$, the center of mass G is located at point M_1. As $\mu(t)$ increases, the center of mass approaches point M_2, merging with it at $t = T$. The segment M_1M_2 is always perpendicular to the vector **OG**.

It is thus clear that the absolute motions of bodies M_1 and M_2 are along spirals. The bodies traverse each turn of the spiral in time 2π (the revolution period of the bodies in the relative motion). The unwinding spiral changes to a winding spiral and vice versa after the time interval T (the time of complete mass transfer).

The separation ρ_i between body $M_i (i = 1, 2)$ and the coordinate origin O (at $\mathbf{a} = \mathbf{b} = 0$) oscillates with a period of $2T$ within the range

$$2\dot{\mu} \le \rho_i \le \sqrt{1 + 4\dot{\mu}^2}. \tag{16}$$

The coordinates and velocities of bodies M_1 and M_2 in the absolute motion can be determined from the formulas

$$\boldsymbol{\rho}_1 = -2\dot{\mu}\dot{\mathbf{r}} - \mu\mathbf{r} + \mathbf{a}t + \mathbf{b}, \tag{17}$$
$$\dot{\boldsymbol{\rho}}_1 = \dot{\mu}\mathbf{r} - \mu\dot{\mathbf{r}} + \mathbf{a},$$

$$\boldsymbol{\rho}_2 = -2\dot{\mu}\dot{\mathbf{r}} + (1 - \mu)\mathbf{r} + \mathbf{a}t + \mathbf{b},$$
$$\dot{\boldsymbol{\rho}}_2 = \dot{\mu}\mathbf{r} + (1 - \mu)\dot{\mathbf{r}} + \mathbf{a}.$$

Note also that the first vector integral in formula (15) is also valid at $e \neq 0$, while the second vector integral is valid only for $e = 0$.

The Trigonometric Law

We assume that the mass variation is defined by the dependence

$$\mu(t) = \frac{1}{2}\left[1 - cos\left(\frac{\pi}{T}t\right)\right], \tag{18}$$

which is also indicated by curve *2* in Fig. 1. The mass variation in this law is also periodic with a period of $2T$, but the rate of mass variation is variable. Since the dependence $d^3\mu/dt^3 = -(\pi/T)^2 d\mu/dt$ holds for this law, double integration by parts transforms integrals (9) and (10) to the form

$$\dot{\boldsymbol{\rho}}_G - \frac{T^2}{T^2 - \pi^2}\left[\left(2 - \frac{\pi^2}{T^2}\right)\dot{\mu}\mathbf{r} + \ddot{\mu}\dot{\mathbf{r}}\right] = \mathbf{a}, \tag{19}$$

$$\boldsymbol{\rho}_G - \frac{T^4}{(T^2 - \pi^2)^2}\left[\left(3 - \frac{\pi^2}{T^2}\right)\ddot{\mu}\mathbf{r} - 2\dot{\mu}\dot{\mathbf{r}}\right] = \mathbf{a}t + \mathbf{b}.$$

Here, the absolute motions of the center of mass G and bodies M_1 and M_2 differ from those for the linear law in that, in addition to all of the motions considered, point G together with the segment M_1M_2 execute another motion (dependent on $\ddot{\mu}$), the oscillation of the radius ρ_G with a period of $2T$. In Fig. 2, this means that the distance of point O from the segment M_1M_2 varies. The vector $\boldsymbol{\rho}_G$ at $\mathbf{a} = \mathbf{b} = 0$ is no longer perpendicular to the segment M_1M_2.

Since the inequality $T \gg \pi$ holds for stellar systems, formulas (19) can be simplified by assuming approximately that $\pi/T = 0$.

The Exponential Law

This law corresponds to exponent 1 in Jeans' law. In contrast to the two previous laws, here we consider a monotonic increase in mass $\mu(t)$ from 0 to 1 on an infinite time interval, from $-\infty$ to $+\infty$, in the form

$$\mu(t) = \begin{cases} e^t/2 & \text{at } t \le 0 \\ 1 - e^{-t}/2 & \text{at } t \ge 0. \end{cases} \quad (20)$$

This law is indicated by curve *3* in Fig. 1. Since $d^3\mu/dt^3 = d\mu/dt$, we obtain from Eqs. (9) and (10) by double integration by parts

$$\dot{\boldsymbol{\rho}}_G - \frac{1}{2}(3\dot{\mu}\mathbf{r} + \ddot{\mu}\dot{\mathbf{r}}) = \mathbf{a}, \quad (21)$$
$$\boldsymbol{\rho}_G - \frac{1}{2}(2\ddot{\mu}\mathbf{r} - \dot{\mu}\dot{\mathbf{r}}) = \mathbf{a}t + \mathbf{b}.$$

We thus see that the motions of the center of mass and bodies M_1 and M_2 in stationary space are similar to those for the trigonometric law, but the periodic oscillations of the radius ρ_G with a period of $2T$ are replaced with its monotonic increase over an infinite time interval, from $-\infty$ to $+\infty$.

The circular motions of the center of mass in the examples considered above become elliptical if the relative motions of the bodies are along ellipses.

BINARY STELLAR SYSTEMS

If only the forces of mutual attraction between the stars in a close binary system are taken into account and if the mass transfer is assumed to be instantaneous and without any loss of the total mass, then the conservative two-body problem considered above can be used to study the relative and absolute motions of the stars. The mass variation law can be arbitrary.

It is occasionally assumed that, during mass transfer from the larger star to the smaller star, the two stars approach each other, because the angular momentum changes. However, this reasoning is valid only for the ejection of some mass from one of the stars, i.e., when the system of stars loses its mass. If, however, the ejected mass is assumed to (instantaneously) reach the other star, then the stars do not approach each other, because the angular momentum remains constant in the relative motion. This is the well-known area integral for Eqs. (2). Equations (2) define a trajectory (a conic section) that does not change for any law of mass transfer from one star to the other.

Therefore, if we are interested only in the relative motion of the stars in a close binary system with mass transfer (but with the total mass of the system conserved), then the formulas of the two-body problem with constant masses can be used absolutely rigorously (by assuming the absence of other forces).

If, however, the absolute motions of the stars are of interest, then the motion of the center of mass of the system must be taken into account. This can be done using formulas (12) in this paper. It is inappropriate to use the barycentric coordinate system, because it is no longer inertial.

Let us estimate the difference between the absolute motions of the stars derived in this paper and those calculated by assuming that the stellar masses are constant, i.e., for uniform and rectilinear motion of the center of mass.

This requires estimating the terms on the right-hand sides of Eqs. (9) and (10) or, which is easier, Eqs. (15), (19), and (21). The vectors $\mathbf{a}$ and $\mathbf{b}$ are similar to those in the constant-mass problem. They define the motions of the stars along helical lines or cycloids. Excluding this motion (i.e., setting $\mathbf{a} = \mathbf{b} = 0$), we can use estimate (11). If we assume that the mass of the stellar system is $\sim 10 M_\odot$, the mass outflow rate is $\sim 10^{-3} M_\odot$/yr, and the revolution period of the stars is $\sim 5 \times 10^{-2}$ yr, then we obtain an estimate of $\rho_G/p \approx 10^{-6}$ for the ρ_G/p ratio, where p is the focal parameter of the relative orbit of the stars.

Therefore, it can be argued that the deviation of the center of mass G from the coordinate origin O at $\mathbf{a} = \mathbf{b} = 0$ is negligible. When calculating the orbits of specific stars, we can assume, to within 10^{-6}, that, just as for systems with constant masses, the origin O of the stationary coordinate system (at $\mathbf{a} = \mathbf{b} = 0$) coincides with the barycenter G of the stellar system. We must only take into account the displacement of the barycenter due to the mass transfer. The absolute coordinates and velocities of the stars can then be determined from the following approximate formulas:

$$\boldsymbol{\rho}_1 \simeq \mathbf{a}t + \mathbf{b} - \mu(t)\mathbf{r}, \quad (22)$$
$$\dot{\boldsymbol{\rho}}_1 \simeq \mathbf{a} - \mu(t)\dot{\mathbf{r}},$$
$$\boldsymbol{\rho}_2 \simeq \mathbf{a}t + \mathbf{b} + [1 - \mu(t)]\mathbf{r},$$
$$\dot{\boldsymbol{\rho}}_2 \simeq \mathbf{a} + [1 - \mu(t)]\dot{\mathbf{r}}.$$

Of course, these estimates will change if we consider sharp (explosive) mass variations.

CONCLUSIONS

In the two-body (and even many-body) problem with constant masses, the integrals of center-of-mass motion always exist and do not depend on whether we know the relative motions or not.

As we see from the above discussion, the integrals of center-of-mass motion (along with the absolute motions of the bodies) in the two-body problem with variable masses can be determined

only when the relative motions of the bodies have already been determined. This applies not only to the problem considered, but also, in general, to the Gylden–Mestschersky problem and Mestschersky's equations.

The conservative two-body problem considered is of interest, first of all, because it is integrable in quadratures for any law of mass transfer between the bodies, including intermittent, jumplike, and reverse mass transfer. Moreover, the equations of relative motion are completely independent of the mass variation law and closely match the thoroughly analyzed equations for the two-body problem with constant masses. This situation for the two-body problem with variable masses is extremely rare, because these problems are mostly not integrable, and one has to use approximate methods to study them. The absolute motions of the bodies in the Gylden–Mestschersky problem have been studied inadequately. Therefore, the result obtained is of theoretical importance.

Second, the conservative problem has clear astronomical applications. It can be used to study the motions of the stars in close binaries with mass transfer from one component to the other, including the possibility of a second mass transfer.

In addition, the problem considered can be used to construct a new model problem, the restricted three-body problem with variable masses, which will make it possible to refine the Roche lobes and the entire process of mass transfer in the case of variable masses. I will consider this problem in the future.

REFERENCES

1. B. E. Gelfgat, Byull. Inst. Teor. Astron. **7**, 354 (1959).
2. H. Gylden, Astron. Nachr. **109**, 1 (1884).
3. J. Hadjidemetriou, Adv. Astron. Astrophys. **5**, 131 (1967).
4. J. H. Jeans, Mon. Not. R. Astron. Soc. **85**, 912 (1924).
5. A. Kruszevski, Adv. Astron. Astrophys. **4**, 233 (1966).
6. L. G. Luk'yanov, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **24**, 62 (1983).
7. I. W. Mestschersky, Astron. Nachr. **132**, 129 (1893).
8. I. W. Mestschersky, Astron. Nachr. **159**, 229 (1902).
9. S. L. Piotrovskiĭ, Astron. Zh. **44**, 241 (1967) [Sov. Astron. **11**, 191 (1967)].
10. E. N. Polyakhova, Uch. Zap. LGU, No. 424, 104 (1989).

Translated by A. Dambis

Astronomy Letters, Vol. 31, No. 8, 2005, pp. 569–574. Translated from Pis'ma v Astronomicheskiĭ Zhurnal, Vol. 31, No. 8, 2005, pp. 634–640.

Automatic Guide for the Raduga Astronomical Spectrograph

S. A. Potanin*

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

Received February 10, 2005

Abstract—An automatic mirror guide has been designed and made for the Raduga fiber-optic echelle spectrograph. The new device was built into one of the parts of the spectrograph and allows the work of observers to be facilitated significantly. The automatic guide efficiently removes stellar image oscillations at frequencies of 0–2 Hz, which compensates almost completely for errors in setting the polar axis of a telescope and in its clockwork drive. The guide can be used on any telescope with a focal length of more than 5 m and has operated on two different telescopes. Over two observing seasons, several hundred stellar spectra were taken with the Raduga spectrograph using the automatic guide. © *2005 Pleiades Publishing, Inc.*

Key words: *automatic guide, wavefront, astronomical observations.*

INTRODUCTION

The efficiency of astronomical observations is known to depend not only on the quality of scientific instruments, but also on the quality of a telescope's mechanics and, to a greater extent, on the seeing conditions at the place where the studies are performed.

About ten years ago, the Raduga fiber-optic echelle spectrograph was created at the Sternberg Astronomical Institute (Shatskiĭ 1999). Observations with this spectrograph are performed on telescopes with a primary mirror diameter of ~1 m. Since the mechanics of these telescopes is imperfect, an exposure longer than 60 s with a 5″ aperture cannot be taken on any of them without manual correction (guiding), while exposures of 3600 s or longer are required for faint (10^m–12^m) objects. A device that allows the displacements of a star to be corrected manually using a tilting plane-parallel plate was built into the Raduga spectrograph from the outset. However, the development of an automatic guide that keeps a star on the aperture of the spectrograph was begun in 2002.

A star is displaced in the focal plane because the wavefront from the star is tilted with respect to the plane of the telescope's entrance pupil. Thus, the task of the automatic guide is to remove this tilt; therefore, these devices are called wavefront tilt compensators or correctors.

The wavefront tilt compensators can be divided by their frequency range into two types.

Low-frequency (0–10 Hz) devices are used to eliminate the shortcomings of a telescope's mounting and its clockwork drive, the construction flexures, the imperfection of the primary mirror load reliefs, the temperature variations, and the lowest-frequency component of the atmospheric wavefront tilts.

High-frequency (10–100 Hz) devices are designed to compensate for the fast variations in the wavefront tilts caused by atmospheric turbulence and are used in systems of adaptive optics.

Since the problem was solved for a specific spectrograph, the principles of its operation were dictated by the adapter, the part of the Raduga spectrograph attached to the telescope.

In this paper, we describe an optical compensator and a modulation detector operating at frequencies 0–5 Hz, since they are most suitable for the formulated task. The modulation detector takes away part of the useful light, but, as we show below, these losses are negligible compared to other unavoidable losses that arise from poor seeing. The various systems of automatic guides can be familiarized with in many publications (Argue *et al.* 1977; Bouchy and Connes 1999; Craven and Fegan 1977; Duchateau 1986; Fegan 1977; Jelley 1980; Kibrich and Robinson 1987; Ohtani *et al.* 1983).

OPTICAL LAYOUT OF THE COMPENSATOR

A peculiarity of the design of the Raduga spectrograph is the transmission of light from the telescope into the device through an optical fiber. Since the

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

1063-7737/05/3108-0569$26.00

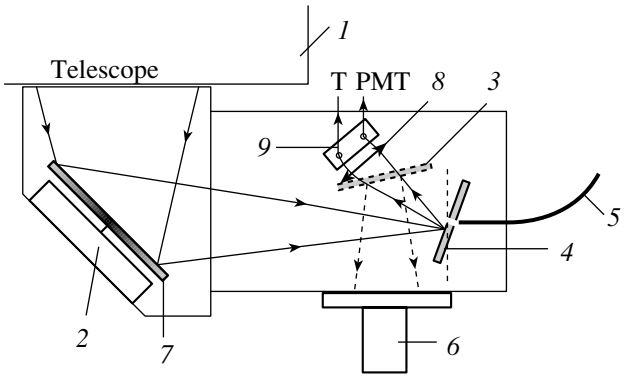


Fig. 1. Modified optical layout of the adapter (top view). The convergent beam from telescope *1* is reflected from active flat mirror *7* and falls on mirror aperture *4* (with optical fiber *5* lying behind it). Mirror *3* reflecting light into the eyepiece can be tilted downward by a special tie-rod and lets the light fall on collecting lens *8*. The light collected by the lens is reflected from mirror *9* upward and falls on a photomultiplier tube. Mirror *7* can independently turn around two axes through the bending of elastic rod *2*.

beam divergence at the exit end of the optical fiber is virtually independent of how the entrance end is illuminated, the spectrograph can be used on telescopes with various focal lengths, and the automatic guide enables observations with it on various telescopes.

A range of correction angles wide enough to keep a star on the aperture for a long time must be ensured for the device to operate efficiently at long exposures. The amplitudes of the systematic displacements of a star (in linear units) depend (in addition to the quality of the telescope's mounting) on the image scale, i.e., on the equivalent focal length of the telescope. The Crimean 1.25-m ZTE reflector has the largest focal length (21.25 m) among the telescopes on which observations with the spectrograph have to be performed. For the faintest accessible objects on this telescope, an exposure time of 1 h was set for Raduga. It was established empirically that the systematic displacement of a star is no more than 20 arcsec in a time of ~1 h. Accordingly, the automatic guide ensures a range of correction angles ±20 arcsec on a telescope with a focal length of 21.25 m.

Since an increase in the exposure time leads to the accumulation of dark electrons on a CCD, causing the signal-to-noise ratio to decrease, the wavefront tilt (star position) detector should not take away more than 10% of the useful light.

It was necessary to approach the development of a new device from the above standpoint; it is this standpoint that dictated the choice of the design and the principles of operation of the detection, analysis, and compensation systems.

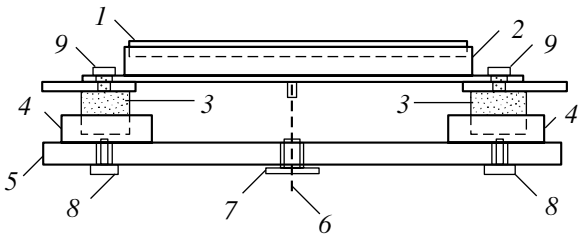


Fig. 2. Active mirror unit (side view): *1*, mirror; *2*, mirror frame; *3*, coil; *4*, magnetic core; *5*, base; *6*, elastic rod; *7*, collet grip; *8*, *9*, clamping screws. The second pair of coils (*3*) and magnetic cores (*4*) is not shown.

REALIZATION

In accordance with the formulated task, it was decided to build the automatic guide into the adapter of the Raduga spectrograph attachable to a telescope. Thus, the new device is not attached to a specific telescope. As was mentioned above, the task of the automatic guide in this case is to keep a star on the entrance aperture of the spectrograph; therefore, it would be appropriate to use a wavefront tilt detection method with modulation. Modulation must be produced by the system's active optical element. A mirror turning around two axes was used as the active element. To reduce the dimensions of the device and the losses of light in it, it was decided to make both the compensation and the modulation by the same active mirror.

A small hole (~0.5 mm in diameter) in the mirror reflecting light into the eyepiece is the entrance aperture of the spectrograph (see Fig. 1). For the recording system, the light reflected from the mirror edges of the aperture is collected and recorded with one photomultiplier tube (PMT).

Figure 1 shows a modified optical layout of the adapter (top view). As we see, the active element (*7*) turns the beam through an angle of 45°. The mirror turns through the bending of an elastic rod at the center of the mirror frame. Thus, the point around which the mirror rotates is not far from its surface.

When the active mirror turns, the image not only is displaced, but also is defocused. If the point around which the mirror rotates lies on its reflecting surface, then the error of the focus can be calculated using the formula

$$a = R\frac{1-\cos\varphi}{\cos\varphi} \approx \frac{R\varphi^2}{2}, \qquad (1)$$

where R is the distance from the mirror to the focal plane and φ is the mirror rotation angle relative to the equilibrium position.

When the mirror turns through 1°, the error of the focus at a distance of $\mathbf{R} = 240$ mm is about 0.04 mm. At a telescope's focal length of ~20 m and a seeing of

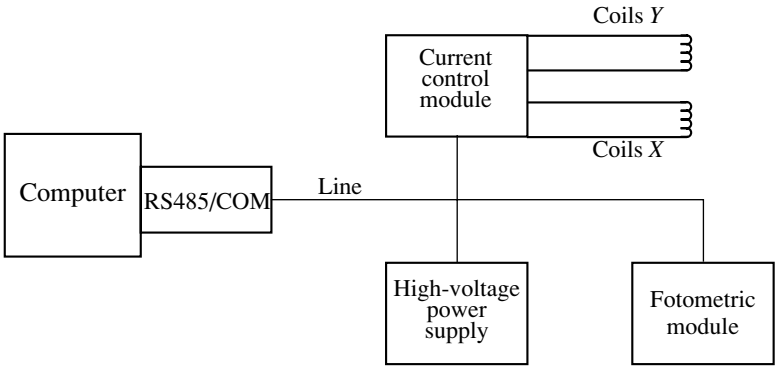


Fig. 3. Operation scheme of the control electronics.

no better than 1″, this error of the focus plays virtually no role.

The motions of the active element must be fairly fast and without any backlashes and hysteresis, if possible. Since the necessary rotation angles of the active mirror do not exceed 1°, electromagnets are used to produce these rotations. Figure 2 schematically shows the arrangement of magnetic coils (*3*) and cores (*4*). The coils are switched on in pairs; opposite currents arise in the pair of coils, which doubles the torque that turns the mirror. Figure 4 shows a photograph of the finished active element.

Standard automation modules of various devices in the Laboratory of New Photometric Methods at the Sternberg Astronomical Institute were used as the control electronics. The connection and interaction scheme for the electronic modules is shown in Fig. 3.

The digital line is connected through an RS485/COM converter to a COM-port of the computer. Three modules are connected to the line in parallel. These include the PMT high-voltage control module, the amplifier and counter controller (photometric module), and the compensator coil current control module. These devices have the following characteristics:

(1) The high-voltage module: sets a high voltage on the PMT in the range 0–1500 V.

(2) The photometric module: the amplifier and counter controller sets a discrimination threshold in the range 0–3 mV and a microexposure time from 10^{-3} s; since the counter has a dead time of 20 ns, the nonlinearity was disregarded. The PMT: a compact HAMAMATSU R1635-02 photomultiplier tube (the high voltage is in the range 0–1.5 kV). The PMT has good performance. The operating point was chosen by a classical method.

(3) The coil current control module: changes the current in the coils in discrete steps. It has two independent channels, X and Y.

Special-format commands are sent to each device using a control program. The detection system is organized as follows. The active mirror is used to modulate the stellar image according to the scheme shown in Fig. 5. The large circle in this figure is the aperture of the spectrograph; the shaded circles are the stellar images (for clarity, they were reduced; actually, the star occupies almost the entire aperture).

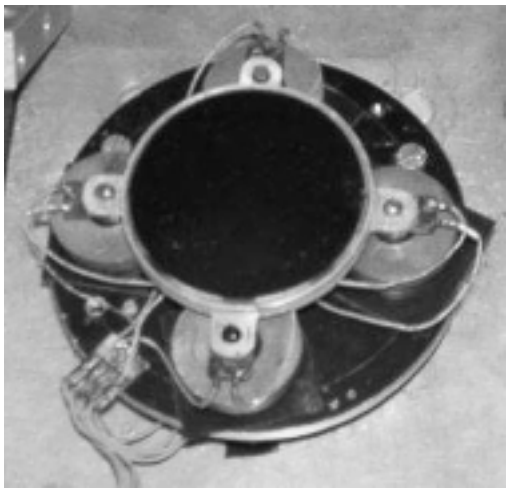


Fig. 4. Active element unit (top view).

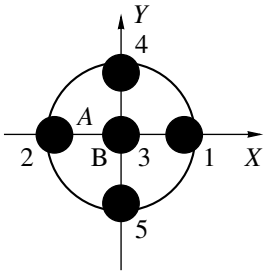


Fig. 5. Sequential positions occupied by a star relative to the aperture during modulation. A is the modulation amplitude. See also the text.

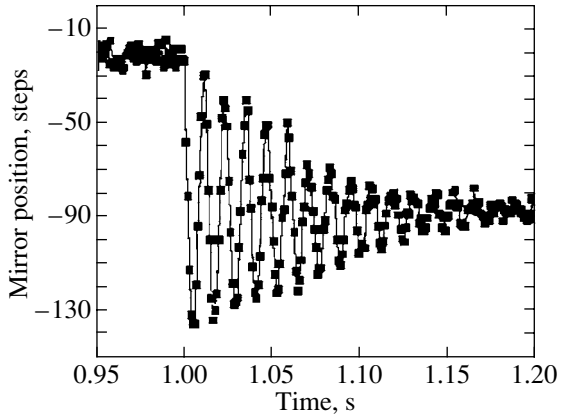


Fig. 6. Mirror position versus time for a displacement by 70 steps without any correction of the free oscillations.

The image moves along a cross-shaped trajectory, with the star sequentially occupying positions 1–2–3–4–5–6. At all these positions (except the sixth one), the flux is recorded on the PMT. The modulation amplitude (A) is chosen in such a way that the fluxes at the vertices of the cross do not exceed a given percentage of the free flux from the star. The free flux is the flux from the PMT measured when the star is far outside the aperture (at a distance of two diameters). These measurements are performed periodically after a given number of modulation cycles.

The modulation amplitude is a multiple of the minimum displacement, which can have 5 or 10 steps, depending on weather conditions. The value of A is chosen so as, on the one hand, to satisfy the requirement of not exceeding 10% of the loss of light and, on the other hand, to reliably detect the position of the star.

Thus, five fluxes are obtained after one such cycle: one at the center and four at the vertices of the cross. Since the light reflected from the mirror edges of the aperture is recorded, when the star is properly located relative to the aperture, the central flux must be lower than the exterior fluxes, and the latter must be equal, within the error limits. When the position of the object does not coincide with the center of the aperture, an imbalance between the fluxes at the vertices of the cross will appear. If the maximum of the four exterior fluxes exceeds the minimum flux by a given percentage (triggering threshold), then the wavefront is considered to have tilted.

Next, *a priori* information about the distribution of light in the stellar image can be used to determine the drift. However, this is not needed. Indeed, when the star is properly located relative to the aperture, the flux at the center of the modulation cross must be at a minimum; therefore, we can fit all five fluxes by a paraboloid of revolution and find the coordinates of its minimum (Janossy 1965). If this is done in the frame of reference with the origin at the center of the modulation cross, then the coordinates of the minimum must be the relative displacements ΔX and ΔY through which the mirror must be rotated to compensate for the wavefront tilt. These quantities are rounded off to values that are multiples of the minimum step. Subsequently, the control program turns the active mirror. This is how the compensation is done. A Linux-based personal computer is used to analyze the data in this system. A special program was written in C++ to control the electronics of the automatic guide; it was initially debugged in DOS and subsequently adapted to Linux.

The program establishes a link between all modules and performs the necessary computations. In addition, it continuously monitors such parameters as the loss of light, seeing, and the presence of cloudiness (by the drop in free flux), and the operating regime of the program is automatically adjusted to these parameters.

STUDIES OF THE ACTIVE ELEMENT

The active element of the optical compensation system is its most important part, since how correctly the errors in the wavefront tilt are compensated for depends on its properties. In addition, it should be known that the active mirror itself does not introduce any errors; therefore, some of its properties must be studied well. The mirror frame is fixed on an elastic rod, implying that it is an oscillating system in which free oscillations can be excited. Since this effect can be a serious hindrance during modulation, it must be removed, where possible, which requires studying it well. It should also be known how the optical compensation system displaces the images in the focal plane: whether it introduces any distortions.

For the subsequent discussion, we need to introduce a system of notations by giving the necessary explanations. The active mirror can turn independently around two perpendicular axes through small angles. It is turned by changing the current in the coils. The electronics allow the current to be changed in discrete steps by 255 steps in each direction from the equilibrium position; therefore, the mirror is turned in the same way. In what follows, it is convenient to use step units in place of angular units. Since there are two perpendicular axes, it would be natural to introduce two quantities, X and Y, that specify two mirror rotation angles (its position) in steps. Below, X and Y are called the mirror coordinates everywhere. Given the distance from the center of the mirror to the focal plane and the mirror-produced image displacements, the steps

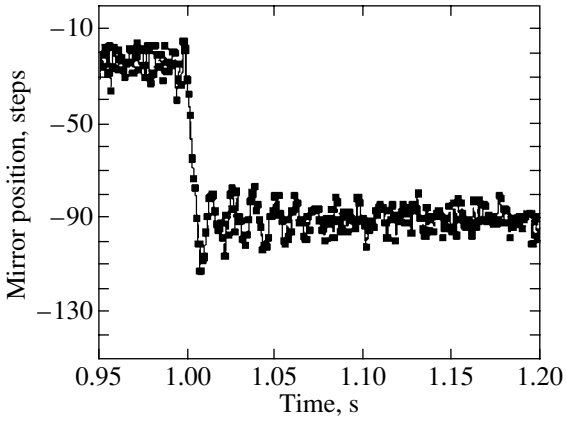


Fig. 7. Mirror position versus time for a displacement by 70 steps with the correction of the free oscillations.

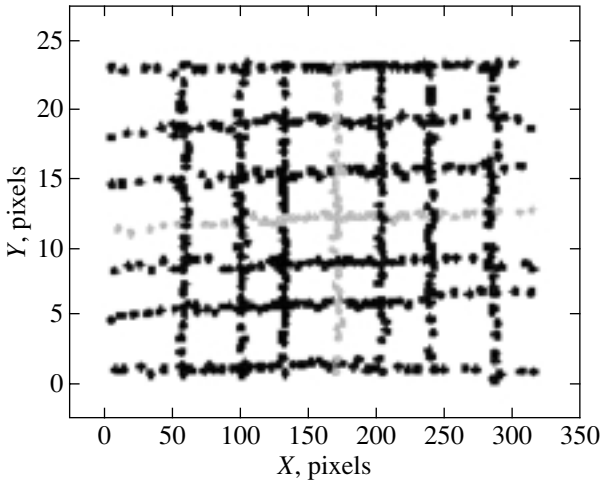


Fig. 8. Motion of the image of the illuminated aperture over the surface of the CCD array at constant X and Y.

can be transformed to angular or linear units for the displacements of a star in the focal plane.

Free oscillations are actually observed in the system under study. They were investigated using a special test bench on which the mirror rotation was transformed to a linear response of the luminous flux.

Thus, the mirror position was plotted against time (Fig. 6). Rapid damped oscillations are seen after a displacement.

These data were used to determine the parameters of the system. The decay time is $\tau = 0.07 \pm 0.01$ s; the oscillation period is $P = 12.0 \pm 0.2$ ms. To suppress the free oscillations, we used the following method that was used previously in designing the radial-velocity meter (Tokovinin 1987). When the current in the coils changes abruptly from one value to another, oscillations are excited because the mirror frame passes through the equilibrium position (at which it must be located for a new value of the current). If we set the current equal to the value necessary for an equilibrium at this position at the time the mirror frame, having passed through the equilibrium position, stops to begin its backward motion, then no oscillations will arise. This means that the current in the coils must change in two stages. This possibility was built into the control electronics. Even the first experiments showed that this method efficiently removes the excitation of free oscillations when the mirror turns (Fig. 7).

The CCD array from a Connectix WEB camera was used to study the displacements of a star in the focal plane of the telescope.

The measurements were performed as follows. The mirror frame was tilted by ten steps, and the image from the CCD array was then written to a hard disk, etc. Figure 8 shows the lines along which the light spot moves over the surface of the CCD array when the current in one pair of coils changes, while the current in the other pair is constant. To learn whether they will remain straight at nonzero values of the other coordinate, the field was covered with a rectangular grid. As we see, no serious distortions are observed.

The total amplitude of the star's motions is 4.5 mm at a distance of 240 mm from the mirror to the CCD. For a telescope with a focal length of 20 m, this corresponds to $\sim 46''$. Thus, the minimum angular step is found to be $46''/512 = 0''.09$. A step of 0.2 of the stellar profile half-width is more than enough for the operation of the automatic guide. Even for a $1''$ seeing, which is an extremely rare occurrence, motions by almost three steps are needed, while for a seeing of $3''$ or more, motions by ten steps will suffice.

OBSERVATIONS

The first observations with the automatic guide were performed in the summer of 2002 on the ZTE telescope at the Crimean Station of the Sternberg Astronomical Institute. It immediately became clear that the device successfully copes with the task of keeping a star on the aperture of the spectrograph. It became possible to take a continuous exposure as long as 1 h without any manual correction. More than 100 spectra of stars with magnitudes from 2 to 9 were obtained over one observing season with the compensator. In the fall of 2004, the spectra of objects as faint as $12^m.5$ were obtained on the same telescope with the automatic guide. Further studies showed that the device operates successfully to magnitudes as faint as 14^m.

In addition, the data on the correction signals written in a log file make it possible to study the error spectrum of the telescope's clockwork drive. The bulk of the spectral power (about 95%) lies within the

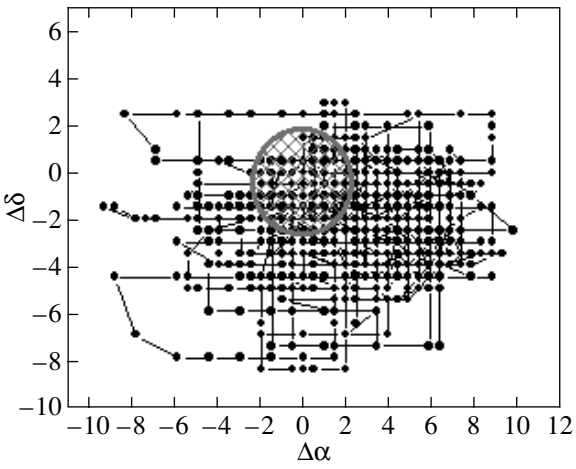


Fig. 9. Motion of a star in the telescope's focal plane.

frequency range 0–0.2 Hz along both axes. The frequency limit of the automatic guide was determined later using modulation on the test bench; it is about 2–5 Hz.

The motion of a star in the telescope's focal plane for one of the actual observations is illustrated by Fig. 9. The circle denotes the aperture of the spectrograph. As we see, the star would be outside the aperture much of the time (in the absence of correction), which would reduce appreciably the efficiency of spectroscopic observations. The maximum displacement amplitude of the star is almost 10 arcsec.

It proved to be also possible to determine the seeing in arcseconds and to compute the losses of light (both the total ones and those related to the operation of the automatic guide) using the log files of the correction signals.

Both the geometrical (the blocking of light by the edges of the aperture) and the forced losses depend on the seeing and the guiding quality. At a 3″ seeing (which was a fairly common occurrence), the losses would be about 30% for ideal guiding. Under such conditions, the compensator takes away another ~1.5% of the useful light. This value is negligible compared to the total losses. At poorer seeing (about 7″), the compensation losses were 2.5%, and the total losses were more than 70%! At the best seeing (0.″9), the total losses were about 8%, and the guiding losses were 1.3%. Thus, the compensator always takes away approximately the same fraction of light (about 2%). The main losses in guiding are caused by poor seeing.

CONCLUSIONS

The designed and made device completely satisfies the formulated task, as confirmed by numerous observations. It meets the reliability and simplicity requirements and is suitable for observations on any telescopes with a focal length larger than 5 m. At the same time, the automatic guide has functional limitations related to its production for a specific spectrograph. Devices of such a kind can be successfully used in other types of spectrographs and in certain types of photometers.

ACKNOWLEDGMENTS

I wish to thank V. Kornilov and A. Tokovinin for valuable discussions, N. Shatskii for help with observations, and N. Slabkaya for help in making optical components.

REFERENCES

1. A. N. Argue, P. G. Craven, J. H. Fruin, *et al.*, Astron. Astrophys. **58**, 27 (1977).
2. F. Bouchy and P. Connes, Astron. Astrophys., Suppl. Ser. **136**, 193 (1999).
3. P. G. Craven and D. J. Fegan, J. Phys. E **10**, 516 (1977).
4. M. Duchateau, ESO Messenger (ISSN 0722-6691), p. 27 (1986).
5. D. F. Fegan, Irish Astron. J. **13** (3/4), 128 (1977).
6. L. Janossy, *Theory and Practice of the Evaluation of Measurements* (Clarendon Press, Oxford, 1965; Mir, Moscow, 1968).
7. J. V. Jelley, Q. J. R. Astron. Soc. **21**, 14 (1980).
8. R. Kibrich and L. Robinson, Astron. Soc. Pac. **99**, 1014 (1987).
9. H. Ohtani, T. Ichikawa, T. Sasaki, *et al.*, Kyoto Univ. Mem. Fac. Sci., Kyoto Univ., Ser. Phys., Astrophys., Geophys., Chem. **36**, 291 (1983).
10. N. I. Shatskiĭ, Candidate's Dissertation in Mathematical Physics (Mosk. Gos. Univ., Moscow, 1999).
11. A. A. Tokovinin, Astron. Zh. **64**, 196 (1987) [Sov. Astron. **31**, 98 (1987)].

Translated by V. Astakhov