SOUKUP, JUDITH ELLEN

THE HIGH LATITUDE EXTENSIONS OF GALACTIC SPIRAL ARMS

City University of New York PH.D. 1979

THE HIGH LATITUDE EXTENSIONS OF GALACTIC SPIRAL ARMS

by

JUDITH E. SOUKUP

A dissertation submitted to
the Graduate Faculty in Physics
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy,
The City University of New York.

1979

This manuscript has been read and accepted for the Graduate Faculty in Physics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9-14-1979
date

C. Yuan
Chairman of Examining Committee

9-17-79
date

Frank Moscati (Jr.)
Executive Officer

Leon Rubin

W. B. Burton

Richard Mathews

Supervisory Committee

The City University of New York

ACKNOWLEDGEMENTS

Throughout my studies, I have received constant intellectual, moral and material support from countless members of the academic as well as administrative and technical staffs of all institutions I attended. To each of them go my heartfelt thanks.

The following provided me with crucial assistance and support in carrying out the research for my thesis:

Professor Chi Yuan gave me penetrating advice, unwavering encouragement, and invaluable insights into the intricacies of fluid flow.

Dr. Robert Jastrow, Director of the Goddard Institute for Space Studies, granted me the use of the Institute's computer and other facilities.

The National Radio Astronomy Observatory in Charlottesville, Virginia, admitted me to an apprenticeship in observational astronomy as a member of its Summer Student Program, and granted me additional observing time and the use of their facilities beyond the expiration of the summer program. The NRAO staff members gave me invaluable assistance and cooperation in learning and applying the techniques of reduction of observational data.

Professor W. Butler Burton, then at NRAO, became my true mentor in observational astronomy, giving unstintingly of his time, good advice, and friendship.

The members of my Supervisory Committee followed the writing of my thesis with interest and understanding.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (continued)

CHAPTER 1: DEFINING THE PROBLEM

1.1 Historical Background: Density Wave Theory of Spiral Galaxies

The basic problem encountered by any theory describing spiral galaxies is the so-called "winding dilemma." If the spiral arms are always composed of the same material, the differential rotation of the galaxy would soon change their shape, winding them up (or unwinding them) within only a few galactic rotation periods. Since the type of spiral structure (i.e., openness or tightness of the arms) is strongly correlated with other physical properties of the galaxy which cannot change rapidly in time, such as relative gas content or frequency of young objects (Roberts, Roberts and Shu 1975), the spiral pattern must be at least quasi-permanent.

The density wave theory, first proposed by Lin and Shu (1964), resolves the "winding dilemma" by postulating that the spiral arms are a density wave pattern. This density wave is assumed neutral (not growing or decaying in time), is shaped like an m-armed spiral, and rotates about the center of the galaxy with a constant angular velocity or pattern speed Ω_p.

The stars and interstellar gas move according to the galactic rotation curve, at velocities varying with distance from

the galactic center, and pass through the density wave. Thus the spiral arms do not continually consist of the same material; at different times, different gas and stars represent the local maxima of the density wave. The constancy of the geometrical shape of the arms is assured since the wave rotates as a rigid body.

Since the density wave is created and maintained by purely gravitational forces, a completely self-consistent treatment would include the self-gravity of both the gas and the stars in producing the spiral perturbation potential. However, since the stars contribute most of the mass of the galaxy, the small changes in the gravitational potential due to the resultant redistribution of the gas are usually ignored, and the spiral field is assumed to be produced solely by the stars. In this thesis, we are concerned with the gaseous response to an imposed background spiral gravitational field supplied by the stars.

The stellar distribution is basically rotationally symmetric except for the small non-axisymmetric component in the spiral arms. (There is only about 5% more mass in the arms than between them.) The total gravitational potential is therefore the axisymmetric potential plus a non-axisymmetric perturbation. Because of the small velocity dispersion of the gas compared to the stellar dispersion velocities, the gaseous response to this weak non-axisymmetric stellar potential is highly nonlinear, resulting in large-scale "galactic shock waves."

A good deal of work has been done in calculating these

galactic shocks in the gas flow in a two-dimensional (flat) galaxy. This two-dimensional approximation is justified as a starting point by the observation that the radius of our Galaxy is about 50 times its thickness, and other galaxies appear to be at least as flattened as ours. Both Shu (1968) and Vandervoort (1970a, 1970b, 1971) have considered the effects of the finite thickness of the Galaxy, but they assumed that the gas behaves as a uniform slab, with no vertical motions.

With these approximations, plus the assumption of a tightly wound spiral pattern, a steady-state, two-armed spiral shock solution to the gas flow in the plane was demonstrated by Roberts (1969, 1972) and Shu, Milione and Roberts (1973) for the case of an isothermal gas, and extended to the two-phase model of the gas by Shu, et al. (1972). Woodward (1973, 1975) calculated the time-dependent shock solution.

The resulting gas flow in the plane of the galaxy is periodic, along narrow, nearly concentric streamlines, passing through large-scale standing shock waves formed near the potential minima of the spiral pattern (see figures 1 and 2).

With the existence of the shock waves in the gas, the density wave theory is able to explain many observed phenomena and even predict others. The galactic shock as a triggering mechanism for star formation can explain why the minor mass perturbations of the arms should be so optically prominent, traced out by young, bright stars, HII regions and narrow dust lanes. The relative lo-

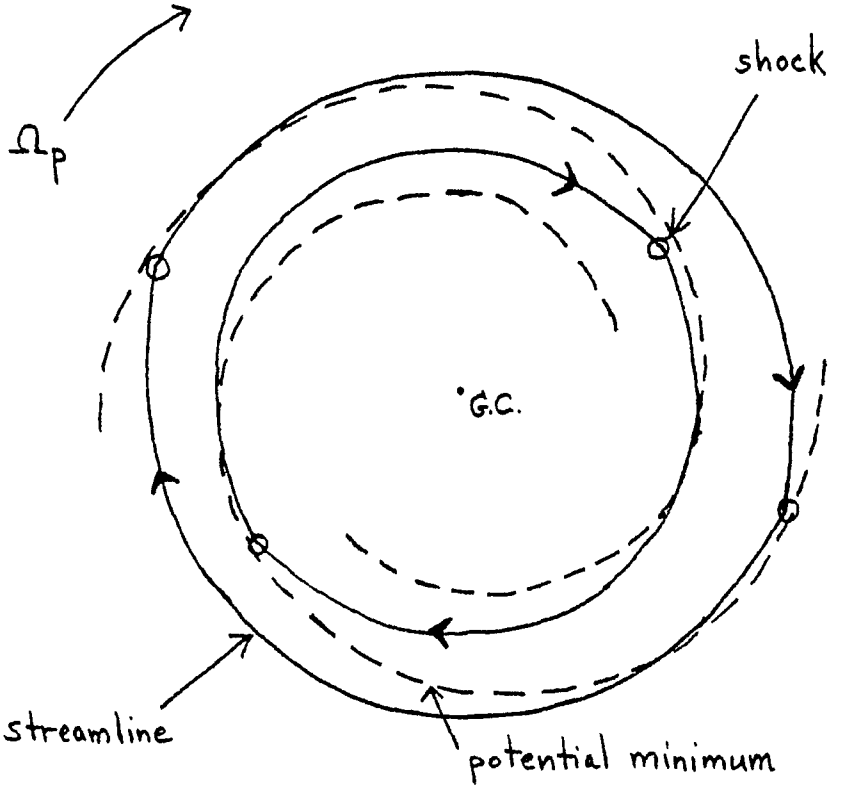


FIGURE 1: Spiral potential minima (dashed lines), streamlines (solid lines), and shock locations (o). From Roberts' (1969) calculations for a tightly wound spiral galaxy.

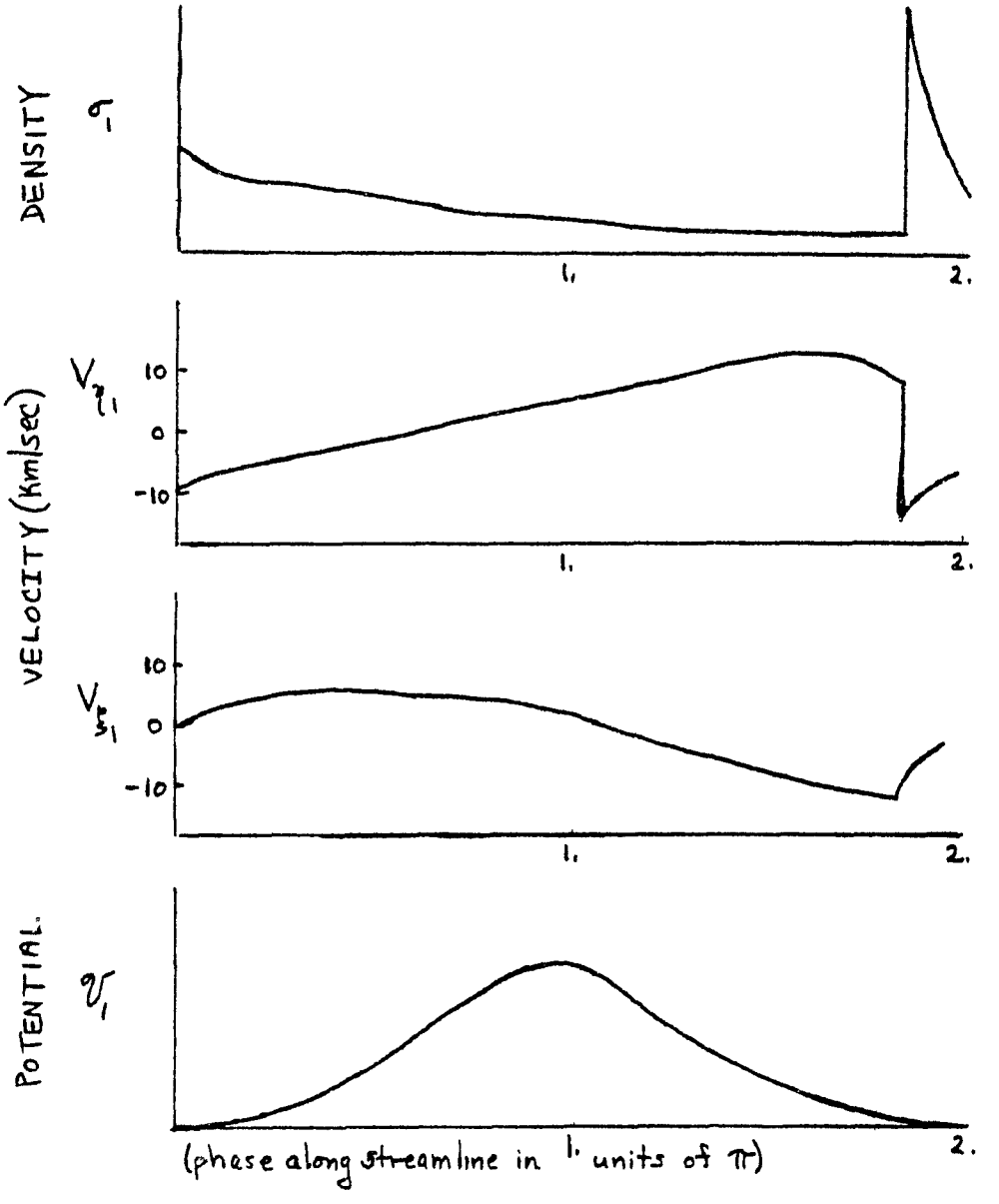


FIGURE 2: Typical density, velocity, and spiral potential variations along the streamlines of figure 1.

cations of these tracers, as well as the enhanced synchrotron radiation along the spiral arms (Matthewson, et al. 1972), can also be explained. An excellent review of the density wave theory can be found in Wielen (1974).

While the density wave theory for a flat, two-dimensional galaxy explains many important observed features, more knowledge is needed of the structure of the galaxy in the z-dimension. Some work has already been done on the vertical component of the gas. Tosa (1973) studied three-dimensional steady-state galactic shock waves, taking into account the finite vertical thickness of the gas layer. However, he assumed local hydrostatic equilibrium in the z-dimension, and complete independence of all variables of z. Thus his calculation tells us little about the vertical gas structure.

A linear analysis of the three-dimensional steady-state gaseous flow was carried out by Yuan and Wallace (1973); the resulting vertical flow of the gas is small (less than 2 km/sec), and the streamlines of the gas are virtually parallel to the plane of the Galaxy.

1.2 The Theoretical Problem

This thesis has extended the above work to the nonlinear regime, combining shock formation in time and two space dimensions (along the streamline in the plane, and the vertical [z] direction). Although the stellar dynamics can be calculated quite accurately

with a linear model (Lin, Yuan and Shu 1969), the gas response to even a weak perturbation is expected to be nonlinear: the response is roughly proportional to c^{-2}, where c is the velocity dispersion in the case of the stars or the sonic velocity in the case of the gas. Since $c \approx 40$ km/sec for the stars, but for the gas $c \approx 10$ km/sec, we expect a highly nonlinear response in the gas that may be quite different from the linear motions.

There are a number of general questions about the vertical motion and distribution of the gas, such as:

(1) What are the streamlines of the gas? What are the magnitudes of its vertical motion? This information is needed in three-dimensional models of the Galaxy used in constructing theoretical profiles and contour maps. Also, the answer to this question might throw some light on the problem of the high velocity clouds.

(2) How does the density of the gas vary with distance from the plane?

(3) How does the theoretical thickness of the gas layer change along the streamline?

This thesis is interested specifically in the vertical extent of the galactic shock associated with the spiral arms and in any shock-induced vertical motions of the gas, in order to see the effect of the shock on the vertical thickness of the gas layer -- i.e., the height of the spiral arm including the high latitude ("high-z") extensions.

1.3 The Observational Problem

The "high-z" extensions are large-scale concentrations of neutral hydrogen gas extending perpendicular to the galactic plane (z-direction), with velocities corresponding to those of the gas in the spiral arms in the galactic plane. They were first observed by Kepner (1970) in the Perseus and the Outer arm. Her data seemed to indicate that these concentrations are relatively smooth continuations of the spiral arms into intermediate latitudes (corresponding to heights of up to 2 kpc above the plane).

Such high latitude extensions of the gas also seem to exist in the Sagittarius arm and the Scutum arm, as the observations of Henderson (1972) have suggested. Some optical observations (Kilkenny 1973) also indicate that the spiral arms may extend to 2 kpc from the plane.

Since the acceptance of this phenomenon has been based primarily on Kepner's observations, independent confirmation of their existence and properties by more sensitive and more extensive observations was needed.

Such observations, made by W. B. Burton and myself with the NRAO 140-ft. and 300-ft. radio telescopes, confirmed their existence, and investigated their extent and continuity in space and their velocity range.

1.4 Structure of the Thesis

Chapter 2 describes the telescope procedures and presents the observations of the hydrogen gas in our Galaxy.

Chapter 3 develops the theoretical model used to calculate the motions and spatial distribution of the gas in response to a given spiral gravitational potential. It discusses how the gas is modeled, and the simplifications that reduce the problem to two spatial dimensions.

Chapter 4 is a relatively technical description of the numerical method used to solve the mathematical problem. It also discusses many of the subtle details involved in using the calculation scheme.

Chapters 5 and 6 break up the full two-dimensional problem into two separate one-dimensional problems, in order to get a better understanding of the numerical method and its numerous pitfalls.

Chapter 5 treats the problem with the x-dimension suppressed: essentially a static atmosphere. Section 5.4 discusses a technical stability problem related to the numerical method.

Chapter 6 suppresses the z-dimension and looks at one-dimensional flow into a spiral potential. This is analogous to the flow through a converging-diverging nozzle, and yields an excellent understanding of the entire problem -- both for details of the numerical scheme and for a physical and intuitive understanding of transonic flow and the formation of a shock.

Chapter 7 treats the full two-dimensional problem. It presents the results obtained using the two basic models of the gas (Model A for an isothermal gas, and Model B for a gas whose temperature varies with $|z|$), and interprets them with regard to the observations.

CHAPTER 2: OBSERVATIONS OF THE HIGH-Z EXTENSIONS

2.1 Introduction

The only observations specifically investigating the high-z extensions were made by Kepner (1970); the extensions can also be seen in the galactic survey by Weaver and Williams (1974) and in Henderson (1972). A new, more extensive investigation of the phenomenon was needed.

Observations of the 21-cm (1420 MHz) emission from the spin-flip transition of neutral hydrogen in our Galaxy were taken by W. B. Burton and myself during July - November, 1975 on the 140-ft. and 300-ft. radio telescopes at the National Radio Astronomy Observatory in Green Bank, West Virginia.

The data obtained is substantially better in angular resolution and velocity resolution. It also includes observations below the galactic plane. In addition, it has a much higher signal-to-noise ratio than that of most galactic structure surveys.

2.2 Apparatus

The 140-ft. telescope is a completely steerable dish with an angular resolution at 21-cm, measured at the half-power level

(i.e., half-power beam width, HPBW), of 20' of arc. The 300-ft. telescope is a meridian transit antenna with a HPBW at 21-cm of 10' of arc. Since its steerability is restricted to the declination coordinate, it cannot track a specific sky position for very long, but it can track galactic longitude or latitude as the sky drifts through the beam, by changing the declination.

The system receiving the incoming signal, the autocorrelator, is the equivalent of a multi-channel spectrum analyzer which measures the power spectrum over a selected bandwidth (frequency coverage) and center frequency (Shalloway, Mauzy and Greenhalgh 1972). It produces a one bit time correlation function of the signal, and then performs an inverse Fourier transform to produce the power spectrum. The autocorrelator has a total of 384 channels which can be divided up into as many as four separate sections, called receivers. The receivers can be operated either in series, as one 384-channel receiver, or in parallel, as four independent 96-channel receivers whose signals can then be combined, or in some other combination. Each receiver can have any bandwidth, independent of the other three receivers.

For accurate spectral measurements, a source spectrum is compared to a reference spectrum, which removes or minimizes instrumental effects. The observed signal is collected for a time τ, called the integration time; the proportion of signal to reference time is specified by the observer. The reference spectrum can be obtained by position switching or frequency switching. In position

switching the reference spectrum is obtained by periodically moving the telescope off the source position. Frequency switching generates a comparison spectrum by shifting the frequency of the receivers to an adjacent bandwidth while the telescope remains stationary.

Position switching usually produces a more stable signal across the filterbank. However, frequency switching must be employed when observing neutral hydrogen in the Galaxy, since the 21-cm emission is so ubiquitous that finding emission-free comparison regions is impossible. In either switching mode, the reference signal is subtracted from the source signal, and then divided by the reference to correct for gain variation across the bandwidth. This output of the autocorrelator is proportional to the intensity of the source.

Radio astronomers normally use units of temperature to measure intensity of radiation. In the range of radio wavelengths, where $h\nu \ll kT$, the Rayleigh - Jeans radiation law holds:

$$I = \frac{2kT_B\nu^2}{c^2} \quad , \tag{2.1}$$

where T_B is the brightness temperature corresponding to the intensity measured. However, since the telescope is never free of losses, the actual antenna temperature T_A is less than T_B by some efficiency factor ε. (It is this factor ε that makes it difficult to compare data from different telescopes.)

In addition, radio astronomers convert frequencies to ve-

locities with respect to the local standard of rest (using the non-relativistic Doppler formula). The velocity resolution corresponding to 1420 MHz is given by the relation

$$1 \text{ km/sec} = 4.74 \text{ kHz} \quad . \tag{2.2}$$

This is a much more physically meaningful way to look at the data, since it tells us how accurately we can resolve components of gas with different velocities.

A typical example of the autocorrelator output is shown in figure 3a. This is called a line profile: antenna temperature T_A as a function of velocity, for a position given by galactic longitude l and latitude b. The regions on the sides of the signal show the receiver fluctuations due to receiver noise, scattered radiation, and the nonlinear response of the "filters" to input. To eliminate them, a baseline (shown in figure 3a as the dotted line) is fitted to the side regions of the profile and subtracted from the spectrum. Figure 3b shows the reduced line profile.

An estimate of the fraction of the observed signal that could be due to sidelobe radiation was obtained by doing moon occultations with the 300-ft. telescope, thus blocking the 10' of arc subtended by the main beam with the 30' of arc subtended by the moon. The moon occultations were done on five successive days. The sky position of each occultation, or "on" scan, was observed on another day as an "off" scan, to see what the total signal from that position is.

FIGURE 3a

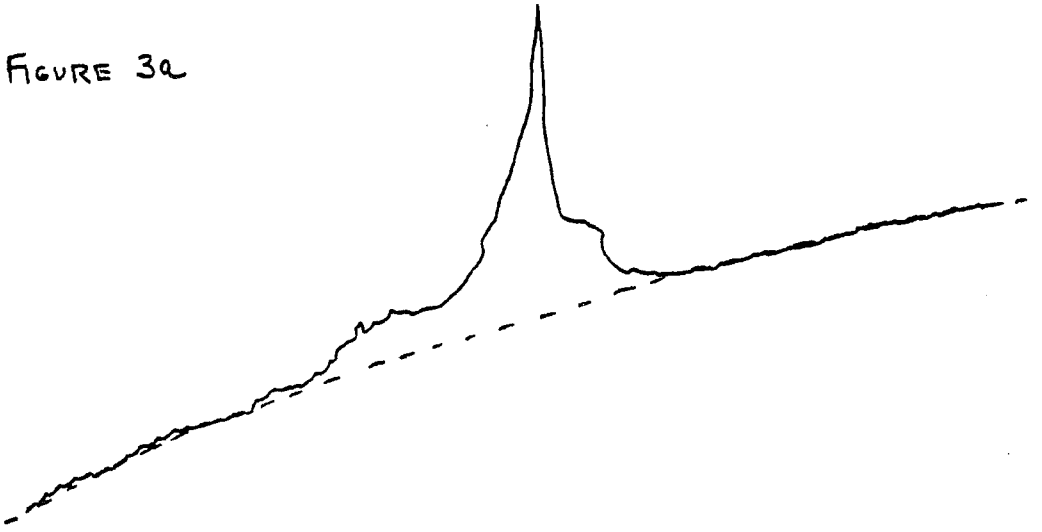


FIGURE 3b

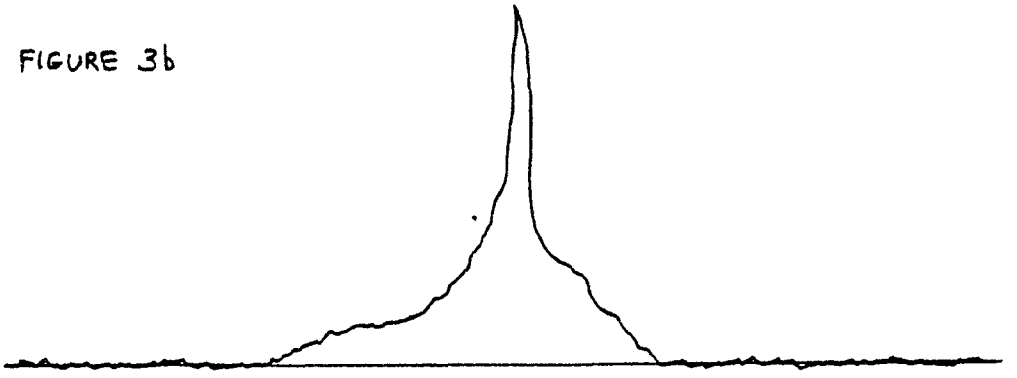


FIGURE 3a: Unreduced spectrum or line profile. The dotted line is the baseline to be removed.

FIGURE 3b: Reduced line profile.

The profiles from one pair of "on" and "off" scans are shown in figures 4a and 4b. The "off" scan has a maximum antenna temperature of 14.0°K, and the "on" scan has a maximum of 4.2°K. Thus, 30% of the signal was received from the far sidelobes. The data for the five occultations is listed in table 1.

As table 1 shows, the radiation in the sidelobes varied from 18% to 37%, depending on the sky position observed. Therefore, only antenna temperatures of 0.6°K and larger were used in describing the features, regardless of the high sensitivity of the observations in terms of receiver fluctuation (average rms noise of 0.09°K for the 140-ft. telescope and from 0.10°K to 0.25°K for the 300-ft. telescope).

2.3 Telescope Procedures

For most observations presented here, the autocorrelator was used in the mode of one receiver of 384 channels with a total frequency bandwidth of 2.5 MHz; the corresponding total velocity coverage at 21-cm is 527 km/sec. The velocity and frequency resolution per channel was 1.37 km/sec and 6.49 kHz.

Four galactic longitudes, 30°, 60°, 80°, and 100°, were scanned in the mode of two independent receivers of 192 channels each, covering the same frequency range; the spectra were averaged for better signal-to-noise ratio. The velocity and frequency resolution per channel was 2.74 km/sec and 12.98 kHz.

All spectra were frequency switched by 2.5 MHz; as a

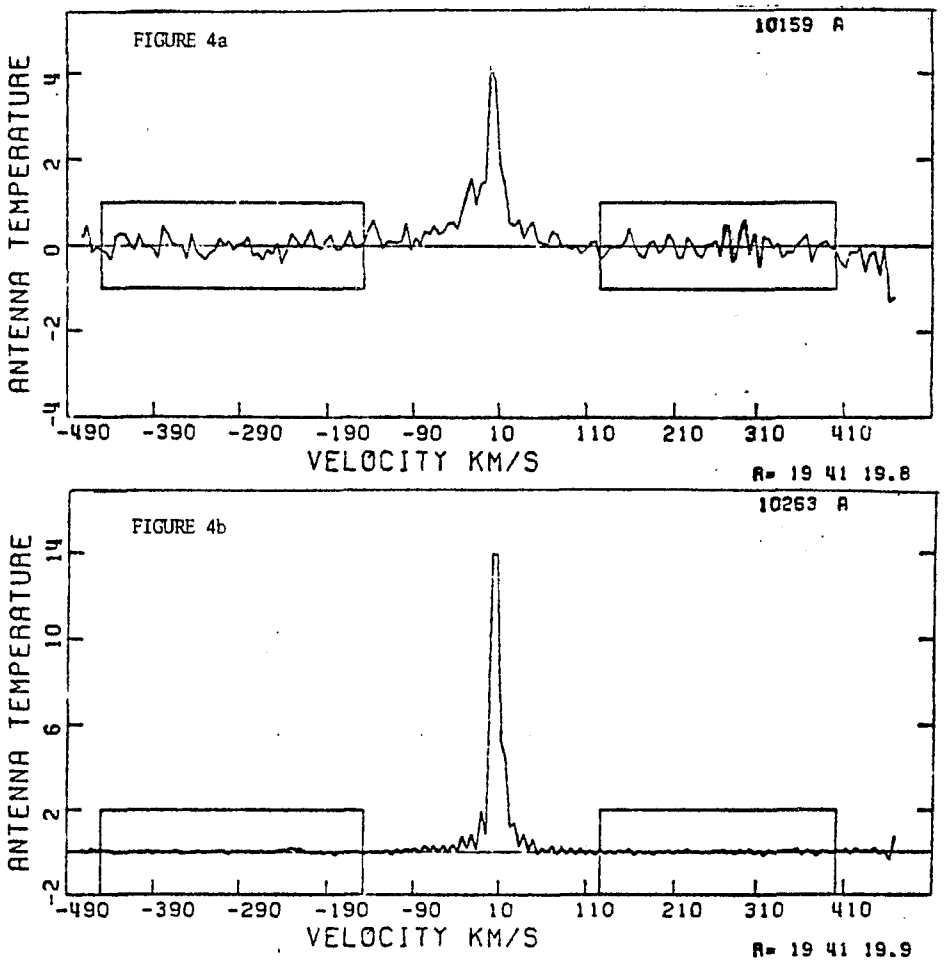


FIGURE 4a: "On" profile for the moon occultation.
Source: Table 1, Position 1.

FIGURE 4b: "Off" profile for the moon occultation.
Source: Table 1, Position 1.

TABLE 1

RESULTS OF FIVE OCCULTATIONS OF
THE 300-FT. TELESCOPE BY THE MOON

Position	T_A(°K) "On"	T_A(°K) "Off"	% Difference
1	4.2	14.0	30
2	2.2	5.9	37
3	3.3	10.3	32
4	2.5	13.2	19
5	1.9	10.5	18

check against possible receiver bias or emission in the switched signal, the middle tuning frequency was switched above and below the signal's center frequency on alternate days. No significant differences were seen in the profiles.

The bandwidth was chosen by inspecting galactic survey maps of the regions to be observed, and choosing a velocity coverage large enough to include adequate regions on both sides of the emission for fitting baselines. Since the center of the velocity range of the emission varies with position, the center velocity of the receiver was changed correspondingly.

For the longitudes observed with the 140-ft. telescope, each sky position was tracked for a time τ, usually 10 minutes.

In tracking longitude and latitude with the 300-ft. telescope, the integration time was 10 seconds. Subsequently, different numbers of these 10-second scans (depending on the specific sky position) were averaged to get a profile of an angular size of 10' or 20' of arc. For some positions that drift through the beam very rapidly, observations were taken on more than one day and the spectra were averaged to get a better signal-to-noise ratio.

All profiles were reduced individually, without reference to profiles of neighboring points. If some sudden irregularity on the scale of 1 - 2 profiles showed up in the preliminary contour maps, the reduction of those profiles was re-checked using the neighboring profiles as a guide. This procedure assumed that the large-scale galactic structure is continuous and smooth on the

scale of 10' to 30' of arc. Only about 3% of the profiles were reduced a second time.

2.4 Observations

Two longitudes, 140° and 240°, were observed on the 140-ft. telescope. The observations at $l = 140°$ are spaced 40' of arc apart, and those at $l = 240°$ are spaced 20' of arc apart. The latitude ranges covered are:

$$l = 140° \qquad -17.3° \leq b \leq 27.3°$$

$$l = 240° \qquad -13.0° \leq b \leq 24.0°$$

On the 300-ft. telescope, observations were made at four longitudes, 30°, 60°, 80° and 100°. Furthermore, four regions were observed at 1° intervals of longitude: $l = 118° - 120°$; $l = 147° - 149°$; $l = 152° - 157°$; and $l = 225° - 231°$. The latitude ranges for each longitude were approximately -20° to +20°, the specific values depending on the longitude.

In addition, constant latitude observations were taken on the 300-ft. telescope at 8°, 10°, 15°, 16° and -7°, covering most of the longitudes observed.

Sample profiles for several sky positions are shown in figure 5. The ordinate is the antenna temperature in °K; the abscissa is the line-of-sight velocity with respect to the local standard of rest, in km/sec.

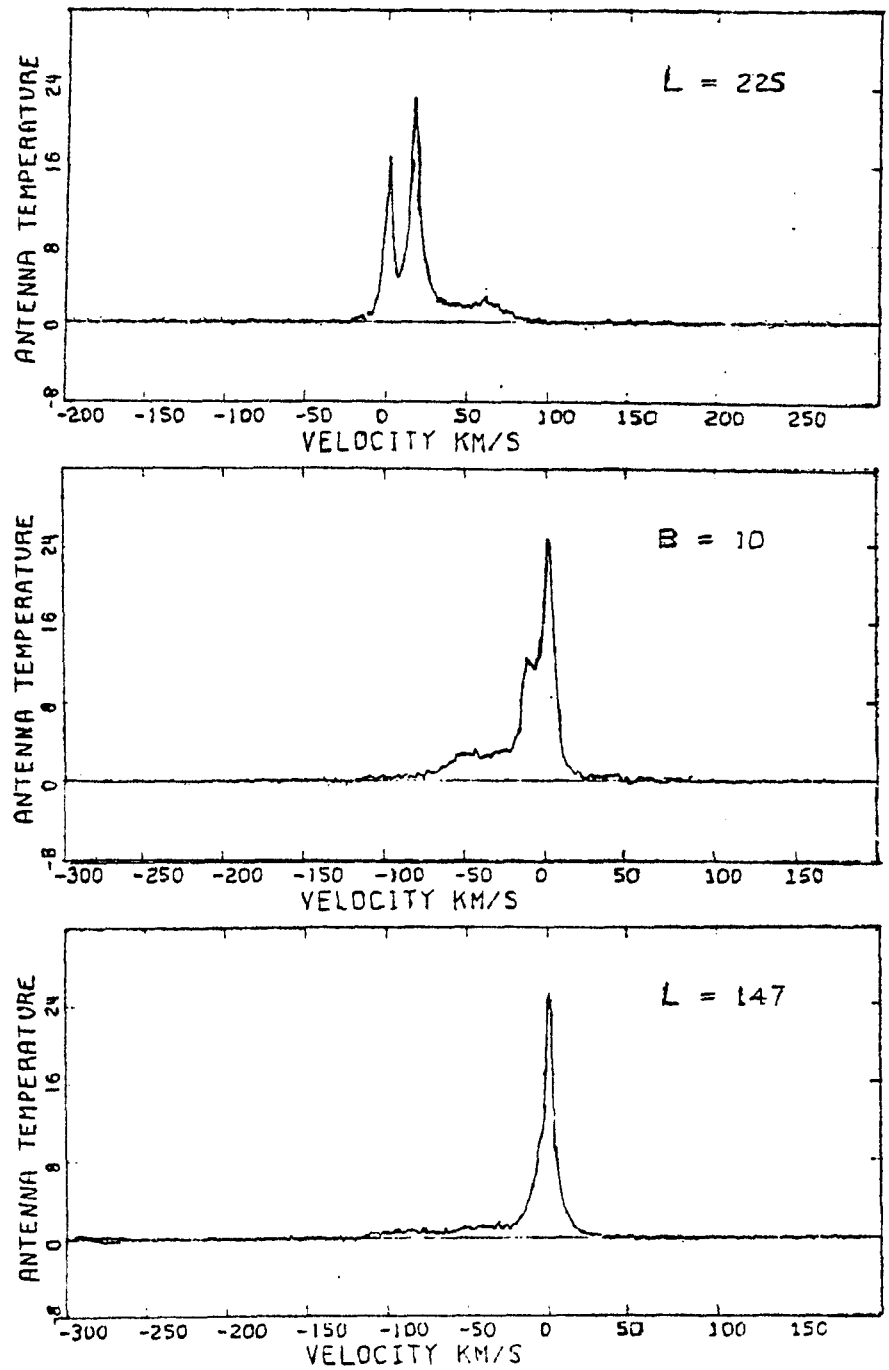


FIGURE 5: Sample profiles.

All observations are summarized in contour maps prepared from the profiles, figures 6 - 34. The maps in figures 6 - 30 present contours of equal antenna temperature in velocity-latitude coordinates. Figures 31 - 34 are at constant latitude and show antenna temperature contours in velocity-longitude coordinates.

The profiles used to produce the contour maps were smoothed; i.e., the signal in each frequency channel was averaged with the adjacent channels with a 1:2:1 weighting. This produced maps in which large-scale galactic structures are easier to recognize. However, some information is lost by this smoothing, as well as by the finite intervals between contour levels. Therefore, all quantitative information was obtained from the profiles themselves.

The contour maps of spectra from the 300-ft. telescope have an angular resolution of 20' of arc, to make them comparable with other investigations as well as to smooth out the irregularities and lower the rms noise. But all quantitative information based on the 20' of arc profiles was checked by randomly remeasuring some results from the 10' of arc spectra. No significant differences were found.

2.5 Discussion

(a) The existence of the high-z extensions

Identifying spiral structure from 21-cm observations (which contain no direct distance information) is largely a matter

FIGURES 6 - 30: Contour maps showing antenna temperature as a function of velocity and latitude at fixed longitudes. The lowest level contours ($< 1°T_A$) vary from map to map, but all have contour drawn at antenna temperatures of 1, 2, 4, 7, 12, 20, 30, 40, ... °K. Hatched contours enclose regions of relatively low antenna temperature.

FIGURES 31 - 34: Contour maps showing antenna temperature as a function of velocity and latitude at fixed latitudes. All details are the same as in figures 6 - 30.

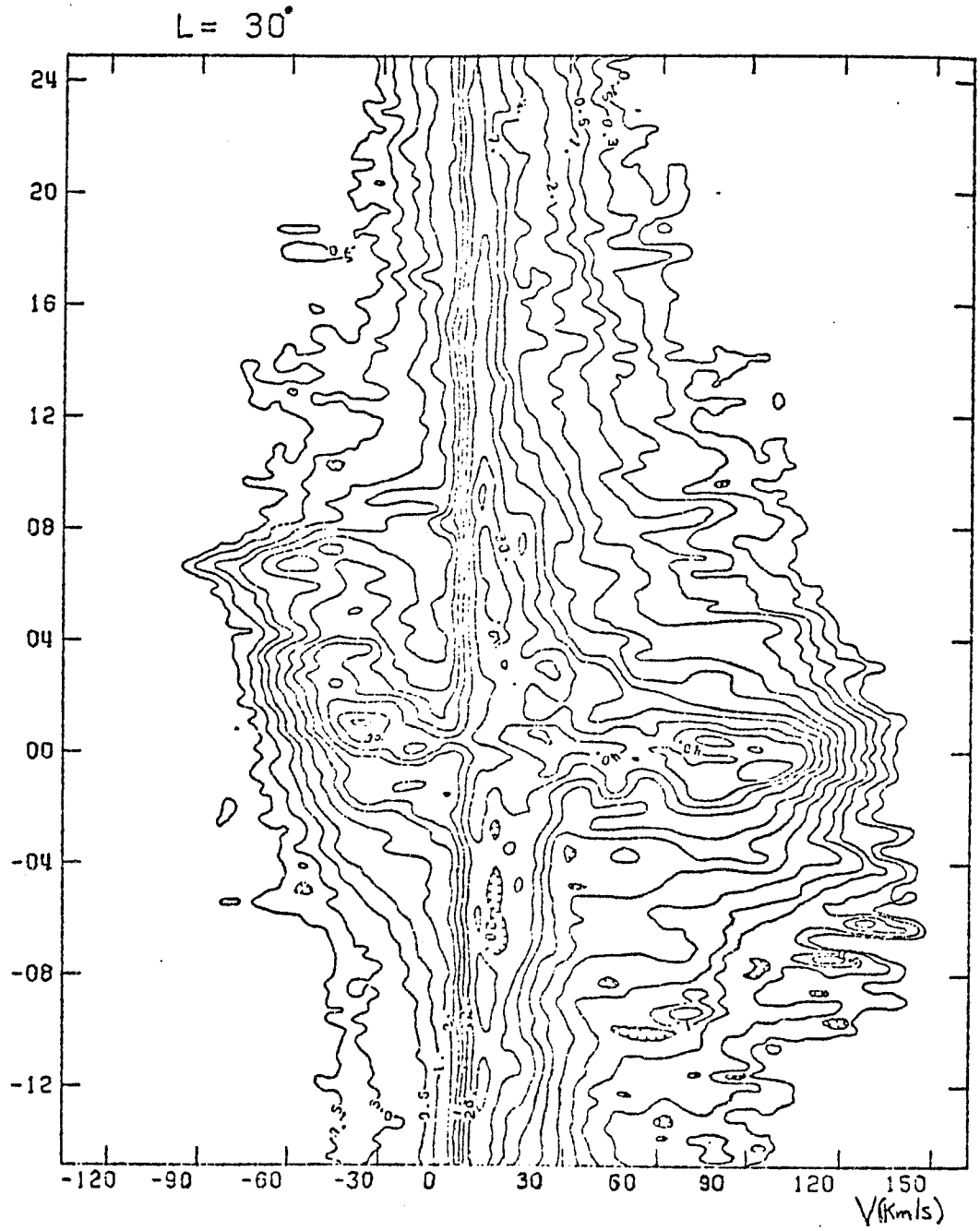


FIGURE 6

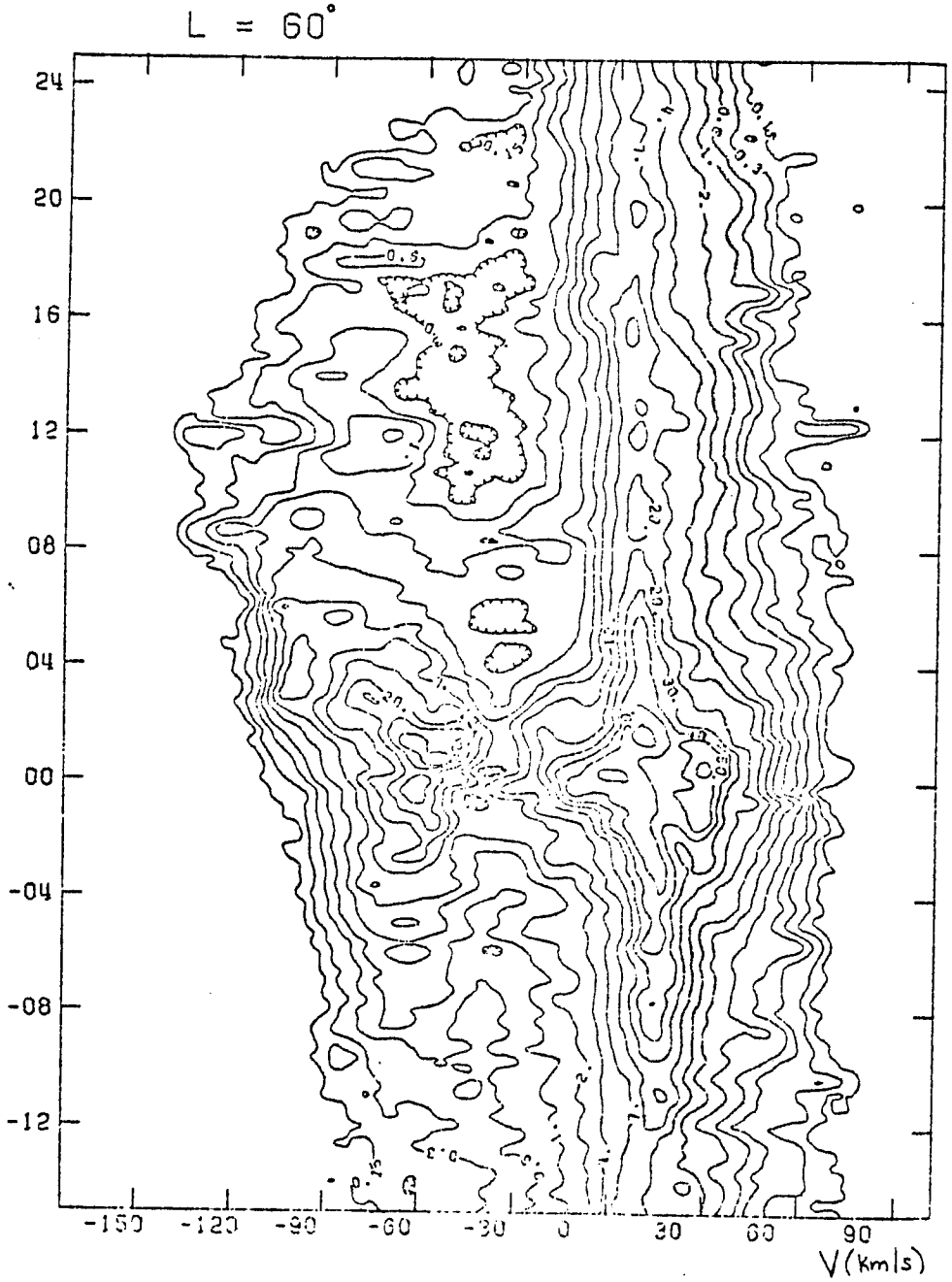


FIGURE 7

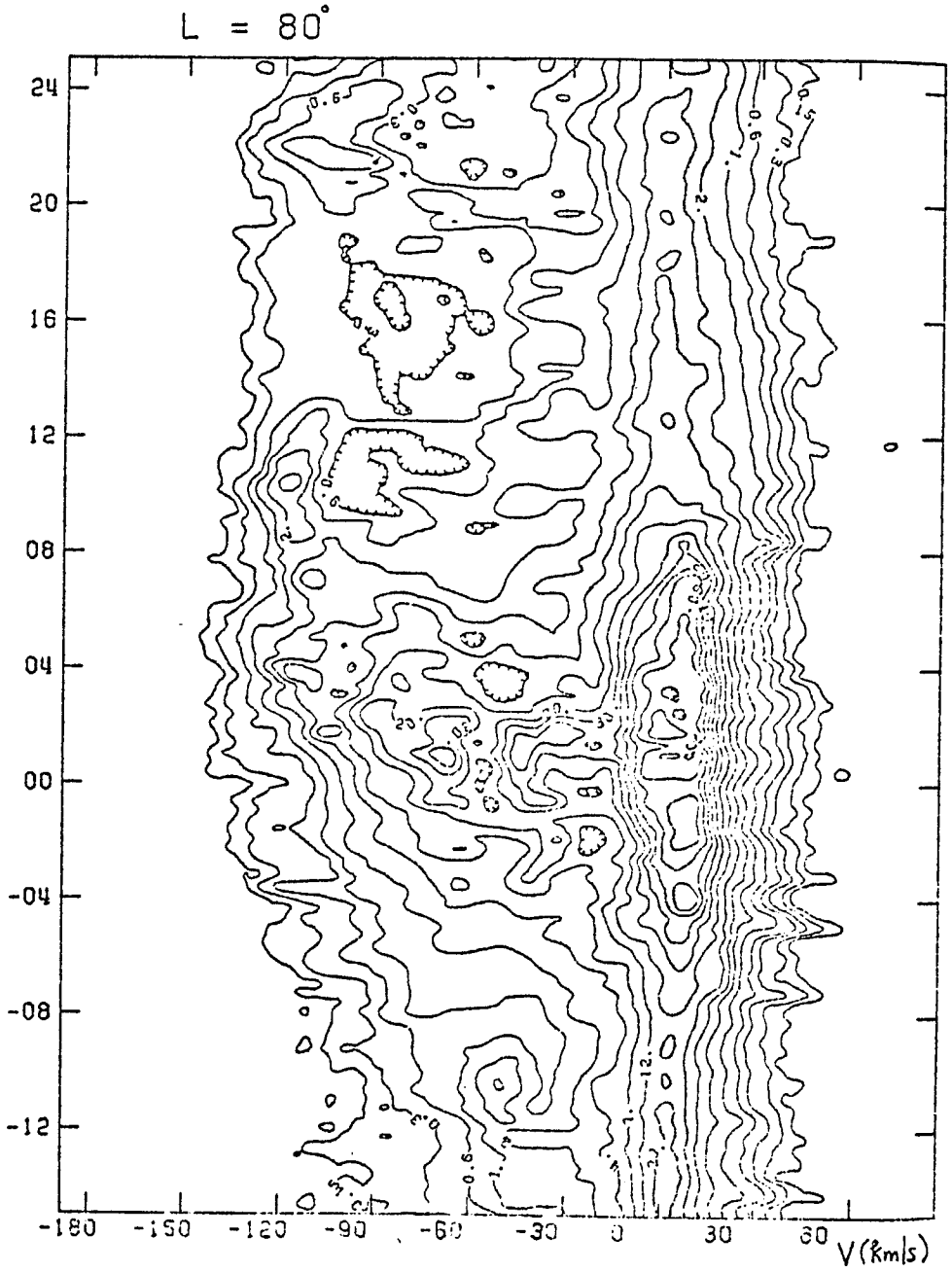


FIGURE 8

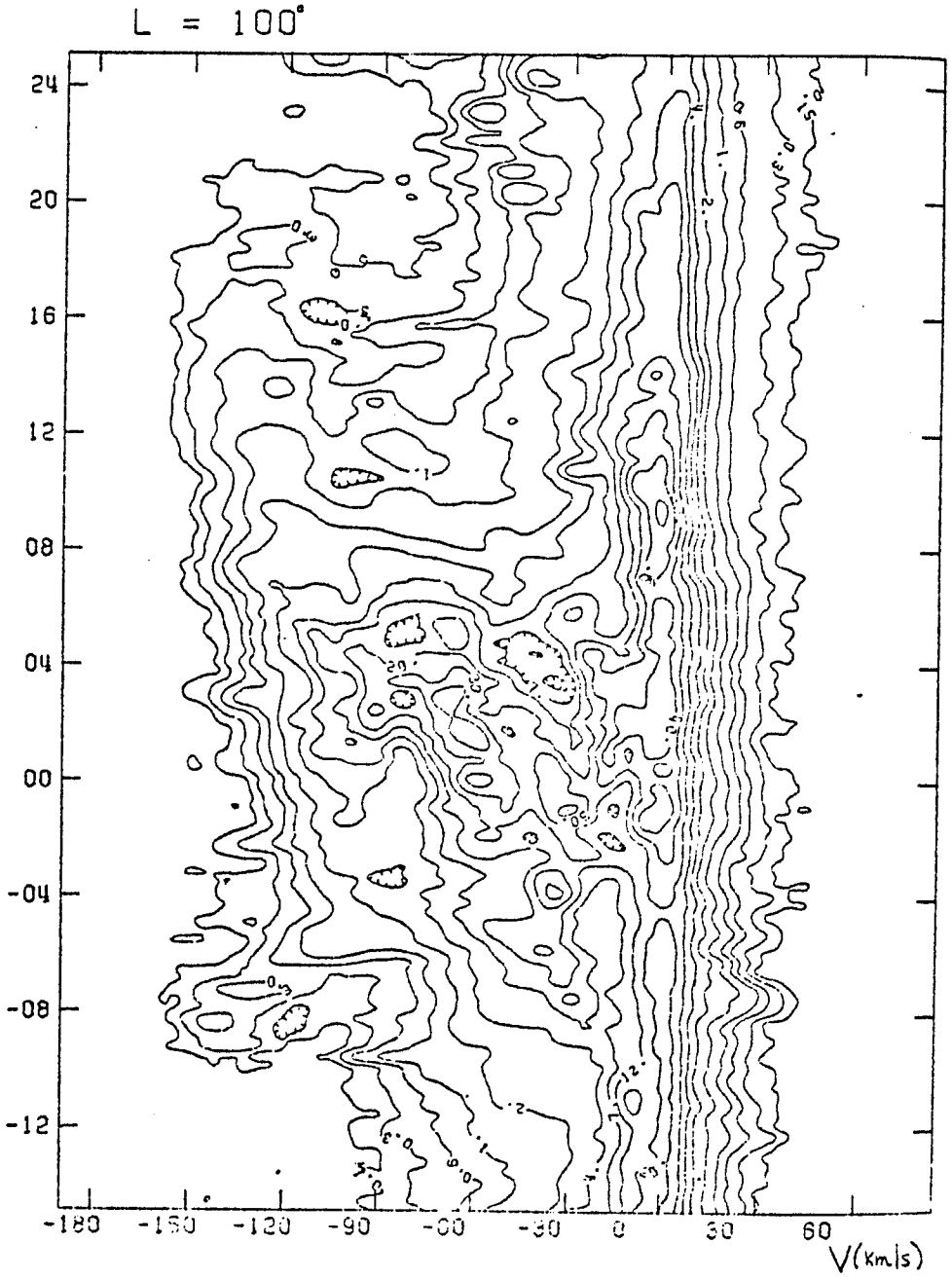


FIGURE 9

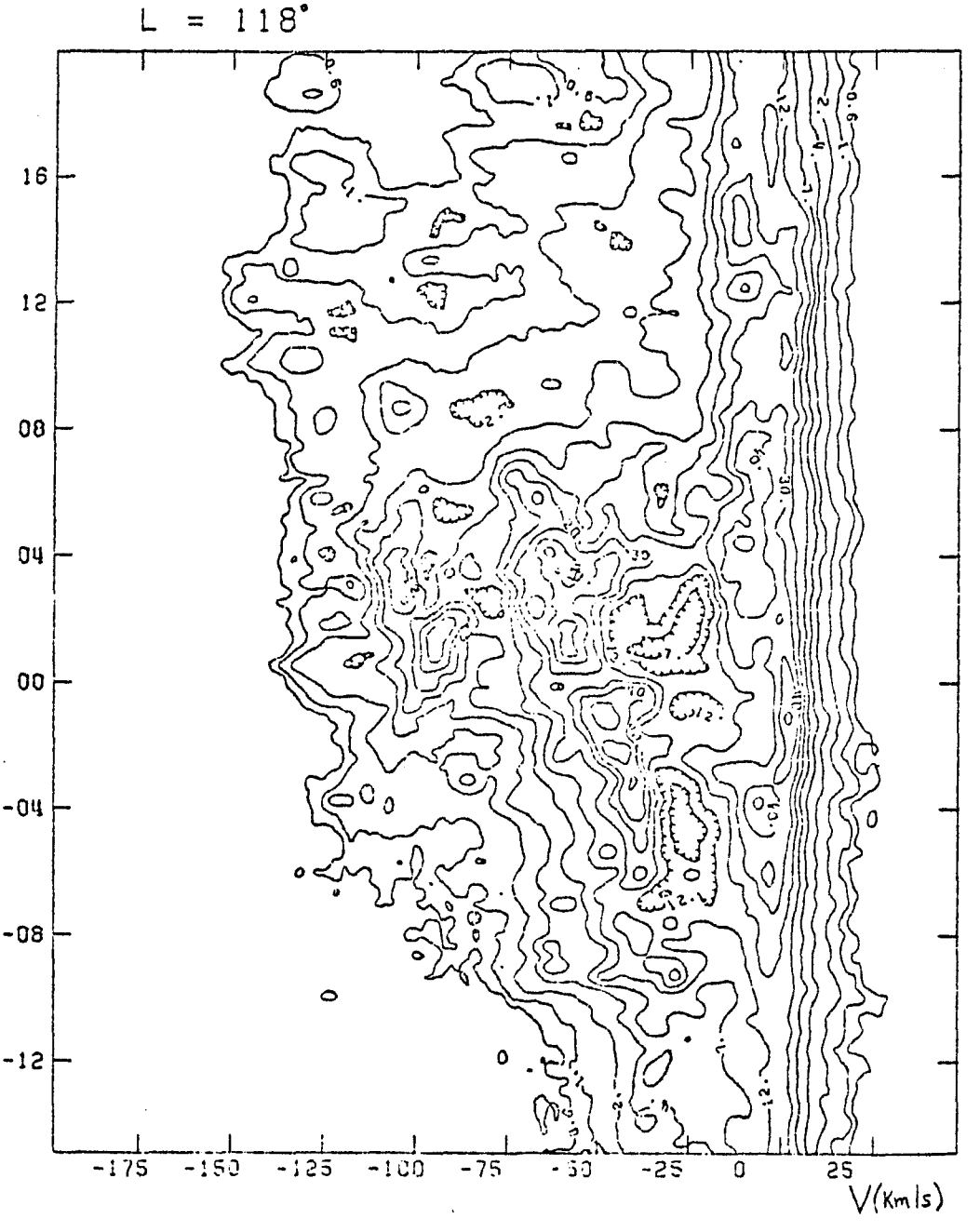


FIGURE 10

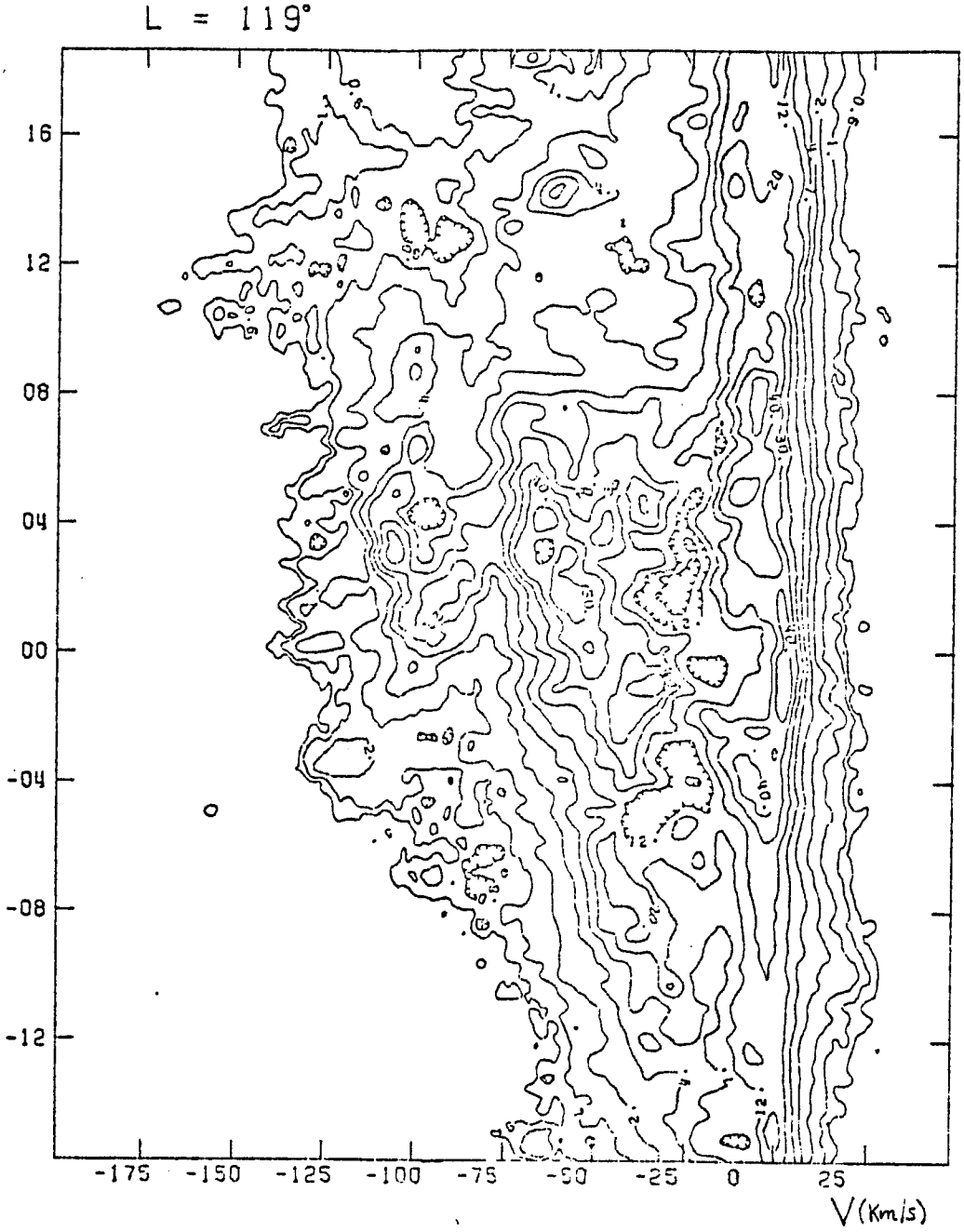


FIGURE 11

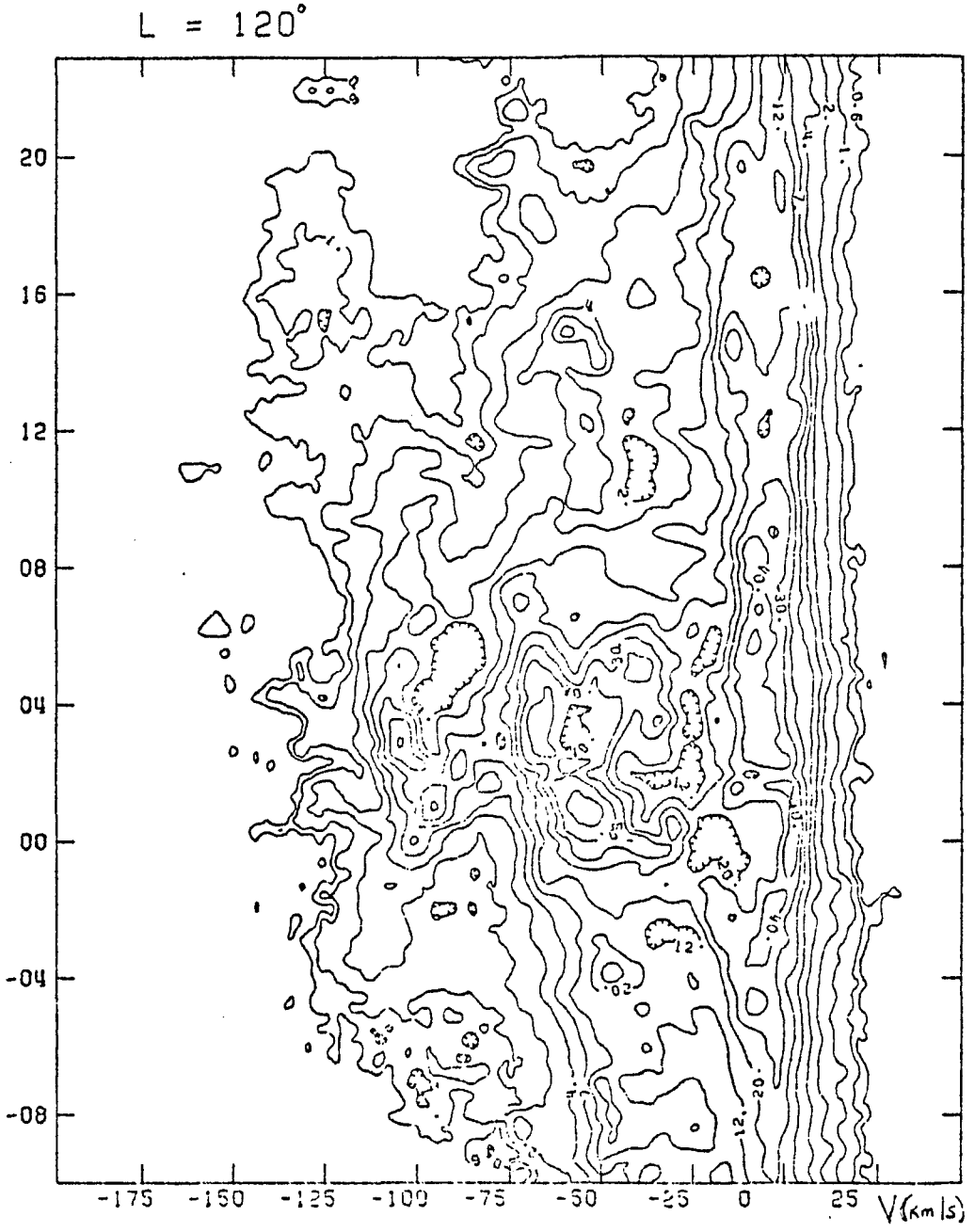


FIGURE 12

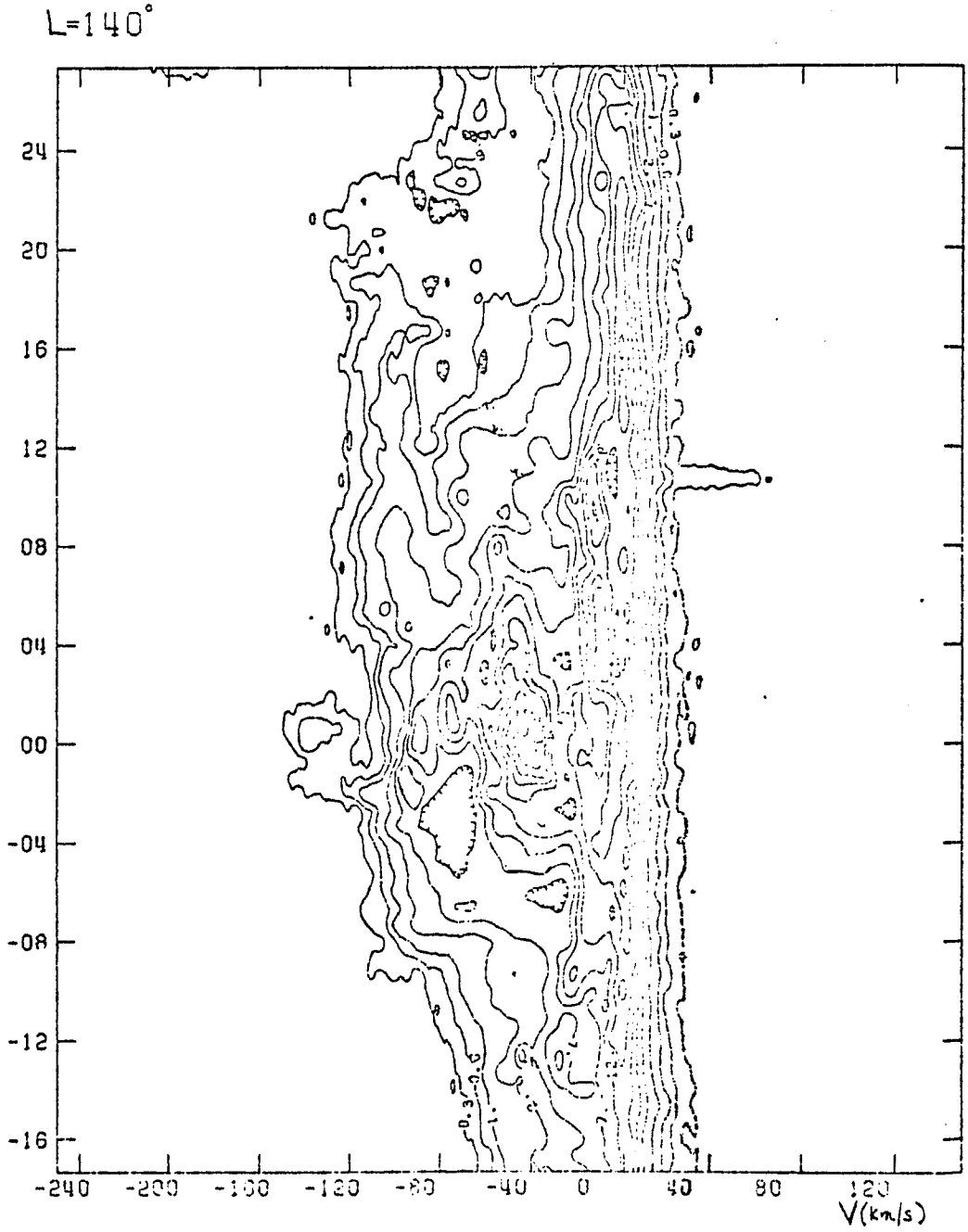

L=140°
24
20
16
12
08
04
00
-04
-08
-12
-16
-240
-200
-160
-120
-80
-40
0
40
80
120
V(km/s)

FIGURE 13

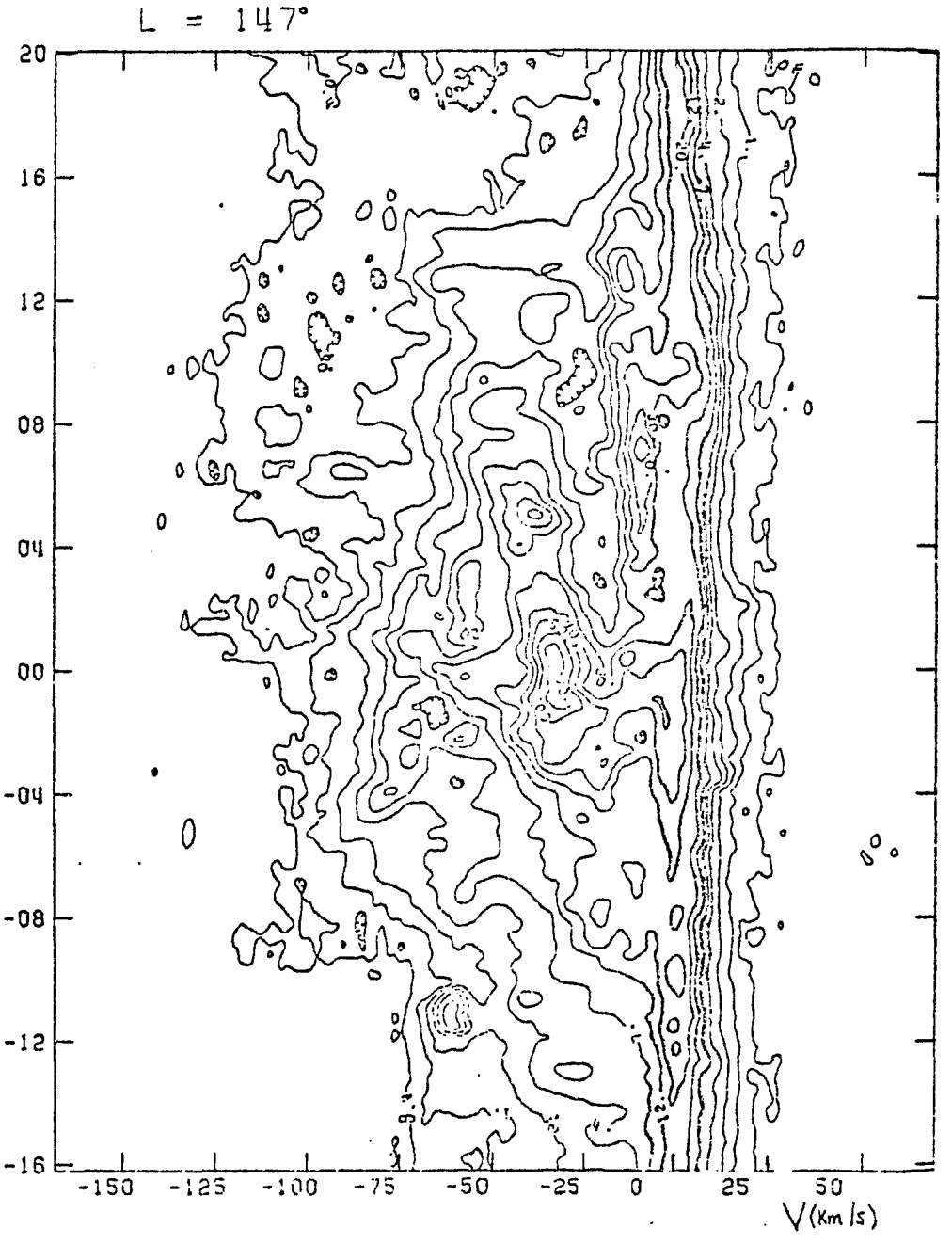


FIGURE 14

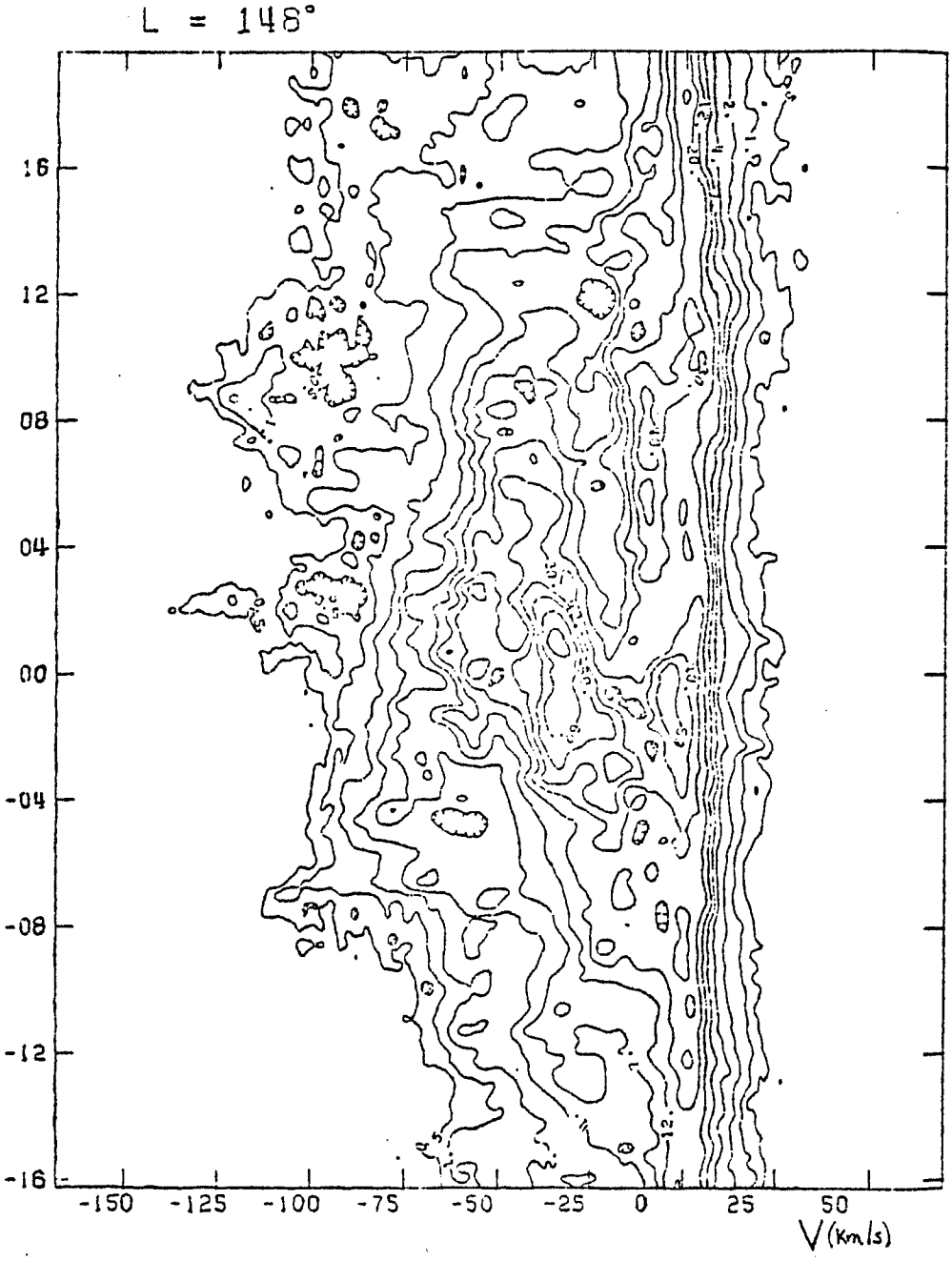


FIGURE 15

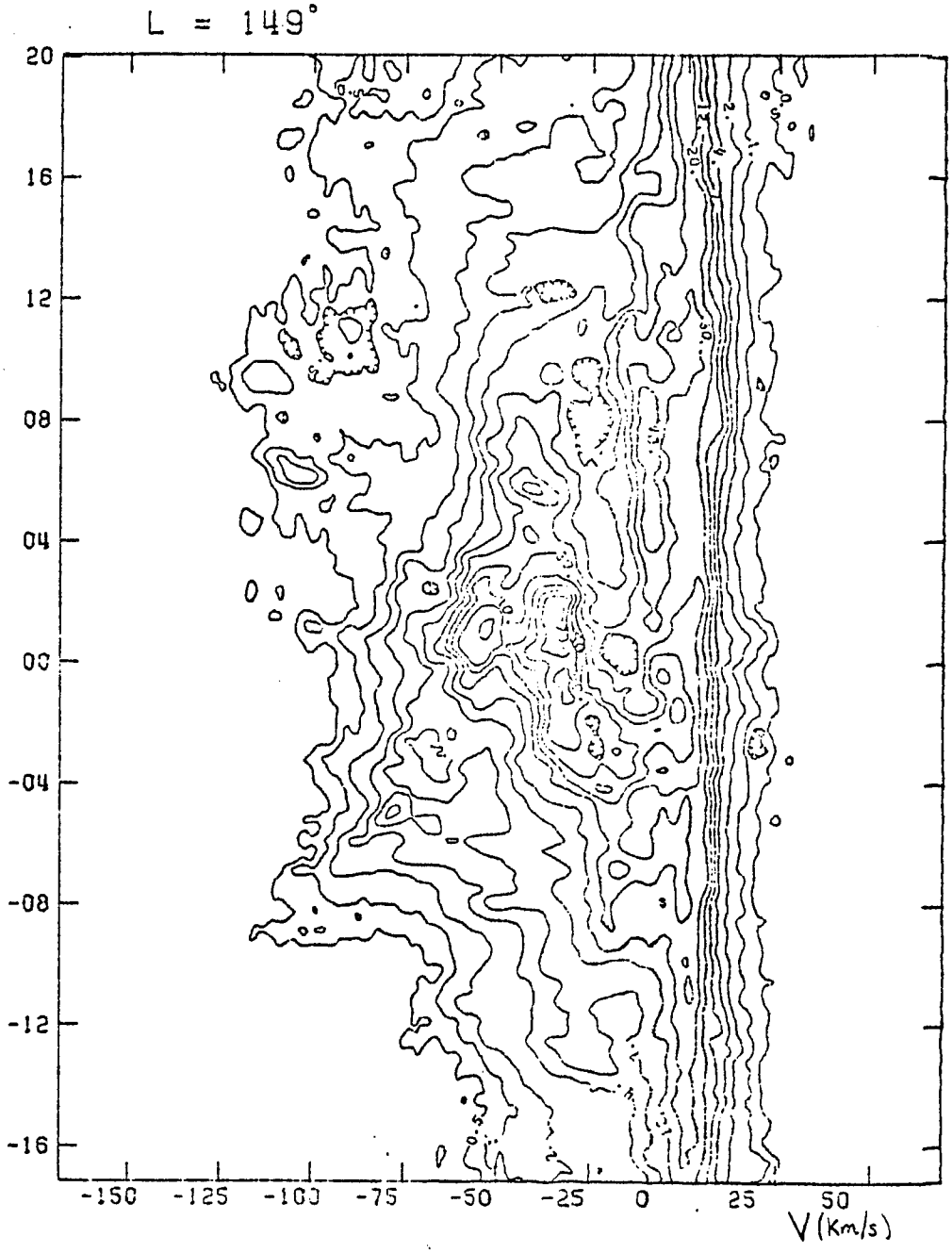


FIGURE 16

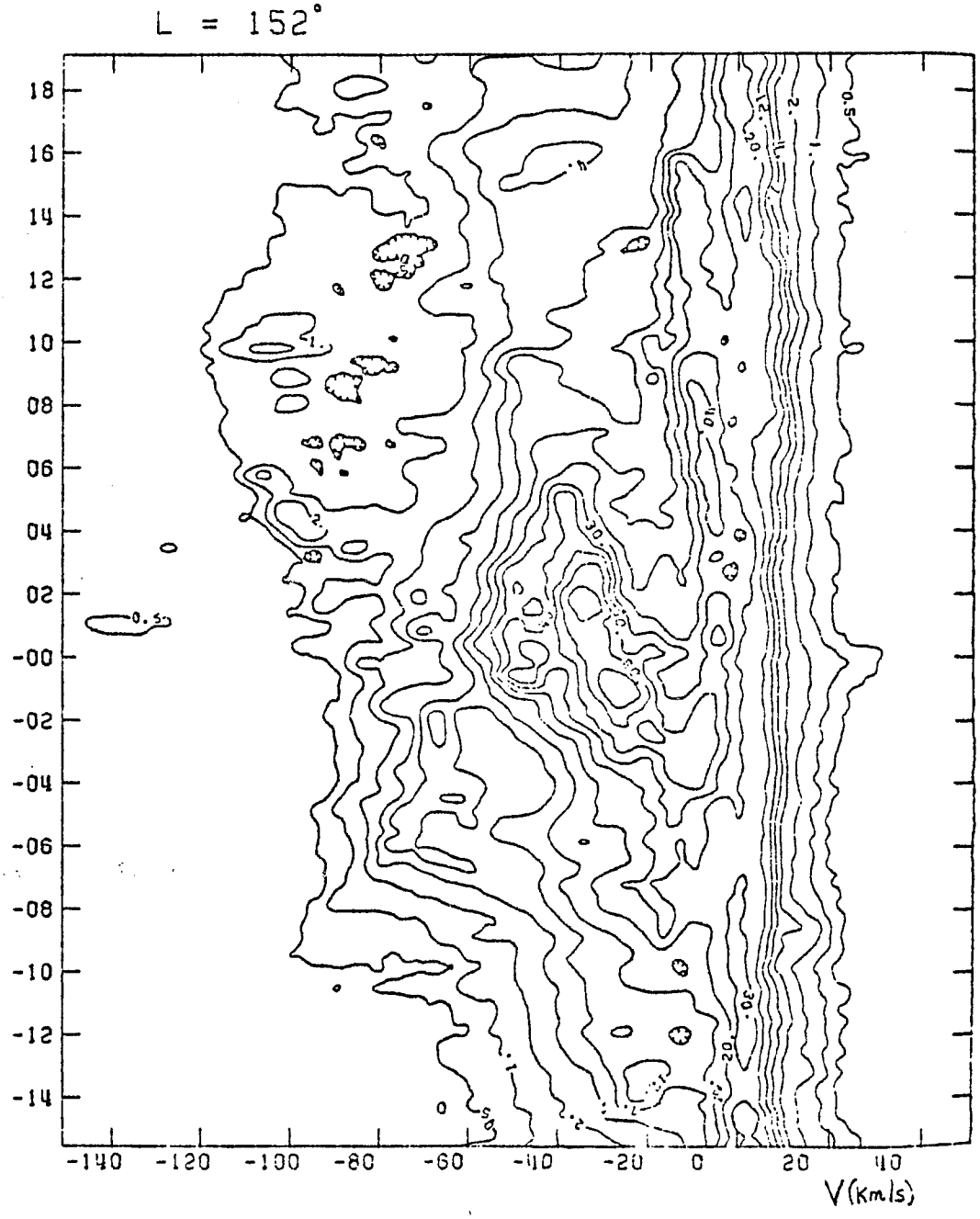

L = 152°
18
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-12
-14
-140
-120
-100
-80
-60
-40
-20
0
20
40
V(Km/s)

FIGURE 17

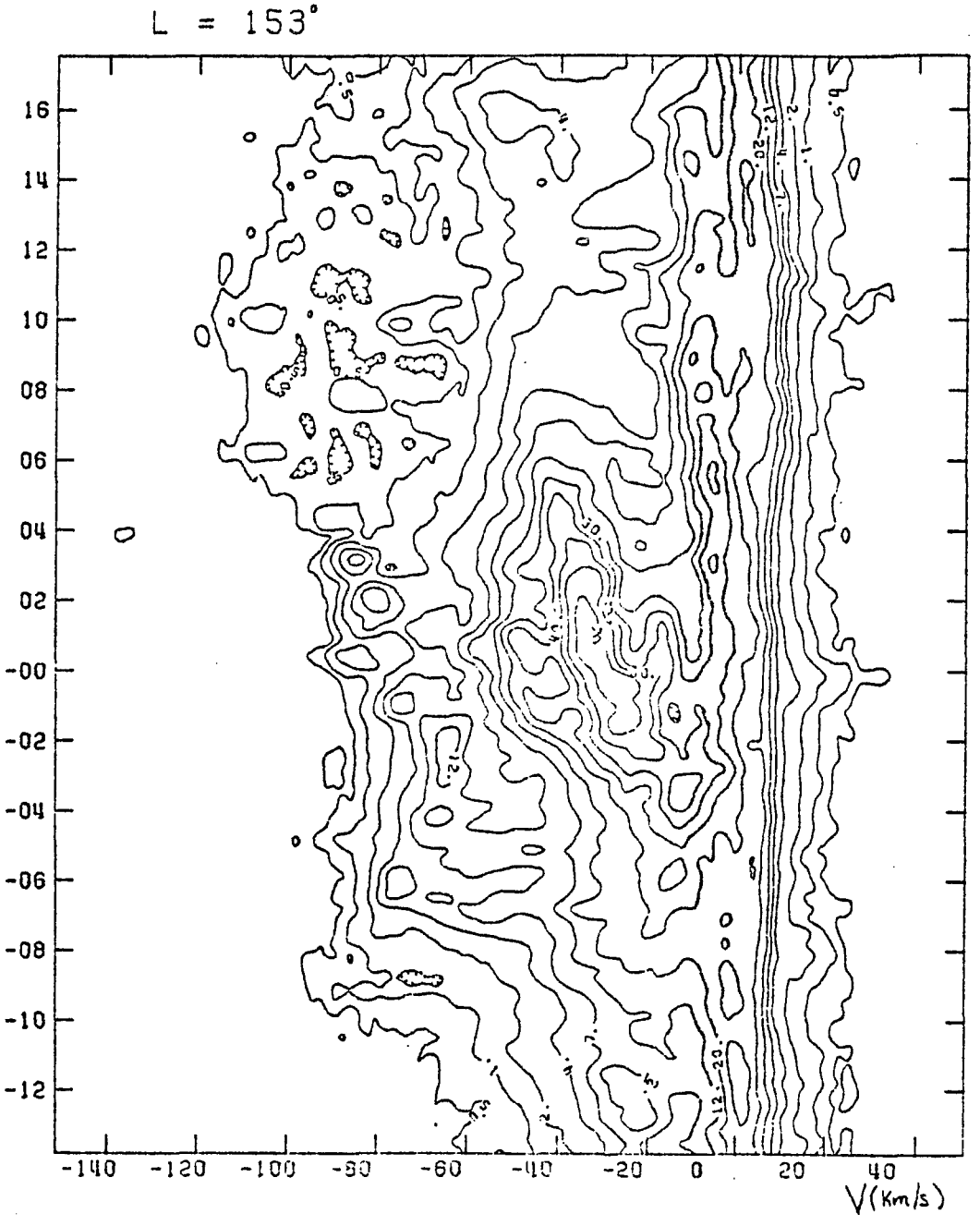

L = 153°
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-12
-140
-120
-100
-80
-60
-40
-20
0
20
40
V(Km/s)

FIGURE 18

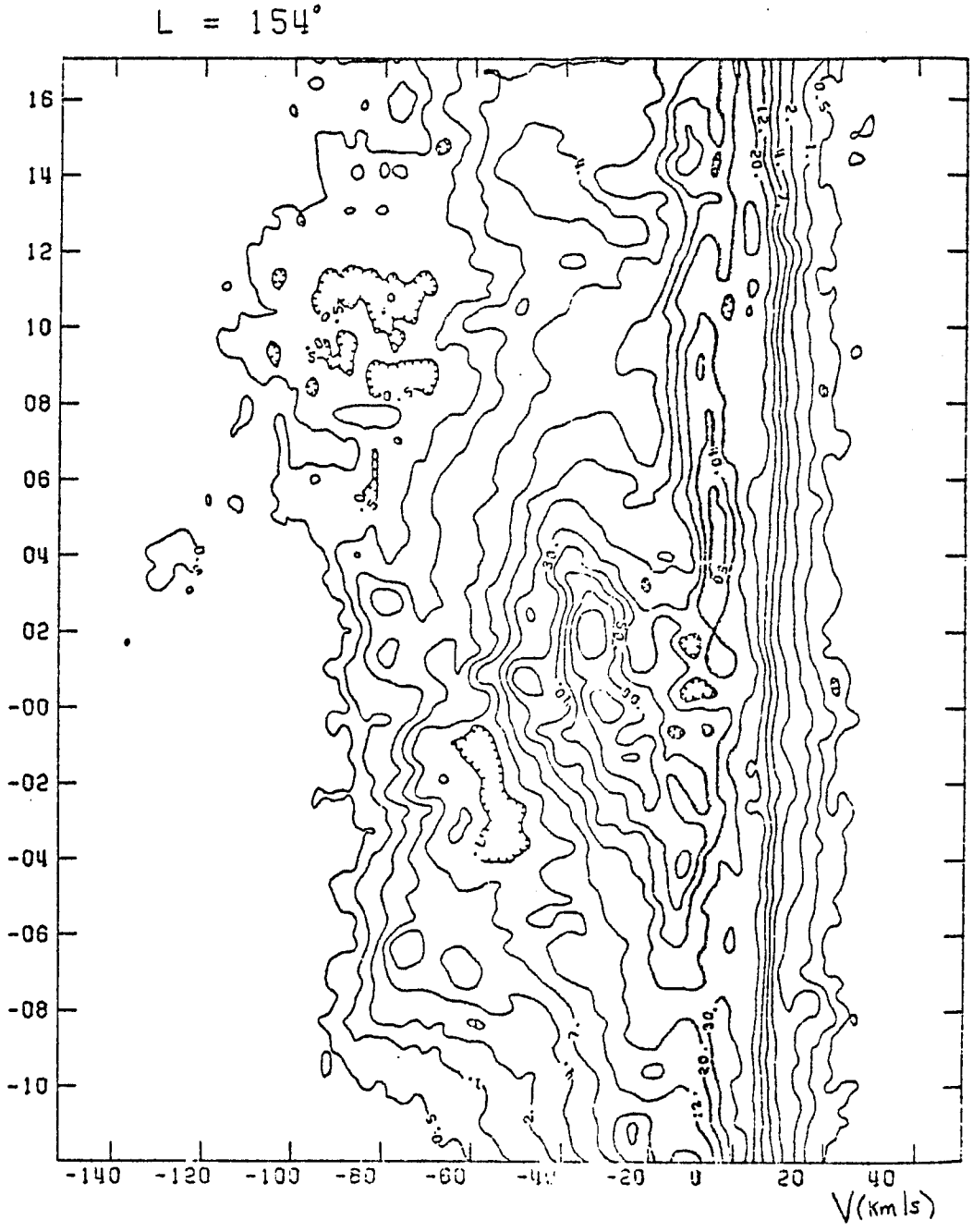

L = 154°
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-140
-120
-100
-80
-60
-20
0
20
40
V(Km/s)

FIGURE 19

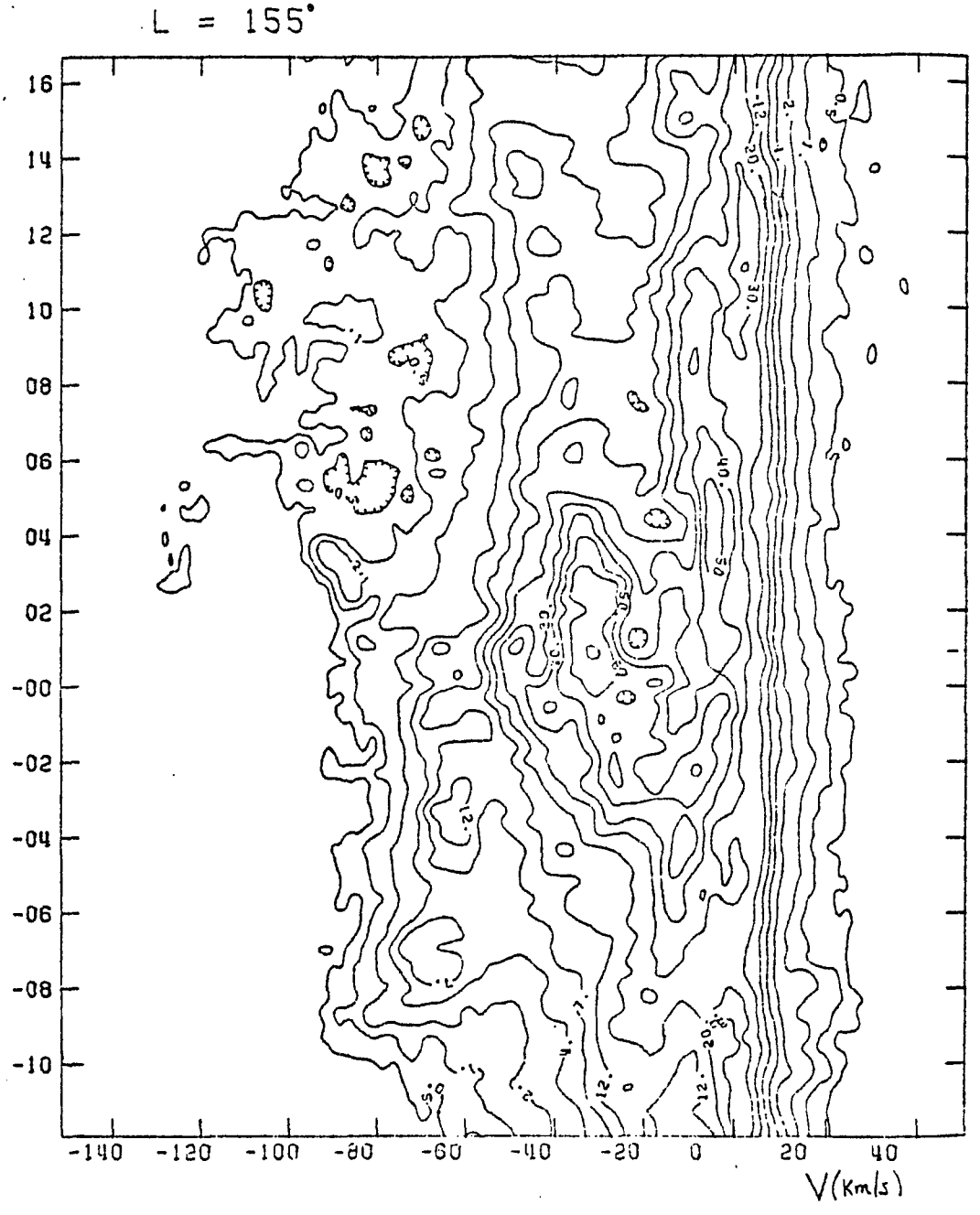

L = 155°
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-140
-120
-100
-80
-60
-40
-20
0
20
40
V(Km/s)

FIGURE 20

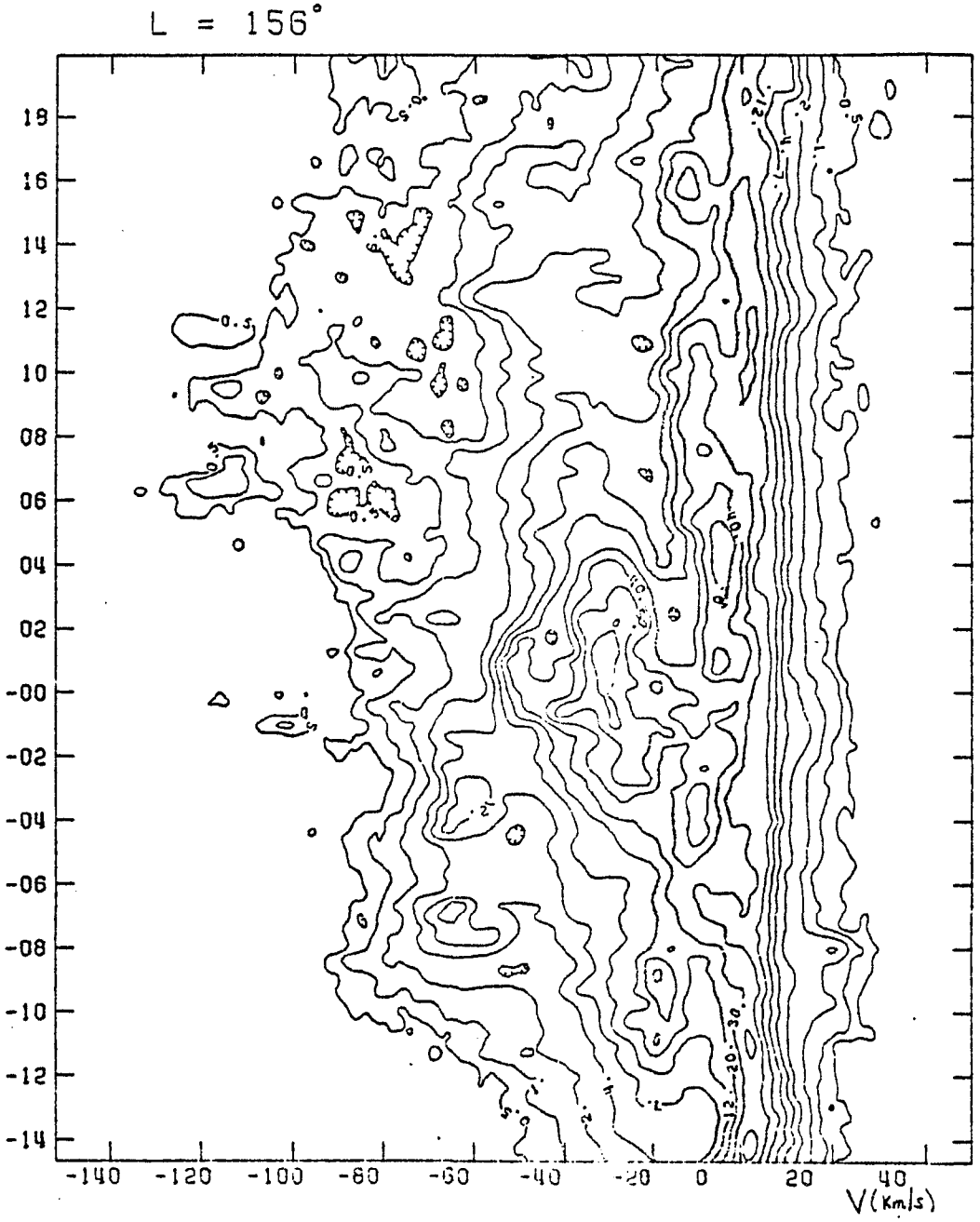

L = 156°
18
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-12
-14
-140
-120
-100
-80
-60
-40
-20
0
20
40
V(Km/s)

FIGURE 21

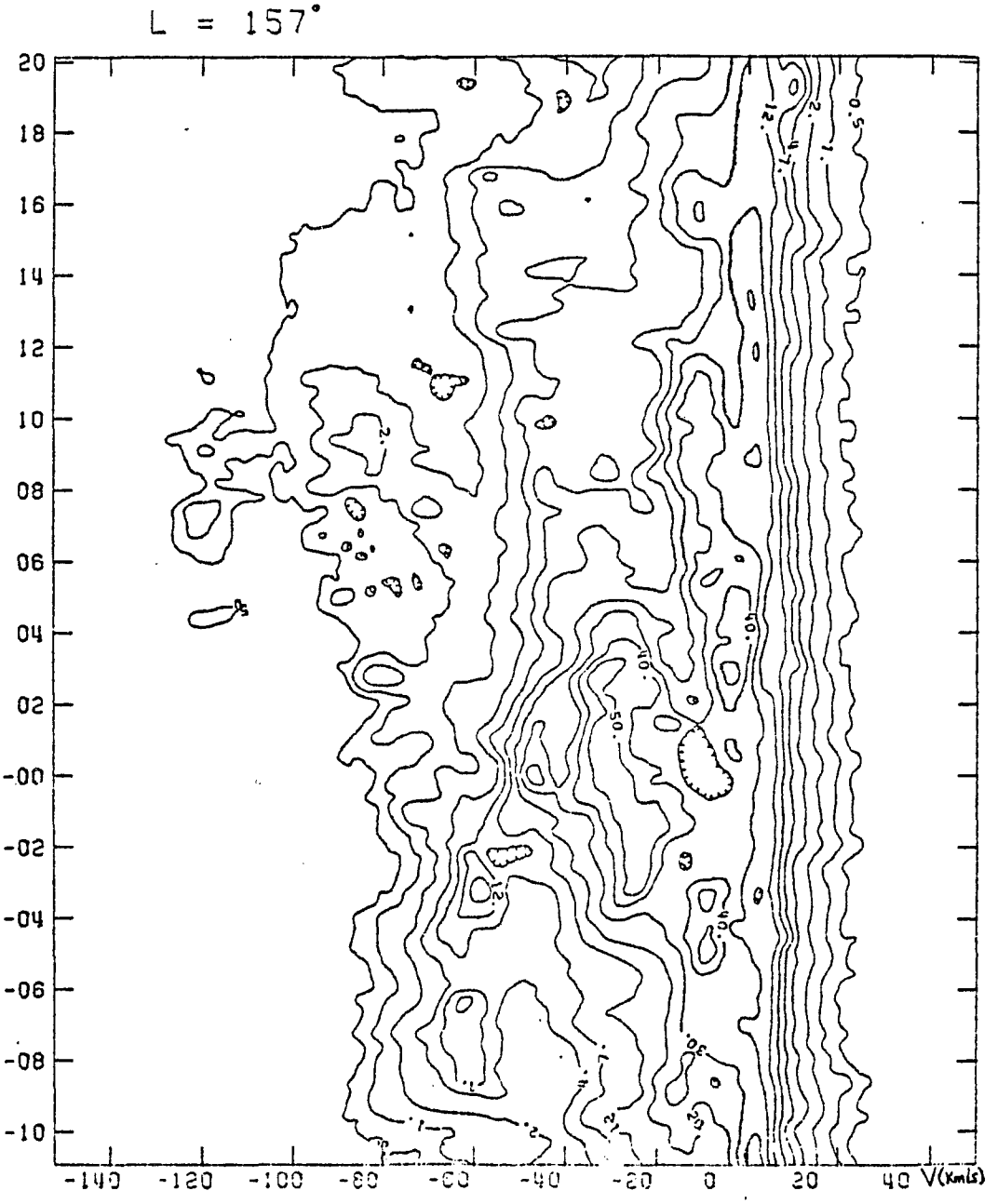

L = 157°
20
18
16
14
12
10
08
06
04
02
-00
-02
-04
-06
-08
-10
-140
-120
-100
-80
-60
-40
-20
0
20
40
V(Km/s)

FIGURE 22

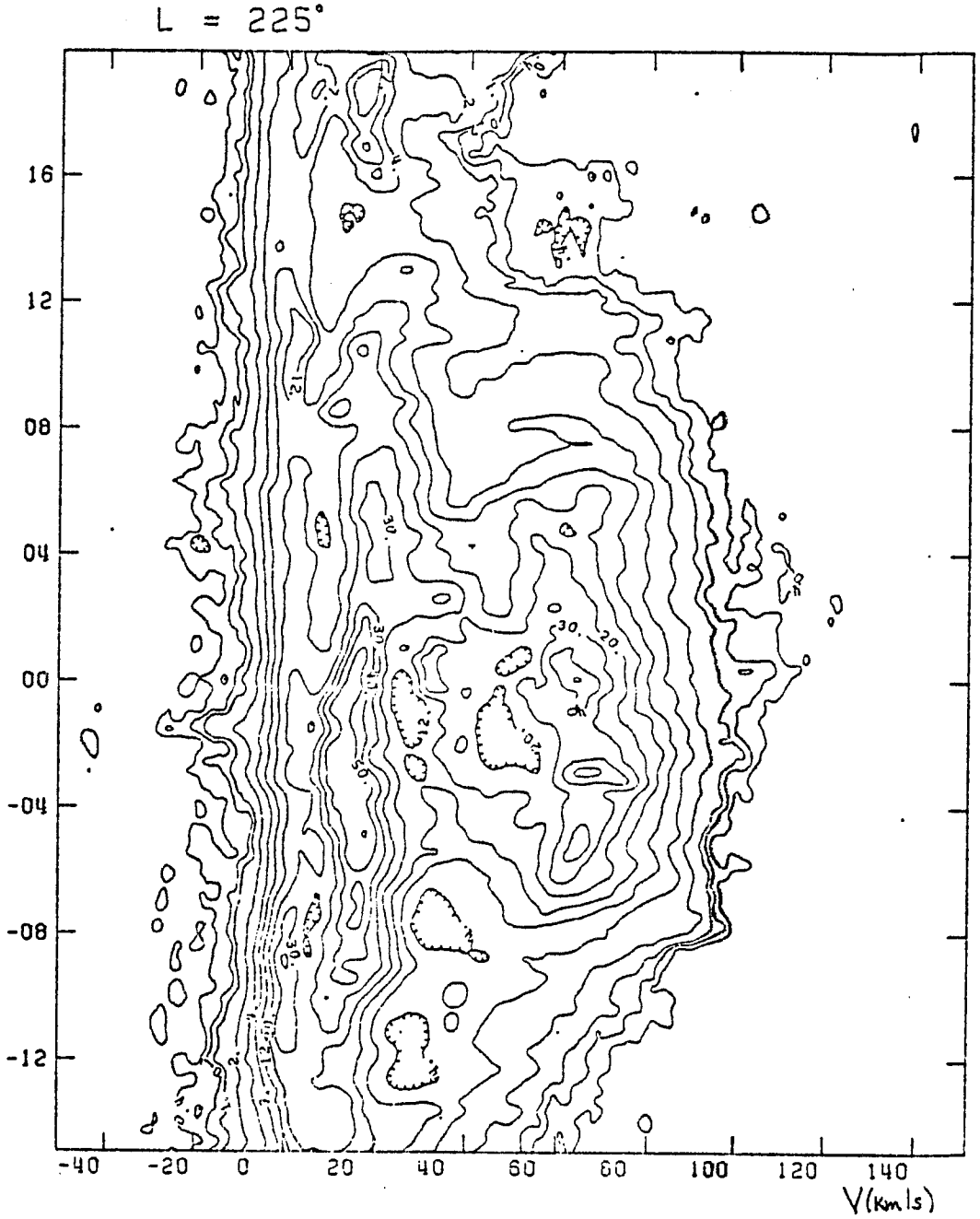


FIGURE 23

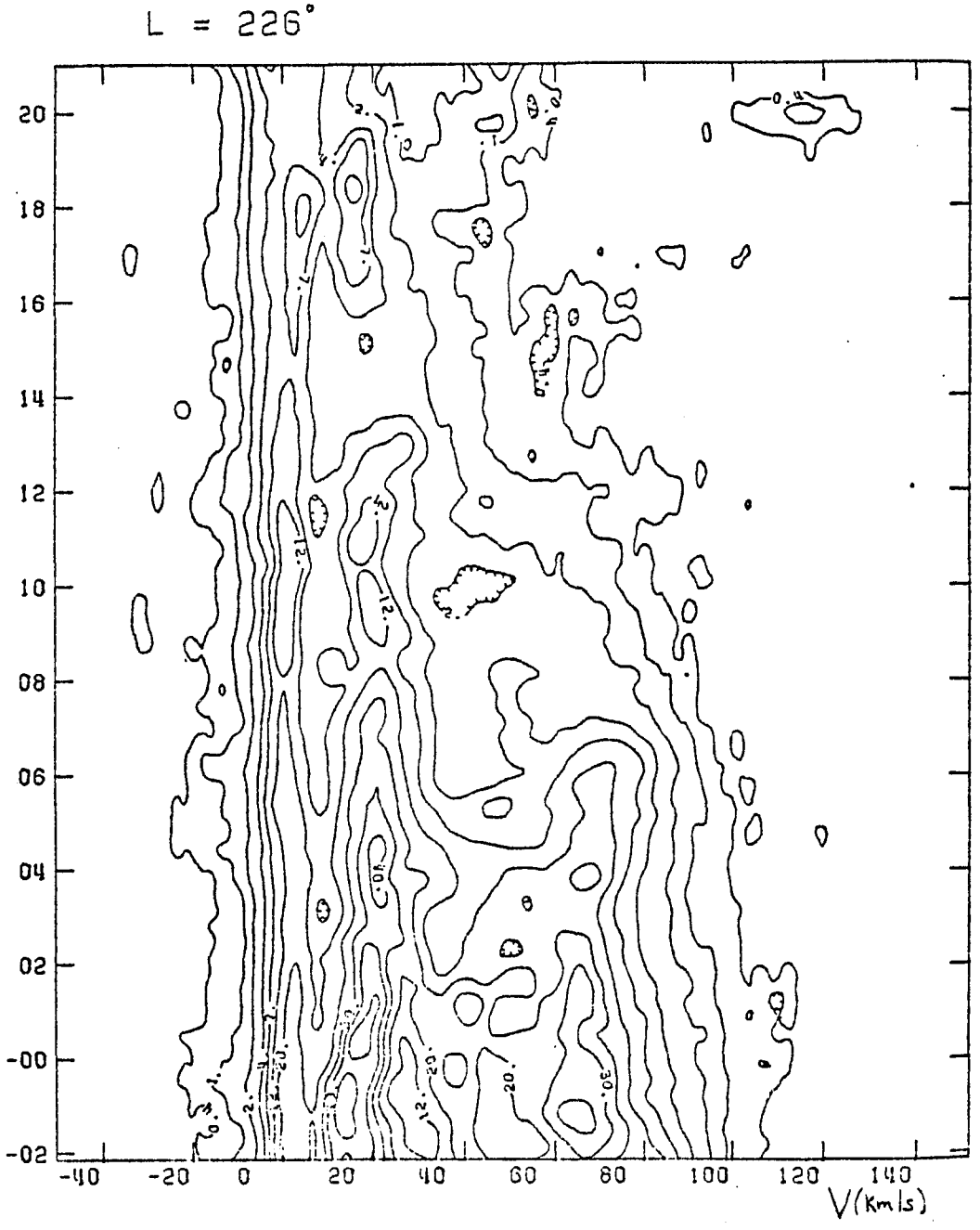

L = 226°
20
18
16
14
12
10
08
06
04
02
-00
-02
-40
-20
0
20
40
60
80
100
120
140
V(km/s)

FIGURE 24

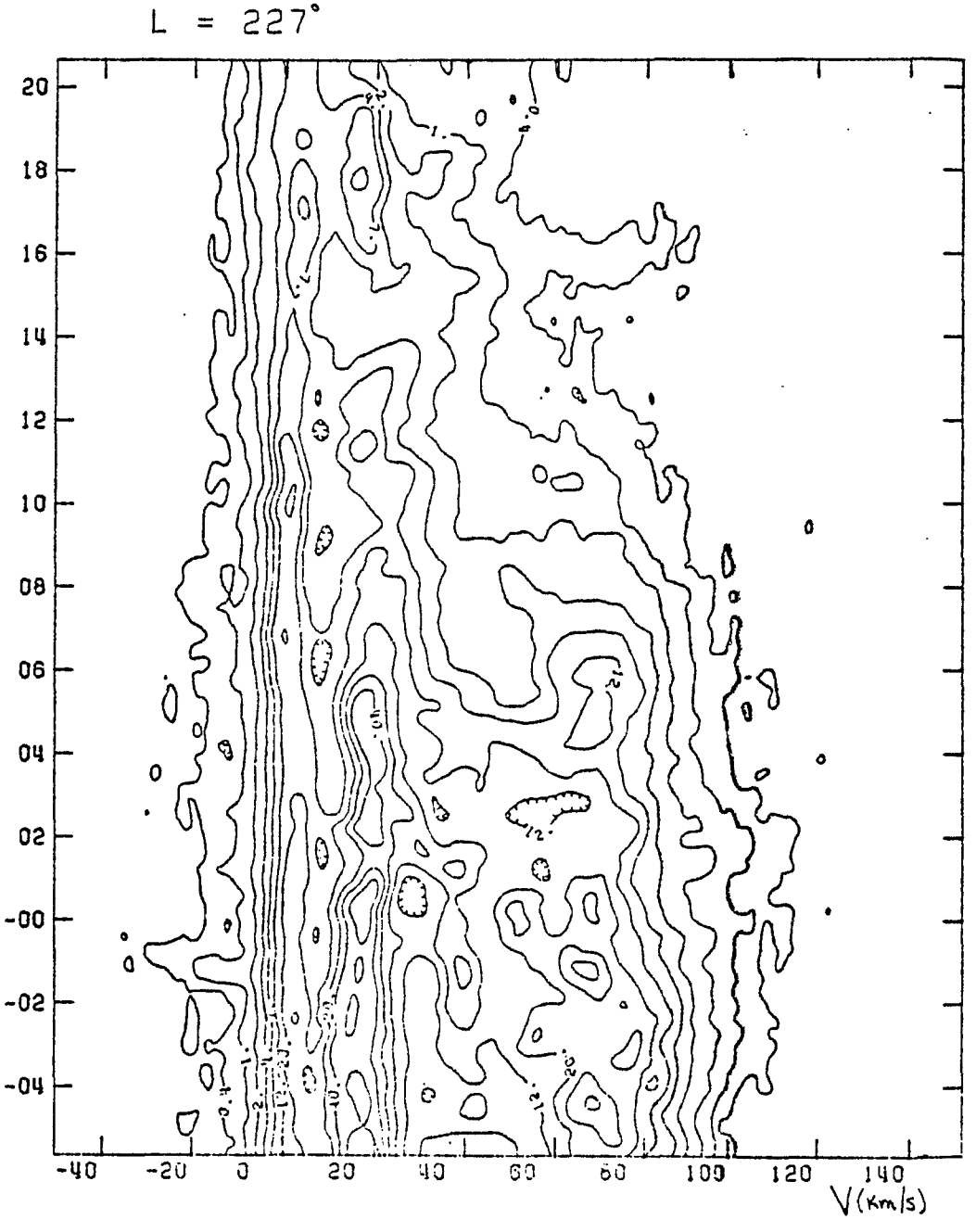

L = 227°
20
18
16
14
12
10
08
06
04
02
-00
-02
-04
-40
-20
0
20
40
60
80
100
120
140
V(Km/s)

FIGURE 25

L = 228°

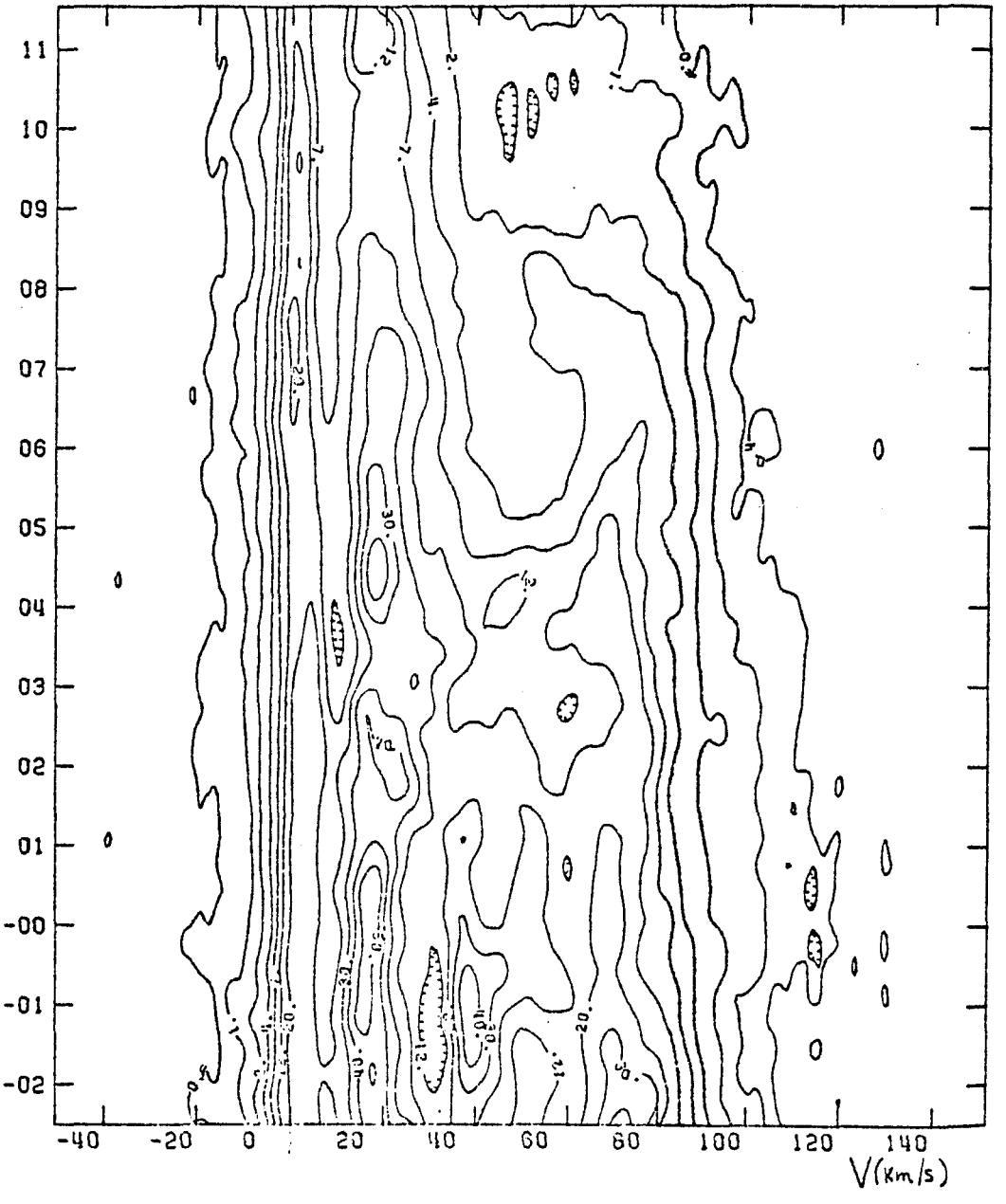


FIGURE 26

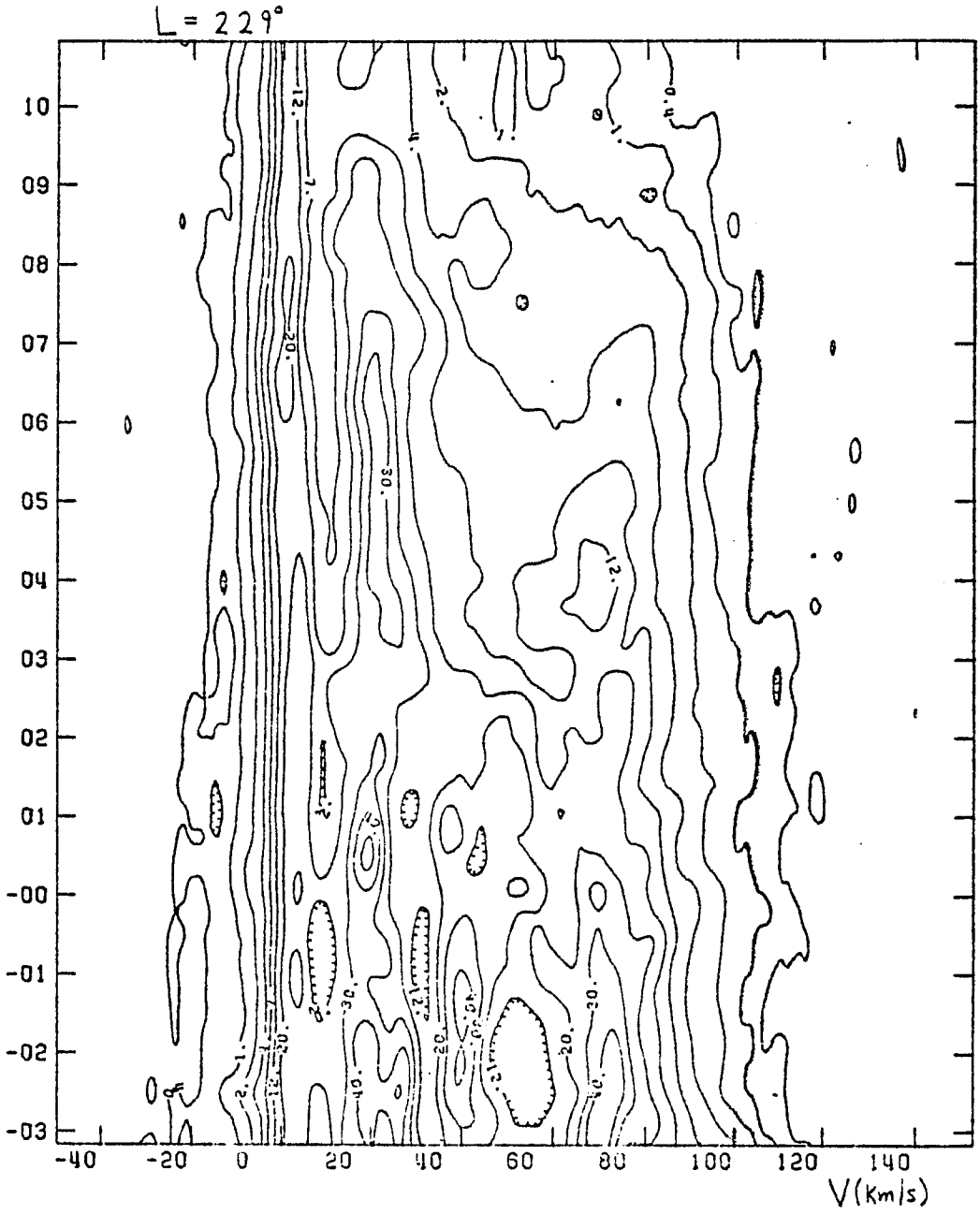

L = 229°
10
09
08
07
06
05
04
03
02
01
-00
-01
-02
-03
-40
-20
0
20
40
60
80
100
120
140
V(Km/s)

FIGURE 27

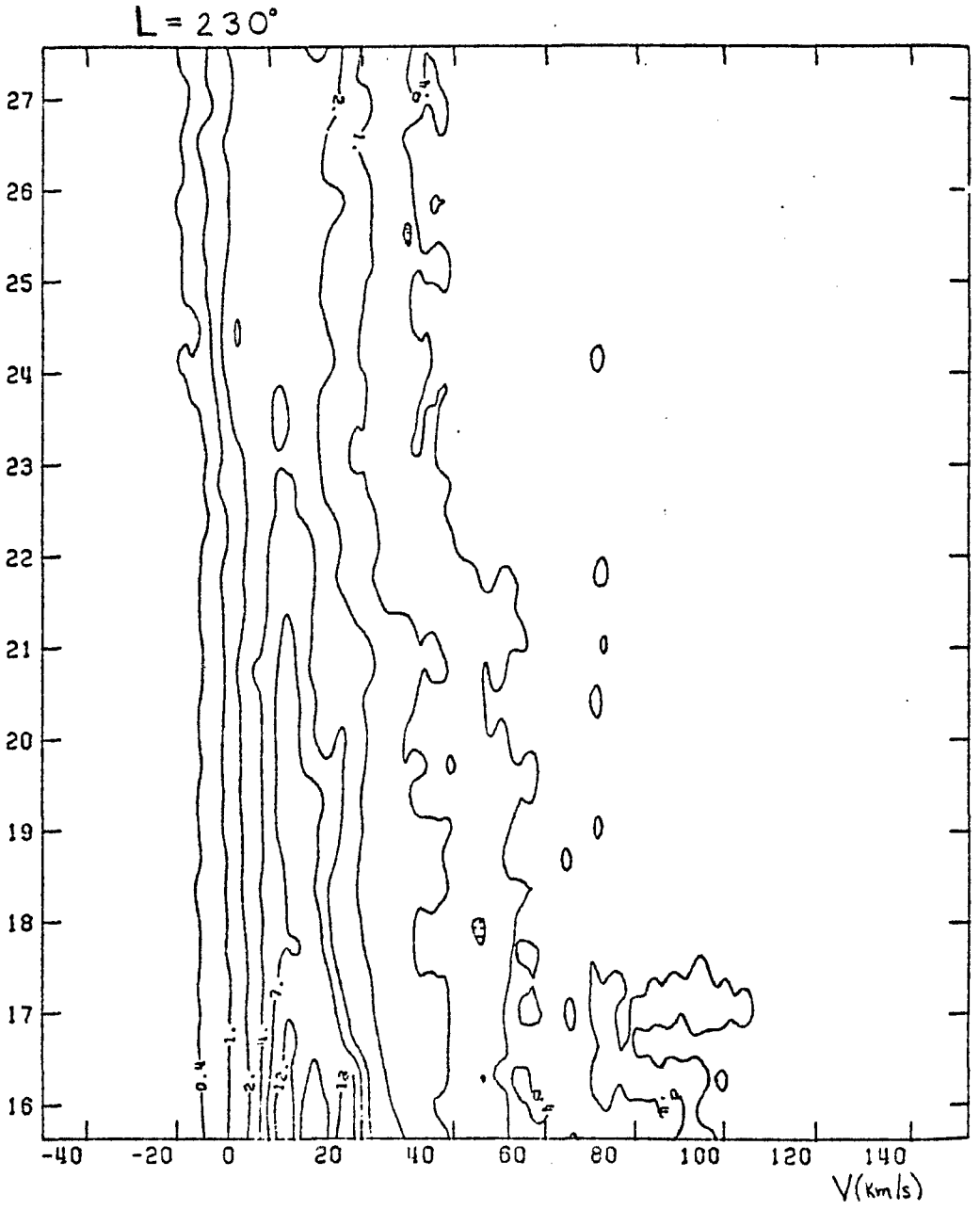

L = 230°
27
26
25
24
23
22
21
20
19
18
17
16
-40
-20
0
20
40
60
80
100
120
140
V(km/s)

FIGURE 28

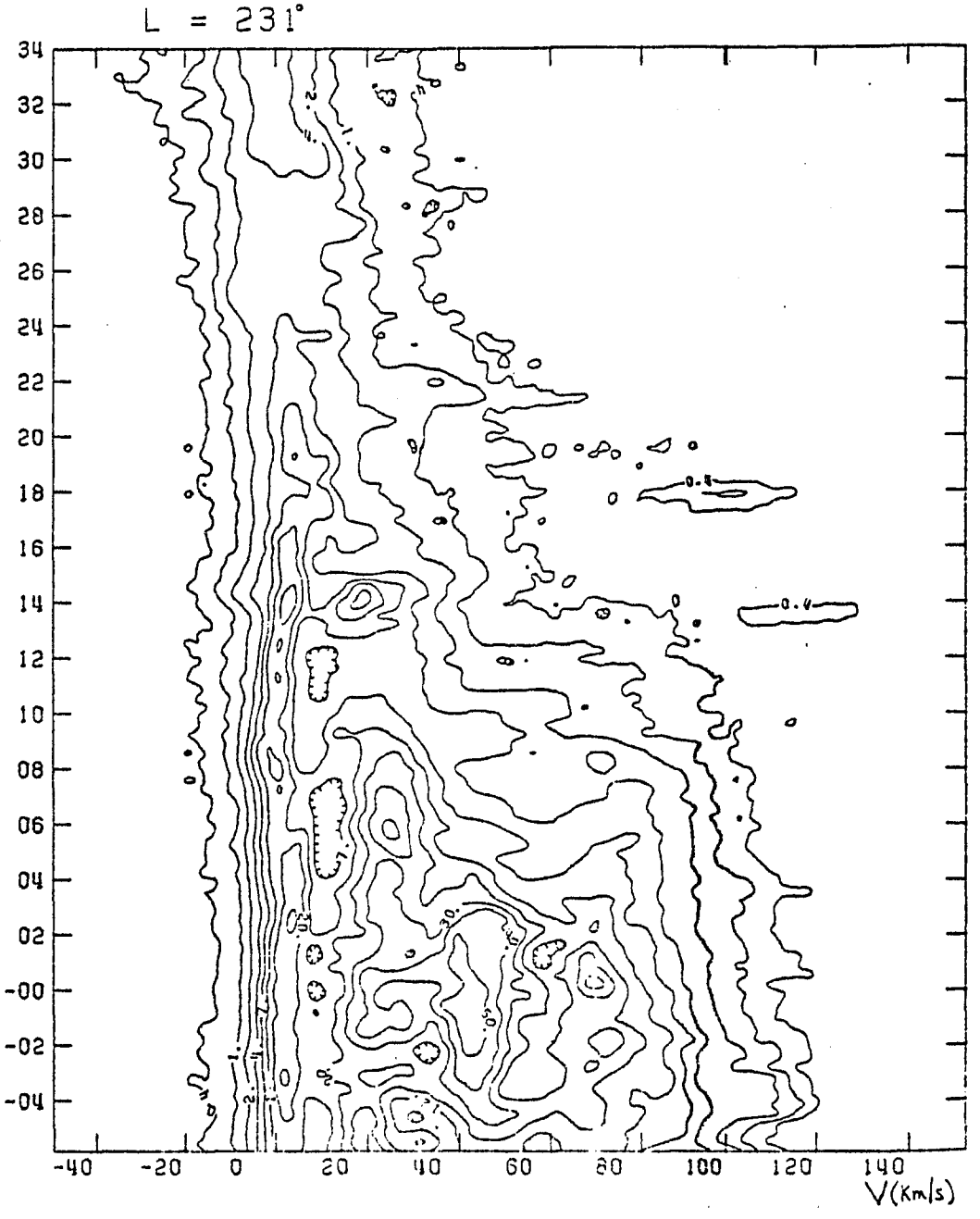

L = 231°
34
32
30
28
26
24
22
20
18
16
14
12
10
08
06
04
02
-00
-02
-04
-40
-20
0
20
40
60
80
100
120
140
V(Km/s)

FIGURE 29

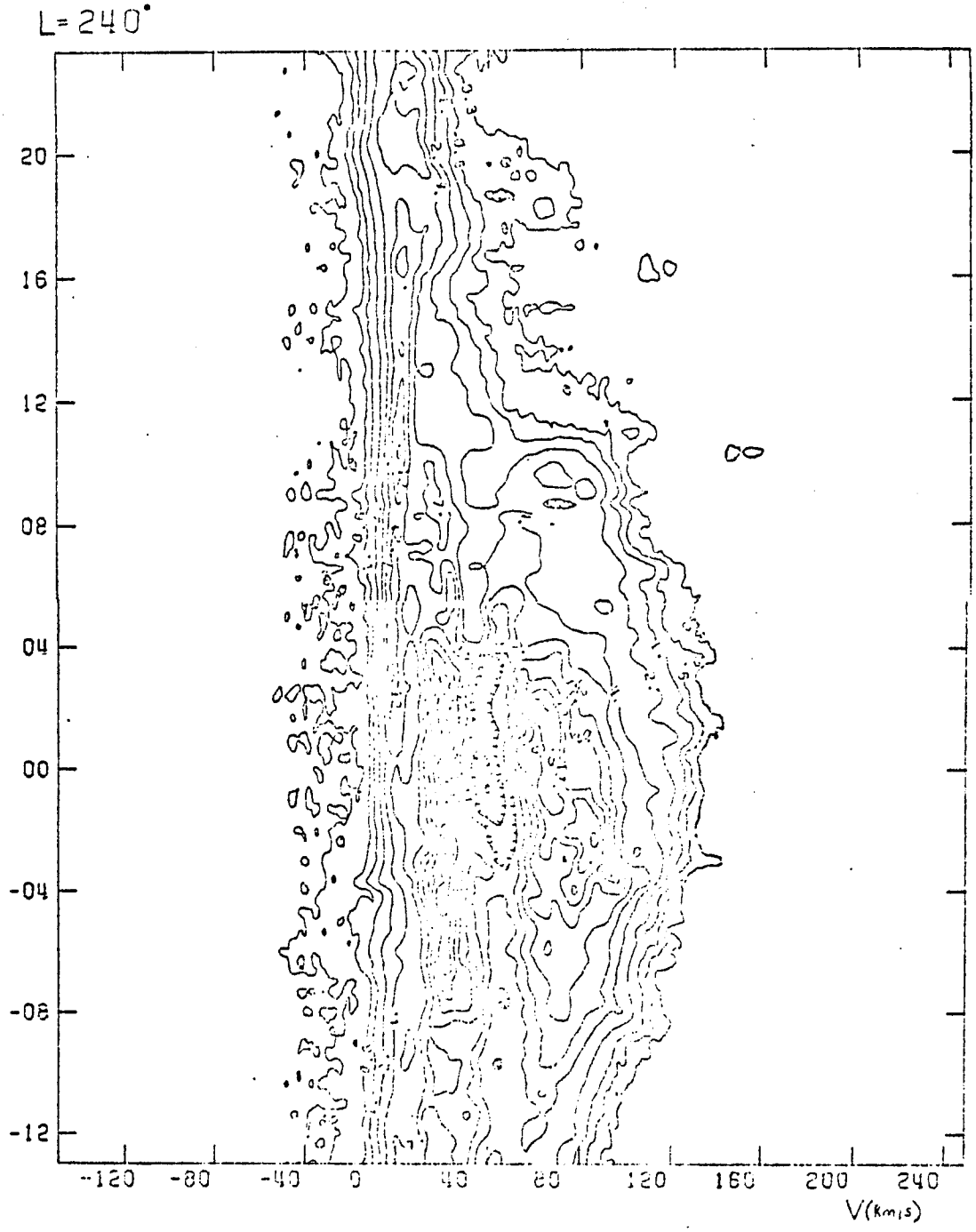

L=240°
20
16
12
08
04
00
-04
-08
-12
-120
-80
-40
0
40
80
120
160
200
240
V(km/s)

FIGURE 30

B = 8°

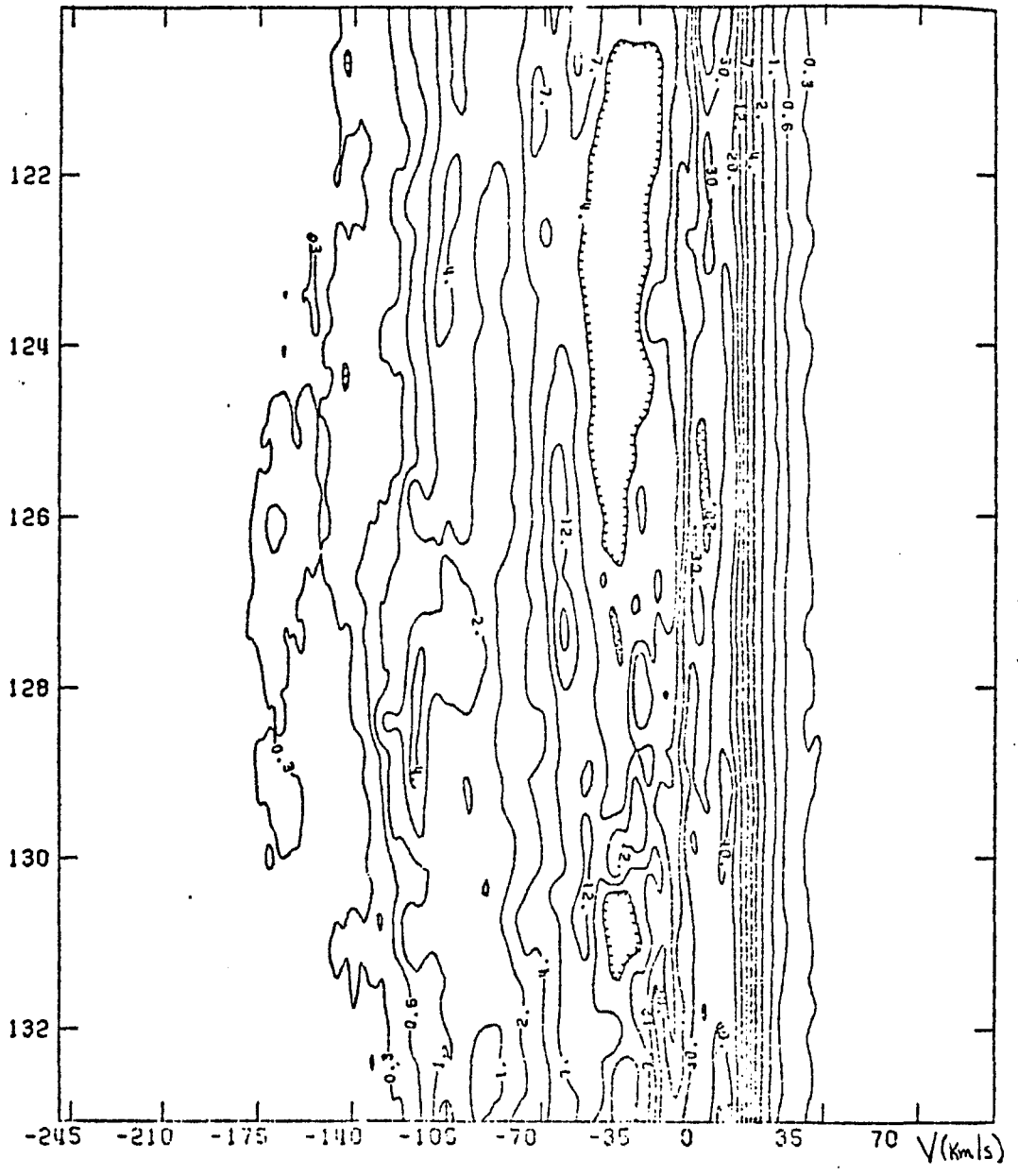


FIGURE 31

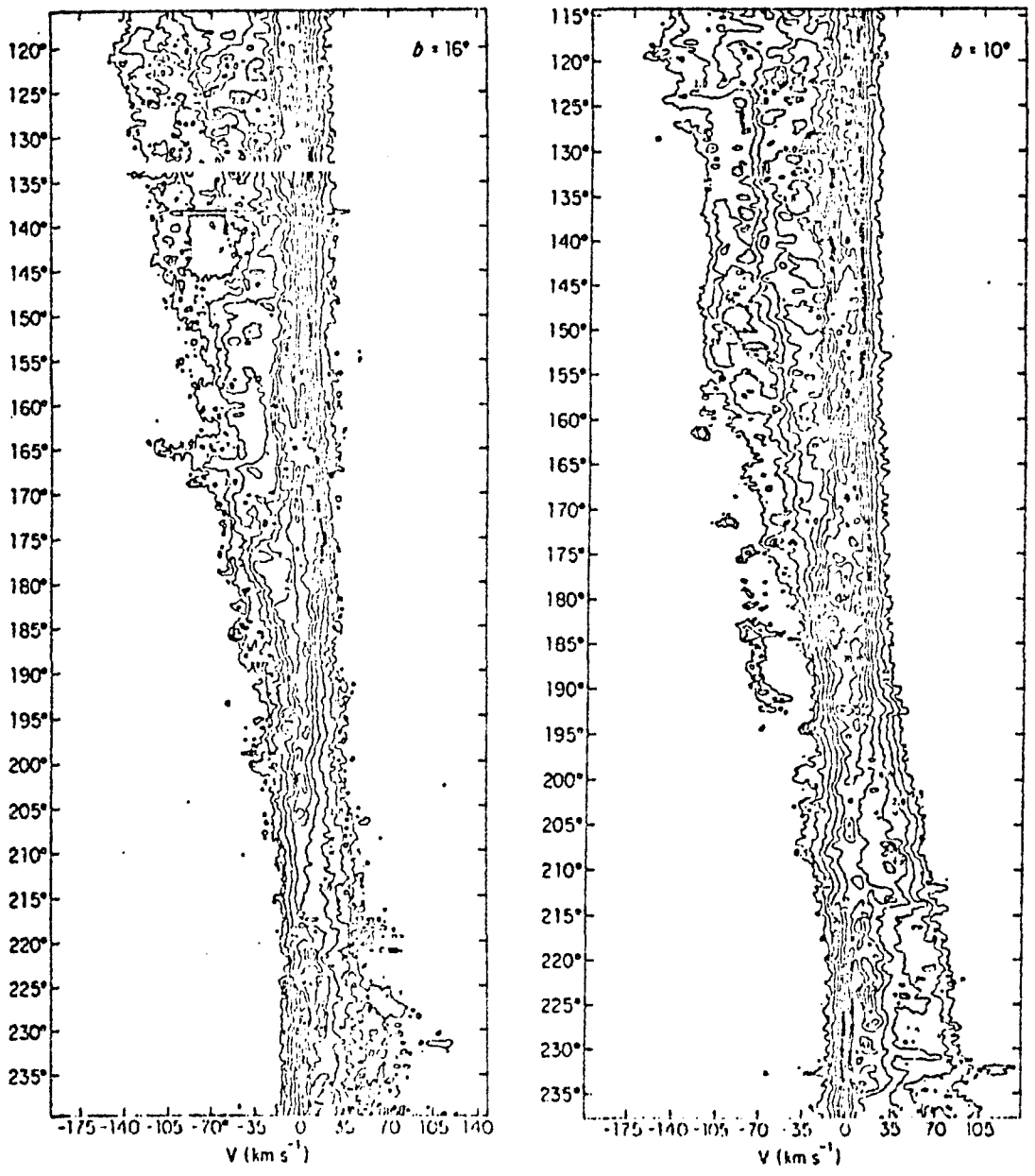


FIGURE 32

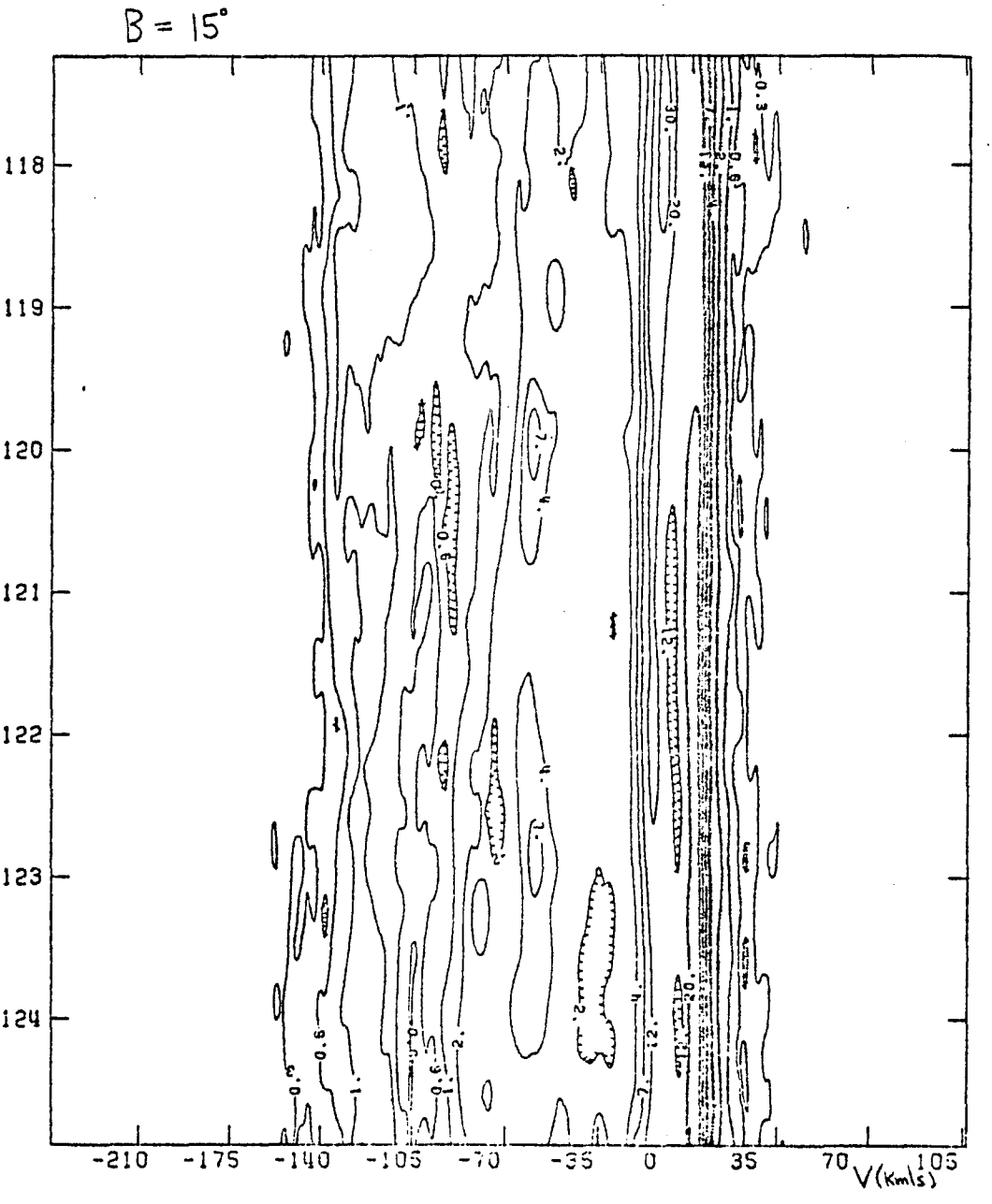

B = 15°
118
119
120
121
122
123
124
-210
-175
-140
-105
-70
-35
0
35
70
105
V(Km/s)

FIGURE 33

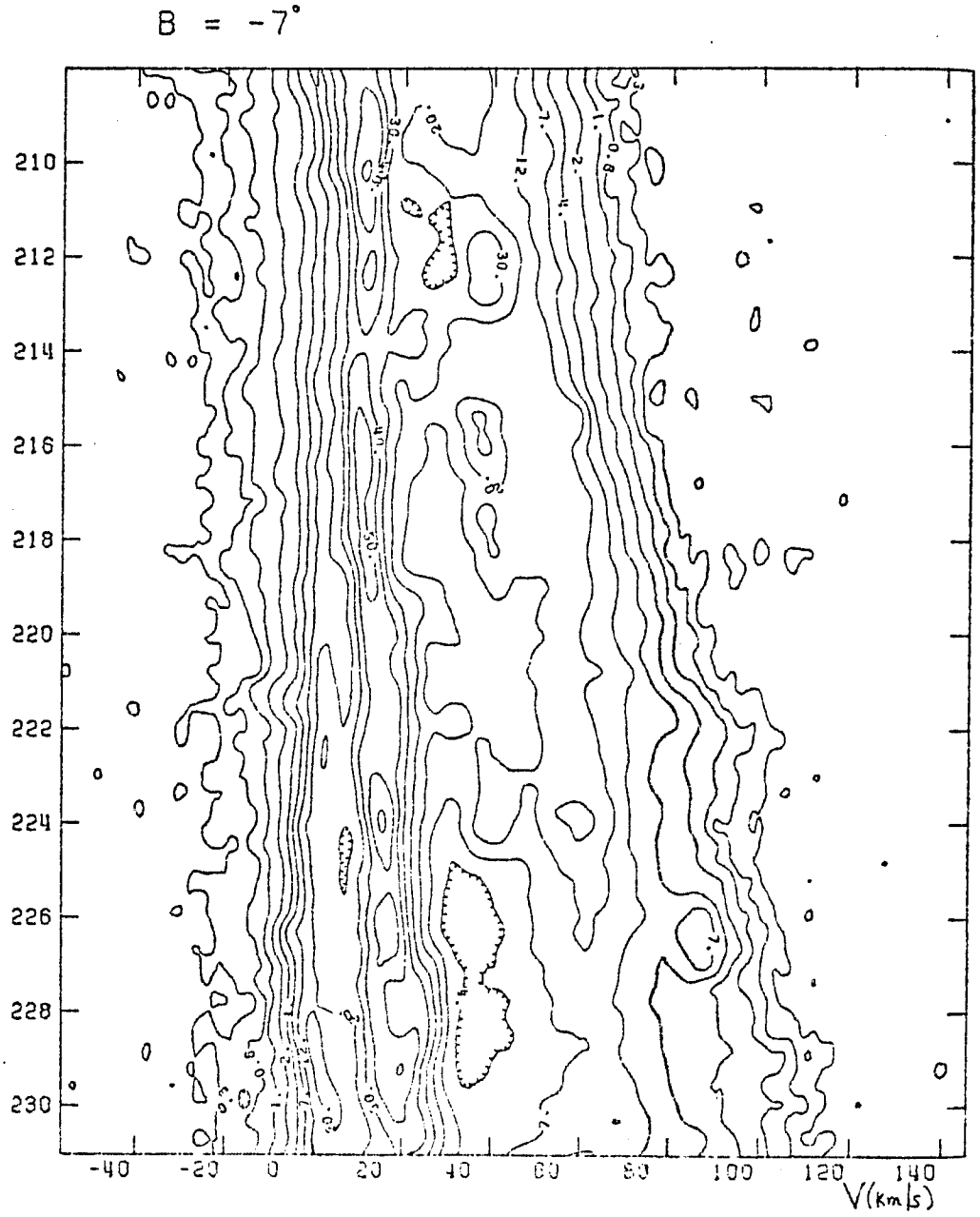

B = -7°
210
212
214
216
218
220
222
224
226
228
230
-40
-20
0
20
40
60
80
100
120
140
V(km/s)

FIGURE 34

of recognizing coherent features that are distributed continuously over some range in latitude, longitude and velocity. If concentrations of gas seen at high latitudes are related to the spiral features in the plane, they should be traceable in both velocity and latitude, and form a smooth distribution in longitude as well.

Although a contour map may be deceiving because of the smoothing and interpolation between observed positions and because information is lost between the drawn contour levels, the continuous nature of the extensions seen in the maps is confirmed by the emission profiles themselves (i.e., antenna temperature versus frequency or velocity at fixed longitude and latitude).

The contour maps at constant longitude show an unambiguous connection between the features in the plane and away from it; the emission varies smoothly over a large latitude range. In addition, the concentrations at intermediate latitudes do not appear isolated in velocity space; their radial velocities can be traced smoothly down to those of the features in the plane.

There is some very weak emission, seen at all latitudes with large (negative) velocities. At many longitudes it extends vertically in b like an envelope of gas. The main features show no real tendency to merge with this weak, ubiquitous emission, which is possibly radiation detected in the sidelobes (see section 2.2).

Figures 31 - 34 show the continuity of these extensions in longitude; a comparison with longitude-velocity maps at $b = 0°$

(Burton 1970; Westerhout 1973) shows that the features away from the plane have similar velocities and similar longitude ranges as the features in the plane.

All high-z extensions show certain characteristics:

(i) The "wings" of the features extend normally to latitudes of 15°. In those cases where they reach as high as 20° or 21°, the emission there is very weak; very little gas is up there. Whether there is more gas there than would be normally expected from the increased density in the plane, will be discussed in section 2.5(d).

(ii) The apparent "rolling motions" of the spiral arms (i.e., the velocity gradient with respect to b across the arm [see Yuan and Wallace 1973]) can be seen near the plane, but in the intermediate positive latitude gas the velocity gradient is greatly reduced. This is difficult to interpret: there is no way to distinguish an increase in the velocity parallel to the plane with $|z|$ from that of the velocity normal to the plane with $|z|$; both would result in the same observed line-of-sight velocity. Moreover, the bending of the plane at large R complicates the velocity field. (This nonlinear effect is not studied in this thesis; we are really interested in [i], why the high-z gas is seen there at all.)

(iii) There is a basic asymmetry with respect to the galactic plane. In the first and second quadrants, the spiral arm

wings extend to much larger positive than negative latitudes. This asymmetry persists even when the galactic warp is taken into account; the gas extends further above the latitude of maximum emission than below.

Our observations in the third quadrant do not cover sufficient latitude range to see if the opposite is true, but observations by others (e.g., Weaver and Williams 1974) show that it is not. There is still more emission above the maximum, even when that maximum is below $b = 0°$. The asymmetry is not as strong as in the first and second quadrants, but the explanation cannot be merely the existence of the galactic warp.

(b) The heights of the extensions

To transform galactic latitude, b, into distance from the plane, z, the distance from the observer to the emitting region must be known. Unfortunately, there is no general agreement on a spiral pattern for the Galaxy. The present calculations adopted the spiral pattern shown in figure 35 (Yuan and Wallace 1973).

Using this pattern we can derive the maximum distances from the plane, $\pm z_m$, to which we can trace each arm. These quantities are summarized in tables 2 and 3, which also contain, for each longitude, the velocity and latitude of the maximum emission. In addition, for each of the regions observed with close spacing in longitude, an average height $\langle z_m \rangle$ has been calculated.

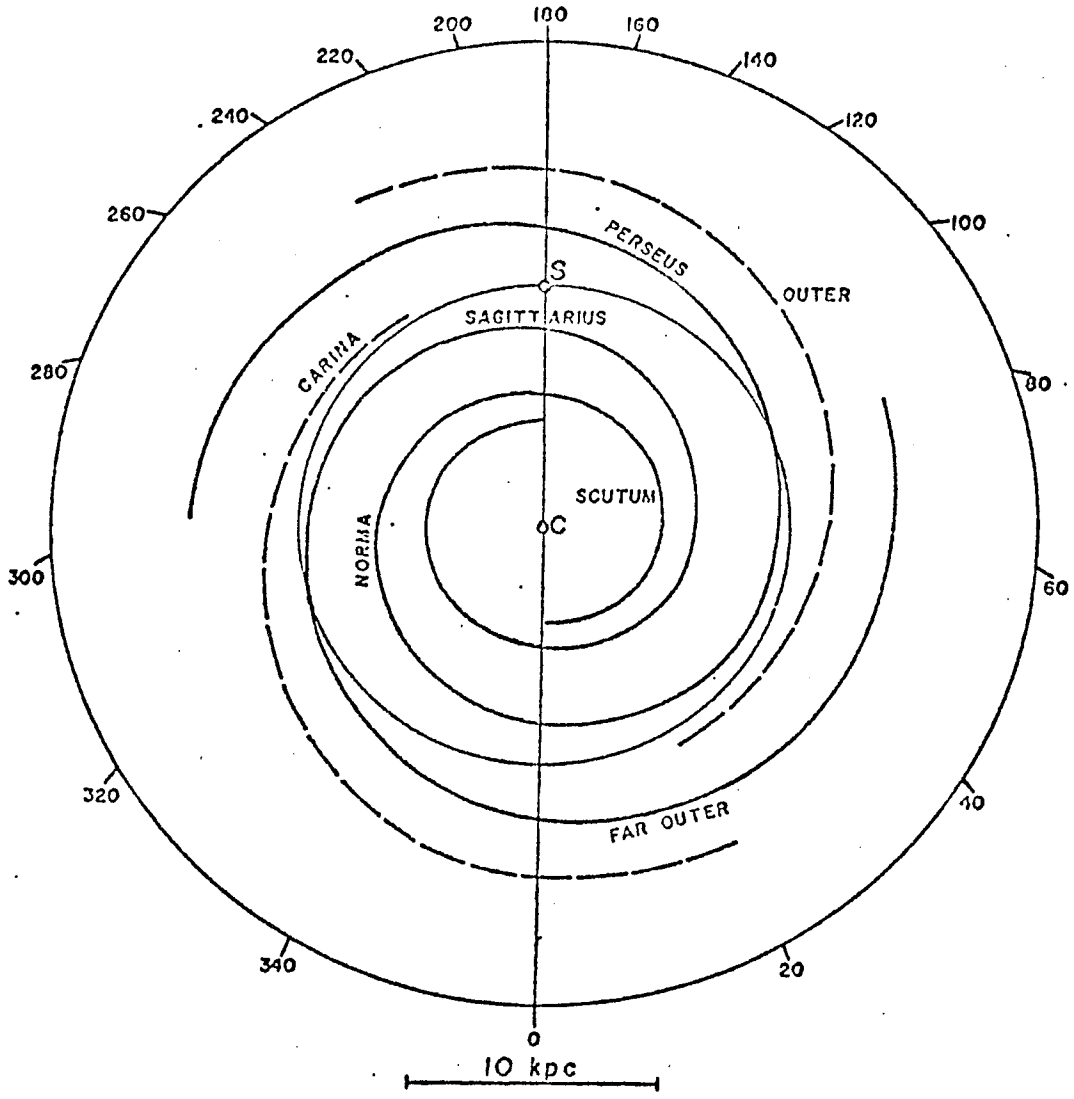


FIGURE 35: Spiral arm pattern adopted for the study of the high-z extensions (from Yuan and Wallace 1973)

TABLE 2a

OBSERVED HIGH-Z EXTENSIONS IN THE PERSEUS ARM

ℓ	V_o (km/sec)	R (kpc)	r (kpc)	b_o	$z_m - z_o$ (kpc)
140°	-40	11.9	2.4	0.9°	0.5 -0.3
147	-33	11.9	2.2	-0.5	0.6 -0.2
148	-31	11.9	2.2	1.0	0.6 -0.3
149	-31	11.9	2.2	0.5	0.7 -0.5
152	-29	11.9	2.1	0.2	0.7 -0.4
153	-30	11.9	2.1	0.2	0.7 -0.4
154	-33	11.9	2.1	1.1	0.7 -0.4
155	-31	11.9	2.1	1.0	0.7 -0.4
156	-31	11.9	2.1	0.5	0.7 -0.5
157	-29	11.9	2.0	0.0	0.7
225	67	12.3	3.0	-2.7	0.7
226	64	12.3	3.0	-1.3	0.4
229	72	12.3	3.2	-2.7	0.4
231	68	12.3	3.2	0.0	0.3
240	68	12.1	3.5	-1.1	0.8 -0.6

ℓ is galactic longitude; R is galactocentric radius; r is distance to the Sun; V_o is velocity of the maximum emission; b_o is latitude of the maximum emission; $z_m - z_o$ is maximum height to which the arms can be traced (taking b_o into account).

TABLE 2b

OBSERVED HIGH-Z EXTENSIONS IN THE OUTER AND FAR OUTER ARMS

ℓ	V_o (km/sec)	R (kpc)	r (kpc)	b_o	$z_m - z_o$ (kpc)
60°	-59	12.0	13.3	0.7°	3.5 -2.3
80	-65	13.5	11.0	0.7	1.5 -1.6
100	-90	14.0	8.2	2.3	2.9 -1.3
118	-93	14.2	6.4	1.3	1.8 -0.9
119	-96	14.2	6.3	1.7	1.9 -0.7
120	-98	14.2	6.3	2.1	2.1 -0.8
140	-77	14.2	5.0	0.8	1.7 -0.8
147	-56	14.2	4.7	2.0	-0.6
148	-55	14.2	4.7	0.4	-0.4
149	-53	14.2	4.7	1.0	-0.7
30	-35	13.0	20.7	1.0	2.5 -1.2
80	-113	14.1	16.3	3.8	2.7

ℓ: galactic longitude

R: galactocentric radius

r: distance to the Sun

V_o: velocity of the maximum emission

b_o: latitude of the maximum emission

$z_m - z_o$: maximum height to which the arms can be traced (taking b_o into account)

TABLE 3

AVERAGE HEIGHTS OF
OBSERVED HIGH-Z EXTENSIONS

Longitude	Average R (kpc)	Average Height (kpc)
118° - 120°	14.2	1.9
		-0.8
147° - 149°	11.9	0.6
		-0.3
147° - 149°	14.2	-0.5
152° - 157°	11.9	0.7
		-0.4
225° - 231°	12.3	0.5

Source: Tables 2a, 2b.

Calculation of the heights of the arms ignored those features that (a) cannot be related unambiguously to an arm in the spiral pattern of figure 35, or (b) cannot be traced from the plane to intermediate latitudes. For example, at l = 100°, where the Perseus arm has two branches with uncertain distances and a great deal of blending of components, only the Outer arm feature was considered. At some longitudes, notably l = 227°, 228° and 230°, the arms extended below the latitudes observed. At others, such as l = 60°, it was not possible to follow the feature more than two degrees below the plane.

The heights of the arms at different positions in the Galaxy vary widely, from 0.6 kpc at l = 148° to 3.5 kpc at l = 60°. There is a pattern (with a great deal of scatter) of height increasing with galactic radius, as the plot of $z_m - z_o$ versus R in figure 36 shows clearly for positive latitude extensions (notwithstanding the apparent anomaly at R = 12.0). There is no similar trend for extensions below the plane, but we do not expect it because of the marked asymmetry.

(c) Comparison with Kepner's results

A comparison of our results with the observations by Kepner (1970) is shown in table 4. For the sake of the comparison, the latitudes were reconverted into heights using her (different) spiral pattern.

Many components that Kepner's wide grid spacing in latitude showed as related, our finer latitude spacing revealed as se-

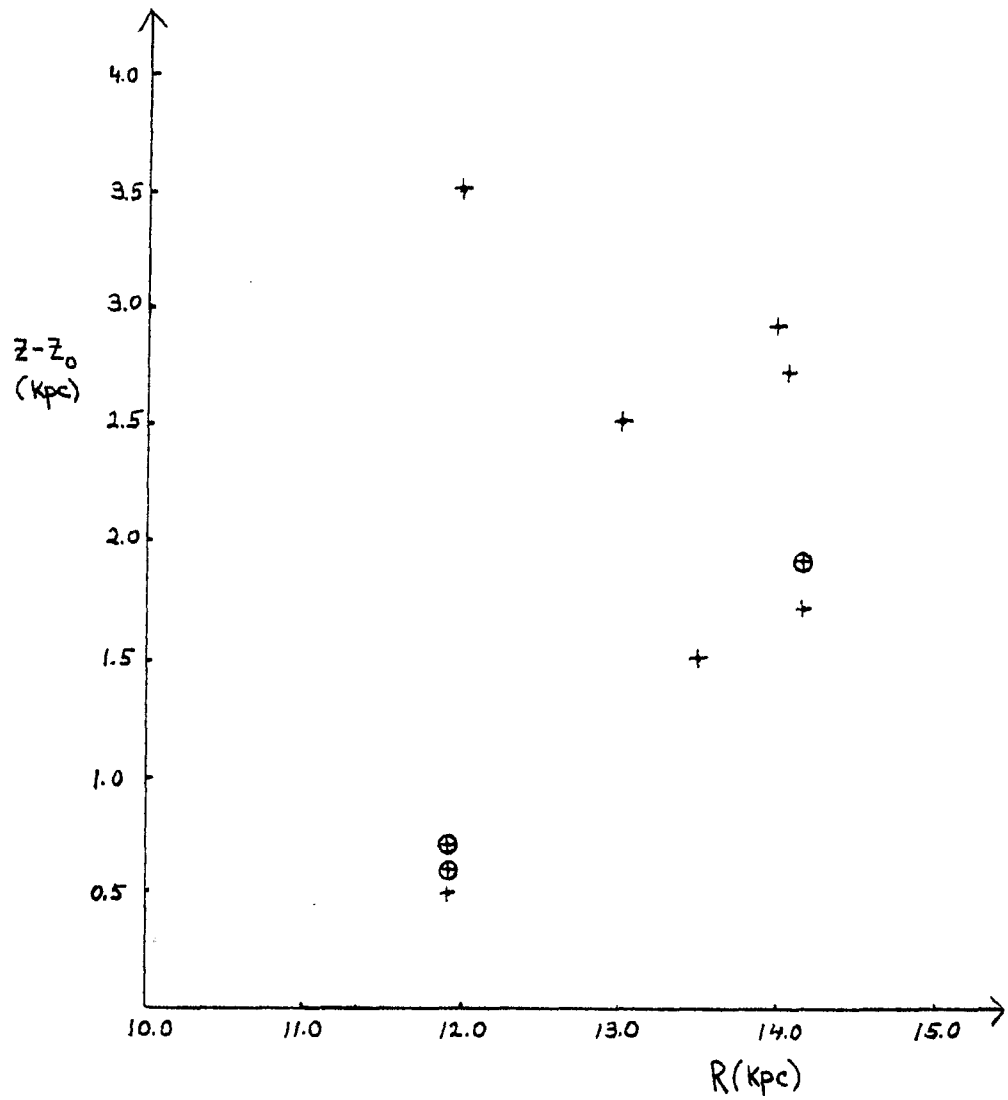


FIGURE 36: Heights of spiral arms above the maximum emission versus galactocentric radius R. Circled points are the average heights of the three closely sampled regions.

TABLE 4

COMPARISON OF OBSERVED HEIGHTS OF HIGH-Z EXTENSIONS

Observer	Longitude	b_o	z_o (kpc)	$z_m - z_o$ (kpc)
Kepner	60°	2.0°	0.5	3.2
Burton-Soukup		0.7	0.2	4.0
Kepner	80	2.0	0.4	3.4
Burton-Soukup		0.7	0.1	1.5
Kepner	80	0.5	1.2	3.4
Burton-Soukup		3.8	0.9	2.2
Kepner	140	0	0	1.0
Burton-Soukup		0.9	0.04	0.7
Kepner	140	2.0	0.2	1.4
Burton-Soukup		0.8	0.1	2.2
Kepner	152	0	0	1.5
Burton-Soukup		0.5	0.04	1.3

b_o is the latitude of the arm emission maximum.

z_o is the height above the galactic plane of b_o.

$z_m - z_o$ is the maximum distance to which the arm can be traced.

parate. For example: at l = 152° Kepner sees a wing at an average velocity of -42 km/sec, and associates it with a distance of 4.4 kpc from the Sun. Our observations make it clear, however, that this wing is associated with the near feature in the plane at V = -30 km/sec, r = 2.1 kpc.

Our higher velocity and spatial resolution did not produce profiles drastically different from Kepner's; thus the number of components identified by her Gaussian analysis was essentially correct. However, we could not trace the spiral arms to the low levels of emission that Kepner did, before they blended with other emission. Therefore, table 4 compares the maximum heights of the features without any reference to intensity. The comparison can be only qualitative, since Kepner's latitude resolution was only 2° and her velocity resolution only 10.55 km/sec. But it does show a basic agreement on the existence of the same features at about the same velocities.

(d) The z-distribution of the emission

The typical trend of the gas distribution with respect to height above the galactic plane is shown for three longitudes in figures 37, 38 and 39. The antenna temperature (as a fraction of the maximum emission) is plotted versus $z - z_o$, where z_o is the height of the maximum temperature. Only the distributions for positive z are analyzed.

Two theoretical curves as well as the observational data are drawn in each figure on the same set of axes. One is the Gaus-

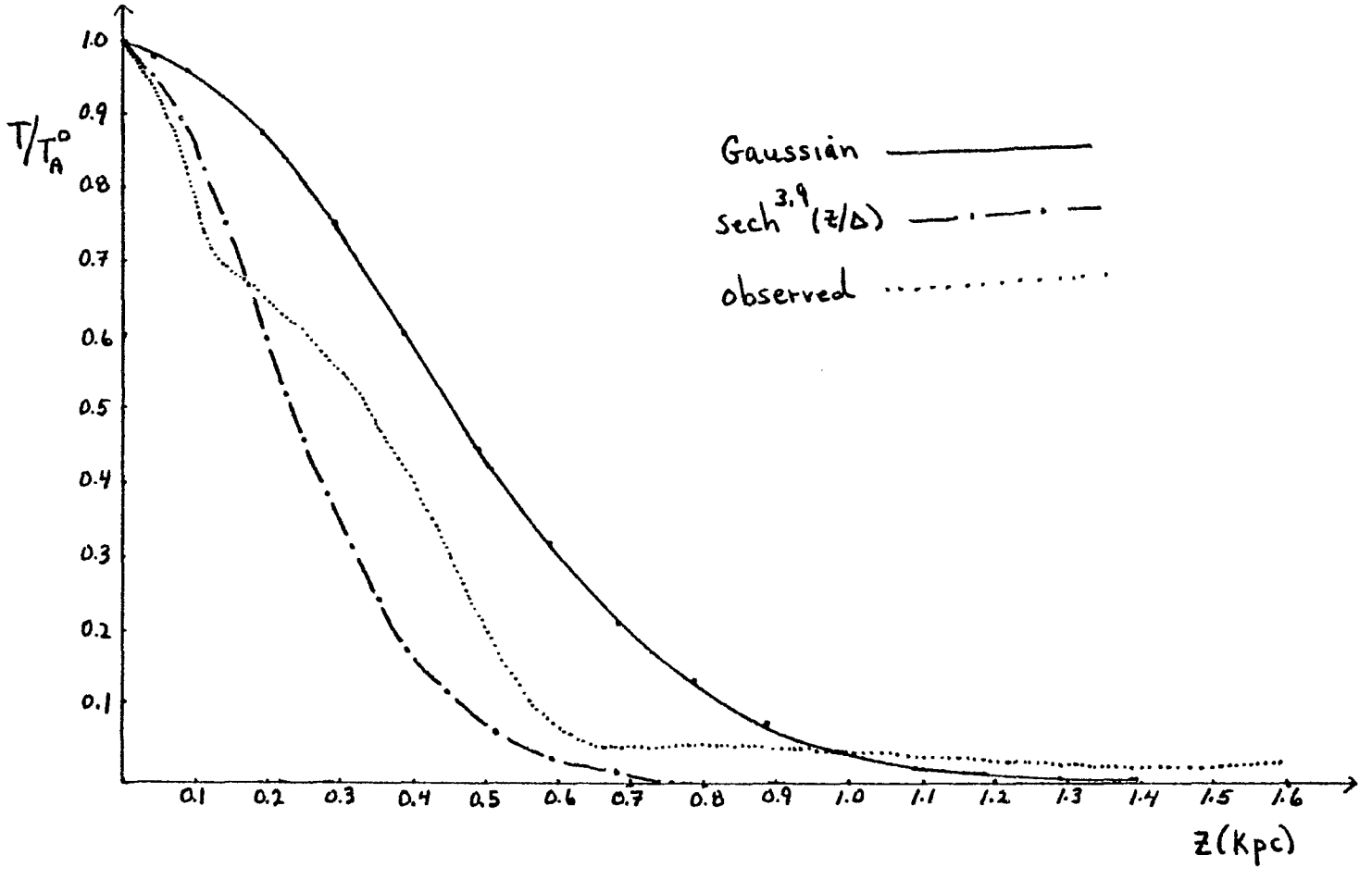


FIGURE 37: 1 = 100°.

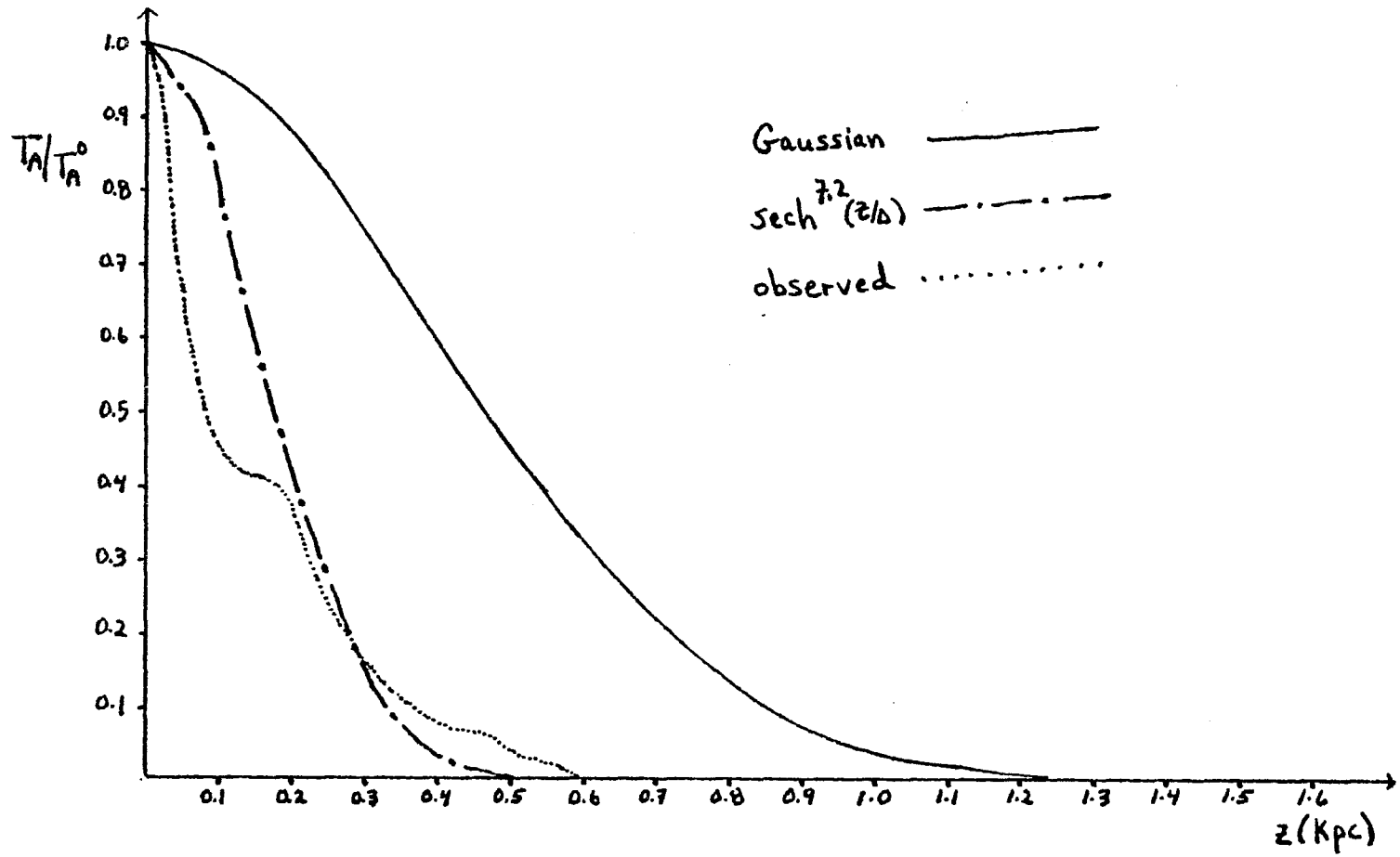


FIGURE 38: 1 = 148°.

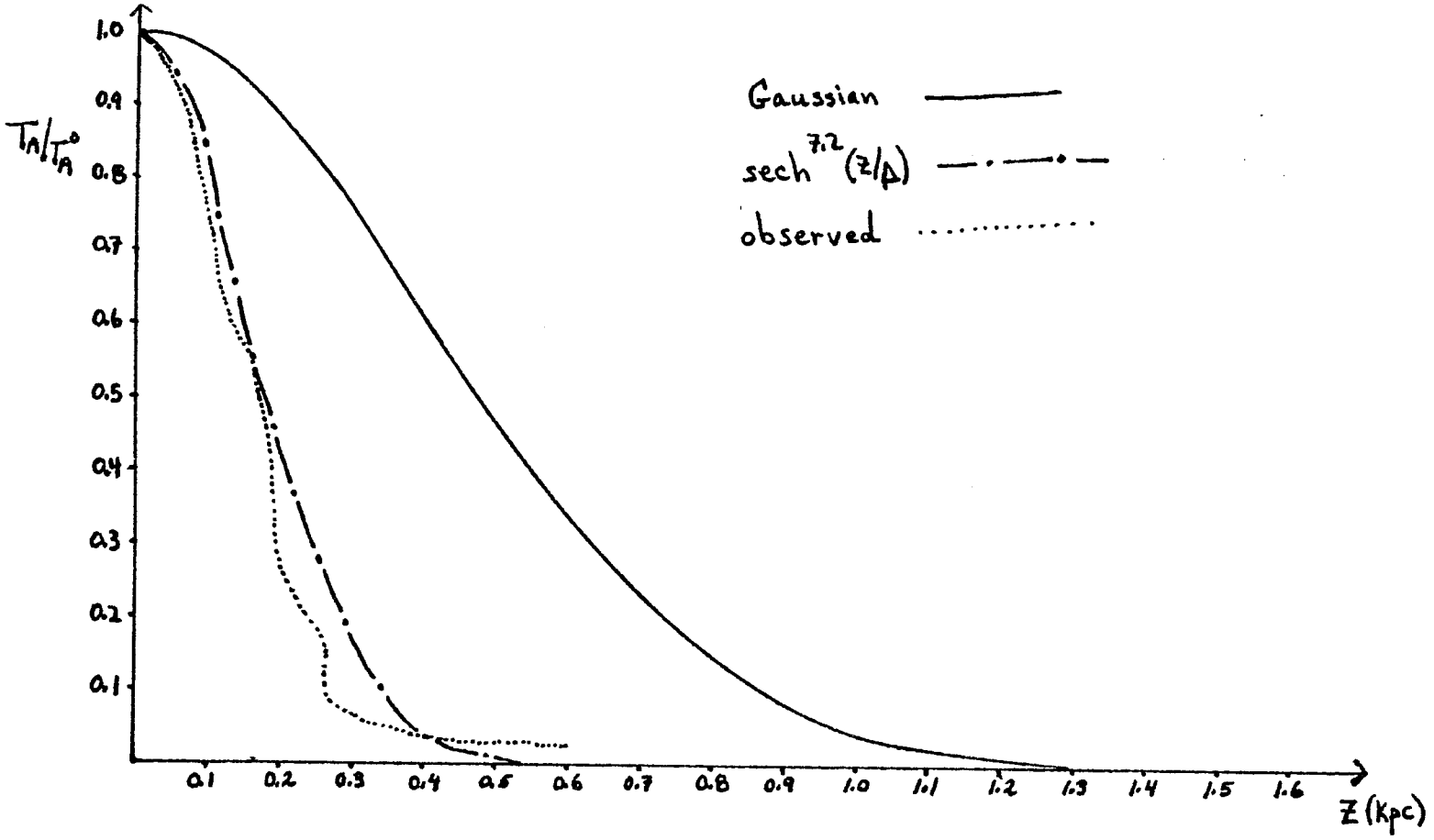


FIGURE 39: 1 = 153°.

sian distribution, $\exp(-z^2/2\Delta^2)$, and the other is $\mathrm{sech}^{\beta_o^2}(z/\Delta)$. The value of Δ is 400 pc for both, and the value of β_o^2 depends on R and on the sonic speed a chosen, namely,

$$\beta_o^2 = \frac{(2\pi G \sigma_*)}{a^2}$$

where $\sigma_*(R)$ is the stellar surface density. The values of σ_* are taken from Vandervoort (1970a).

The hydrostatic solution for the density distribution, $\mathrm{sech}^{\beta_o^2} z/\Delta$, is discussed in Chapter 5. The Gaussian curve is often used as a rough approximation to the density drop-off. The figures show that the $\mathrm{sech}^{\beta_o^2} z/\Delta$ is a much better fit to the observational data than the Gaussian, although there seems to be more gas at high z than either curve predicts. Chapter 7 will compare the hydrostatic and dynamic predictions for the density distributions using two thermodynamic models of the gas.

CHAPTER 3: THE MODEL OF THE GAS FLOW

3.1 Introduction

The problem under consideration is the nonlinear three-dimensional flow of neutral hydrogen gas in the Galaxy, with emphasis on the vertical structure of that flow.

We neglect the self-gravity of the gas and assume that the gravitational potential seen by the gas is produced solely by the stars, and that it is not significantly changed by the resultant redistribution of the gas. Our problem is to calculate the gaseous response to a given stellar gravitational potential.

The fundamental equations of motion for the gas flow written in cylindrical coordinates (r, θ, z) are:

the equation of continuity (conservation of mass)

$$\frac{\partial \rho}{\partial t} + \frac{1}{r}\frac{\partial (r V_r \rho)}{\partial r} + \frac{1}{r}\frac{\partial (V_\theta \rho)}{\partial \theta} + \frac{\partial (V_z \rho)}{\partial z} = 0 \qquad (3.1a)$$

the three equations of motion (conservation of momentum)

$$\frac{\partial V_r}{\partial t} + V_r \frac{\partial V_r}{\partial r} + \frac{V_\theta}{r}\frac{\partial V_r}{\partial \theta} + V_z \frac{\partial V_r}{\partial z} - \frac{V_\theta^2}{r}$$

$$= -\frac{1}{\rho}\frac{\partial P}{\partial r} - \frac{\partial \mathcal{V}}{\partial r} + \text{viscous terms} \qquad (3.1b)$$

$$\frac{\partial V_\theta}{\partial t} + V_r \frac{\partial V_\theta}{\partial r} + \frac{V_\theta}{r} \frac{\partial V_\theta}{\partial \theta} + V_z \frac{\partial V_\theta}{\partial z} + \frac{V_r V_\theta}{r}$$

$$= -\frac{1}{\rho r} \frac{\partial P}{\partial \theta} - \frac{1}{r} \frac{\partial \mathcal{V}}{\partial \theta} + \text{viscous terms} \qquad (3.1c)$$

$$\frac{\partial V_z}{\partial t} + V_r \frac{\partial V_z}{\partial r} + \frac{V_\theta}{r} \frac{\partial V_z}{\partial \theta} + V_z \frac{\partial V_z}{\partial z}$$

$$= -\frac{1}{\rho} \frac{\partial P}{\partial z} - \frac{\partial \mathcal{V}}{\partial z} + \text{viscous terms} \qquad (3.1d)$$

the energy equation (conservation of entropy)

$$\frac{\partial S}{\partial t} + \frac{V_r}{r} \frac{\partial (rS)}{\partial r} + \frac{V_\theta}{r} \frac{\partial S}{\partial \theta} + V_z \frac{\partial S}{\partial z}$$

$$= \text{heating and cooling terms} \qquad (3.1e)$$

and the equation of state

$$P = P(\rho, S) \qquad (3.1f)$$

where t denotes the time, $\rho(r,\theta,z,t)$ is the gas density, (v_r, v_θ, v_z) are the velocities, $\mathcal{V}(r,\theta,z,t)$ is the total stellar gravitational potential, $P(r,\theta,z,t)$ is the pressure, and S is the entropy per unit mass. These equations are derived in any fluid dynamics textbook (e.g., Landau and Lifshitz 1959).

To simulate the physical flow with these equations accurately, assumptions made about each quantity must be clearly understood.

Assuming that the gas is inviscid, we can omit the viscous forces in the momentum equations and go on to consider the gravitational forces on the system.

3.2 The Gravitational Potential

The gas flow in the Galaxy is governed primarily by the smoothed axisymmetric gravitational forces of the stars. The basic equilibrium state of the gas is one of pure circular rotation in which the centrifugal forces are exactly balanced by this axisymmetric gravitational force field. The gas density of this base state is $\rho_0(r,z)$ and the gravitational potential is $\mathcal{V}_0(r,z)$. For circular flow, the velocities are $(v_r, v_\theta, v_z) = (0, v_{\theta_0}[r,z], 0)$, and the balance of forces may be written as

$$\frac{v_{\theta_0}^2}{r} = \frac{\partial \mathcal{V}_0(r,z)}{\partial r} \quad . \tag{3.2}$$

The non-axisymmetric portion of the gravitational field is a two-armed spiral perturbation $\mathcal{V}_1$, superimposed on the axisymmetric potential $\mathcal{V}_0$, so we may write

$$\mathcal{V}(r,\theta,z,t) = \mathcal{V}_0(r,z) + \mathcal{V}_1(r,\theta,z,t) \quad . \tag{3.3}$$

The forms for both portions of $\mathcal{V}$ are taken from Vandervoort (1970b), where they are given in a particularly convenient mathematical form. For the axisymmetric part we have

$$\mathcal{V}_0(r,z) = \mathcal{V}_\infty(r,o) + (2\pi G\sigma_*)\,\Delta \ln(\cosh[z/\Delta]) \tag{3.4}$$

where $\mathcal{V}_\infty(r,o)$ is the potential due to an infinitesimally thin stellar disk of surface density σ_*, and the second term is due to a finite disk of scale height Δ.

The spiral portion is

$$\mathcal{V}_1(r,\theta,z,t) = A(r)\,\mathrm{sech}^{|k|\Delta}(z/\Delta)\cos(\omega t - 2\theta - \bar{\Phi}(r)) \qquad (3.5)$$

where A(r) is the slowly varying amplitude and $\bar{\Phi}(r)$ is the phase term that specifies the shape of the spiral pattern. The wave number $|k| \equiv \left|\frac{d\bar{\Phi}(r)}{dr}\right|$ determines the local radial spacing between neighboring spiral arms. We see that the spiral pattern rotates around the galactic center with a constant angular velocity $\omega/2$.

In the plane of the Galaxy, z = 0, this potential reduces to the tightly wound two-armed spiral potential used in the shock calculations of Roberts (1969). Section 3.4 shows that we do not need to know the form of $\mathcal{V}_\infty(r,o)$.

3.3 Reduction of the Spatial Dimensions

The gas dynamical equations (3.1) are time-dependent, three-dimensional equations, and very complicated to solve. To simplify them, we follow Roberts (1969) for the two-dimensional flow in the plane of the galaxy. We assume $v_z = 0$ and z = 0 in equations (3.1), and consider ρ to be a surface density σ.

(a) Flow in the galactic plane

Since the spiral pattern rotates as a rigid body with

angular velocity $\omega/2$, we transform into a rotating coordinate system $(r, \psi = \theta - [\omega/2]t)$. In this frame, the potential is now stationary and all velocities are relative to the spiral pattern.

In addition, we change from (r,ψ) to an orthogonal, curvilinear set of coordinates (ξ,η), fixed in the rotating frame, that are parallel and perpendicular to the spiral equipotential curves defined by $2\psi + \bar{\Phi}(r) = \text{constant}$. The coordinate ξ is along the spiral arm and η is perpendicular to it (see figure 40).

In the specific case of a two-armed logarithmic spiral pattern with a constant inclination angle i, we have

$$\bar{\Phi}(r) = \frac{-2}{\tan i} \ln\left(\frac{r}{r_o}\right) \tag{3.6}$$

where r_o is a constant.

Since

$$\xi = -\ln\left(\frac{r}{r_o}\right)\sin i + (\theta - \Omega_p t)\cos i \tag{3.7a}$$

and

$$\eta = \ln\left(\frac{r}{r_o}\right)\cos i + (\theta - \Omega_p t)\sin i \quad , \tag{3.7b}$$

the expression for the spiral gravitational potential becomes

$$\mathcal{V}_1 = A(\xi,\eta)\cos\left(\frac{-2}{\sin i}\eta + \beta\right) \quad . \tag{3.8}$$

The gravitational potential is thus seen to be oscillatory in η.

All variables are now written as sums of an axisymmetric contribution and a spiral perturbation, denoted by subscripts o and 1:

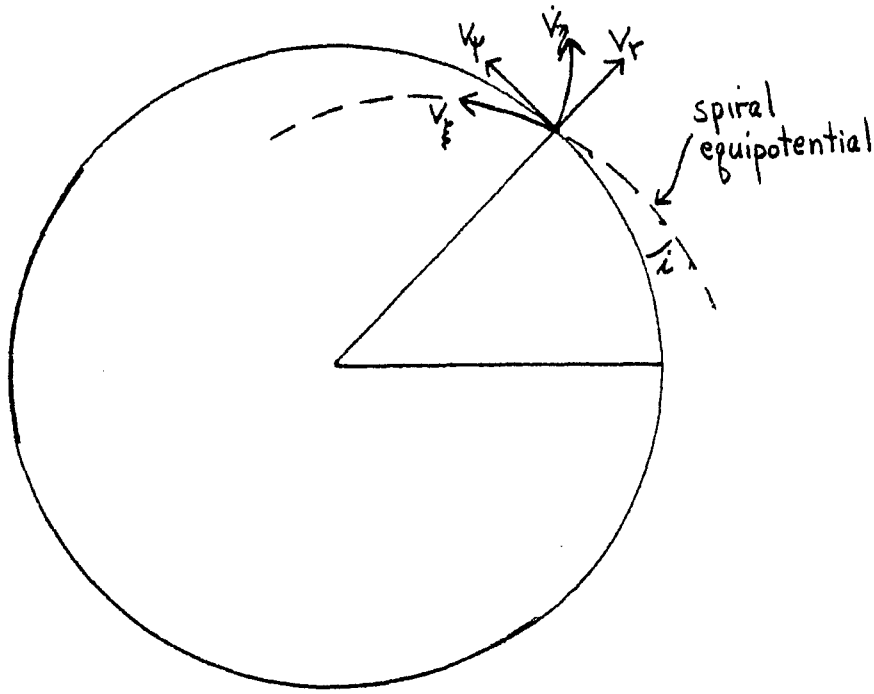


FIGURE 40: The (r,ψ) and (ξ,η) coordinate systems for the galactic disk. ξ and η run parallel and perpendicular to the spiral equipotential curves.

$$\sigma = \sigma_0 + \sigma_1$$
$$V_\xi = V_{\xi_0} + V_{\xi_1} \qquad (3.9)$$
$$V_\eta = V_{\eta_0} + V_{\eta_1}$$

In the asymptotic approximation of a tightly wound spiral, $A(\xi,\eta)$ is a slowly varying function, and therefore we expect the perturbation quantities to vary slowly with respect to ξ, but rapidly with respect to η. In other words, for tightly wound spirals the gradients in the ξ direction can be neglected compared to the gradients in the η direction.

Thus the steady-state equations for the perturbation quantities (neglecting terms of 2nd order or higher) are:

$$\frac{\partial V_{\eta_1}}{\partial \eta} = \frac{-(V_{\eta_0}+V_{\eta_1})(2\Omega r V_{\xi_1} + rf)}{a^2 - (V_{\eta_0}+V_{\eta_1})^2} \qquad (3.10a)$$

$$\frac{\partial V_{\xi_1}}{\partial \eta} = \frac{-\left(\frac{\kappa^2}{2\Omega}\right) r V_{\eta_1}}{V_{\eta_0} + V_{\eta_1}} \qquad (3.10b)$$

$$(V_{\eta_0} + V_{\eta_1}) \frac{\partial \sigma_1}{\partial \eta} + (\sigma_0 + \sigma_1) \frac{\partial V_{\eta_1}}{\partial \eta} = 0 \qquad (3.10c)$$

where $\frac{\kappa^2}{2\Omega}$ and Ω are functions of r that depend on the rotation curve, and f is the perturbation force in the η direction, defined as

$$f = \frac{-1}{r}\frac{\partial V_1}{\partial \eta} = \frac{2A}{r \sin i}\sin\left(\frac{-2}{\sin i}\eta + \beta\right). \qquad (3.11)$$

(Some extra terms of secondary importance were retained in the calculation, but are left out of these expressions.)

To solve these equations, we must integrate numerically along only one spatial coordinate, η. The well-known solutions to these equations are narrow, nearly concentric streamlines, containing two periodically located shock waves which lie just before the minima of the spiral gravitational potential (figures 1 and 2). The gas leaving an arm is subsonic, speeds up through the sonic point to become supersonic, and returns to subsonic speeds via a shock in the next spiral arm.

Equations (3.10a) and (3.10b) are coupled equations for v_{ξ_1} and v_{η_1}. Once we solve them for the velocities, we can determine the surface density σ_1 by solving (3.10c). However, (3.10a) and (3.10b) are coupled only through the Coriolis force $2\Omega v_{\xi_1}$ in the numerator.

It is instructive to write equation (3.10a) as

$$\frac{\partial v_{\eta_1}}{\partial \eta} = \frac{-(v_{\eta_o} + v_{\eta_1})\, rf'}{a^2 - (v_{\eta_o} + v_{\eta_1})^2} \qquad (3.12)$$

where $f' = (2\Omega v_{\xi_1} + f)$ is the effective force seen by the gas, which includes the Coriolis force. This effective force is still periodic but no longer sinusoidal.

Now the coupling between equations (3.10a) and (3.10b) is hidden in f', but it is easier to see what role f' plays in de-

termining the position of the sonic point on the streamline. When $v_{\eta_0} + v_{\eta_1} = \pm a$, the denominator of equation (3.12) vanishes; so that $\frac{\partial v_{\eta_1}}{\partial \eta}$ would be finite there, the numerator must also vanish. Thus, the gas can pass continuously through a sonic point only where $f' = 2\Omega v_{\xi_1} + f = 0$. (The role of f' in equation (3.12) is explored more thoroughly in Chapter 6, which shows how this problem corresponds to the flow in a nozzle.)

(b) Adding the z-dimension

Being able to treat the flow in the plane of the Galaxy as a one-dimensional flow greatly simplifies our problem. We now add the z-dimension to this flow to get a set of two-dimensional time-dependent equations in η and z.

The main effect of the Coriolis force is to determine the exact position of the sonic point, and therefore the position of the shock along the streamline in the plane. Since we are interested primarily in the vertical extent and structure of the shock, which occupies a relatively narrow region of about 50 pc along that streamline (Shu et al. 1972), we can omit the Coriolis force and formulate the problem in rectangular Cartesian coordinates. This will reproduce all the important physics of the problem, as shown below.

We take as our gravitational potential the original expression described in section 3.2, equations (3.4) and (3.5), but now we omit $V_\infty(r,o)$, since we have eliminated the base state of circular rotation. The streamline (η) direction will now be

called x, and the vertical direction z. The spiral gravitational perturbation is now written with a phase that depends on x rather than η, and a coefficient, now a constant, called N_o:

$$\mathcal{V}(x,z) = (2\pi G\sigma_*)\Delta \ln\left(\cosh(z/\Delta)\right) + N_o \operatorname{sech}^{1/2}(z/\Delta)(1-\cos\alpha x). \quad (3.13)$$

The spiral portion of this potential, $N_o(1 - \cos\alpha x)$, is drawn in figure 41a.

The true potential seen by the gas in one-dimensional spiral flow (i.e., with the Coriolis force included) is shown in figure 41b. This potential, $\mathcal{V}_{eff}$, was calculated from the effective force f' in equation (3.12); knowing the streamlines and the values of the variables along them, f' was integrated numerically to get the effective potential -- the solid line in figure 41b. It is <u>not</u> periodic even though the <u>shape</u> is repeated (i.e., the curve from A to B is the same as from C to D). We can make it periodic by changing $\mathcal{V}_{eff}$ by a constant at the shock location, as shown by the dashed line in the figure. This may be done because the potential is one term in the Bernoulli quantity, which is constant along a streamline, for continuous flow:

$$\mathcal{V}_{eff} + \frac{1}{2}v^2 + \int\frac{dP}{\rho} = \text{constant} \quad , \qquad (3.14)$$

where v is the velocity. This constant changes across a shock because of the discontinuity in velocity.

Figure 41 shows that omitting the Coriolis force changes some details of the potential, but not its general shape, and

FIGURE 41a

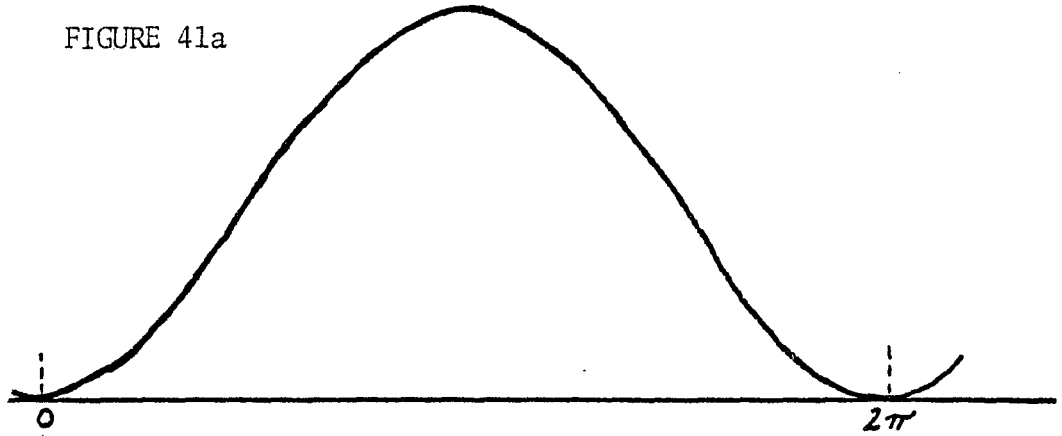


FIGURE 41b

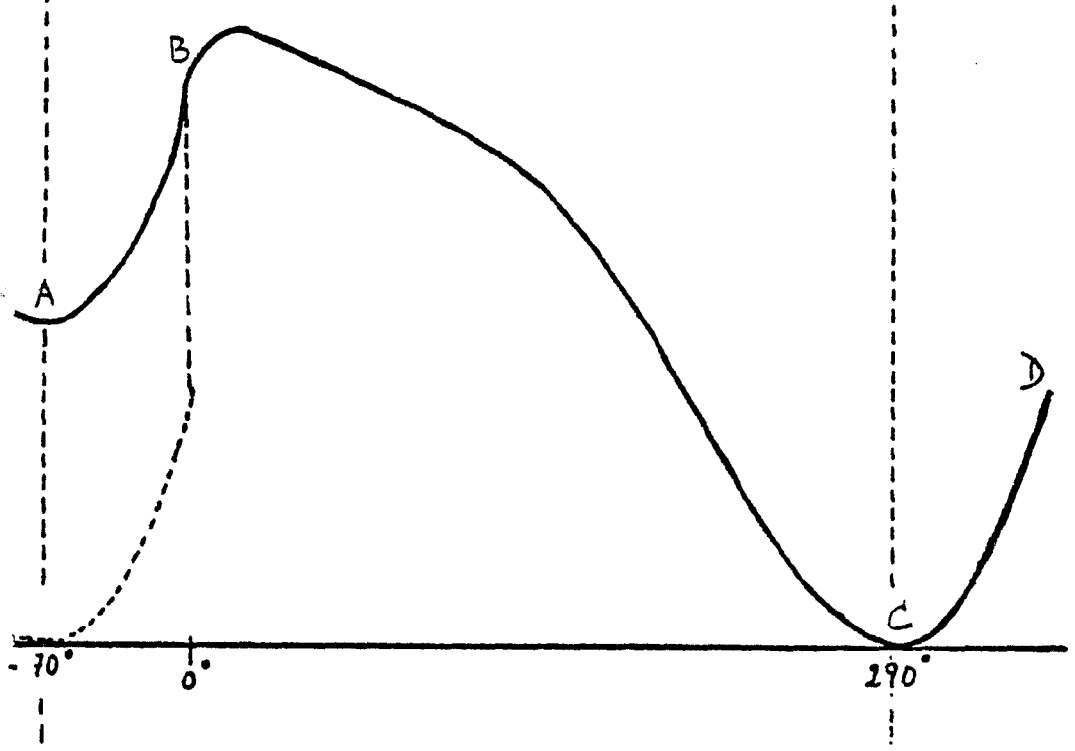


FIGURE 41a: The spiral portion of the gravitational potential, $N_o(1 - \cos \alpha x)$, as a function of the phase αx.

FIGURE 41b: The effective potential (solid line) seen by the gas moving along a streamline, as a function of the phase. The curve from A to B is the same as from C to D. The dashed line at the shock position B shows how the potential is made periodic by the change in the Bernoulli constant.

therefore not the vertical structure of the shock.

Although the problem is usually treated as one of the gas flowing into a potential well (the spiral arm), viewing it as flowing into a potential bump (i.e., out of the spiral arm into the interarm region) makes the problem analogous to the physical nozzle flow, discussed in Chapter 6, which elucidates the problem of shock formation in supersonic flow.

For this simplified model, the governing equations for the time-dependent flow are:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{3.15a}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} = -\frac{1}{\rho}\frac{\partial P}{\partial x} - \frac{\partial V}{\partial x} \tag{3.15b}$$

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + w\frac{\partial w}{\partial z} = -\frac{1}{\rho}\frac{\partial P}{\partial z} - \frac{\partial V}{\partial z} \tag{3.15c}$$

$$\frac{\partial S}{\partial t} + u\frac{\partial S}{\partial x} + w\frac{\partial S}{\partial z} = \text{heating} + \text{cooling terms} \tag{3.15d}$$

$$P = P(\rho, S) \tag{3.15e}$$

where $\rho(x,z,t)$ is the density, $u(x,z,t)$ is the velocity in the x direction, $w(x,z,t)$ is the vertical velocity, S is the entropy per unit mass, P is the pressure and V is the gravitational potential of equation (3.13).

Since we are no longer treating a spiral-type problem, the meaning of α in the expression for V changes. To simulate the coordinate along a streamline, α was chosen to be $2\pi/20$, so that the distance covered by one cycle of the potential would be 20 kpc. The values of σ_* and Δ are taken from Vandervoort (1970a,

1970b); Δ is a constant, 400 pc, and σ_* varies from 110 to 35 $M_\odot/pc^2$, corresponding to values of R from 10 to 14 kpc.

The values of $|k|\Delta$ used vary from 0.5 to 2.0, corresponding to radial distances between arms of 1.2 to 4.8 kpc for Δ = 400 pc. Since the preliminary calculations show that the steady-state flows were not sensitive to the exact value within this range, the results presented in Chapter 7 used $|k|\Delta = 1.0$ (spacing of 2.4 kpc).

The magnitude of the spiral potential is chosen so that the spiral force corresponds to a fixed fraction F of the mean axisymmetric gravitational force in the galactic plane, which is the centrifugal force $v_{\sigma_o}^2/r$. The radial spiral force, from equation (3.5), is $|k|A(r)$ (for slowly varying A(r)). Writing

$$|k|A(r) = F\frac{v_{\sigma_o}^2}{r} \tag{3.16}$$

for values of V_{σ_o} and r at R = 12 kpc and $|k|\Delta = 1.0$, we have

$$A(r) = 1880\ F\ . \tag{3.17}$$

This is the value used for the constant N_o in the spiral potential for the rectangular problem. The fraction F varies from 2% to 7%, but we can expect it to be about 5% (Yuan 1969).

3.4 Physical Properties of the Gas

In order to solve the set of equations (3.15), we must assume physical properties for the neutral hydrogen gas we are modeling.

We have already assumed that the gas is not self-gravitating. We now assume it is compressible. In addition, we may consider turbulence, magnetic fields and cosmic ray particles.

Roberts and Yuan (1970) showed that the magnetic field influences the gas dynamics through the Maxwell stresses, largely in the form of a pressure. The resulting solutions for density and velocity are very similar; in general, the inclusion of magnetic effects makes the gas compression a little smaller, the shocks weaker, and the systematic motions not as large.

Since turbulence and cosmic ray particles also act as a "pressure," we will assume the gas to have a "mean equivalent" sonic speed. This speed represents the combination of turbulence, kinetic velocity, and contributions due to cosmic rays and magnetic fields. Since all these factors increase the overall dispersion speed of the gas, this effective sound speed is larger than the thermal sound speed, thus decreasing the response of the gas to the given potential field. The value of this effective sound speed is taken as 10 km/sec in the galactic plane.

The interstellar medium is often considered to be a "two-phase" medium, consisting of cold, dense gas clouds in rough pressure equilibrium with a hot, rarefied intercloud medium (Field, Goldsmith, Habing 1969; Habing and Goldsmith 1971; Schwarz et al. 1972). The clouds vary in size and temperature, but an average cloud has a density of about 10 H atoms/cm^3 and a temperature of about 70°K. The intercloud medium has a density of about

0.1 H atoms/cm^3 and a temperature of about 10^4°K.

These phases can be identified with the observed cold dense clouds at temperatures 20°K - 150°K and with an unobserved hot, rarefied medium at temperatures of perhaps 10^4°K. The two phases interact by phase transitions (i.e., condensation and evaporation) in order to maintain the pressure equilibrium as the temperature of either phase changes.

Shu et al. (1972) studied the gas flow in the density wave theory based on this two-phase concept, and found that galactic shocks are initiated by the intercloud medium. There are phase transitions between the two phases along the streamlines. However, the dynamical solution for the intercloud flow is very similar to the results of Roberts (1969), discussed above in section 3.3(a), who used a one-component model of the interstellar medium. Roberts' interstellar gas consisted of clouds with velocity dispersion of 10 km/sec, which act as molecules of a "gas" when viewed on a scale much greater than the cloud diameter. The "gas" density was the smeared-out density of the clouds. The results of Shu et al. (1972) refer to the intercloud medium rather than the clouds, and his sonic speed is associated with the gas kinetic temperature rather than with the velocity of random cloud motions. Since the two sonic velocities are numerically very close and the two phases are in pressure equilibrium, the large-scale dynamics for the two models are similar in many respects, especially where no phase transitions occur.

Thus, as long as we limit ourselves to discussion of large-scale phenomena (length scales $\gtrsim 100$ pc) (Shu et al. 1973), we can model the two-phase interstellar medium adequately with a one-phase equation of state $P = P(\rho,S)$ and an "effective sound speed" a, where

$$a^2 \equiv \left(\frac{\partial P}{\partial \rho}\right)_S \cdot \tag{3.18}$$

In this model the pressure and sonic speed have their usual meanings for the intercloud medium.

Last, the choice for the equation of state of the gas must be examined. We have already assumed that the gas has no explicit viscosity, although Chapter 4 shows that the numerical scheme adds an "artificial viscosity" to the equations in order to stabilize them. We will treat the interstellar medium as an ideal gas (sometimes called a perfect gas); that means its equation of state can be written as

$$P = \rho RT \tag{3.19}$$

and its internal energy is a function of temperature alone. If we assume, in addition, that the ratio of specific heats $\gamma \equiv \frac{C_P}{C_V}$ is a constant, it can be shown (Shapiro 1953, p. 43) that the equation of state can be written as

$$P = \text{constant} \cdot \rho^{\gamma} e^{\frac{\gamma S}{C_P}} \tag{3.20}$$

where S is the entropy of a fluid particle. From this, the sonic

speed a is given by

$$a^2 = \text{constant} \cdot \rho^{\gamma-1} e^{\frac{\gamma S}{c_p}} \tag{3.21a}$$

or

$$a^2 = \gamma RT \quad . \tag{3.21b}$$

Comparing equations (3.19) and (3.20) we see that, for an ideal gas with a constant ratio of specific heats, the entropy can be written as

$$S = \text{constant} + \frac{R}{\gamma - 1} \ln\left(\frac{P}{\rho^{\gamma}}\right) \tag{3.22a}$$

or

$$S = S_o + C_v \ln\left(\frac{T}{T_o}\right)\left(\frac{\rho_o}{\rho}\right)^{\gamma - 1} \quad . \tag{3.22b}$$

The choice of the equation of state of the gas now depends on which properties we consider constant or definite. The "variables" in the equations of state (3.19) and (3.20) are γ, S and T. Some typical assumptions are:

(i) Isothermal flow: the temperature is a constant <u>everywhere</u>.

$$P = \text{constant} \cdot \rho \tag{3.23a}$$

$$a^2 = \gamma RT = \text{constant}. \tag{3.23b}$$

If $\gamma = 1$, S is a constant everywhere.

(ii) Isentropic flow: there is no heat conduction or viscosity, so the entropy of each fluid particle remains constant as it moves about. In this case, we have

$$P = C(S) \cdot \rho^{\gamma} \tag{3.24a}$$

where C(S) is a constant of proportionality that depends on S and differs for each particle. If $\gamma = 1$, T is a constant for each fluid particle. Again, we have

$$a^2 = \gamma RT \quad , \tag{3.24b}$$

but now T may differ from particle to particle. The most important isentropic flow is steady-state flow, in which the entropy is constant along the streamlines.

(iii) Homoentropic flow: isentropic flow where the entropy everywhere in the fluid has the same value.

$$P = \text{constant} \cdot \rho^{\gamma} \quad . \tag{3.25a}$$

If $\gamma = 1$, T = constant everywhere. Once more, we have

$$a^2 = \gamma RT \quad . \tag{3.25b}$$

Isentropic or, more often, homoentropic flow can often be an accurate approximation to adiabatic flow, since for the latter $\Delta S \geq 0$, where the upper sign corresponds to an irreversible process and the lower sign to a reversible one (i.e., an adiabatic reversible process is an isentropic one). This approximation is especially valid for the flow we are interested in, since we have assumed an inviscid gas (and therefore reversible processes). Thus, if the gas is modeled as adiabatic, it can be considered isentropic.

Unfortunately, there is no unequivocal answer to the

question, what will most accurately describe the gas? Physical arguments based on dynamical and thermal time scales suggest that, for constant heating and cooling rates, the intercloud medium is nearly isothermal in the plane of the Galaxy. The gas is heated and partially ionized by soft x-rays, low energy cosmic rays, etc. (Silk 1973), and is cooled radiatively from inelastic collisions (Field, Goldsmith and Habing 1969; Spitzer and Scott 1969; Dalgarno and McCray 1972). The thermal time scale is usually less than 10^6 years (Shu et al. 1972; Field 1974b), the dynamical time scale in the galactic plane is about 10^8 years, and the time scale for atomic collisions is 10^2 - 10^3 years (Shu et al. 1972). Thus, the gas has adequate time to adjust its local temperature. If the gas cools below its equilibrium temperature, the ambient heating sources soon warm it up again, and if it gets heated up too much, it can rapidly radiate the extra energy away.

Therefore, virtually all one-phase descriptions of the interstellar medium in the galactic plane consider it an isothermal gas. This is the simplest description as well, since then we have $P = a^2\rho$, with a = constant, and we do not need to consider the energy equation (for S) in our calculations. Assuming that the flux of the heating sources falls off with z at the same rate that the density does, the temperature of the gas is independent of z as well (Shu et al. 1972). Our Model A uses this description of the gas to calculate the flow.

However, another reasonable model for the gas is neither

isothermal nor isentropic, but one in which the temperature of the gas rises as $|z|$ increases. This is not inconsistent with the arguments for an isothermal gas, since the conclusion that the temperature T is independent of z depends critically on how fast the heating sources drop off with distance from the plane.

To look at the heating and cooling of the gas in more detail: The radiative loss rate Λ is due to inelastic binary collisions, and therefore is proportional to n^2 (where n is the number density). The heating rate Γ is proportional to n. If we write

$$\Gamma = Gn \tag{3.26a}$$

$$\Lambda = L(T)n^2 \tag{3.26b}$$

where L is a function of temperature, the equilibrium temperature (away from boundary layers and shock waves) is given by

$$\Gamma = \Lambda$$

or

$$G = nL(T) \quad . \tag{3.27}$$

The function L(T) can be calculated (Dalgarno and McCray 1972) as a function of ionization. The final equilibrium temperature of the gas depends on G/n: on how rapidly the heat sources change with height compared to the decrease in n.

Assuming that the primary sources of ionization are produced mainly within the galactic plane, and letting G scale with the density n, we get an isothermal gas -- Model A. Unfortunately, the distribution of these sources in the Galaxy is very poorly

known. The assumption that their flux falls off with increasing $|z|$ would be incorrect even qualitatively if the sources are extragalactic.

The assumption that G is independent of $|z|$ is consistent with observations, although obviously not unique, and has been used before (Falgarone and Lequeux 1973; Field 1974b). If G is constant as n falls, L(T), and with it T, increase. Actually, for T to increase with $|z|$, we only need G to decrease less rapidly with $|z|$ than the density does. Electron dispersion measures imply that the scale height of the thermal electron disk is larger than that of the neutral gas. If the thermal electrons are produced by ionization due to low-energy radiation, this implies that the scale height of G is larger than that of the HI gas.

Other arguments suggested for the plausibility of a temperature gradient in the gas include.

(i) the stellar scale height is larger than the gaseous scale height, so hot stars at high z will help heat up the gas.

(ii) The buoyancy of the gas will push the hotter gas to higher z. Gas heated in the plane (perhaps by supernovae) will rise.

(iii) The cooling rate ($\propto n^2$) of the very tenuous gas at high z will be low compared to its heating rate ($\propto n$), and thus the gas will be hotter (Field 1974a).

(iv) Observations of OVI absorption lines (York 1974; Jenkins and Meloy 1974; Gorenstein and Tucker 1972) suggest that

a hot coronal-type gas ($T \gtrsim 3 \times 10^5$°K) must occupy some fraction of the space between clouds. Any one-phase model should take into account that this fraction, and therefore the mean temperature in any volume of space, increases with increasing $|z|$ (i.e., the filling factor of the colder gas decreases with height).

The gas has time to adjust its temperature as it moves away from the plane. The dynamical time scale for vertical motion is $10^9/v$ years, so for velocities of the order of 10 km/sec, $t_{dynamical} >> t_{thermal}$, and there is no problem.

It is reasonable, therefore, to consider a second model for the gas, in which the temperature is an increasing function of z. Flows calculated for this varying temperature model, called Model B, will be compared to those of Model A.

The temperature gradient is unknown, and the observational evidence is inconclusive. We therefore choose a form for T(z) that is physically reasonable, and analytically convenient.

Since the sonic speed squared is proportional to temperature (see equation 3.21b), both a^2 and T have the same distribution (see figure 42):

$$T(z) = T_H - (T_H - T_o)\,\mathrm{sech}(z/\Delta) \qquad (3.28a)$$

$$a^2(z) = a_H^2 - (a_H^2 - a_o^2)\,\mathrm{sech}(z/\Delta) \qquad (3.28b)$$

where the subscripts o and H denote the values at $z = 0$ and $z = \infty$, respectively. (We can unphysically specify the distribution out

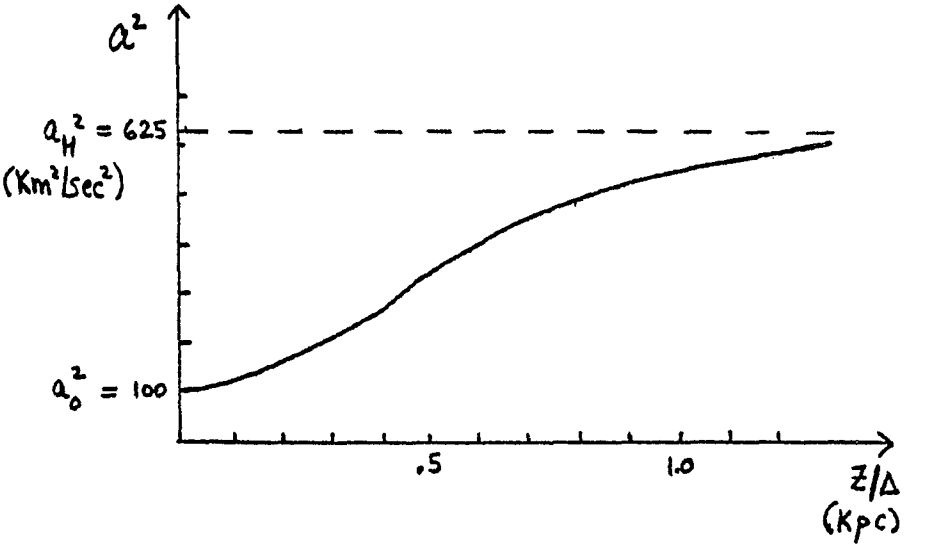


FIGURE 42: Sonic velocity squared as a function of z/Δ. a_o = 10 km/sec, a_H = 25 km/sec, Δ = 0.4 kpc.

to $z = \infty$, since all calculations stop at z = 1.5 kpc.

3.5 The Equations for the Two Models

For both models, the equation of state is

$$P = a^2 \rho \tag{3.29a}$$

where a^2 is the sonic speed squared, given by

$$a^2 = RT \quad . \tag{3.29b}$$

The governing equations for the flow are

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{3.30a}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + w\frac{\partial u}{\partial z} + \frac{1}{\rho}\frac{\partial(a^2\rho)}{\partial x} = -\frac{\partial \mathcal{V}}{\partial x} \tag{3.30b}$$

$$\frac{\partial w}{\partial t} + u\frac{\partial w}{\partial x} + w\frac{\partial w}{\partial z} + \frac{1}{\rho}\frac{\partial(a^2\rho)}{\partial z} = -\frac{\partial \mathcal{V}}{\partial z} \tag{3.30c}$$

where

$$\mathcal{V} = (2\pi G \sigma_*)\Delta \ln\left(\cosh(z/\Delta)\right) + N_0(1-\cos\alpha x)\,\mathrm{sech}^{|k|\Delta}(z/\Delta) \tag{3.30d}$$

and

$$a^2 = \begin{cases} a_0^2 & \text{for Model A} \\ a_H^2 - (a_H^2 - a_0^2)\,\mathrm{sech}(z/\Delta) & \text{for Model B} \quad . \end{cases} \tag{3.30e}$$

We no longer need the energy equation since we put in the temperature distribution rather than calculate it from the poorly known heating and cooling functions. Instead of specifying T, we specify the sonic speed. For both models, we chose $\gamma = 1$ for simplicity.

Model A (Isothermal)

The sonic speed is a constant, a_o, and we take it to be 10 km/sec, following Roberts (1969), Shu et al. (1972), and others. The temperature T_o corresponding to this speed is 1.2×10^4°K. This is an equivalent temperature since a^2 includes contributions from turbulence, cosmic rays and magnetic fields, and is therefore larger than the purely thermal sound speed.

Model B (Varying Temperature)

The sonic speed varies with $|z|$ and we take the value in the plane (a_o) to be 10 km/sec, corresponding to that of Model A. The value for a_H varies in different calculations from 15 km/sec (corresponding to $T_H = 2.7 \times 10^4$°K) up to 40 km/sec (corresponding to $T_H = 19.4 \times 10^4$°K). Again, these are effective, not purely thermal temperatures.

CHAPTER 4: THE NUMERICAL METHOD

4.1 Introduction

The equations to be solved are coupled, nonlinear time-dependent equations in two space dimensions, for an inviscid, compressible ideal gas, given by equations (3.30). We are interested in the flow in a finite region, 20 kpc along the gas streamline in the galactic plane, and extending about 1.0 kpc above and below the plane. We have to formulate an initial state for the gas in this region at time $t = 0$, upstream and downstream boundary conditions, boundary conditions at large $|z|$, and then use the equations to advance the fluid variables in time by an amount Δt for each iteration.

Multidimensional, time-dependent flows "provide a formidable task for computation" (Burstein 1964); that task is made even more difficult because the nonlinearity of the equations often results, for supersonic flow, in the appearance of discontinuities (i.e., shock waves) in the fluid, which cannot be integrated over.

It might be thought that, since we are looking for a steady-state solution, the calculations would be simpler if we dropped the time dependence and treated a pure boundary value prob-

lem. However, from the solutions to the flow in the galactic plane (Roberts 1969), we expect shocks, which means the gas will be supersonic before the shock and subsonic afterward. The equations for subsonic and supersonic flow are very different, both mathematically and physically.

Mathematically, the equations for supersonic flow are classified as hyperbolic, while those for subsonic flow are elliptic; they require very different boundary conditions in order to have a unique solution (or a solution at all!) (Courant and Hilbert 1953; Von Mises 1958). Thus, different mathematical models usually have to be developed for solving the flow problems in the two regimes.

It is easy to see the difference in the physical nature of the flow. Subsonic flow can be thought of in two ways -- either it is flow at any speed in an incompressible fluid (in which the sonic speed is infinite), or it is flow at less than the speed of sound in a compressible fluid. In either case, signals from one region, traveling at the sound speed, can reach and affect every other region of the gas: no region is independent of what happens in any other region.

Supersonic flow, on the other hand, is moving faster than signals can propagate in the fluid, and thus cannot be affected by regions of gas "behind" it. The flow in a region is determined by the flow in only part of the rest of the gas.

Thus, if we assume that the differential equations of the

flow in a region R, together with a knowledge of conditions along the boundaries, are sufficient to determine the flow inside R, we see that subsonic flow is affected by all the boundaries, while supersonic flow is only affected by parts of the boundaries.

Solving a time-dependent initial-value problem rather than a steady-state boundary value problem eliminates the need for treating the flow regimes independently, since the time-dependent equations remain hyperbolic even when the flow is of a mixed type containing subsonic as well as supersonic regions. The steady-state solution is obtained asymptotically as $t \to \infty$. The type and number of boundary conditions to be imposed do not change as the flow changes from subsonic to supersonic, or vice versa. This simplifies the handling of the boundaries considerably.

4.2 The Numerical Difference Scheme

There is a great variety of numerical schemes in computational physics, no one of which is appropriate to all problems in compressible fluid dynamics. The present numerical procedure is a two-step finite-difference method originally formulated by Lax and Wendroff (1960, 1964), and later extended to more complicated equations (with source terms) by Burstein (1967). It is just one of a general class of finite-difference schemes that allow one to treat a shock discontinuity without knowing its position a priori, as in shock fitting methods. (Good general discussions of these methods are found in Richtmyer and Morton 1967, and Potter 1973.)

The method handles the shock by introducing an explicit "artificial viscosity" term into the equations which broadens the shocks so that any shock is at least one mesh step thick. The shock then appears not as a discontinuity, but as a rapid change in density and velocity.

The differential equations for a compressible fluid may be written in a variety of ways and, while these differential forms are exactly equivalent, their analogous finite-difference forms are not. Because the rapid changes in the variables (ρ, u, w) across a shock front are likely to lead to numerical errors, the original differential equations are usually rewritten in terms of the variables (ρ, ρu, ρw). Two of this new set, the mass fluxes ρu and ρw, are conserved across a shock front, even though u and w are no continuous (see appendix A). In this conservation form, the governing equations (3.30) become:

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} + \frac{\partial (\rho w)}{\partial z} = 0 \tag{4.1a}$$

$$\frac{\partial (\rho u)}{\partial t} + \frac{\partial (\rho u^2 + a^2 \rho)}{\partial x} + \frac{\partial (\rho u w)}{\partial z} = -\rho \frac{\partial V}{\partial x} \tag{4.1b}$$

$$\frac{\partial (\rho w)}{\partial t} + \frac{\partial (\rho u w)}{\partial x} + \frac{\partial (\rho w^2 + a^2 \rho)}{\partial z} = -\rho \frac{\partial V}{\partial z} \tag{4.1c}$$

Because it identically satisfies the conservation of mass and momentum, this conservative formulation helps minimize nonlinear numerical effects.

We can write equations (4.1) as a vector equation:

$$\frac{\partial \vec{V}}{\partial t} + \frac{\partial \vec{F}}{\partial x} + \frac{\partial \vec{G}}{\partial z} = \vec{S} \tag{4.2}$$

where F and G are the fluxes. The right-hand side of the equation, $\vec{S}$, is called the source term, since it is the source of mass and momentum. $\vec{V}$, $\vec{F}$, $\vec{G}$, and $\vec{S}$ are three-component vectors. If we call our dependent variables ρ, M ($= \rho u$), and N ($= \rho w$), and write out the source term $\vec{S}$ explicitly, we have

$$V = \begin{pmatrix} \rho \\ M \\ N \end{pmatrix}, \quad F = \begin{pmatrix} M \\ M^2/\rho + a^2\rho \\ MN/\rho \end{pmatrix}, \quad G = \begin{pmatrix} N \\ MN/\rho \\ N^2/\rho + a^2\rho \end{pmatrix}, \tag{4.3abc}$$

$$S = \begin{pmatrix} 0 \\ -\rho\alpha N_0 \sin\alpha x \operatorname{sech}^{|k|\Delta}(\beta/\Delta) \\ -\rho\tanh(\beta/\Delta)\left\{2\pi G\sigma_* - |k| N_0(1-\cos\alpha x)\operatorname{sech}^{|k|\Delta}(\beta/\Delta)\right\} \end{pmatrix} \tag{4.3d}$$

(We will omit all vector signs on V, F, G and S from now on.)

To write this equation in finite-difference form, we subdivide the flow region with a rectangular grid, with spacings Δx and Δz. The grid points are given by $x = i\Delta x$ and $z = j\Delta z$, where i and j are integers. We also quantize time by dividing it into steps of Δt; thus $t = n\Delta t$, where n is an integer. A function $f(x,z,t)$ defined on the (spatial and temporal) grid is then $f(i\Delta x, j\Delta z, n\Delta t)$ and is written as f^n_{ij}.

Since this is an explicit (rather than implicit) finite-difference method, we assume that the variables V^n_{ij} are known at time t for all i and j, and therefore we know $F^n = F(V^n)$,

$G^n = G(V^n)$ and $S^n = S(V^n)$ for all points on the grid. In practice, we start at time t = 0 with reasonable initial values for all variables. (The initial conditions used are discussed in Chapters 5, 6 and 7.)

The straightforward way to get a difference equation is to convert all derivatives in the original equation (4.2) directly into finite differences to get

$$\frac{\Delta V}{\Delta t} + \frac{\Delta F}{\Delta x} + \frac{\Delta G}{\Delta z} = S \quad , \tag{4.4}$$

and therefore

$$V^{n+1} = V^n - \frac{\Delta t}{\Delta x}\Delta F - \frac{\Delta t}{\Delta z}\Delta G + S^n \Delta t \quad . \tag{4.5}$$

In essence, this is what all finite-difference schemes do, but there are numerous ways to calculate ΔF, ΔG, etc., in order to get the most stability and accuracy. The method used here advances the dependent variables from time t to $t + \Delta t$ in <u>two</u> steps rather than one, and uses quite complicated expressions for ΔF, ΔG, S^n and V^n.

For any quantity defined on the grid, f(x,z), we can define a

<u>forward spatial difference</u> in each direction,

$$\Delta_x^+ f(x,z) = f(x + \Delta x, z) - f(x,z) \tag{4.6a}$$

$$\Delta_z^+ f(x,z) = f(x, z + \Delta z) - f(x,z) \quad ; \tag{4.6b}$$

backward spatial differences,

$$\Delta_x^- f(x,z) = f(x,z) - f(x - \Delta x, z) \tag{4.7a}$$

$$\Delta_z^- f(x,z) = f(x,z) - f(x, z - \Delta z) \quad ; \tag{4.7b}$$

a centered spatial difference in each direction,

$$\Delta_x^c f(x,z) = \frac{f(x + \Delta x, z) - f(x - \Delta x, z)}{2} \tag{4.8a}$$

$$\Delta_z^c f(x,z) = \frac{f(x, z + \Delta z) - f(x, z - \Delta z)}{2} \quad ; \tag{4.8b}$$

and an "off-center" average over four points ([x,z] being one of the corner points, not the center of the average),

$$f_{av}(x,z) = \frac{1}{4}\Big\{f(x,z) + f(x + \Delta x, z) + f(x, z + \Delta z) + f(x + \Delta x, z + \Delta z)\Big\} \quad . \tag{4.9}$$

For every mesh point, Step 1 generates temporary variables V^* defined at an intermediate time $t^* < t + \Delta t$, and at intermediate space points "between" the defined mesh points. Following the form of equation (4.5), they are given by

$$V^* = V_{av}^n - \frac{\Delta t}{2\Delta x}\left[\Delta_x^+ F^n(x, j+\Delta j) + \Delta_x^+ F^n(x,j)\right] - \frac{\Delta t}{2\Delta j}\left[\Delta_j^+ G^n(x+\Delta x, j) + \Delta_j^+ G^n(x,j)\right] + \Delta t\, S_{av}^n \tag{4.10}$$

Next, the variables V^* are used to update F, G and S:

$$F^* = F(V^*)$$

$$G^* = G(V^*)$$

$$S^* = S(V^*)$$

Step 2 evaluates V^{n+1} by again using equation (4.5), but with a complex combination of the quantities at t and t^*.

The final value at $t + \Delta t$ is given by

$$V^{n+1} = V^n - \frac{\Delta t}{2\Delta x}\left[\Delta_x^c F^n(x,z) + \frac{1}{2}\left(\Delta_x^- F^*(x,z) + \Delta_x^- F^*(x,z-\Delta z)\right)\right]$$
$$-\frac{\Delta t}{2\Delta z}\left[\Delta_z^c G^n(x,z) + \frac{1}{2}\left(\Delta_z^- G^*(x,z) + \Delta_z^- G^*(x-\Delta x,z)\right)\right]$$
$$+\frac{1}{2}\Delta t\left[S^n(x,z) + S^*_{av}(x,z)\right] + Q^n \qquad (4.11)$$

Thus in V the second step goes directly from t to $t + \Delta t$, but for the fluxes (F, G) and source term (S) it uses a combination of the grid points at t with the intermediate points at t^*. After each complete cycle of two steps, the intermediate values of the variables are discarded.

Proving why this complex difference scheme is more accurate than some others is beyond the scope of this thesis (see Burstein 1964, 1967; Roache 1972). It will suffice to say that the accuracy of a difference scheme is the number of correctly evaluated time derivatives it includes. The intermediate values V^* have only 1st order accuracy, but the second step gives V^{n+1} accurate to 2nd order in Δt.

The Q^n term in equation (4.11) is the "artificial viscosity" term that is added to stabilize the numerical method. Actually, the finite-difference equations contain an <u>implicit</u> viscosity -- the truncation error made in approximating the partial derivatives by finite differences. This is the numerical smoothing intrin-

sic to any computer simulation of a physical problem. The truncation error is present in every equation and smooths out small oscillations in the dependent variables. However, its magnitude is very small and in the case of the present method, is not large enough to stabilize the system. That is why the explicit artificial viscosity term Q^n is added. While some numerical schemes contain this extra artificial viscosity implicitly in the equations (e.g., the "beam scheme"), the advantage of an explicit term Q^n is that its strength can be controlled directly.

An artificial viscosity term is needed for flows that contain shocks because of the nature of a shock. In a gas without viscosity or heat conduction the shock would be a discontinuous change in the fluid's state (although in reality viscous effects always become large within the shock, causing the changes to take place over some small but finite distance). In a finite-difference scheme, no distance smaller than the mesh step length (Δx or Δz here) can be described. Whereas in a real shock, viscosity and heat conduction in the narrow shock region transfer energy of short wavelength disturbances into thermal energy, in a difference solution to the equations, the energy will collect in the shortest wavelength of the mesh -- that of the mesh spacing itself. This can result in very large oscillations between the variables at adjacent mesh points.

The Q^n term must therefore damp these short wavelength oscillations without greatly affecting the solution at large wave-

lengths. A term that is quadratic in the difference between neighboring variables will ensure that the viscosity will influence the solution only at points where rapid variations occur. The practical result of the Q^n term is to broaden the shock on the mesh so that it is spread over several mesh steps.

Many forms for Q^n have been used by various people; fortunately, the results are not very sensitive to the exact form chosen, so that it is necessary only to choose a form appropriate to the particular difference scheme being used. The form used here (Liebovitch 1978) is

$$Q^n = \tfrac{1}{2} B \frac{\Delta t}{\Delta x} \left\{ \left| \Delta_x^+ u(x,z) \right| \Delta_x^+ V^n(x,z) - \left| \Delta_x^- u(x,z) \right| \Delta_x^- V^n(x,z) \right\}$$

$$+ \tfrac{1}{2} B \frac{\Delta t}{\Delta z} \left\{ \left| \Delta_z^+ w(x,z) \right| \Delta_z^+ V^n(x,z) - \left| \Delta_z^- w(x,z) \right| \Delta_z^- V^n(x,z) \right\} \qquad (4.12)$$

where u, w are the x and z velocities, respectively, and B is an adjustable dimensionless constant of order unity.

The value of B was determined essentially by trial and error, and ranges from 0 to 8 for various problems. (It is interesting that occasionally a one-dimensional solution converges to a steady state very rapidly with B = 0, even with a strong shock.) The value of B must be chosen as a compromise between good stability but reduced spatial resolution of the shock (high B) and poor stability but high spatial resolution (low B).

4.3 Stability of the Numerical Method

The addition of an artificial viscosity damps out small-wavelength oscillations. But other kinds of instability can occur; the most fundamental is related to the size of iteration time step, Δt.

For linear one-dimensional finite-difference equations with no artificial viscosity term, it is relatively easy to look at the system's stability by Fourier analysis and demand that the amplitude of each wavelength component be bounded (Richtmyer and Morton 1967). The result is very simple: for stability we must have

$$\Delta t \leq \frac{\Delta x}{V} \tag{4.13}$$

where V is the velocity of propagation of a signal. For most situations

$$V = |v| + a \tag{4.14}$$

where $|v|$ is the velocity of the fluid and a is the sonic velocity. This condition is the famous Courant-Friedrichs-Lewy (CFL) condition, first derived by Courant et al. (1928).

The physical meaning of this limit on the time step is clear: computational information flows in a mesh with spacing Δx and time step Δt at a velocity $\frac{\Delta x}{\Delta t}$. This velocity must be larger than the velocity of propagation of a signal, so that the numerical calculation can update the state of the fluid at each mesh point at least as fast as the state can change (i.e., no signal may be

allowed to cross a cell Δx in a time less than the calculation time step Δt).

Nonlinear multidimensional finite-difference equations cannot be analyzed rigorously for a stability criterion. The maximum allowed time step must be determined from an approximate analysis and a good deal of experience. Since the time step condition depends on the resulting flow velocity, it is necessary to have some idea of the velocities expected before the calculation is done (if Δt is not to be reestimated at each iteration). In addition, since nonlinear flows often develop shocks and thus require an artificial viscosity, a strengthening of the CFL condition is required for stability.

Without an artificial viscosity, Burstein (1964) found analytically that a sufficient stability condition in two-dimensional flow with equal mesh spacings $\Delta x = \Delta z = \Delta$ is

$$\Delta t \leq \frac{\Delta}{\sqrt{8}\,(V + a)} \quad . \tag{4.15}$$

The artificial viscosity adds a factor d to the above equation, where $d \leq 1$. Its exact value depends on the strength of the artificial viscosity term.

Since the mesh sizes Δx and Δz used in this calculation are unequal, the time steps were estimated separately for the flow in the x and z directions, namely,

$$\Delta t_x \leq \frac{d\ \Delta x}{|u| + a} \tag{4.16a}$$

$$\Delta t_z \leq \frac{d\ \Delta z}{|w| + a}\ , \tag{4.16b}$$

and the smaller value was used for Δt. The value $d = 1/3$ was satisfactory for most of the calculations. It was very easy to tell when Δt was too large: negative densities developed within fewer than 100 iterations.

An additional condition in the computation, a spatial mesh restriction that depends on the size of the source term S in equation (4.2), is discussed in Chapter 5.

4.4 Initial Conditions

The numerical calculation starts at time $t = 0$ with given initial values of all dependent parameters, and advances them through a time interval Δt for every iteration or complete cycle of two steps. Although the choice of the initial conditions is said to be more critical in supersonic flow than in subsonic or incompressible flow (Roache 1972, p. 284), for this numerical scheme the steady-state solutions are remarkably independent of the initial state. Tests run with different initial conditions show identical steady-state results.

The initial values were chosen to be the free-stream state (i.e., with no source term present); the source term was then "turned on" impulsively at the first time step. Experiments were made, as in Sanders and Huntley (1976), with letting the source

term increase slowly to its maximum value over a range of 10 to 500 time steps. The gradual introduction of the source term had no effect on the initial oscillations of the system nor did it speed up the convergence to a steady state.

4.5 Boundary Conditions

While the numerical method described so far may seem complex, it is relatively straightforward compared to the problem of finding appropriate boundary conditions for a stable, unique solution. Although the boundary conditions for the differential equations are usually known, and can often be used as a guide in the numerical problem, they cannot be simply transferred whole to the finite-difference equations.

Sometimes the problem is obvious; e.g., the mesh covers only part of the flow. But many boundary problems are more subtle, e.g., reflections off the edge of the mesh, or numerical instabilities of the difference method.

Before the computational problem could be attacked, it was necessary to determine the correct number of parameters or dependent variables to be specified on each boundary, and the sensitivity of the solution to the boundary values specified. The solutions to the boundary problems were developed from various sources and in various ways, including:

(i) Using the differential equations as a guide. As an example, if the differential equations show that a certain

derivative of the velocity is identically zero at a boundary, the difference scheme should specify the velocity on that boundary in such a way that the derivative will automatically vanish.

(ii) Papers on the theoretically appropriate boundary conditions for a specific numerical problem and scheme (Ganz and Serra 1974; Serra 1972). However, occasionally I found that for stability I had to use different boundary conditions than these.

(iii) Extensive reading of the literature to see how similar problems were handled was of limited usefulness because, with the exception of Roache (1972) and Moretti (1969), most published literature describes the method used in detail but omits any discussion of the boundary conditions.

(iv) Breaking up the two-dimensional problem into separate one-dimensional flows having known solutions, and in which the numerical boundary conditions are easier to formulate. This was an enormously useful procedure, helping not only in handling the boundaries but in gaining insight into many aspects of the scheme.

(v) Testing different boundary conditions in actual numerical calculations, while inefficient, was occasionally the only way to decide between two possible boundary conditions, both reasonable.

This discussion of boundary conditions has been very general; the details and actual choices for the boundaries are discussed in the next three chapters, which describe the one-dimensional flows studied, and the full two-dimensional problem.

The method used requires that the values at points $x + \Delta x$ and $x - \Delta x$ be known in order to update the values of the variables at point x (and the same hold for points z). When the area of the flow is covered with a mesh that goes from x_1 to x_N, and z_1 to z_N, we must therefore have a method for determining the variables one mesh length outside the region.

The flow region in these computations is rectangular, with boundaries given by

$$\begin{aligned} x &= 0 \\ x &= x_{max} \\ z &= 0 \\ z &= z_{max} \quad . \end{aligned} \tag{4.17}$$

A row of fictitious mesh points is added outside each of these boundaries.

Since we have assumed symmetry about the plane of the Galaxy, we can get the variables at $z = -\Delta z$ by reflecting the flow across the plane:

$$\begin{aligned} \rho(x,-\Delta z) &= \rho(x,\Delta z) \\ u(x,-\Delta z) &= u(x,\Delta z) \\ w(x,-\Delta z) &= -\, w(x,\Delta z) \end{aligned} \tag{4.18}$$

The equation for ρw implies that the z velocity w must equal zero at z = 0, so we get one of the conditions at z = 0 for free, namely

$$\rho w(z = 0) = 0 \quad . \tag{4.19}$$

The values of ρ, ρu and ρw at $x = -\Delta x$, $x = x_{max} + \Delta x$, and $z = z_{max} + \Delta z$ are given by repeating the values on the boundaries, i.e., for any quantity V(x,z),

$$\begin{aligned} V(-\Delta x,z) &= V(0,z) \\ V(x_{max} + \Delta x,z) &= V(x_{max},z) \\ V(x,z_{max} + \Delta z) &= V(z_{max}) \end{aligned} \tag{4.20}$$

Henceforth, all references to <u>boundaries</u> of the mesh will refer to those defined in equations (4.17), and ignore the extra mesh points outside the boundaries.

CHAPTER 5: ONE-DIMENSIONAL ATMOSPHERE

5.1 Introduction

A seemingly simple problem was chosen as the first test of the numerical method: galactic gas with no horizontal motion in a vertical axisymmetric gravitational potential (i.e., no spiral field) -- the hydrostatic atmosphere of the Galaxy.

The gravitational potential is due to a disk of finite thickness Δ (see equation [3.4]):

$$\mathcal{V} = (2\pi G \sigma_*)\, \Delta \ln\left(\cosh(z/\Delta)\right) . \tag{5.1}$$

The time-dependent governing equations in one dimension (z) for this potential are obtained from the full equations in conservation form (4.1) by setting u = 0 and $\mathcal{V} = \mathcal{V}(z)$. This gives

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{5.2a}$$

$$\frac{\partial(\rho a^2)}{\partial x} = 0 \tag{5.2b}$$

$$\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho w^2 + a^2 \rho)}{\partial z} = -\rho \frac{d\mathcal{V}}{dz} \tag{5.2c}$$

Equation (5.2b) says that ρ is a function of z alone, for Model A

(a = constant) and Model B (a = a[z]), so all x dependence can be ignored.

The equations in the general matrix formulation are:

$$\frac{\partial V}{\partial t} + \frac{\partial G}{\partial z} = S \quad , \tag{5.3}$$

where

$$V = \begin{pmatrix} \rho \\ \rho w \end{pmatrix} \tag{5.4a}$$

$$G = \begin{pmatrix} \rho w \\ \rho w^2 + a^2 \rho \end{pmatrix} \tag{5.4b}$$

$$S = \begin{pmatrix} 0 \\ -\rho(2\pi G \sigma_*) \tanh(z/\Delta) \end{pmatrix} \quad . \tag{5.4c}$$

5.2 The Analytic Solution

The steady-state solution to these equations can easily be found analytically. A solution with w = 0 and $\rho = \rho(z)$ satisfies

$$\frac{d(\rho a^2)}{dz} = -\rho(2\pi G \sigma_*) \tanh(z/\Delta) \quad . \tag{5.5}$$

For Model A, since $a = a_0$, the solution is

$$\rho = \rho_0 \operatorname{sech}^{\beta_0^2}(z/\Delta) \tag{5.6}$$

where

$$\beta_0^2 \equiv \frac{(2\pi G \sigma_*)\,\Delta}{a_0^2} \tag{5.7}$$

and ρ_0 is the value of ρ at $z = 0$.

For Model B, $a^2(z) = a_H^2 - (a_H^2 - a_0^2)\mathrm{sech}(z/\Delta)$, and the solution is

$$\rho = \rho_0 \left(\frac{a_0^2}{a^2(z)}\right)^{\delta+1} \mathrm{sech}^{\delta}(z/\Delta) \tag{5.8}$$

where

$$\delta \equiv \frac{\beta_0^2 a_0^2}{a_H^2} \tag{5.9}$$

and again ρ_0 is the value of ρ at $z = 0$.

These solutions show that the hydrostatic density is highly stratified. For values of σ_* ranging from 35 to 110 $M_\odot/pc^2$ (see section 3.3), we get

$$3.8 \leq \beta_0^2 \leq 11.9 \quad . \tag{5.10}$$

For $15 \leq a_H \leq 40$ km/sec, we have for the corresponding β_0^2 above,

$$1.7 \geq \delta \geq 0.75 \quad . \tag{5.11}$$

5.3 The Numerical Solution

The time-dependent equations can now be solved numerically and compared to the exact solutions, to see how accurate the numerical scheme is, as well as to learn something about the intricacies of its operation. (Although this chapter discusses only Model A (constant a), the results obtained for Model B are exactly the same.)

The equations are rewritten in terms of a new density variable -- the constant coefficient ρ_0 in the exact steady-state so-

lutions. This procedure makes it easier to compare the numerical results with the exact solutions; it also insures that the values of the density do not get too small and cause computational difficulties.

We define the new variable $\hat{\rho}$ such that (see equation [5.6])

$$\rho = \hat{\rho}\ \mathrm{sech}^{\beta_o^2}(\gamma/\Delta) \ . \tag{5.12}$$

Rewriting the governing equations (5.3) in terms of $\hat{\rho}$, we get

$$\frac{\partial \hat{\rho}}{\partial t} + \frac{\partial(\hat{\rho} w)}{\partial \gamma} = \left(\frac{\beta_o^2}{\Delta}\right) \hat{\rho}\, w \tanh(\gamma/\Delta) \tag{5.13a}$$

$$\frac{\partial(\hat{\rho} w)}{\partial t} + \frac{\partial(\hat{\rho} w^2 + a_o^2 \hat{\rho})}{\partial \gamma} = \left(\frac{\beta_o^2}{\Delta}\right) \hat{\rho} w^2 \tanh(\gamma/\Delta) \ . \tag{5.13b}$$

These equations differ from equations (5.3) only by the source term S, which is now

$$S = \left(\frac{\beta_o^2}{\Delta}\right) \tanh(\gamma/\Delta) \begin{pmatrix} \hat{\rho} w \\ \hat{\rho} w^2 \end{pmatrix} \ . \tag{5.14}$$

Since equation (5.13a) is the equation describing conservation of mass, we now have a mass "source" or "sink."

The finite-difference equations corresponding to equations (5.13) are derived in a straightforward manner in appendix B.

The one-dimensional mesh stretches from z = 0 to

$z = 1.0$ kpc, with $\Delta z = 0.08$ kpc for most calculations.

Although other people doing numerical calculations often find they get better stability (Liebovitch 1978) or faster convergence to a steady state by overdetermining the problem, i.e., specifying more variables on the boundary than are necessary, trial calculations showed that the best way of handling the boundaries in this problem is to specify as <u>little</u> as possible. Thus, $w = 0$ for $z = 0$ is required by the symmetry of the problem, and nothing is specified at the top boundary.

The initial conditions were chosen by keeping in mind the expected steady-state solution, $\hat{\rho}$ = constant. The initial density distribution of $\hat{\rho}$ was a linear decrease with height of about 20%. The initial velocity w was usually taken to be zero, but the method was tested by letting $w_{initial} \neq 0$; no difference in the steady-state results was seen.

The Courant condition on the time step Δt for linear equations without artificial viscosity is (in units of 10^6 years)

$$\Delta t \leq \frac{\Delta z}{|w| + a_o} \approx 7 \quad .$$

This was an excellent estimate -- $\Delta t = 6.5$ could be used even with a moderately strong ($B = 4$) artificial viscosity. The value of B was usually not critical; in fact, in many runs the artificial viscosity term could be omitted.

The above results are all expected; the source S presented an unexpected stability problem.

5.4 Source Term Stability

Analysis of the nonlinear finite-difference equations for stability with respect to the source term is exceedingly difficult, and the results are not very revealing. But a simple physical argument gives the limits on source strength for a given mesh.

The equation of continuity is the simplest to look at:

$$\frac{\partial \hat{\rho}}{\partial t} + \frac{\partial(\hat{\rho} w)}{\partial z} = g \hat{\rho} w \tag{5.15}$$

where $g = (\beta_o^2/\Delta)\tanh(z/\Delta)$. Approximating all derivatives by simple differences and rearranging a little gives

$$\frac{\Delta \hat{\rho}}{\hat{\rho}} + \frac{w\left(\frac{\Delta \hat{\rho}}{\hat{\rho}}\right)}{\left(\frac{\Delta z}{\Delta t}\right)} + \frac{\Delta w}{\left(\frac{\Delta z}{\Delta t}\right)} = \frac{w}{\left(\frac{1}{g \Delta t}\right)} . \tag{5.16}$$

The dimensions of each term tell us that $(g \Delta t)^{-1}$ is the "velocity" of the source term.

The CFL condition on Δt says that the computer velocity across a cell, $\Delta z / \Delta t$, must be larger than the velocity of a signal, $|w| + a$. For the source term, however, we must demand the opposite; the source velocity must be <u>larger</u> than the computational velocity. The source term, evaluated essentially at positions between the mesh points, travels from cell to cell at the source velocity, $(g \Delta t)^{-1}$, creating (or destroying) mass as it moves. If the computer velocity is larger than the source velocity, the calculation travels across a cell faster than the mass inside can be

created, as if the source term did not exist for that step. Thus, the source term velocity must be larger than the computer velocity.

This requirement means that

$$\frac{1}{g\,\Delta t} \geq \frac{\Delta z}{\Delta t}$$

or

$$g\,\Delta z \leq 1 \quad . \tag{5.17}$$

This estimate of the source stability condition is fairly accurate. The source term strength depends on both $\beta_o^2/\Delta \propto 2\pi G\sigma_*$ and the position z. On the plane, g = 0 and the source velocity is infinite, causing no problems. At larger z, since σ_* varies with position in the Galaxy, g varies from 9.5 to 30.0 at $z = \infty$.

The mesh size Δz was originally set at 0.08 kpc for the weakest source term. As the value of g was increased, a maximum field strength g_{max} was reached above which the numerical calculation was unstable for that value of Δz. This occurred when $g\,\Delta z \approx 12$, closely corresponding to the theoretical prediction. To increase the field strength further it was necessary to decrease the mesh size, Δz, but much more rapidly than the simple linear analysis above suggests.

The instability shows up as a loss of mass throughout the grid, as if the mass were disappearing between the mesh points. (Since the dependent variable here is the coefficient $\hat{\rho}$, it is this "mass" that is lost.) The mass loss, first appearing at $z = z_{max}$, as we expect since the source term is largest there, rapidly

affects the rest of the gas and the calculation breaks down.

5.5 Results

The resulting steady-state solution, for all values of field strength, is described by

$$\hat{\rho} = \text{constant}$$

$$w = 0 \quad , \tag{5.18}$$

precisely the same as the exact solution (equation [5.6]). For field strengths much smaller than the maximum allowed for a given Δz, (g_{max}), the time-dependent solution converged very rapidly to steady state (less than 400 iterations for $g = 0.5\ g_{max}$), but increased to a few thousand iterations for $g = 0.9\ g_{max}$. The steady-state solution was constant density to 5 decimal places and velocities $|w| \leq 10^{-6}$ km/sec. The same kind of accuracies were obtained for larger strengths when Δz was decreased. Figure 43 shows how the density and the velocity at one grid point approach steady-state values. This oscillatory approach to a steady state is typical of all solutions calculated using this numerical scheme.

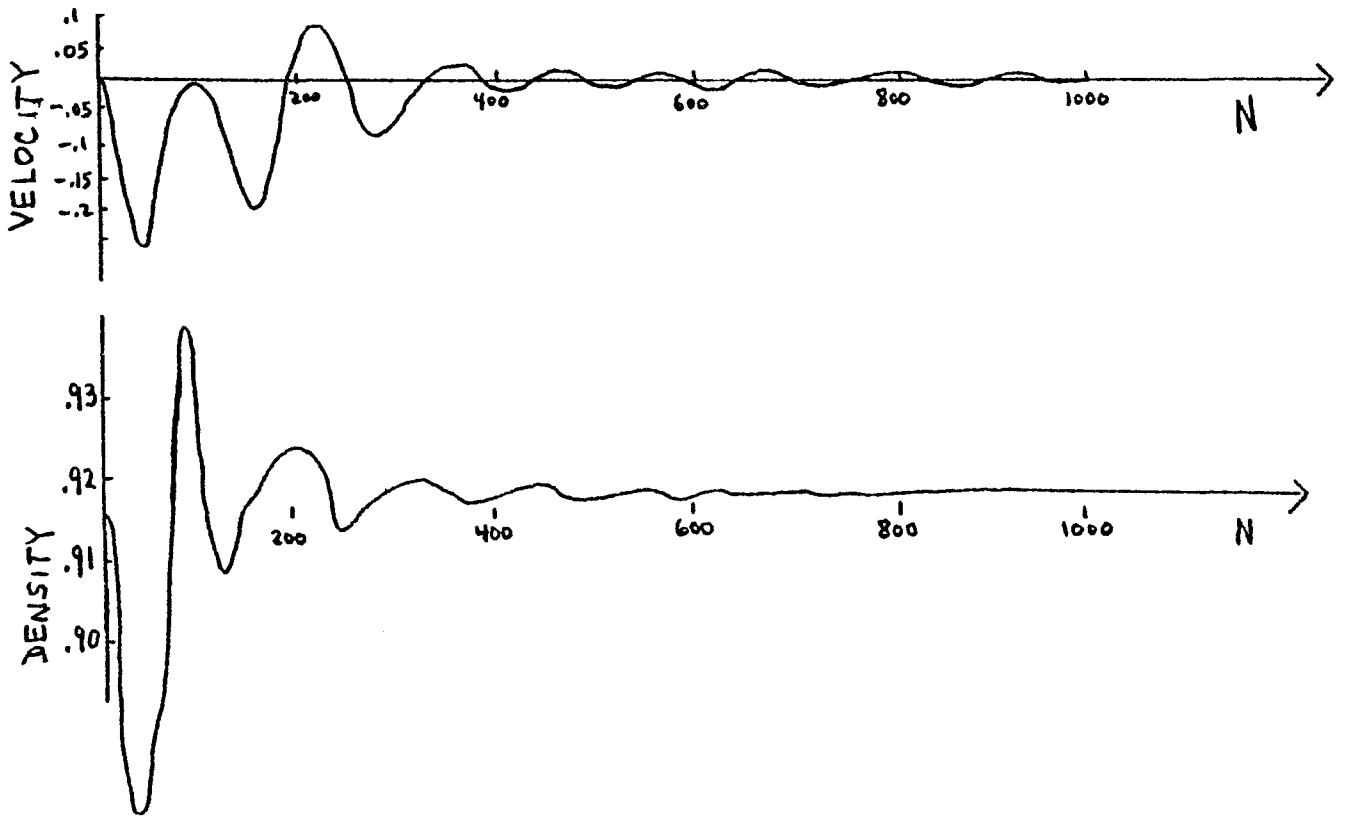


FIGURE 43: Relaxation of density and velocity in time (at z = 0.5 kpc) to steady-state values for a typical calculation of an isothermal atmosphere.

CHAPTER 6: ONE-DIMENSIONAL FLOW INTO A SPIRAL ARM

6.1 Introduction

A nonstatic test of the numerical method was performed by suppressing the z-dimension and calculating one-dimensional flow into a spiral gravitational potential; this turns out to be extremely rewarding in terms of understanding why a shock forms in the gas, as well as what boundary conditions should be imposed in the full two-dimensional calculation.

The spiral gravitational potential (see figure 41a) is

$$\mathcal{V} = N_0(1 - \cos\alpha x) \quad , \quad 0 \leq \alpha x \leq 2\pi \; . \tag{6.1}$$

The value of N_0 is 1880 F, where F is the ratio of the spiral to the axisymmetric field, usually taken as 5% (see section 3.3). The values of x go from 0 to 20 kpc, so α is $\pi/10$.

The time-dependent equations are obtained from equations (3.30) by setting w = 0 and $\mathcal{V} = \mathcal{V}(x)$:

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} = 0 \tag{6.2a}$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2 + a^2 \rho)}{\partial x} = -\rho \frac{d\mathcal{V}}{dx} \; . \tag{6.2b}$$

In matrix form, these equations become:

$$\frac{\partial V}{\partial t} + \frac{\partial F}{\partial x} = S \quad , \tag{6.3}$$

where

$$V = \begin{pmatrix} \rho \\ \rho u \end{pmatrix} \tag{6.4a}$$

$$F = \begin{pmatrix} \rho u \\ \rho u^2 + a^2 \rho \end{pmatrix} \tag{6.4b}$$

$$S = \begin{pmatrix} 0 \\ -\rho \alpha N_0 \sin \alpha x \end{pmatrix} \quad . \tag{6.4c}$$

The finite-difference equations corresponding to equation (6.3) can be found in appendix B.

The entire analysis in the chapter is fully applicable to both Model A and Model B, since it concerns only the x-variation of the variables. We therefore consider $a = a_o$ = constant for this one-dimensional flow, and the gaseous equation of state as $P = a_o^2 \rho$.

6.2 The Analytic Solution (I)

The analytic solution for the steady-state flow satisfies

$$\frac{d(\rho u)}{dx} = 0 \tag{6.5a}$$

$$\frac{d(\rho u^2 + a_o^2 \rho)}{dx} = -\rho \alpha N_0 \sin \alpha x \quad . \tag{6.5b}$$

Equation (6.5a) says that the mass flux J is constant:

$$J = \rho u = \text{constant} \quad . \tag{6.6}$$

We can use this to eliminate ρ in equation (6.5b) to get, after a little manipulation,

$$\frac{du}{dx} = \frac{-u \frac{dV}{dx}}{u^2 - a_o^2} \quad . \tag{6.7}$$

This can be immediately solved for u(x), since the variables are separable. The formal solution is

$$\tfrac{1}{2}u^2 - a_o^2 \ln u = -V(x) + K \tag{6.8}$$

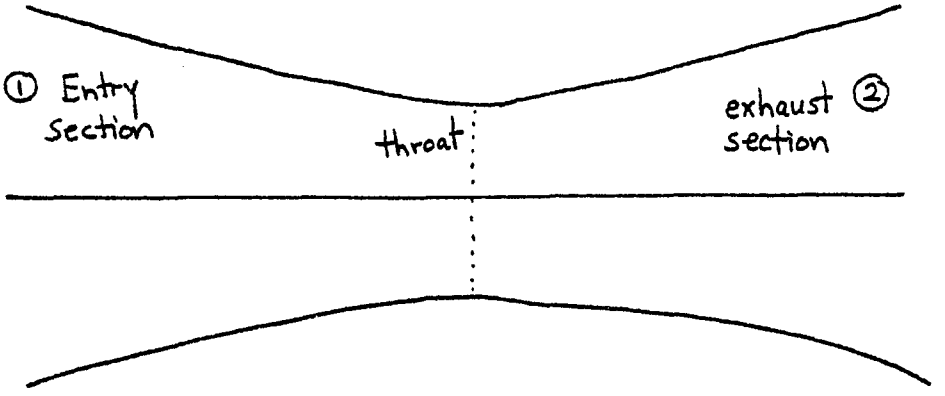
where K is a constant of integration. The steady-state flow is thus determined by equations (6.6) and (6.8); J and K are to be determined by the boundary conditions at x = 0 and 20.

The proper boundary conditions to be imposed for a unique solution are not obvious because, as discussed in section 4.1, the number of parameters to be specified on each boundary depends on whether the flow solution is subsonic or supersonic. A study of one-dimensional nozzle flow, which is the exact analogy to this flow into a potential hill, will make it clear exactly what boundary conditions should be prescribed, as well as when and where the solution will contain a shock.

6.3 The Nozzle Analogy

A de Laval nozzle is a converging-diverging nozzle, shown in figure 44a. It has an entry section 1, a throat (where the

FIGURE 44a



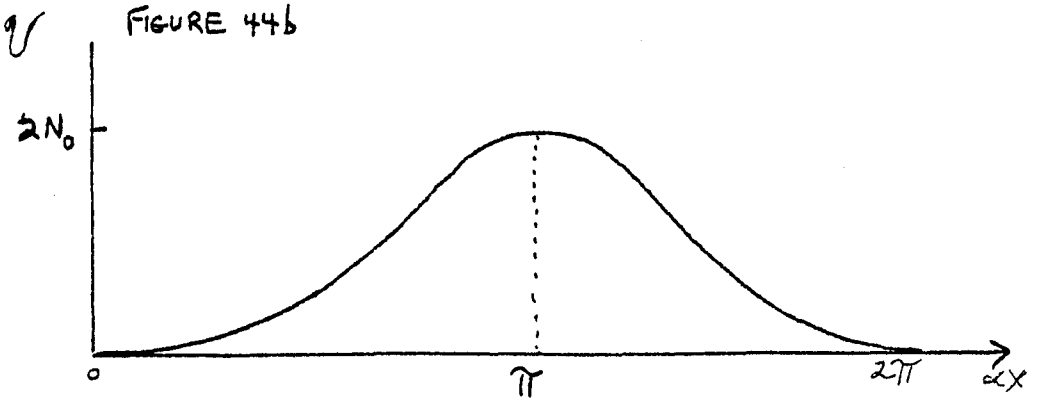


FIGURE 44a: de Laval nozzle.

FIGURE 44b: The spiral gravitational potential, $N_0(1 - \cos \alpha x)$.

cross-sectional area A is a minimum), and an exhaust section 2. A one-dimensional approximation to the real flow in a nozzle can be used under the following conditions:

(i) that the fractional rate of change of area with respect to the distance x along the axis is small, i.e.,

$$\frac{x}{A}\frac{dA}{dx} << 1 \quad ;$$

(ii) that the radius of curvature of the axis (if any) is much larger than the nozzle diameter.

Under this approximation, the steady-state equation for conservation of mass in the nozzle is

$$\rho \cdot uA = \text{constant} \quad , \tag{6.9}$$

and the steady-state equation for conservation of momentum is

$$u\frac{du}{dx} = -\frac{1}{\rho}\frac{dP}{dx} \quad . \tag{6.10}$$

Using the equation of state $P = a_o^2\rho$, equations (6.9) and (6.10) are combined to get

$$\frac{du}{dx} = \frac{a_o^2\frac{u}{A}\frac{dA}{dx}}{u^2 - a_o^2} \quad . \tag{6.11}$$

This equation is exactly the same as equation (6.7), if we consider $\frac{a_o^2}{A}\frac{dA}{dx}$ to be the analogue of $-\frac{d\mathcal{V}}{dx}$. This is the reason that $\mathcal{V}$ was chosen as $N_o(1 - \cos\alpha x)$; the signs of $\frac{dA}{dx}$ and $-\frac{d\mathcal{V}}{dx}$ are the same for a converging-diverging nozzle and a potential "hill"

(rather than a "well"). The behavior of the solution is the same for both, but it is easier to work with the physical nozzle flow, while remenbering that the shape of the nozzle is the direct analogy of the shape of the gravitational potential (see figure 44).

Certain features of the flow can be seen directly from equation (6.11):

(i) For $\frac{du}{dx}$ to be finite when $u = \pm a_o$ (sonic point), $\frac{dA}{dx}$ must be zero.

(ii) Using L'Hôpital's rule we find that, for $\frac{du}{dx}$ to be real when $u = \pm a_o$, A must have a relative minimum at the sonic point.

Therefore, from (i) and (ii), if there is to be a sonic point at all, it can occur <u>only</u> at a relative minimum of A, i.e., at the throat of the nozzle.

(iii) If the sonic point is not reached, the maximum or minimum value of the velocity occurs at the throat of the nozzle.

(iv) Subsonic and supersonic flows behave <u>very</u> differently -- in fact, in opposite manner. The sign of $\frac{du}{dx}$ at any position in the nozzle depends on the sign of $(u^2 - a_o^2)$, i.e., whether it is subsonic or supersonic. For example, subsonic gas entering the converging part of the nozzle, where $\frac{dA}{dx} < 0$, speeds up, while the supersonic gas slows down. This behavior is shown in figure 45.

The behavior described in (iv) is often felt as being

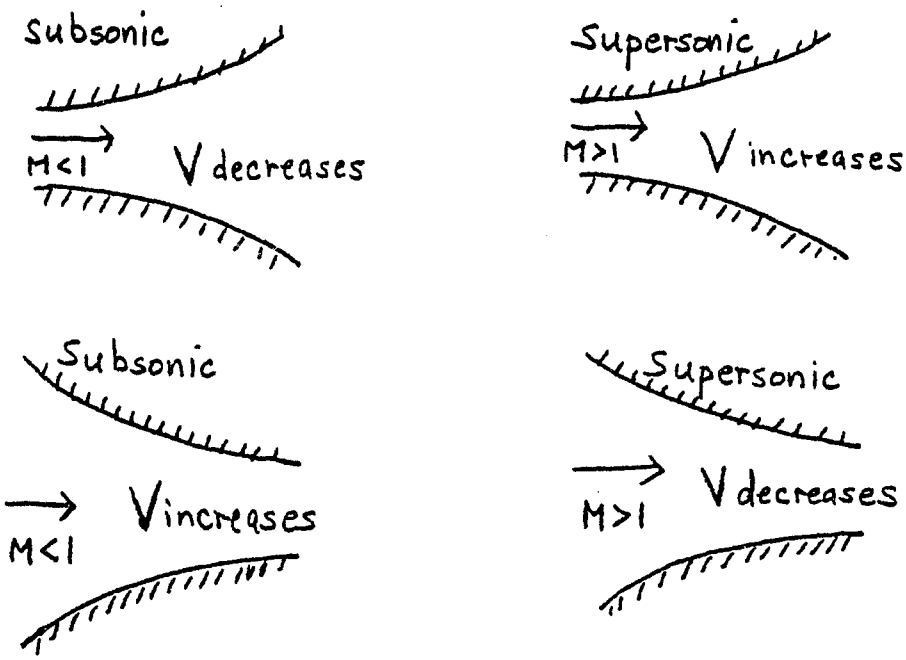


FIGURE 45: Effects of area change on velocity in subsonic and supersonic flows.

"strange." In terms of a gravitational potential, the force on any specific fluid particle due to the potential is given by $-\frac{d\mathcal{V}}{dx}$. If $\frac{d\mathcal{V}}{dx} > 0$, the force is negative, thus acting to <u>slow</u> the fluid particle down. In a nozzle, the converging walls are often thought of as exerting a physical force on the gas to slow it down. Why, then, does supersonic gas slow down as expected but subsonic gas do the opposite??

The apparent paradox is resolved when two things are emphasized: (a) the flow described above has already adjusted itself to a steady-state flow, and (b) a fluid particle is not a projectile. A macroscopic projectile is not affected by the pressure of surrounding particles, so its velocity is affected only by the external forces it feels (e.g., walls of a container, collisions, gravitational potential). A fluid, on the other hand, feels a force due to the pressure gradient, $a_o^2 \frac{d\rho}{dx}$, that can be neglected only when $u^2 >> a_o^2$ (see equation [6.2b]). In other words, only highly supersonic flow acts as a collection of projectiles. The pressure force propagates at the speed of sound. Subsonic flow feels this force from all directions; highly supersonic flow is not affected by it at all.

Consider gas flow in a potential as a function of time. Using equation (6.5a), we can write equation (6.5b) as

$$u \frac{du}{dx} + \frac{a_o^2}{\rho} \frac{d\rho}{dx} = -\frac{d\mathcal{V}}{dx} = f \qquad (6.12)$$

where f is the force.

If the gas starts off with a small but constant velocity and constant density when a negative (to the left) force is turned on, both derivatives are initially zero. If the inertial force is neglected, the pressure gradient becomes negative due to f, and pushes a fluid particle to the right -- speeding it up. If, on the other hand, the pressure gradient is neglected, the resultant negative inertial force will move the fluid particle to the left -- slowing it down. Which tendency wins depends on the relative values of u^2 and a_o^2, as can be seen from dimensional analysis.

If the flow is followed in time, one should be able to see the density gradients build up to result in this seemingly unphysical behavior. In fact, the calculations show that the subsonic gas entering the nozzle slows down and actually moves back toward the entrance for a time, before the density gradients build up enough to move it forward again and accelerate it.

The different types of steady-state nozzle flow occurring under various conditions are well known (Shapiro 1954; Courant and Friedrichs 1948); they are described in appendix C.

6.4 The Analytic Solution (II)

To return to the original problem of one-dimensional gas flow into a spiral gravitational potential: The discussion of steady-state nozzle flow (see appendix C) shows the types of flow to expect. It also tells us which of the possible parameters (up-

stream and downstream density and velocity) should be specified on the flow boundaries for a unique solution to exist.

For wholly subsonic flow, the upstream velocity and the downstream density (or vice versa) should be specified; the flow itself will determine the other two parameters by the requirement that the flow be symmetric.

For transonic flow without a shock, the density is set on the upstream boundary to determine the subsonic flow and nothing is set at the exit, since the supersonic flow cannot feel it. It appears that we have lost one boundary condition for continuous transonic flow, but there is an implicit condition on the velocity -- the sonic point must occur at the "throat" or the potential maximum. This determines the mass flux at the potential maximum and the upstream velocity develops to a steady-state value that is compatible with this mass flux. The supersonic portion is determined solely from the shape of the potential.

If the transonic flow contains a shock at position x_s, an additional boundary condition, setting the downstream density, is added to those of continuous transonic flow. The value, if chosen correctly, will force the flow to be subsonic at the exit, thus forcing a shock to form. The position of x_s and strength of the shock are determined by its exact value. It now seems that there are too many (implicit and explicit) specified parameters for the problem. But the appearance of a shock adds another variable to the equations -- the shock position.

The number of parameters to be specified on the boundaries is the same for wholly subsonic flow and for transonic flow containing a shock; the exact values of these parameters determine whether or not a shock is formed.

The analytic solution for the velocity is given by equation (6.8); the constant of integration is found from the condition that the sonic velocity be reached at $\alpha x = \pi$. Thus the flow up to the shock is governed by

$$\frac{1}{2}(u^2 - a_o^2) - a_o^2 \ln\left(\frac{u}{a_o}\right) = -V(x) + V\left(\frac{\pi}{\alpha}\right) \quad \text{for } x \leq x_s \qquad (6.13)$$

where x_s is the position of the shock. At $x = x_s$, the velocity, u_1, is given by

$$\frac{1}{2}(u_1^2 - a_o^2) - a_o^2 \ln\left(\frac{u_1}{a_o}\right) = -V(x_s) + V\left(\frac{\pi}{\alpha}\right) \quad . \qquad (6.14)$$

The density and mass flux J are determined by specifying the upstream density ρ_L (subscript L denotes the upstream boundary; subscript R denotes the downstream boundary).

This solution cannot be continued across the discontinuity; the flow downstream of the shock is governed also by equation (6.8), but with a different value of the integration constant. We know from the Rankine jump conditions (see appendix A) that the velocity jumps from u_1 to $u_2 = a^2/u_1$ at $x = x_s$; this value of u_2 determines the integration constant, giving for the flow past the shock

$$\frac{1}{2}(u^2 - u_2^2) - a_o^2 \ln\left(\frac{u}{u_2}\right) = -V(x) + V(x_s) \quad \text{for } x \geq x_s \qquad (6.15)$$

The shock location, x_s, is determined from the downstream velocity u_R, obtained from the value of ρ_R (since $J = \rho u$ is continuous across the shock). Equation (6.15), evaluated at $\alpha x = 2\pi$, gives

$$\frac{1}{2}(u_R^2 - u_2^2) - a_o^2 \ln\left(\frac{u_R}{u_2}\right) = -V\left(\frac{2\pi}{\alpha}\right) + V(x_s) \quad . \qquad (6.16)$$

This equation has a solution for x_s between 10 and 20 kpc for only a certain range of u_R (and therefore ρ_R).

6.5 The Numerical Solution

The numerical scheme solves the time-dependent problem rather than the steady-state flow discussed above. But it has been shown (Serra 1972) that the time-like boundary conditions are the same as those for steady-state flow, so the analysis can be used to suggest the boundary conditions for the numerical method.

The nozzle flow analysis says that a different number of boundary parameters should be specified for various types of flow. However, we are primarily interested in wholly subsonic flow or discontinuous transonic flow, and for these the boundary conditions are the same: one parameter each on the upstream and downstream boundaries.

The value of the mass flux J_L was set on the upstream boundary and ρ_R was set on the downstream boundary. The upstream

density was computed from a parabolic extrapolation of the interior flow. For wholly subsonic flow, these conditions alone determine the flow. For transonic flow containing a shock, these, plus the implicit condition on the position of the sonic point, determine the flow.

The same boundary conditions were specified for transonic flow without a shock although it overspecifies the downstream boundary. In real nozzle flow, ρ_R (the outside atmospheric density) is always given; the flow stays supersonic when the value of ρ_R is too low for the flow to adjust to it (via a shock) inside the nozzle. The gas then adjusts to ρ_R by shocking outside the nozzle, in the atmosphere; since the mesh stops at the end of the "nozzle," we do not see the external shock.

Overspecifying the boundary did not affect the solution a great deal. The finite-difference equations do not model the differential equations exactly (Richtmyer and Morton 1967), and tests here showed that the numerical scheme is, in some ways, much more forgiving than the set of differential equations. When a parameter specified on a boundary does not have the correct value, a sharp transition between the boundary values and the interior values of the flow occurs over a few grid points adjacent to the boundary. In this way, the exact values fixed on the boundaries have little effect on the interior flow.

The knowledge that the numerical scheme is stable with overspecified boundaries is very important for two-dimensional

flows, in which the "correct" boundary conditions are not clear and it is occasionally necessary for stability to overspecify a boundary.

The initial state of the gas at time t = 0 is uniform flow u_L with constant density ρ_L. Again, as in the one-dimensional atmosphere described in Chapter 5, the solution was independent of the mode of introduction of the source term.

The time step was chosen using the CFL condition, equation (4.13). The value of the artificial viscosity was taken to be B = 1.0, a low value that gave excellent spatial resolution across the shock front. Any lower value of B gave oscillations near the shock front. The grid spacing was chosen to be $\Delta x = 0.5$ kpc.

6.6 Results

The computations were very stable and the results were not very sensitive to how the boundaries were handled. Different types of flows were calculated: subsonic, continuous transonic and transonic with a shock. The parameters that were varied are:

J_L - the incoming mass flux

ρ_R - the outgoing density

F - the strength of the gravitational spiral field as a fraction of the galactic axisymmetric field.

The sonic velocity was kept fixed at 10 km/sec.

A typical solution relaxed to a steady-state flow fairly rapidly, although the wholly subsonic flows took longer to relax than the transonic solutions. The density in all calculations relaxes even faster than the velocity.

Plots of velocity u versus distance x are shown for six calculations in figures 46 - 51. The first two figures show wholly subsonic flow; the others show transonic flow, both with and without a shock. Figure 52 is for the same calculations as figure 49, but shows the density instead of velocity. Table 5 shows variable values for figures 49 and 52.

Figures 46 and 47 show the smooth, subsonic type of flow. The downstream density ρ_R is set at 1.0; the entering mass flux J_L has a value of 1.0 in figure 46 and 1.28 in figure 47. Both flows show that the maximum velocity is reached at x = 10 kpc, as predicted. The flow is fairly symmetric, but the exact symmetry is broken by the artificial viscosity; as this viscosity is increased, the symmetry of the flow decreases.

The flow shown in figure 47 has almost reached the sonic speed at x = 10; if the incoming mass flux is increased slightly, it will be the maximum that can pass through the "throat" of the potential. Any attempt to increase the mass flux past this critical value results in a readjustment of the flow until the incoming flux is again smaller. To reach supersonic speeds it is necessary to lower the downstream density ρ_R, and thus remove the symmetry of the flow.

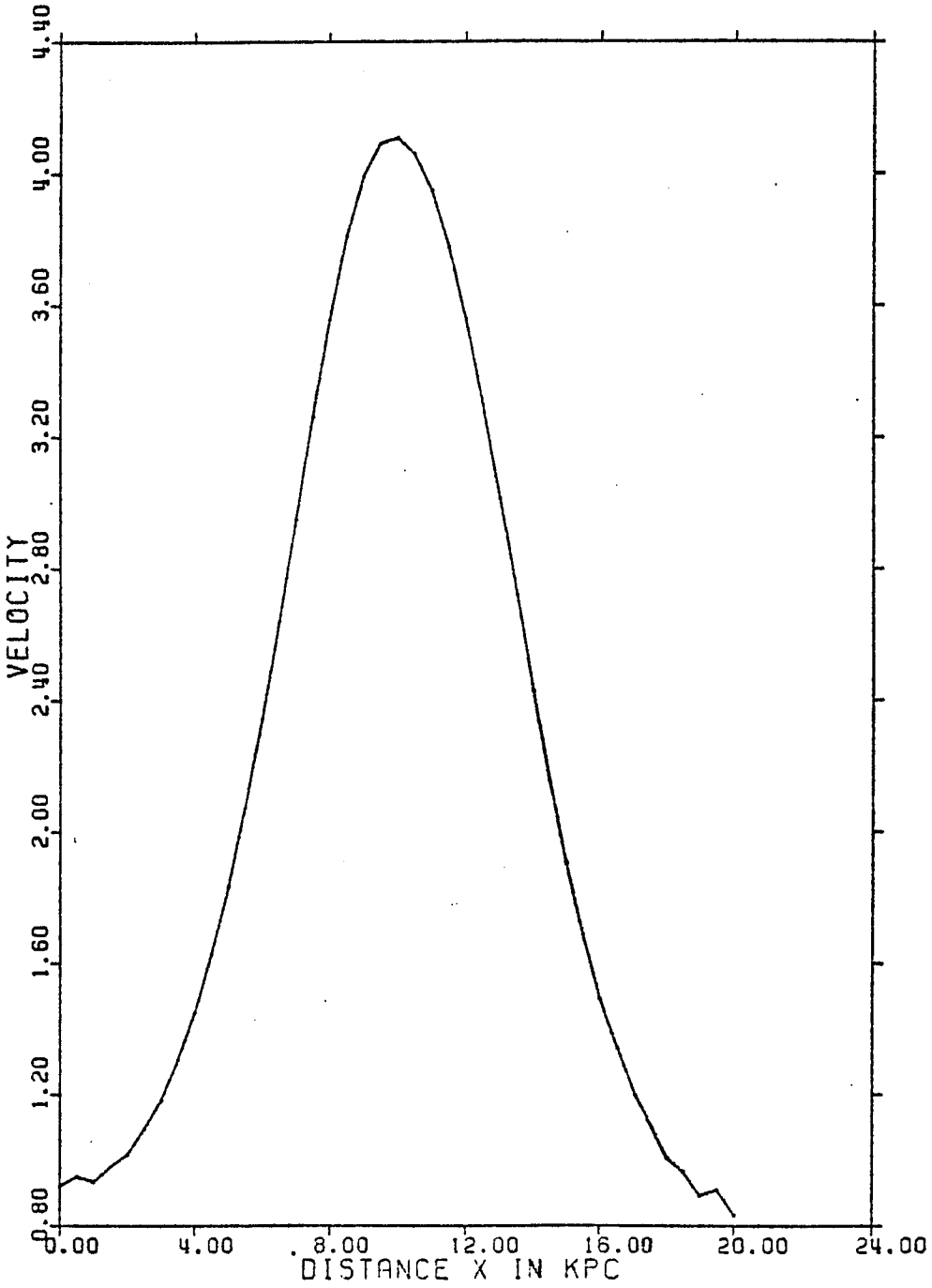


FIGURE 46: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 1.0, ρ_R = 1.0, F = 0.05.

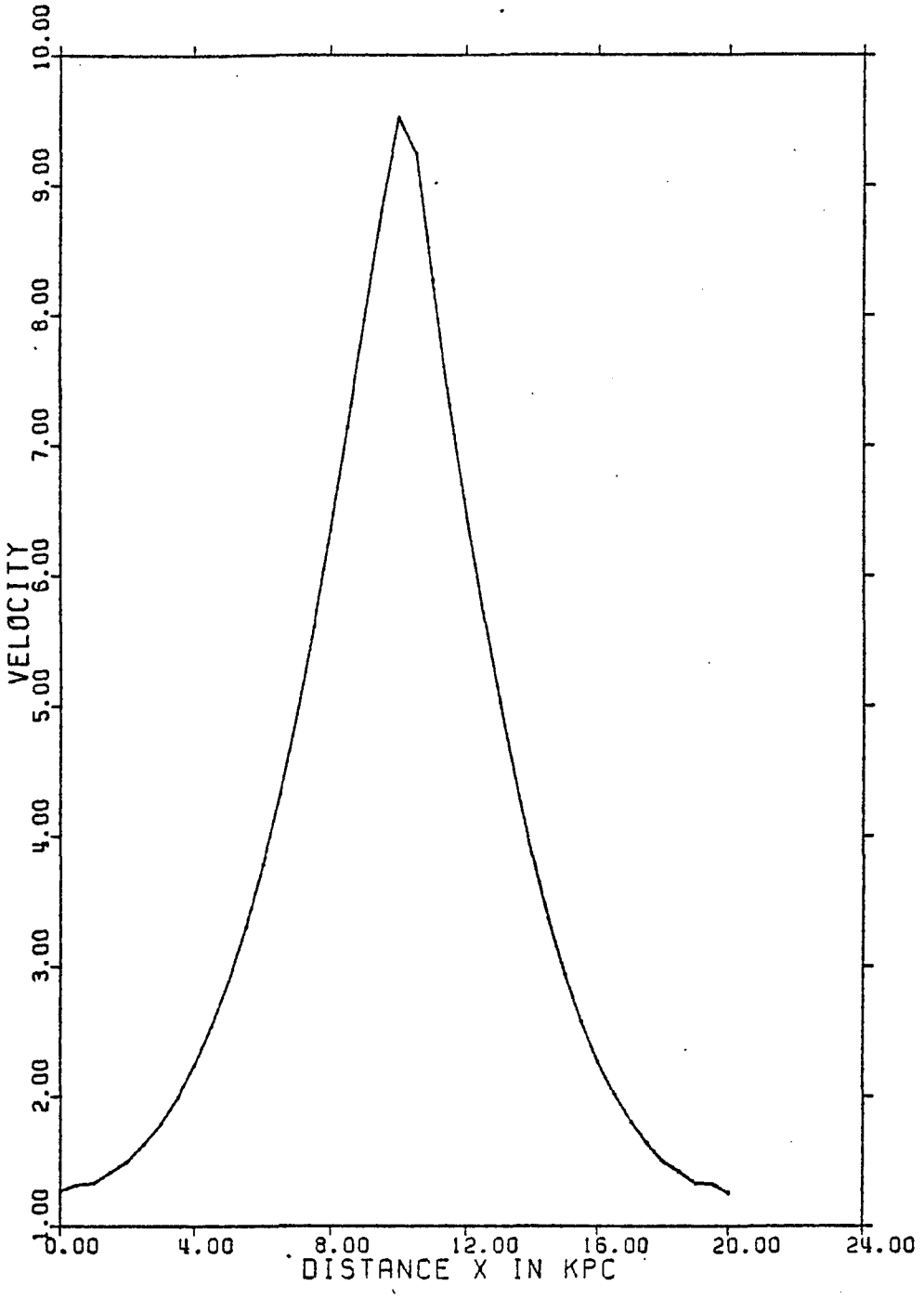


FIGURE 47: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 1.28, ρ_R = 1.0, F = 0.05.

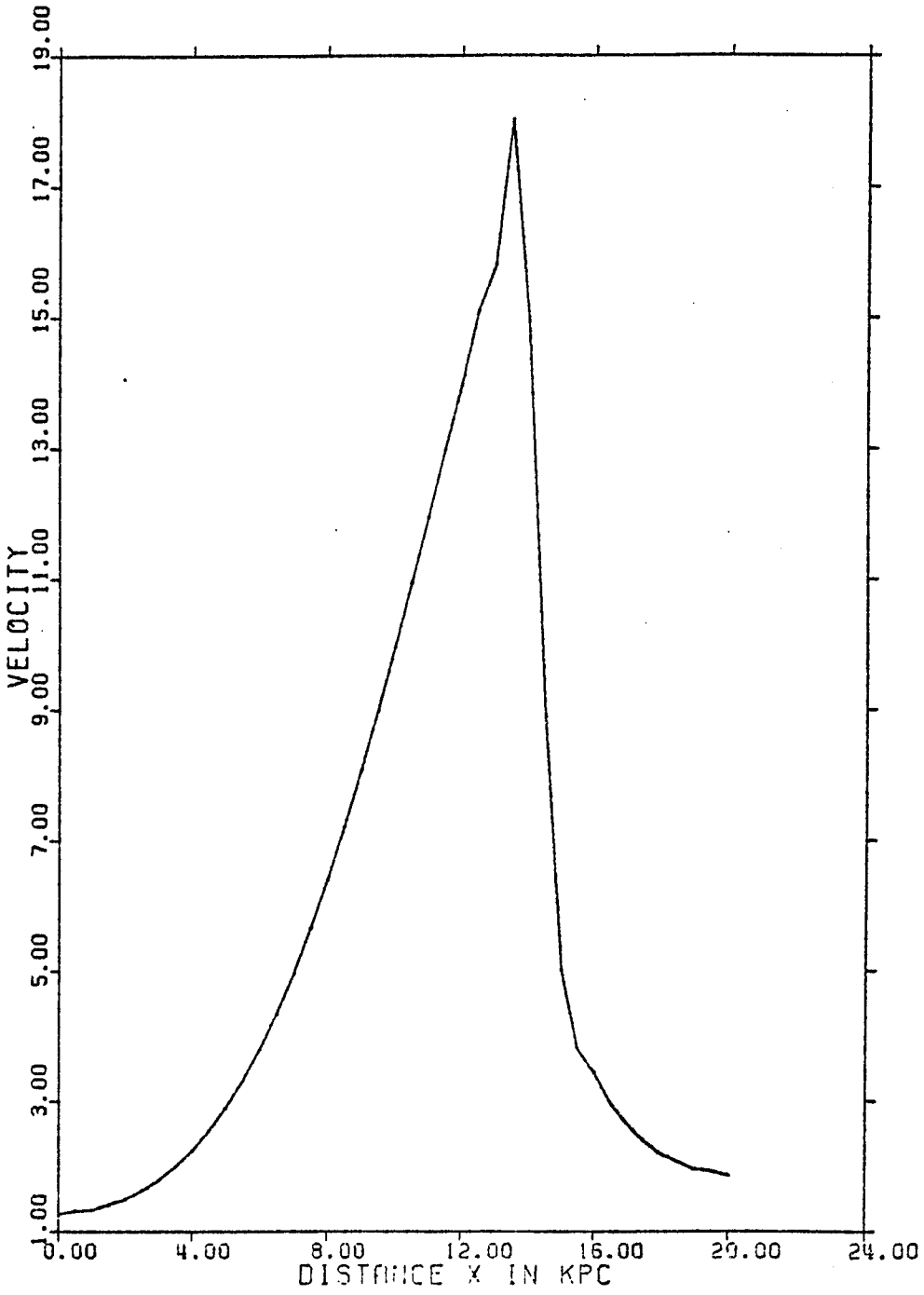


FIGURE 48: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 1.3, ρ_R = 0.7, F = 0.05.

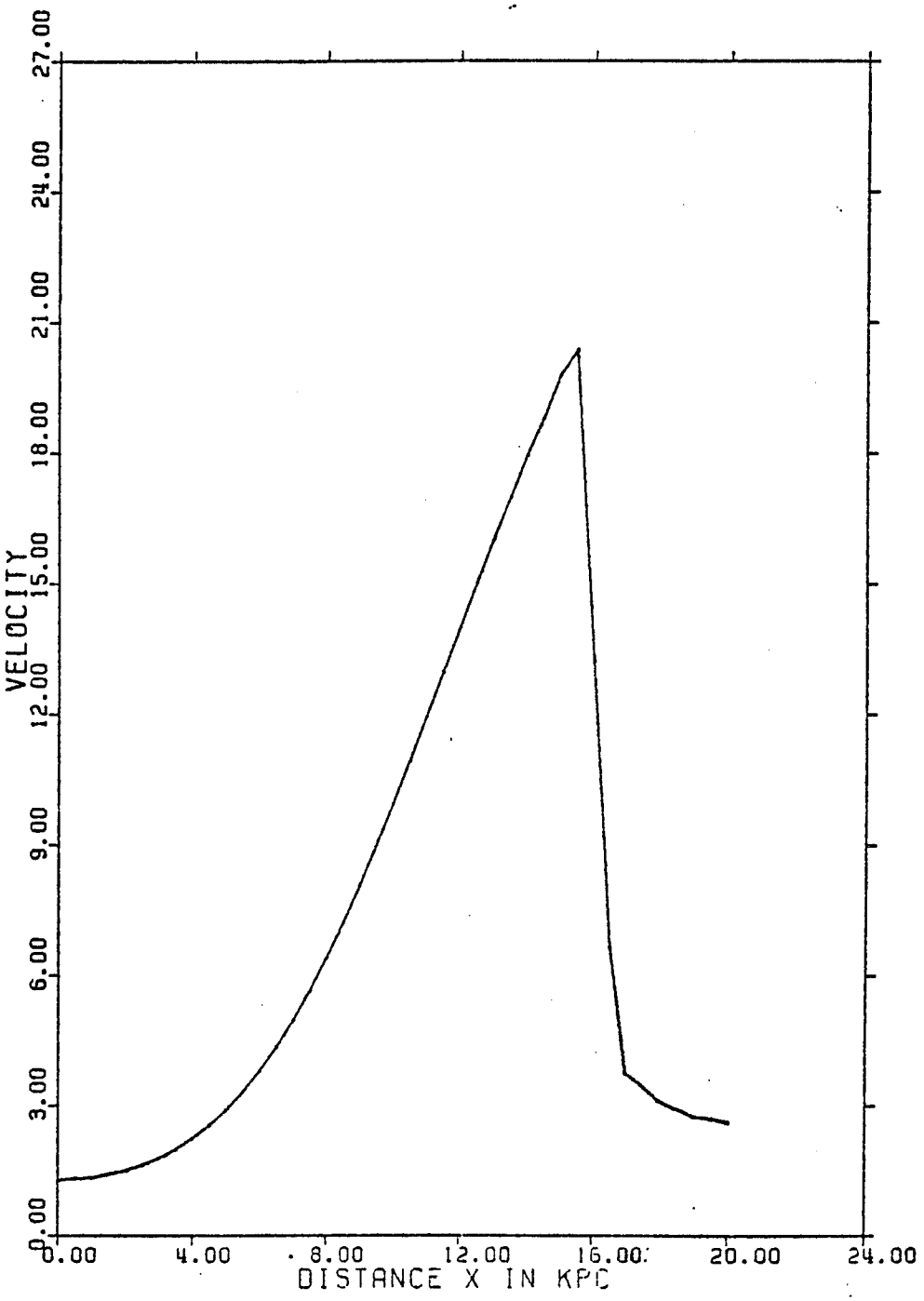


FIGURE 49: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 1.3, ρ_R = 0.5, F = 0.05.

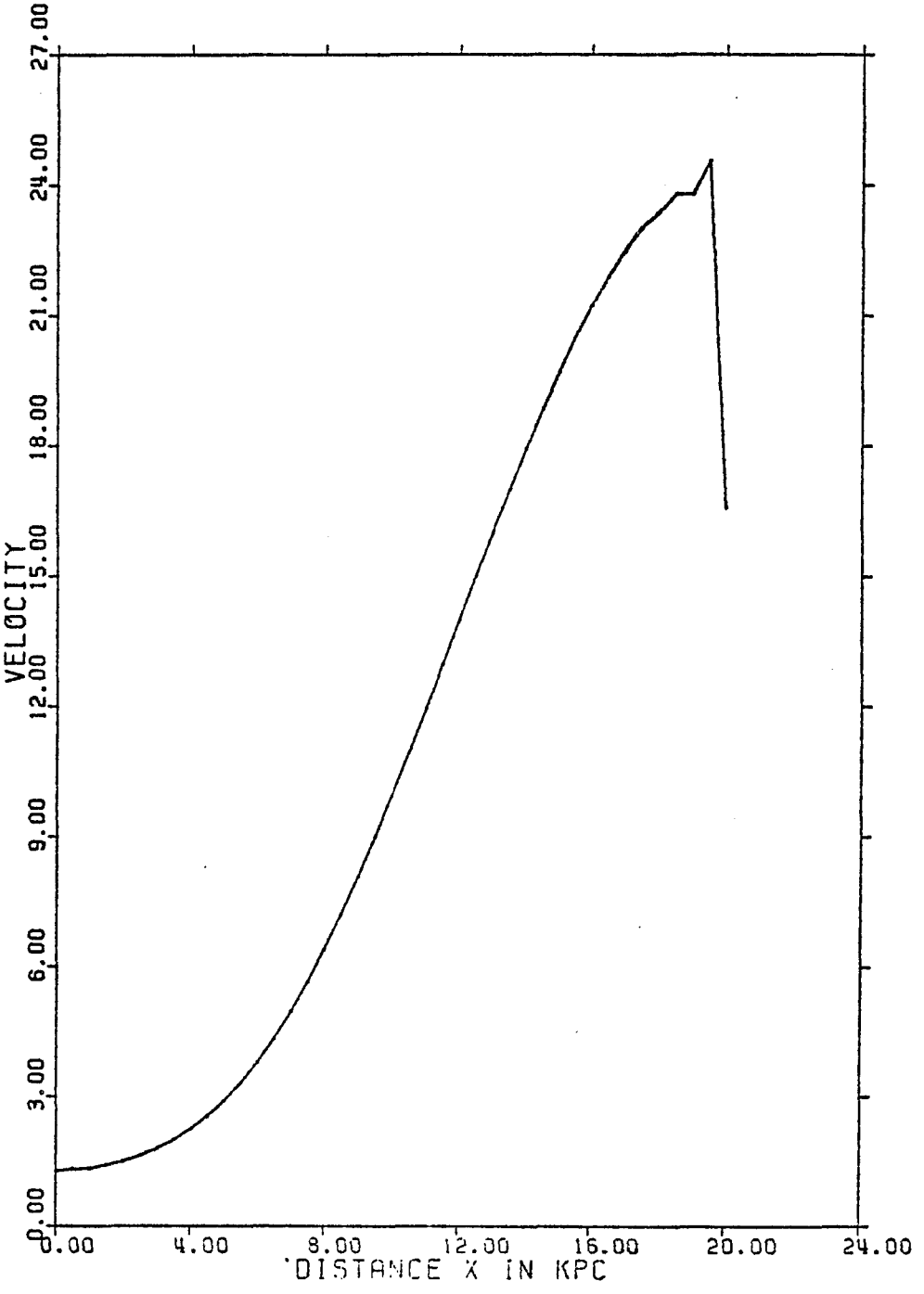


FIGURE 50: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 1.3, ρ_R = 0.1, F = 0.05.

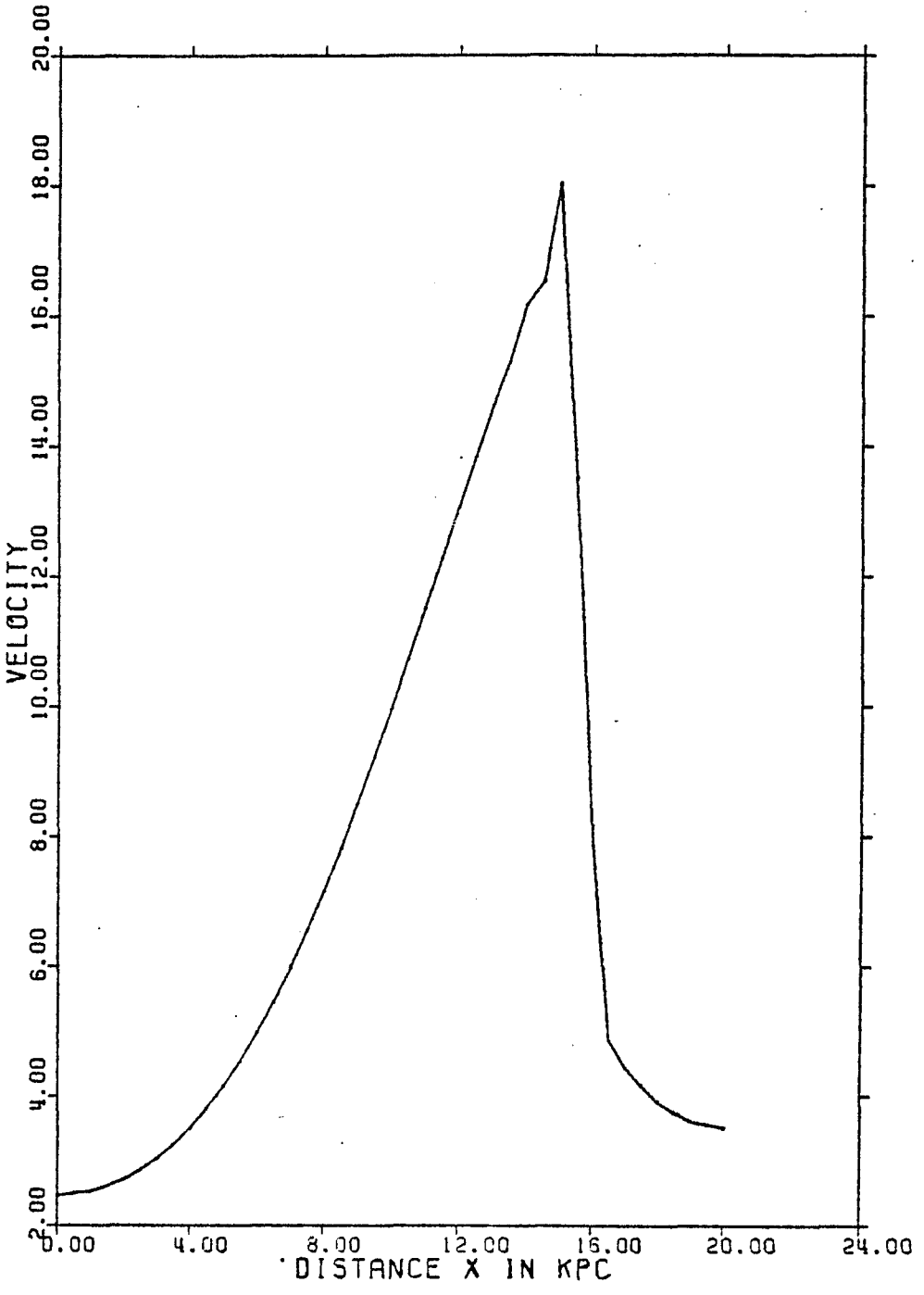


FIGURE 51: Velocity versus distance for one-dimensional flow into the spiral potential. a_o = 10 km/sec, J_L = 2.46, β_R = 0.7, F = 0.03.

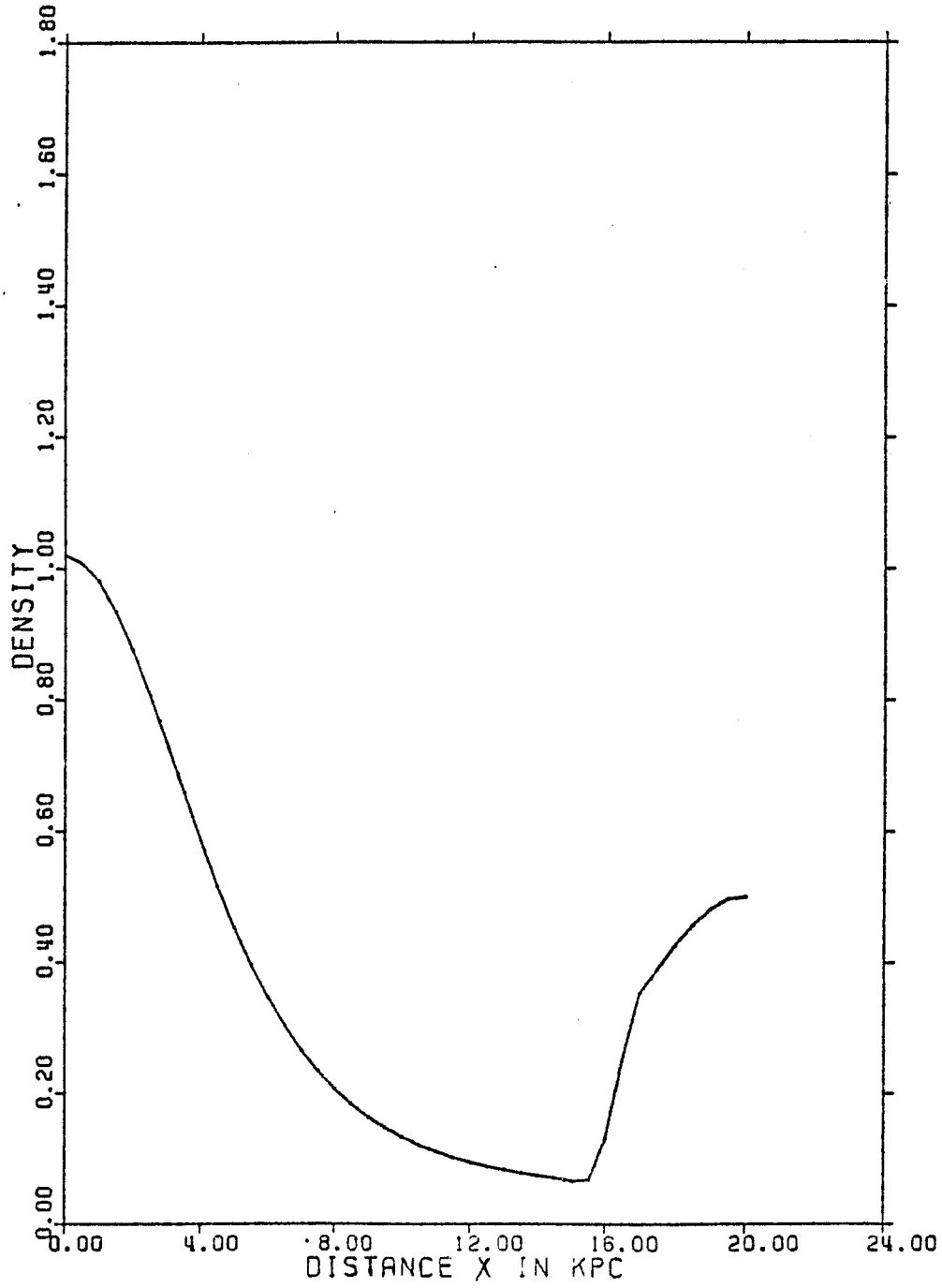


FIGURE 52: Density versus distance, for the flow shown in figure 49.

TABLE 5- VALUES OF VARIABLES FOR THE FLOW OF FIGURE 49

X	DENSITY	MOMENTUM	VELOCITY
Ø.Ø	1.Ø21	1.3ØØ	1.273
Ø.5	1.ØØ8	1.332	1.322
1.Ø	Ø.981	1.31Ø	1.335
1.5	Ø.934	1.325	1.418
2.Ø	Ø.876	1.314	1.5ØØ
2.5	Ø.8Ø7	1.32Ø	1.635
3.Ø	Ø.734	1.315	1.792
3.5	Ø.659	1.316	1.999
4.Ø	Ø.586	1.314	2.245
4.5	Ø.516	1.315	2.546
5.Ø	Ø.453	1.314	2.9ØØ
5.5	Ø.396	1.314	3.317
6.Ø	Ø.346	1.314	3.798
6.5	Ø.3Ø3	1.315	4.346
7.Ø	Ø.265	1.316	4.963
7.5	Ø.233	1.317	5.648
8.Ø	Ø.2Ø6	1.318	6.398
8.5	Ø.183	1.319	7.21Ø
9.Ø	Ø.163	1.32Ø	8.Ø77
9.5	Ø.147	1.321	8.993
1Ø.Ø	Ø.133	1.322	9.95Ø
1Ø.5	Ø.121	1.323	1Ø.938
11.Ø	Ø.111	1.324	11.95Ø
11.5	Ø.1Ø2	1.324	12.974
12.Ø	Ø.Ø95	1.325	14.ØØ1
12.5	Ø.Ø88	1.326	15.Ø21
13.Ø	Ø.Ø83	1.326	16.Ø3Ø
13.5	Ø.Ø78	1.328	16.998
14.Ø	Ø.Ø74	1.325	17.972
14.5	Ø.Ø71	1.334	18.81Ø
15.Ø	Ø.Ø66	1.311	19.8Ø2
15.5	Ø.Ø68	1.377	2Ø.367
16.Ø	Ø.131	1.733	13.234
16.5	Ø.25Ø	1.659	6.649
17.Ø	Ø.353	1.317	3.736
17.5	Ø.389	1.344	3.459
18.Ø	Ø.427	1.324	3.1Ø5
18.5	Ø.457	1.336	2.922
19.Ø	Ø.481	1.32Ø	2.745
19.5	Ø.497	1.339	2.694
2Ø.Ø	Ø.5ØØ	1.3Ø7	2.614

The transonic flows in figures 48 - 50 show the systematic change in the solution as ρ_R is lowered. In all of them, the sonic velocity is reached at x = 10 kpc, and the subsonic flow before this point is independent of ρ_R. The flows have progressively smaller values of ρ_R, of 0.7, 0.5, and 0.1. As ρ_R decreases the shock location moves downstream, because the velocity needed at the exit to conserve the mass flux J increases. Since the subsonic flow after the shock slows down, the shock position must move to the right to end up with the correct density at the exit.

In figure 50, the value of ρ_R is too low to allow the gas to adjust to it <u>via</u> a shock. This is the analogy to nozzle flow which exits into an extremely rarefied atmosphere. The flow starts to slow down as it leaves the potential, unable to adjust to the very low density on the boundary.

The shock strength (S) can be defined as the density contrast across the shock, ρ_2/ρ_1. This is theoretically equivalent to the velocity contrast, since mass is conserved across the shock, but the two ratios do not always agree in the numerical solution. The density is smoothed out across the shock slightly more than the velocity since density, as one of the three dependent variables, is affected directly by the artificial viscosity. This makes it more difficult to derive the density contrast than the velocity contrast. The shock strength S is therefore measured by

$$S \equiv \frac{u_1}{u_2} \quad , \tag{6.17}$$

where u_1 and u_2 are the velocities immediately upstream and downstream, respectively, of the shock. We expect from the Rankine jump conditions (see appendix A) that the product u_1u_2 equals a_o^2. In all cases, we find that u_1u_2 is less than a_o^2 by 10-15%, due primarily to numerical errors.

Figures 48 and 49 show the shock strength increasing from 3.5 to 5.3 as ρ_R is lowered. This is directly related to the location of the shock, since the flow is faster when the shock is further downstream.

Since F, the spiral field strength, affects the critical mass flux needed for a transonic solution as well as the shock location and strength, changes in the value of F have to be accompanied by adjustments in the upstream mass flux. The flow in figure 51 has the same downstream density ρ_R as in figure 48, but F was decreased from 5% to 3% and the incoming mass flux was increased just enough to support a transonic solution. The shock location is further upstream than in figure 48, and the shock strength is smaller due partly to the new shock position and partly to the more gently sloping potential. The shock location is now the same as in figure 49, but the maximum velocity is smaller.

Finally, figure 52 shows the same solution as figure 49, but the density is plotted rather than velocity, showing the more rapid change across the shock in velocity.

These flow solutions show that the numerical method is qualitatively dependable; in addition, its accuracy was checked by

comparing numerically computed values with the analytic values. Calculation of the shock position x_s, the maximum velocity u_1, and the boundary velocities u_L and u_R for transonic flow is easy:

Step 1: Evaluate equation (6.13) at $x = 0$ to get the upstream velocity u_L, using $\mathcal{V} = N_o(1 - \cos \alpha x)$:

$$\frac{1}{2}(u_L^2 - a_o^2) - a_o^2 \ln(u_L/a_o) = 2N_o \quad . \tag{6.18}$$

Step 2: The mass flux $J = \rho_L u_L = \rho_R u_R$ is known, so we solve for the downstream velocity:

$$u_R = \frac{J}{\rho_R} \quad . \tag{6.19}$$

Step 3: Eliminate $V(x_s)$ between equations (6.14) and (6.16), and use the jump condition across the shock, $u_1 u_2 = a_o^2$, to get an equation for u_1:

$$\frac{1}{2}(u_1^2 - a_o^2) - a_o^2 \ln\left(\frac{u_1}{a_o}\right) - 2N_o$$

$$= -\frac{1}{2}\left(u_R^2 - \frac{a_o^4}{u_1^2}\right) + a_o^2 \ln\left(\frac{u_R u_1}{a_o^2}\right) \quad . \tag{6.20}$$

Step 4: Put u_1 back into equation (6.14) to solve for x_s:

$$2N_o - \frac{1}{2}(u_1^2 - a_o^2) + a_o^2 \ln\left(\frac{u_1}{a_o}\right) = N_o(1 - \cos \alpha x_s) \tag{6.21}$$

Table 6 shows the comparison of computed and analytic

TABLE 6

COMPARISON OF ANALYTIC AND NUMERICAL VALUES
AS A CHECK OF SCHEME ACCURACY

		Analytic	Numerical
From Figure 49:			
Upstream velocity	u_L (km/s)	1.3	1.25
Downstream velocity	u_R (km/s)	2.6	2.66
Velocity before shock	u_1 (km/s)	21.5	21.2
Shock location	x_s (kpc)	16.1	15.5 $\pm$0.5*
From Figure 51:			
Upstream velocity	u_L (km/s)	2.5	2.46
Downstream velocity	u_R (km/s)	3.5	3.51
Velocity before shock	u_1 (km/s)	18.3	18.0
Shock location	x_s (kpc)	15.7	15.0 $\pm$0.5*

Source: Figures 49 and 51.

* The error due to the finite mesh size is 0.5 kpc.

values for the flows in figures 49 and 51. The agreement, even with the errors due to the finite mesh size, boundary effects, and the smearing due to the artificial viscosity, is excellent.

CHAPTER 7: THE FULL TWO-DIMENSIONAL CALCULATION

7.1 Introduction

The results of the one-dimensional problems discussed in Chapters 5 and 6 show excellent agreement with the known solutions. The numerical scheme has proved so far to be both accurate and stable, and the spatial resolution of the shock is good. In addition, a great deal of knowledge has been gained about how to handle the initial conditions and the boundary conditions, about the effects of the artificial viscosity term, and some of the difficulties that may arise from having a strong source term in the equations. This experience is indispensable in handling the two-dimensional flow.

Let us review the problem to be investigated. The gas enters a region in which the gravitational potential is given by an axisymmetric portion and a spiral portion, namely,

$$\Psi = (2\pi G\sigma_*)\,\Delta \ln\left(\cosh(\gamma/\Delta)\right) + N_0(1-\cos\alpha x)\,\mathrm{sech}^{|k|\Delta}(\gamma/\Delta)\,. \qquad (7.1)$$

This is a rectangular region, bounded by x = 0, x = 20 kpc, z = 0 and z = 1.0 kpc, and covered with a computational mesh whose grid points are a distance Δx or Δz apart. In most of the calcula-

tions, $\Delta x = 1.0$ kpc and $\Delta z = 0.1$ kpc. The gas enters at $x = 0$ (the left-hand boundary, L) from a region in which the spiral potential is zero, and leaves at $x = 20$ kpc (the right-hand boundary, R) into a similar region. The flow is assumed symmetric with respect to $z = 0$, so we do not need to extend the mesh below the plane.

7.2 The Two Models

Two different thermodynamic models for the gas are considered, as described in section 3.4. Model A is isothermal, with a constant sonic velocity $a_o = 10$ km/sec; Model B has a temperature, and therefore sonic velocity, distribution that depends on z. This distribution is chosen to be

$$a^2 = a_H^2 - (a_H^2 - a_o^2)\,\mathrm{sech}(z/\Delta) \quad . \tag{7.2}$$

The equations governing the flow for both models are

$$\frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{7.3a}$$

$$\frac{\partial(\rho u)}{\partial t} + \frac{\partial(\rho u^2 + a^2\rho)}{\partial x} + \frac{\partial(\rho u w)}{\partial z} = -\rho \frac{\partial \mathcal{V}}{\partial x} \tag{7.3b}$$

$$\frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho u w)}{\partial x} + \frac{\partial(\rho w^2 + a^2\rho)}{\partial z} = -\rho \frac{\partial \mathcal{V}}{\partial z} \quad . \tag{7.3c}$$

The gas entering the spiral potential field at $x = 0$ is assumed to be the steady-state flow in the axisymmetric gravitational potential. If we also assume that the incoming flow is par-

allel to the plane (i.e., w = 0) and independent of x, we can solve equations (7.3) for the hydrostatic solution. The solution for the velocity, for both models, is

$$u = f(x) \tag{7.4}$$

where f(x) is an arbitrary function of x.

The solution for the density was already obtained in section 5.2; the results are:

$$\rho = \rho_0 \operatorname{sech}^{\beta_0^2}(z/\Delta) \qquad \text{for Model A} \tag{7.5a}$$

$$\rho = \rho_0 \left(\frac{a_0^2}{a^2(z)}\right)^{\delta+1} \operatorname{sech}^{\delta}(z/\Delta) \qquad \text{for Model B} \tag{7.5b}$$

where ρ_0 is a constant,

$$\beta_0^2 = \frac{(2\pi G \sigma_*)\,\Delta}{a_0^2} \tag{7.6a}$$

and

$$\delta = \frac{\beta_0^2 a_0^2}{a_H^2} \quad . \tag{7.6b}$$

Because the hydrostatic density is stratified, it is more convenient to change the density variable from ρ to $\hat{\rho}$ (just as in section 5.3), where

$$\rho = \hat{\rho}\ \mathrm{sech}^{\beta_0^2}(\zeta/\Delta) \qquad \text{for Model A} \qquad (7.7a)$$

$$\rho = \hat{\rho}\left(\frac{a_0^2}{a^2(\zeta)}\right)^{\delta+1} \mathrm{sech}^{\delta}(\zeta/\Delta) \qquad \text{for Model B.} \qquad (7.7b)$$

Equations (7.3) can now be rewritten in terms of the new variables $\hat{\rho}$, $\hat{\rho}u$, and $\hat{\rho}w$. These new equations have the same form as the originals, except that the right-hand side (the source term) is now different:

$$\frac{\partial\hat{\rho}}{\partial t} + \frac{\partial(\hat{\rho}u)}{\partial x} + \frac{\partial(\hat{\rho}w)}{\partial\zeta} = S_1 \qquad (7.8a)$$

$$\frac{\partial(\hat{\rho}u)}{\partial t} + \frac{\partial(\hat{\rho}u^2+a^2\hat{\rho})}{\partial x} + \frac{\partial(\hat{\rho}uw)}{\partial\zeta} = S_2 \qquad (7.8b)$$

$$\frac{\partial(\hat{\rho}w)}{\partial t} + \frac{\partial(\hat{\rho}uw)}{\partial x} + \frac{\partial(\hat{\rho}w^2+a^2\hat{\rho})}{\partial\zeta} = S_3 \qquad (7.8c)$$

The new source term is given, in vector form, by

$$\begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} \frac{\beta_0^2}{\Delta}\,\hat{\rho}w\tanh(\zeta/\Delta)\,[P(\zeta)] \\ \frac{\beta_0^2}{\Delta}\,\hat{\rho}uw\tanh(\zeta/\Delta)\,[P(\zeta)] - \hat{\rho}\alpha N_0 \sin\alpha x\ \mathrm{sech}^{|k|\Delta}(\zeta/\Delta) \\ \frac{\beta_0^2}{\Delta}(\hat{\rho}w^2+a^2\hat{\rho})\tanh(\zeta/\Delta)[P(\zeta)] \\ \qquad -\hat{\rho}\tanh(\zeta/\Delta)\left[\frac{\beta_0^2}{\Delta}a_0^2 - |k|N_0(1-\cos\alpha x)\mathrm{sech}^{|k|\Delta}(\zeta/\Delta)\right] \end{pmatrix} \qquad (7.9)$$

where

$$P(\zeta) = \frac{a_o^2}{a_H^2} + \left(\frac{a_o^2}{a_H^2} + \frac{1}{\beta_o^2} \right) \frac{(a_H^2 - a_o^2)}{a^2(\zeta)} \operatorname{sech}(\zeta/\Delta) . \qquad (7.10)$$

For Model A, since $a(z) = a_o = a_H$, we have $P(z) \equiv 1$. The values of P(Z) for Model B depend on both a_H and z; e.g., for $a_H = 30$, P(Z) ranges from 2.2 at z = 0 to 0.16 at z = 1.0 kpc. Since $\tanh(z/\Delta)$ is zero on the plane, the source terms are identical at z = 0, and differ more and more as z increases. The first two components, S_1 and S_2, are smaller in Model B than in Model A; the third component, S_3, is much larger at large z in Model B.

This gives a hint of future problems, since, as we saw in Chapter 5, the Lax-Wendroff Two-Step method (like most finite-difference methods) is not able to handle large source terms very well. If a problem arises, this analysis tells us it is more likely to occur in Model B than in Model A, and at large z. The component S_3 in Model B is largest at large z for x = 0 or 20. In fact, a problem does arise, in the predicted area, for large values of the source term. It is solved in the same manner as for the hydrostatic atmosphere -- by decreasing the mesh size Δz.

7.3 The Numerical Method

We have removed the hydrostatic non-spiral portion of the density solution, and are now concerned only with the coefficient $\hat{\rho}$. All further references to density in this chapter will therefore refer to the coefficient $\hat{\rho}$ unless it is explicitly stated that the values refer to the "real densities."

The finite-difference equations for the new variables $\hat{\rho}$, $\hat{\rho}u$ and $\hat{\rho}w$ are obtained from equations (7.8) and (7.9) in the manner described in section 4.2, and will not be repeated here.

As in most numerical calculations of transonic flow, the boundary conditions pose the most difficult problems. The boundary conditions to be imposed on two sides of the flow region (see [i], [iii] below) are fairly straightforward; they follow directly from the knowledge of the one-dimensional flows discussed in Chapter 5 and 6. The other two boundaries (see [ii], [iv]) are more difficult.

(i) On the galactic plane ($z = 0$), since we assumed a symmetric flow, as in the one-dimensional atmosphere, we specify $w = 0$.

(ii) On the upstream boundary, the one-dimensional nozzle flow suggests that we should not specify all variables for incoming subsonic flow. Ganz and Serra (1974) showed that for subsonic flow entering an axisymmetric nozzle, the density and only a linear combination of the velocity components should be specified.

However, for the sake of stability it was necessary to use different boundary conditions. For Model A, we specified $w = 0$ and the incoming mass flux $J_L = \hat{\rho}_L u_L$ = constant (as in the one-dimensional flow into a spiral potential). The density $\hat{\rho}_L$ was calculated by extra-

polation from the interior flow; it turned out to be almost independent of z.

For Model B, which is more prone to instabilities because of the varying sound speed over the mesh, the boundary was overspecified by setting $\hat{\rho}_L = 1.0$, u_L = constant and $w = 0$. If u_L was set at too different a value from the interior solution, the flow adjusted itself within one or two mesh points to the interior solution. This kind of behavior gives confidence that the overspecification on the boundary does not critically alter the interior flow solution.

(iii) On the downstream boundary, we again follow the nozzle flow conditions and specify only the density $\hat{\rho}_R$ = constant. This is, of course, the solution for two-dimensional flow parallel to the plane in an axisymmetric gravitational field, and it is not specified here that the gas leaving the spiral potential must be horizontal. However, the numerical results confirm that the magnitudes of the exit z-velocities are small.

(iv) The "top" boundary at $z = z_{max}$ cannot be handled by referring to the one-dimensional problems. There are many possible ways to handle it, but the best way found (i.e., most stable and accurate) for Model A was not to specify anything at all, but let the interior flow determine the

values of the variables at $z = z_{max}$. This approach was confirmed by tests in which one or both of the velocities u and w were specified on the top boundary. It was found that either the calculation blew up immediately (i.e., velocities became exceedingly large), or the interior solution ignored the specified values, changing very abruptly (within 1 mesh length Δz) from the interior to the boundary values. This sharp transition was very striking; often the interior flow was highly supersonic and the imposed boundary values were subsonic. This shows the stability and flexibility of the numerical scheme with respect to the imposed boundary conditions.

Again, Model B is less stable than Model A, and we need to overspecify the top boundary by setting the values of u there to be subsonic (the z-velocity w was kept free). This kept the problem at the boundary under control long enough for the interior flow to reach steady state, although the calculation was never stable enough to run several thousand iterations.

The initial values of the variables were chosen with the one-dimensional solutions in mind; i.e., that $J = \hat{\rho} u$ = constant. As in the other problems, increasing the spiral force gradually offered no advantage.

The time step Δt was chosen as described in section 4.3; separate linear time conditions were estimated for the x and z di-

rections, and the smaller of the two was chosen as Δt. The value used was 3×10^6 years.

The strength of the artificial viscosity, measured by B, was chosen to be B = 4 as a compromise between stability and poor spatial resolution (high B) and instability with high spatial resolution (low B). The instability at low B values showed up as oscillations in w, with a wavelength of one mesh spacing.

7.4 Results

A typical solution for both Model A and Model B converged to a steady state within 800 to 1000 time steps. Model A was exceptionally stable, even at large values of z, and remained at the steady-state values (with oscillations of less than 2%) for as long as the calculation was allowed to run. The calculation for Model B, on the other hand, was much more unstable and would not remain at the steady-state values for more than about 300 time iterations; at this point an instability at the top boundary would invariably develop and spread downward until it destroyed the steady-state flow.

The flow depends on a number of parameters, whose effects on the flow vary considerably. The value of $|k|\Delta$, as mentioned in section 3.3, can vary from 0.5 to 2.0. However, since the final steady flow is only changed quantitatively by about 2% for a 50% change in $|k|\Delta$, and not changed qualitatively at all, the results shown were all computed with $|k|\Delta = 1.0$.

The value specified at the left-hand boundary for the incoming mass flux J_L (for Model A) or incoming density $\hat{\rho}_L$ and velocity u_L (for Model B) did not affect the interior solution very much. As long as the values chosen were not so large as to "choke" the flow (see section 6.6), the flow adjusted itself within 2 mesh points to the interior steady-state solution. For all the solutions shown below, the values were kept constant at the minimum needed to produce a shock solution.

The value of the axisymmetrical gravitational force depends on σ_*, which varies throughout the Galaxy. Some computations were made with large values of σ_*, and they showed, surprisingly, that its effect is small. The steady-state flow changes by only 10% when σ_* changes by 50%. We have already discussed (sections 5.4 and 7.2) the numerical problems associated with a large value of σ_*, which are solved by decreasing the mesh size Δz. This demands several compensations: (a) to cover the same region, the number of mesh points must increase; (b) the allowed time step Δt decreases as Δz decreases (from the Courant condition, equation [4.13]); and (c) usually, the artificial viscosity must be made larger to prevent oscillations of the system, which further decreases the allowed time step. Thus, the computer time needed for the computation rapidly becomes prohibitive. Therefore, all results shown are for a constant, relatively small value: $\sigma_* = 50\ M_\odot/pc^2$, corresponding to R = 13 kpc.

The parameters that <u>do</u> affect the flow substantially are

the specified outgoing density $\hat{\rho}_R$, the strength of the spiral field as measured by F, and for Model B, the temperature distribution with z as measured by the sonic velocity at $z = \infty$, a_H. Flows were calculated with different values of these parameters; representative results of the computations are presented in figures 53 - 55 for Model A and in figures 56 - 60 for Model B.

(a) Model A (Isothermal)

Figure 53 shows a typical solution for an isothermal gas: figure 53a is a plot of the Mach number at every mesh point, and figure 53b shows the (real) density as a function of x for selected z values; figure 53c is a sketch of the streamlines in which the z-velocities were exaggerated (their maximum values are less than 2 km/sec).

Figures 54 and 55 are Mach number plots for flows in which the outgoing density is lower, and the spiral field strength is increased, respectively.

(b) Model B (Varying Temperature)

Figure 56 shows a typical solution for a gas whose temperature increases with z. Again, the z-velocities were exaggerated in the streamlines in figure 56c, since the maximum value is again less than 2 km/sec. The flows shown in figures 57 and 58 have lower outgoing density $\hat{\rho}_R$ and stronger spiral field F, respectively. Figures 59 and 60 show the results of raising or lowering the value of a_H.

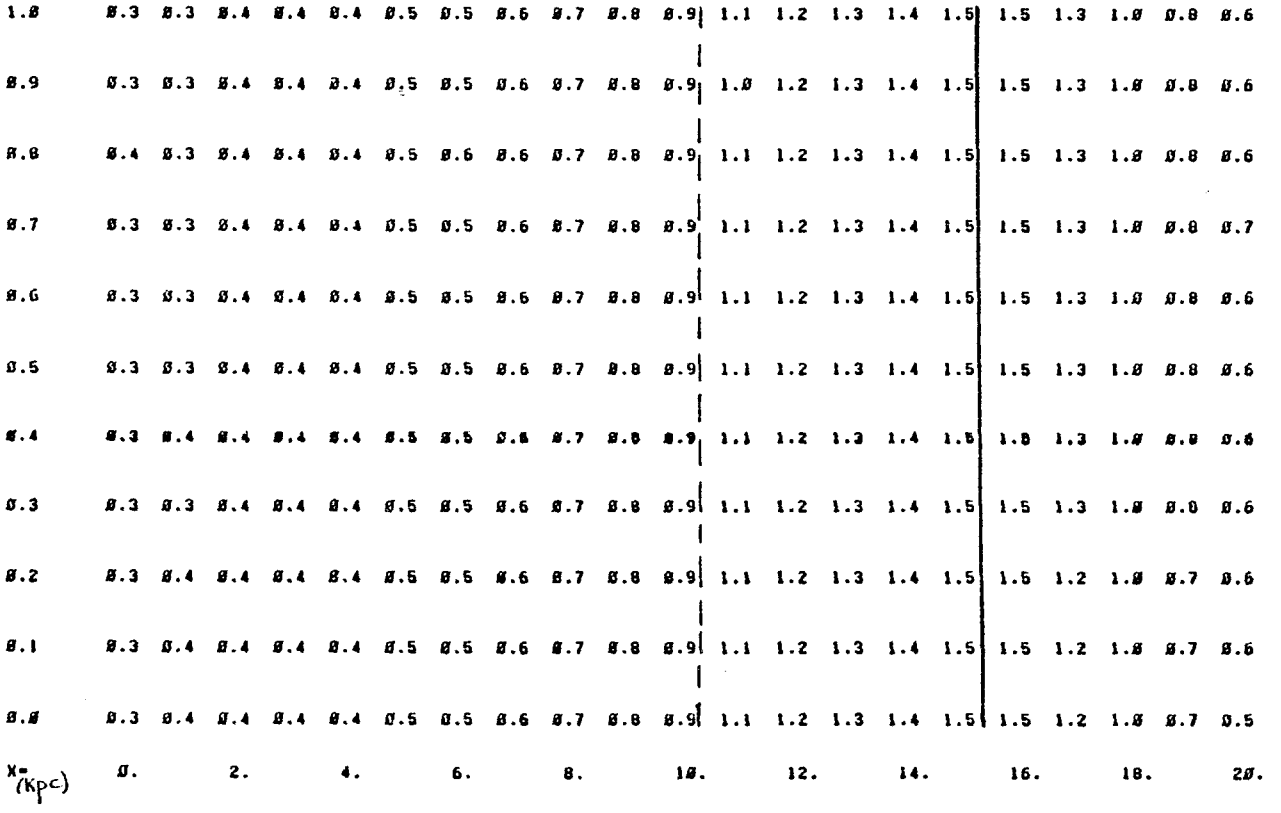
FIGURE 53 a.

MODEL A: $\hat{\rho}_R = 0.5$; $F = 0.05$

sonic line: - - - -

shock front: ——

Mach number as a function of x and z



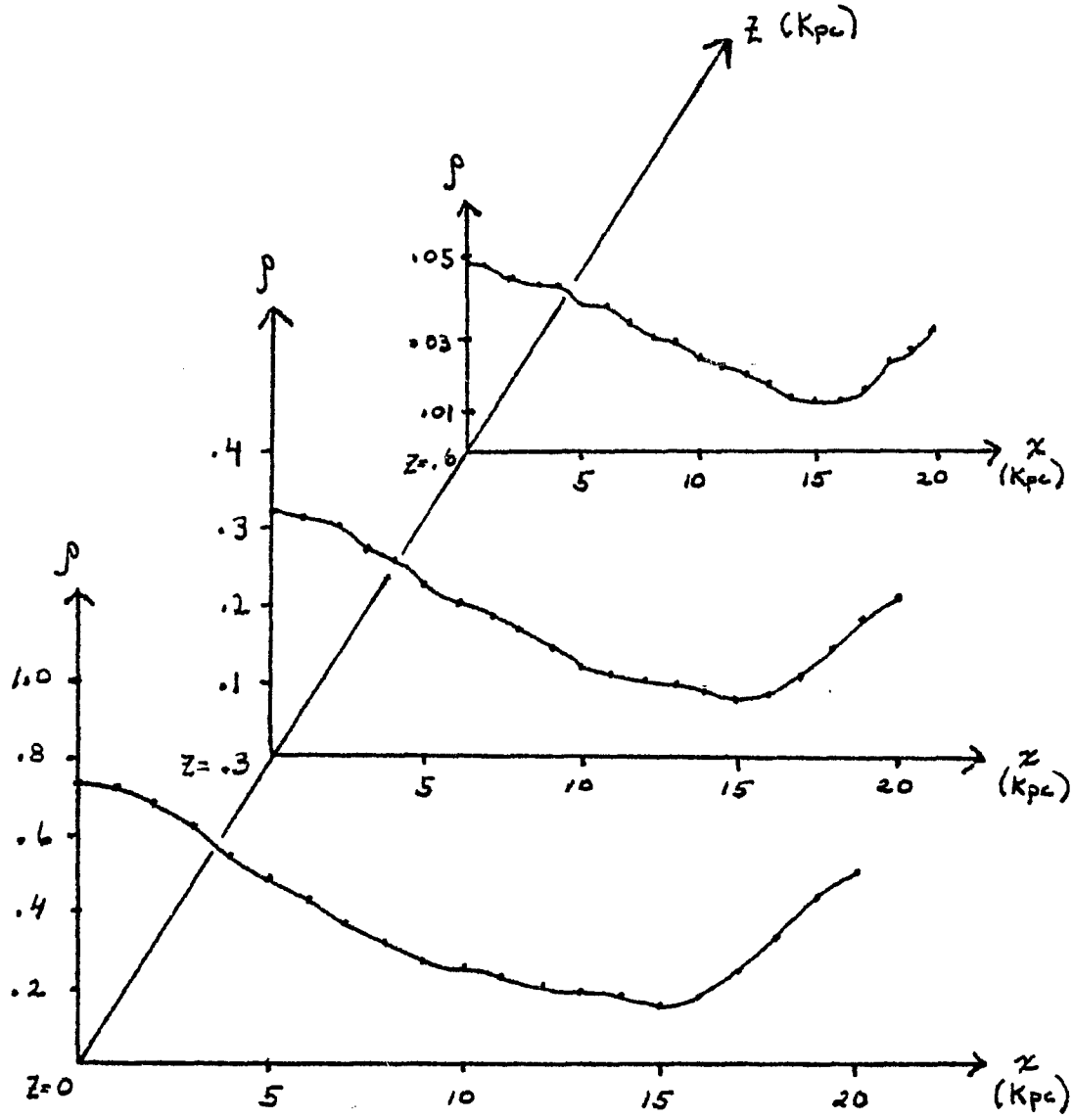


FIGURE 53 b: Real densities as a function of x and z, for flow shown in figure 53a.

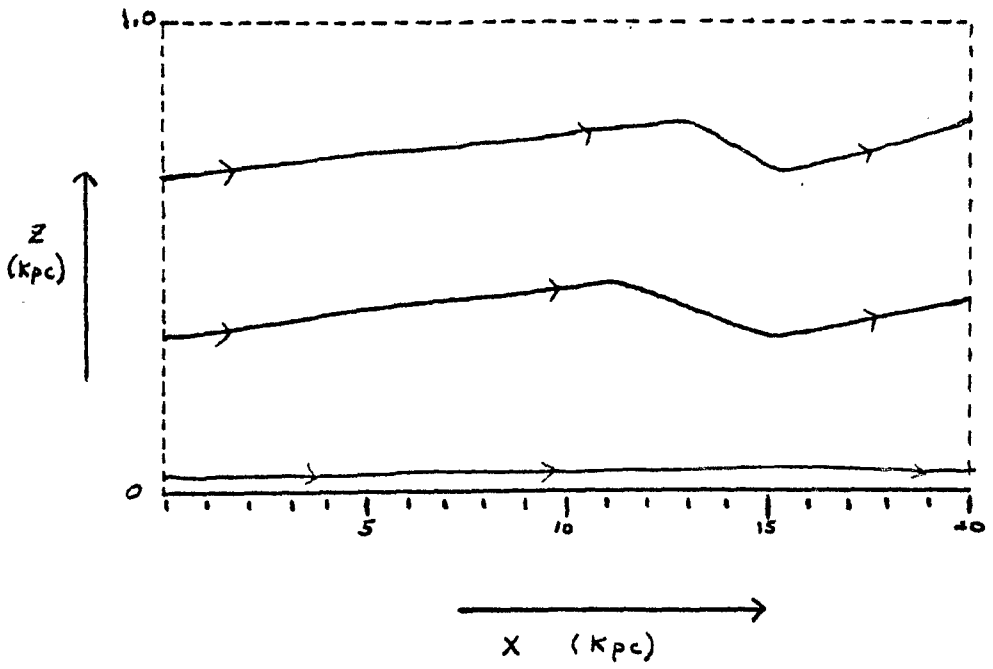


FIGURE 53c: Streamlines for the flow in figure 53a. The z-motions have been exaggerated.

FIGURE 54

MODEL A: $\hat{\rho}_R = 0.4$; $F = 0.05$

sonic line: - - - -

shock front: ——

Mach number as a function of x and z

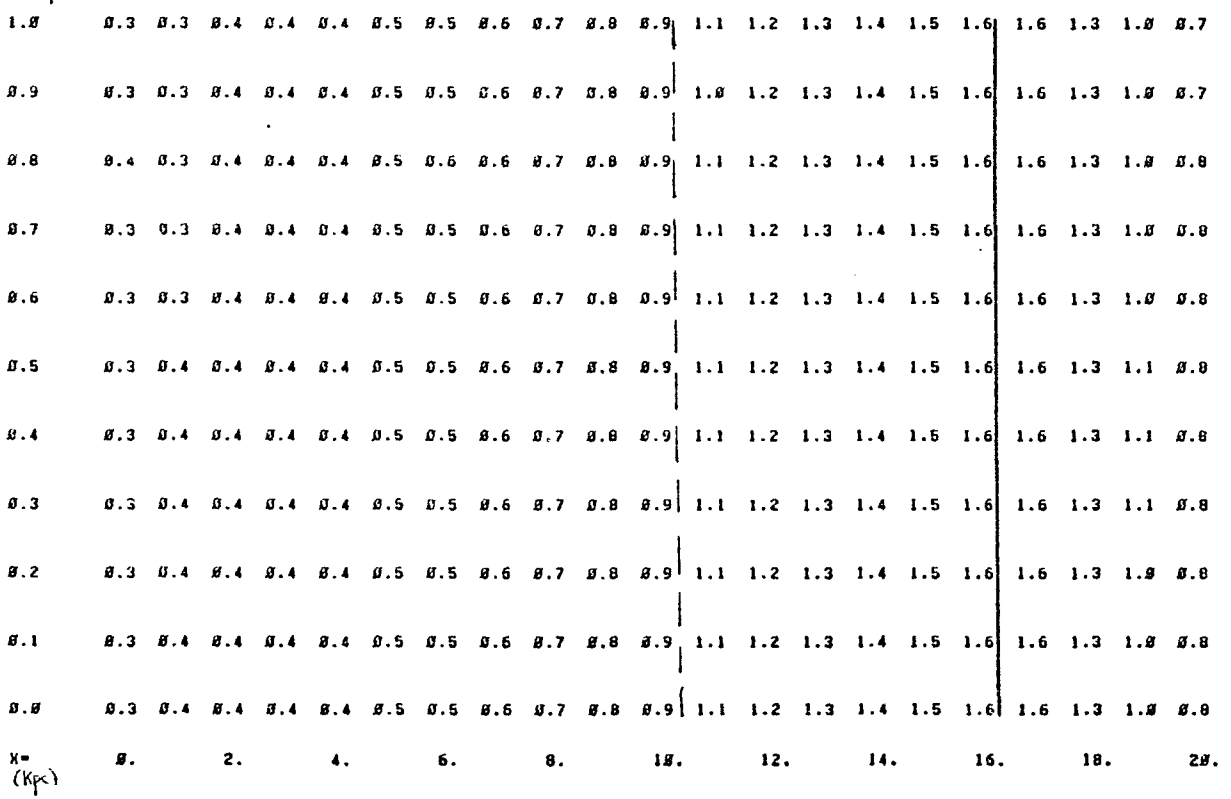


FIGURE 55

MODEL A: $\hat{\rho}_R = 0.5$; $F = 0.07$

sonic line: — — —
shock front: ———

Mach number as a function of x and z

z(Kpc)																					
1.0	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.6	0.6	0.8	0.9	1.1	1.2	1.4	1.5	1.6	1.5	1.3	1.3	1.0	0.9
0.9	0.2	0.3	0.3	0.3	0.4	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.1	1.0	0.9
0.8	0.3	0.3	0.3	0.3	0.4	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.0	0.9	0.8
0.7	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.0	0.9	0.8
0.6	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.0	0.8	0.7
0.5	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.0	0.8	0.7
0.4	0.3	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.5	1.2	1.0	0.8	0.7
0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.4	1.2	0.9	0.7	0.6
0.2	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.4	1.1	0.9	0.6	0.5
0.1	0.2	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.4	1.1	0.8	0.5	0.4
0.0	0.3	0.3	0.3	0.3	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.1	1.2	1.3	1.5	1.6	1.4	1.1	0.7	0.5	0.3
x= (Kpc)	0.		2.		4.		6.		8.		10.		12.		14.		16.		18.		20.

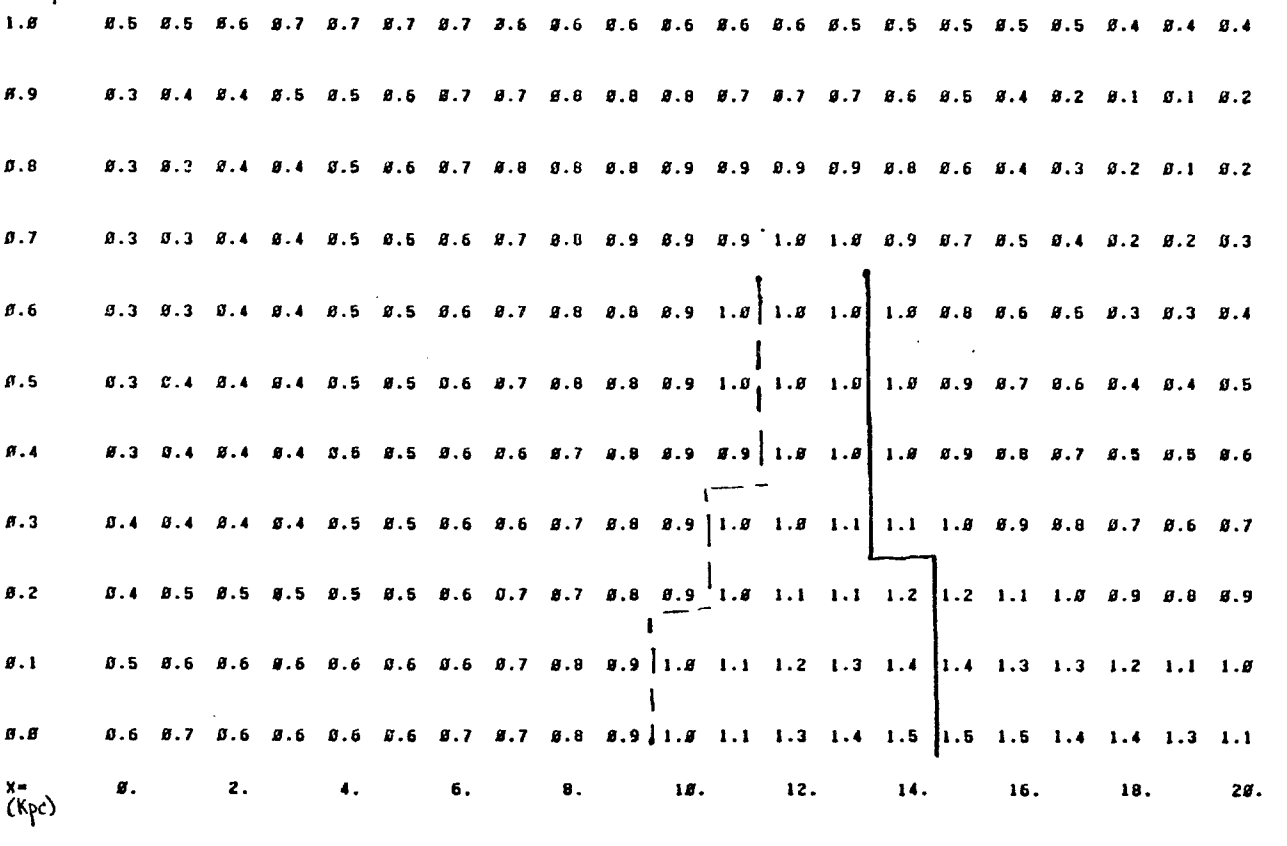
FIGURE 56 a.

MODEL B: $a_o = 10$ km/s ; $a_H = 25$ km/s ; $\hat{\rho}_R = 0.8$; $F = 0.05$

sonic line: - - - -

shock front: ————

Mach number as a function of x and z



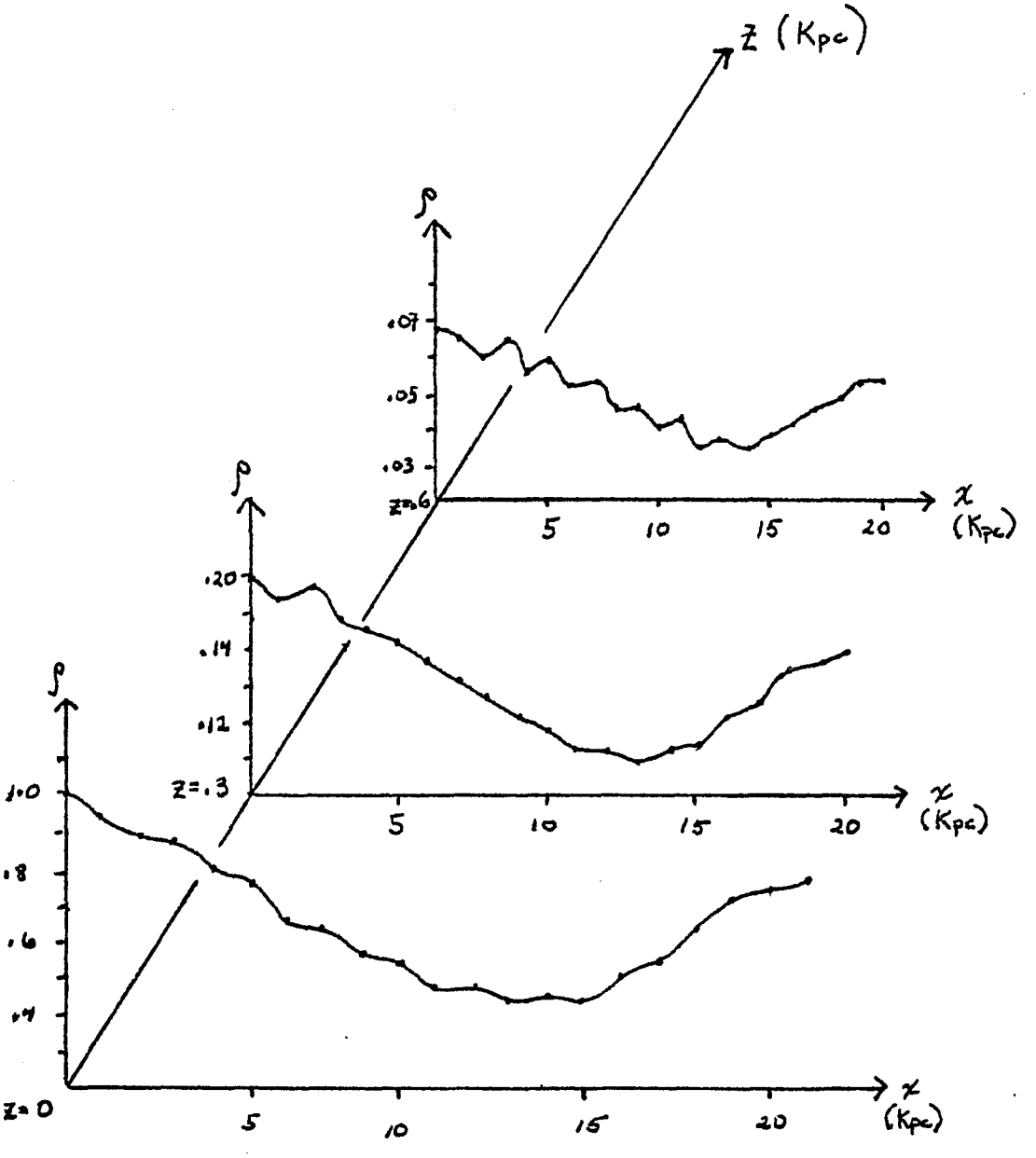


FIGURE 56b: Real densities as a function of x and z, for the flow shown in figure 56a.

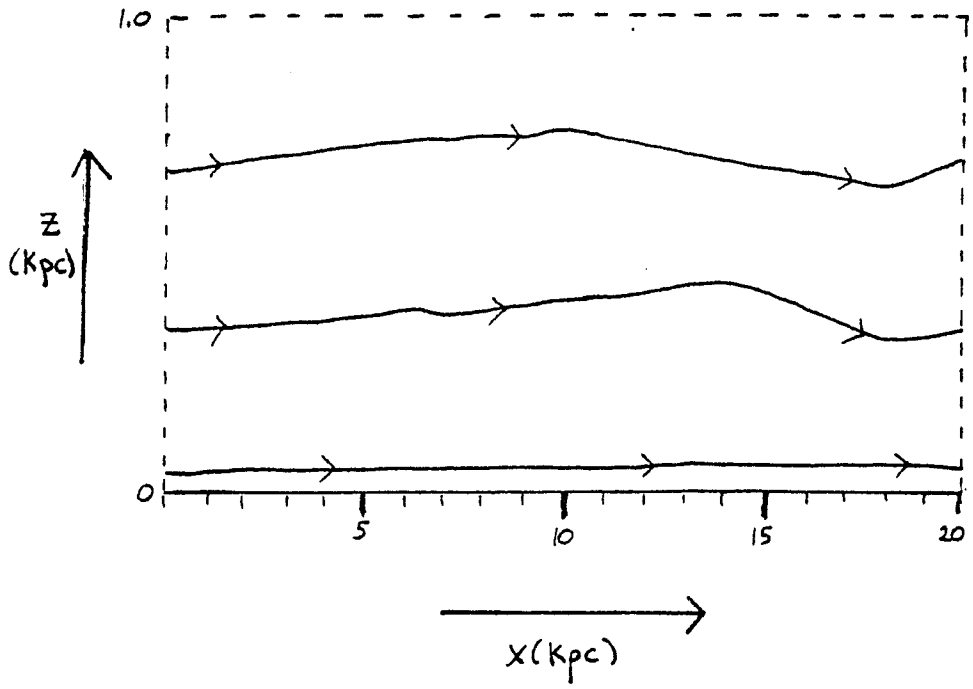


FIGURE 56c: Streamlines for the flow in figure 56a. The z-motions have been exaggerated.

FIGURE 57

MODEL B: $a_o = 10$ km/s ; $a_H = 25$ km/s ; $\hat{\rho}_R = 0.7$; $F = 0.05$

sonic line — — —
shock front ———

Mach number as a function of x and z

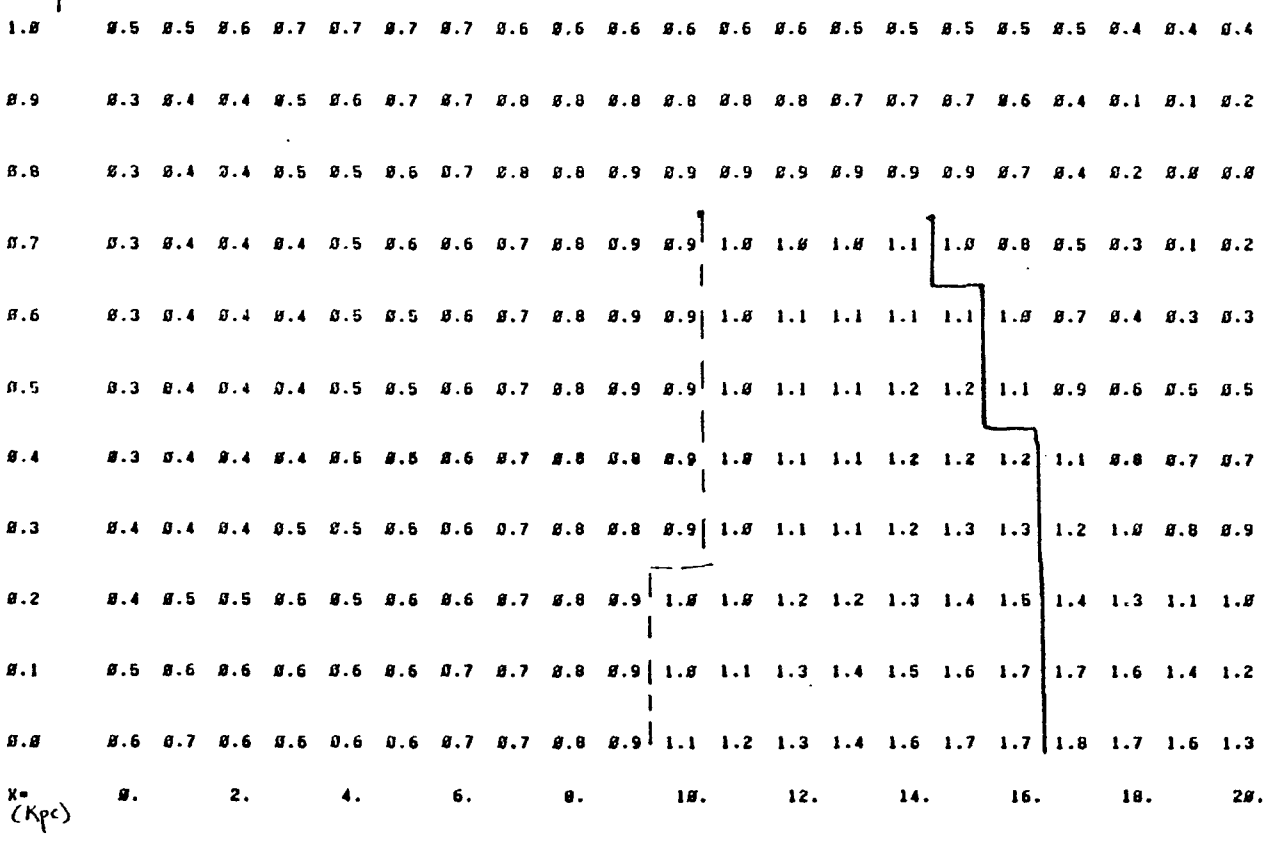


FIGURE 58

MODEL B: $a_0 = 10$ km/s; $a_{11} = 25$ km/s; $\hat{\rho}_R = .08$; $F = 0.07$

Mach number as a function of X and z

sonic line – – – –

shock front ———

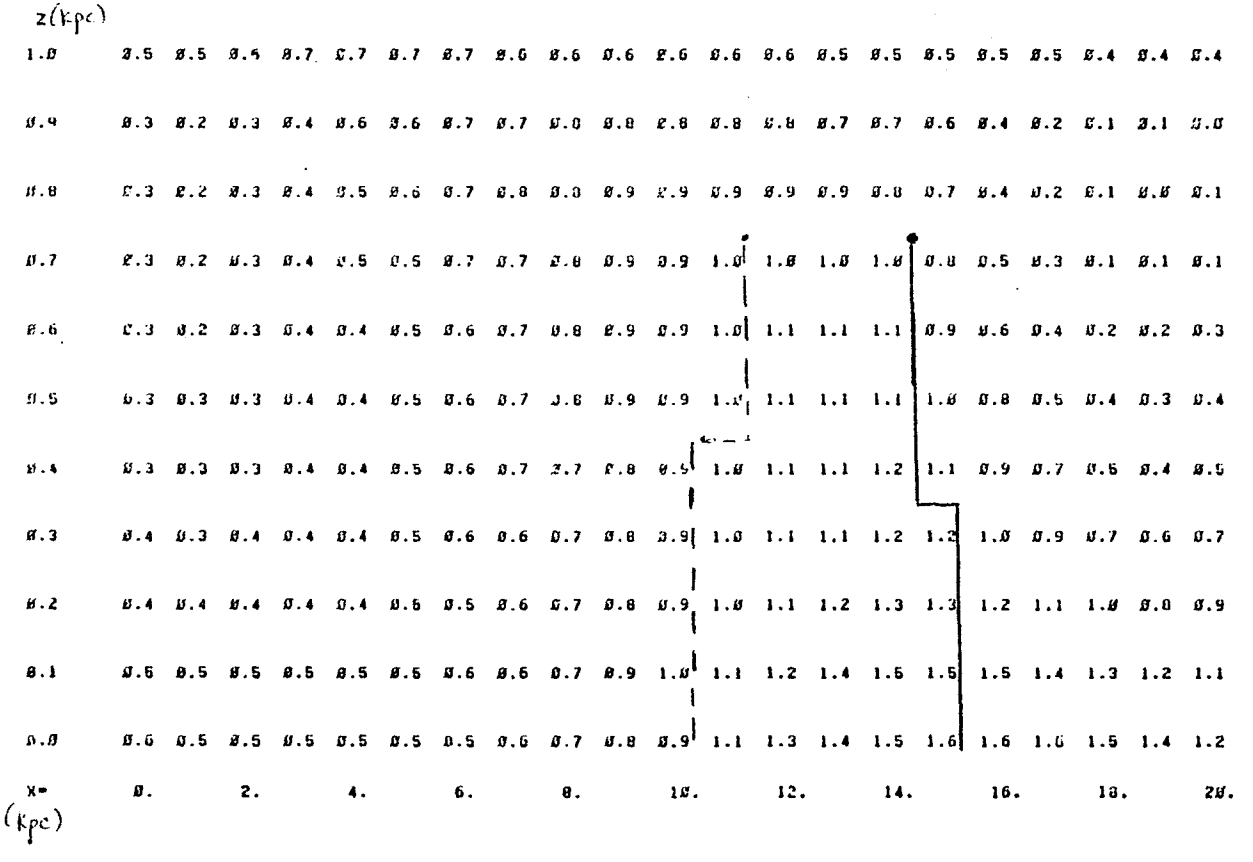


FIGURE 59

MODEL B: $a_o = 10$ km/s ; $a_H = 30$ km/s ; $\hat{f}_R = 0.8$; $F = 0.05$

Mach number as a function of x and z.

sonic line — — — —

shock front ————

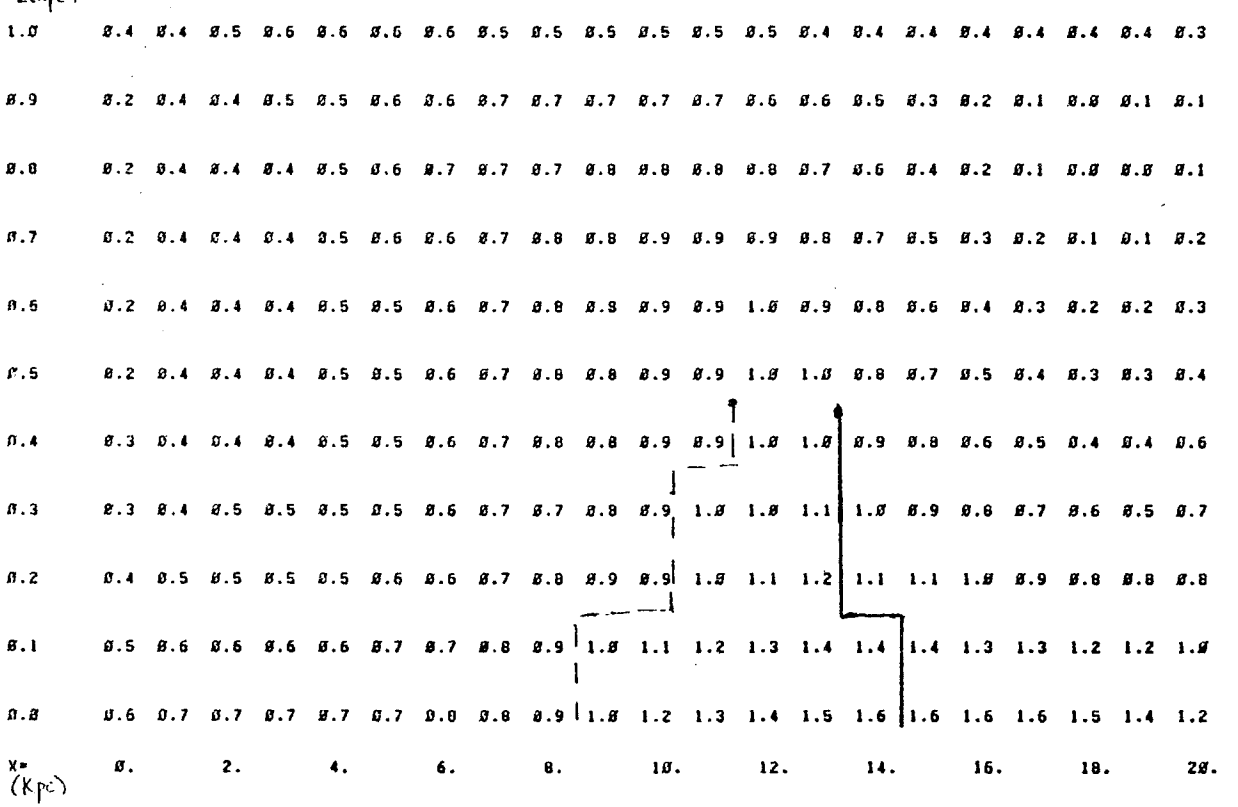
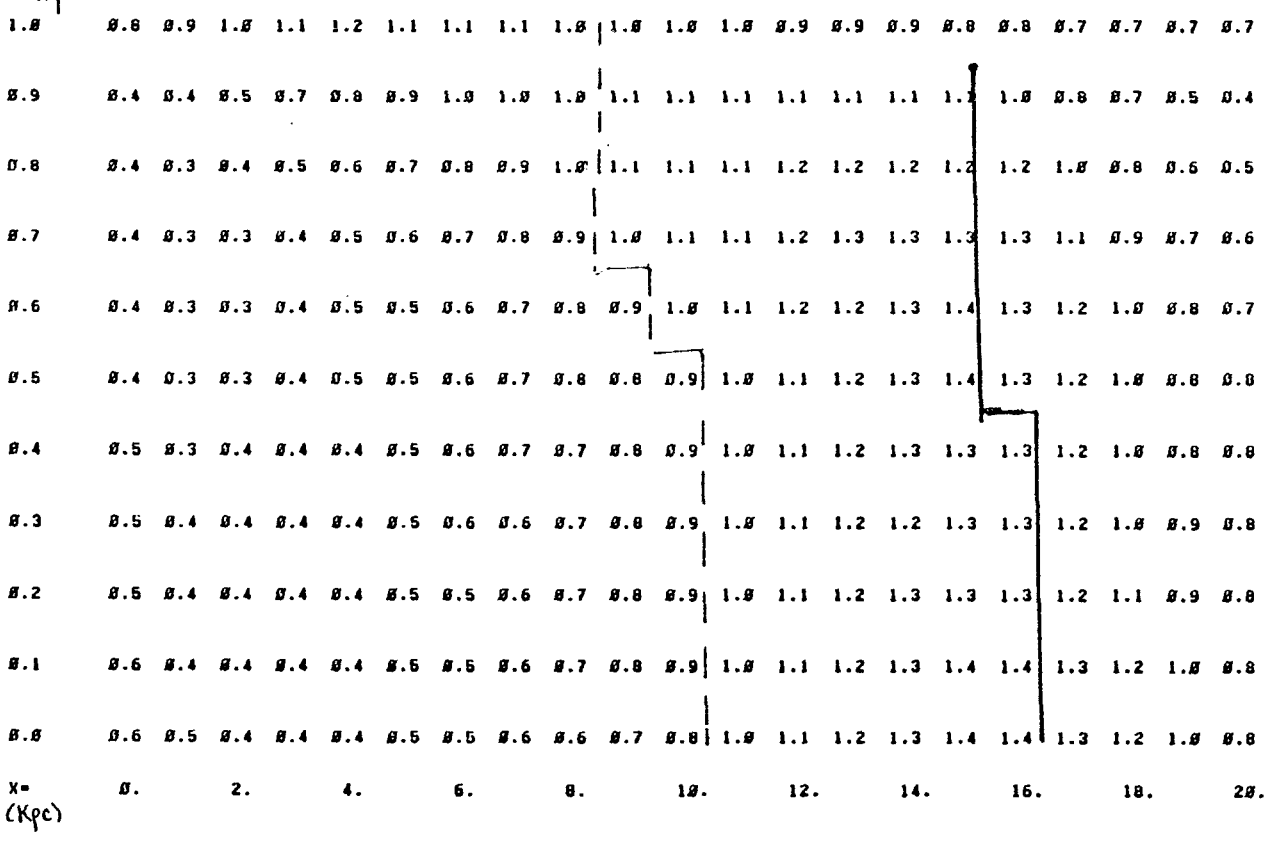


FIGURE 60

MODEL B: $a_o = 10$ km/s ; $a_H = 15$ km/s ; $\hat{\rho}_R = 0.8$; $F = 0.05$ sonic line — — —

Mach number as a function of x and z shock front ——



7.5 Discussion

(a) Model A (Isothermal)

The outstanding feature in the steady-state flow shown in figure 53 is the speed-up of the gas through a sonic point and the formation of a stationary shock. The sonic line in figure 53a is the line connecting all mesh points with Mach number M = 1.0 (it happens to fall between the grid points), and the shock front is the line connecting the maximum values of M. In figure 53b the shock occurs where the density is a minimum, which is at the same x value as the maximum velocity.

Both the sonic line and the shock front are vertical and extend all the way up to the end of the mesh at z = 1.0 kpc, i.e., the gas at large z forms a shock at the same position as the gas in the plane. (If the mesh is extended to larger z's, the vertical shock continues to exist at z values up to 1.8 kpc -- as high as the expression for potential is valid for.)

The one-dimensional shock strength S was defined in Chapter 6 as $S = u_1/u_2$, where u_1 and u_2 are the velocities before and after the shock. We can define shock strength in two-dimensional flow in terms of velocities along the streamline, but it is much more useful to define it as

$$S = \frac{\hat{\rho}_2}{\hat{\rho}_1} \tag{7.11}$$

where $\hat{\rho}_2$ and $\hat{\rho}_1$ are the densities after and before the shock, respectively. Using this definition, we see from figure 53b that

the shock strength decreases from 2.99 at z = 0 to 2.06 at z = 1.0 kpc. This behavior is typical of all the isothermal calculations -- S decreases by about 25% over the first 600 to 800 pc.

The thickness of the shock is constant with z; this is expected since the width is proportional to the viscosity and inversely proportional to the sonic speed, both of which are constant here.

Although this is a fully nonlinear two-dimensional calculation, the resulting flow is surprisingly like the one-dimensional flow. The absolute value of the vertical velocities never gets larger than 2 km/sec, and therefore the streamlines are very nearly parallel to the plane. The gas, entering the calculation region parallel to the plane, rises away from the plane in the interarm region. As it passes x = 10 kpc it dips toward the plane a little but it moves away again after the shock.

The gas is constrained to flow parallel to the plane at z = 0, so very near the plane the streamlines are straight. The gas further away can move in the z-direction, but there is a height for maximum deviation from parallel flow, as seen in figure 53c. The horizontal region over which the gas dips toward the plane becomes smaller as z increases.

Figures 54 and 55 show the effect of the values of $\hat{\rho}_R$ and F on the flow. As $\hat{\rho}_R$ is lowered from 0.5 to 0.4, the sonic line remains in the same position near x = 10 kpc, but the location of the shock is moved downstream, and its strength increased cor-

respondingly. Otherwise, the general appearance of the flow is very much like that in figure 53.

If we increase the strength of the spiral potential F to 0.07 (figure 55), for the original value $\hat{\rho}_R$ in figure 53, the sonic line and the shock are at the same locations (within the mesh limitations) but its strength increases.

Changing $\hat{\rho}_R$ and F affects the flow solutions in the same way as in the one-dimensional flow into a spiral arm. In this, and in the surprisingly small z-velocities, the flow is very much like one-dimensional flow. This seems to justify previous treatments of the z-dimension, in which the gas distribution was considered as a series of thin layers and each layer was analyzed independently (Tosa 1973). That the shock location is independent of z can also be understood from one-dimensional flow, in which the shock position is fixed by the downstream density $\hat{\rho}_R$ (see section 6.5).

But it is important to note that the flow solution cannot be completely explained as a set of one-dimensional solutions layered upon each other. The spiral potential field varies as $\text{sech}^{|k|\Delta}(z/\Delta)$ (see equation [7.1]), so the effective amplitude, $F_{eff} = F\ \text{sech}^{|k|\Delta}(z/\Delta)$ decreases rapidly with z. For $|k|\Delta = 1.0$, $F_{eff} = F/10$ at z = 900 pc. For a spiral potential strength of 0.005 and the parameters of figure 53, the one-dimensional flows have no shock; in fact, the velocities are subsonic for $F \leq 0.025$, which corresponds to z = 500 pc in the two-dimensional flow.

The existence of the shock at such large values of z is not likely to be an effect of the numerical scheme. The shock exists independently of any boundary conditions imposed on the top, refusing to become subsonic for even one row of mesh points. An additional reason for confidence in the solution is found in Tubbs (1979) who calculated the z-motions for an isothermal gas using a radically different numerical method and obtained essentially the same results -- a vertical shock extending to z = 500 pc (the maximum height of his calculation).

The crossing time for the gas to move through the spiral potential is on the order of 1 kpc/km/sec. The time for a signal to travel from the plane to z = 1.0 kpc is only about 0.1 kpc/km/sec. Therefore, the gas near the galactic plane has plenty of time to affect the flow of the gas high above it, pulling it along to supersonic velocities, and forcing it to form a shock even where the spiral field is too low by itself to force it into a shock. This kind of interaction was not predicted by any treatment of the gas as independent layers.

(b) Model B (Varying Temperature)

Since, for this model, the sonic velocity increases with z, we expect it to be more difficult for the gas to reach supersonic speeds and to form a shock far away from the plane. This is verified by the solution shown in figure 56, in which a_o = 10 km/sec and a_H = 25 km/sec: the stationary shock that forms now extends only up to 600 pc.

This solution differs also in other ways from isothermal (Model A) solutions. The sonic line (see figure 56a) is no longer vertical; the gas at higher z must travel a longer x distance in order to reach its larger sonic velocity. It appears, in addition, that the shock front is not quite vertical, but test runs with greater spatial resolution confirm that this is a result of the relatively coarse mesh; when a finer mesh is used the shock front is vertical.

Figure 56b shows that the density is no longer correlated perfectly with the velocity; i.e., the minimum density is sometimes one mesh point away from the location of the maximum velocity. This effect persists in computations with better spatial resolutions; this appears to be caused by small-amplitude oscillations in the density.

As in the isothermal case, the shock strength decreases with increasing z, but more gradually. In figure 56b, S decreases from 1.82 at z = 0 to 1.56 at z = 600 pc. This is a decrease of only 14%, compared to a typical decrease of 25% for Model A.

The shocks are somewhat weaker than for the isothermal model, partly because the width of the shock is very large at z = 0; using a smaller artificial viscosity helps, but leads to oscillations that make it difficult to see the flow. The shock width decreases with height, as expected, since it is inversely proportional to the sonic velocity, which increases with z.

The z-velocities are very small in these flows, just as

for the isothermal gas -- less than 2 km/sec. The streamlines in figure 56c are slightly different than for isothermal flow; the region over which the gas moves toward the plane increases for larger z.

Figure 57 shows the effect of changing the value of $\hat{\rho}_R$. As $\hat{\rho}_R$ is lowered from 0.8 to 0.7, the shock location moves further downstream and its strength increases. In addition, the shock now extends up to 700 pc.

If the spiral strength F is increased to 0.07, with the original value of $\hat{\rho}_R$ as in figure 56, we get the flow shown in figure 58. The shock is now in the same position as in figure 56, but it is stronger and extends up to 700 pc, instead of 600 pc.

For Model B we have an additional parameter to be varied: a_H, the sonic velocity at $z = \infty$. The value of a_H was 25 km/sec for the flow in figure 56; figures 59 and 60 show the effects of raising it to 30 km/sec and lowering it to 15 km/sec, respectively. When a_H = 30 km/sec, the shock extends only to 400 pc. When a_H = 15 km/sec, the shock reaches up to 1.0 kpc, very much like a totally isothermal gas. Thus there seems to be a relatively small range of values of a_H that result in a shock that extends to at least 400 pc but does not reach beyond 1.0 kpc.

The flows of Model B have the same kind of paradoxical behavior as those of Model A: in terms of their response to changes in $\hat{\rho}_R$ and F and their small z-velocities, they can be con-

sidered as layered one-dimensional solutions. This explains why the shock front here is also vertical: for each layer of gas, the shock position is fixed by the value of the downstream density $\hat{\rho}_R$.

But the decrease in the effective strength of the spiral field F_{eff} with z, and the increase of sonic velocity with z, are expected to prevent the formation of a shock at high z. One-dimensional flows with sonic velocities and effective spiral field strengths corresponding to different heights were calculated: the results imply that if the gas were a set of independent layers, the shock would not extend past 400 pc for a_H = 25 km/sec. That it does is due to interactions of the gas at different heights, as in the isothermal case.

7.6 Conclusions

We have seen in section 2.5(d) (figures 37 - 39) that the observed density distribution in z could be approximately fitted by the Model A hydrostatic solution, but that there is more gas at large z than the hydrostatic solution predicts. The Model B hydrostatic solution has a larger density at large z, but still not enough to fit the observed distribution.

Figures 61 and 62 show the density distributions predicted by Models A and B, respectively, comparing the hydrostatic and the dynamic solutions. The latter are the densities just past the shock position. For both models, the scale height of the dynamic distribution is only about 5% larger than that of the hydrosta-

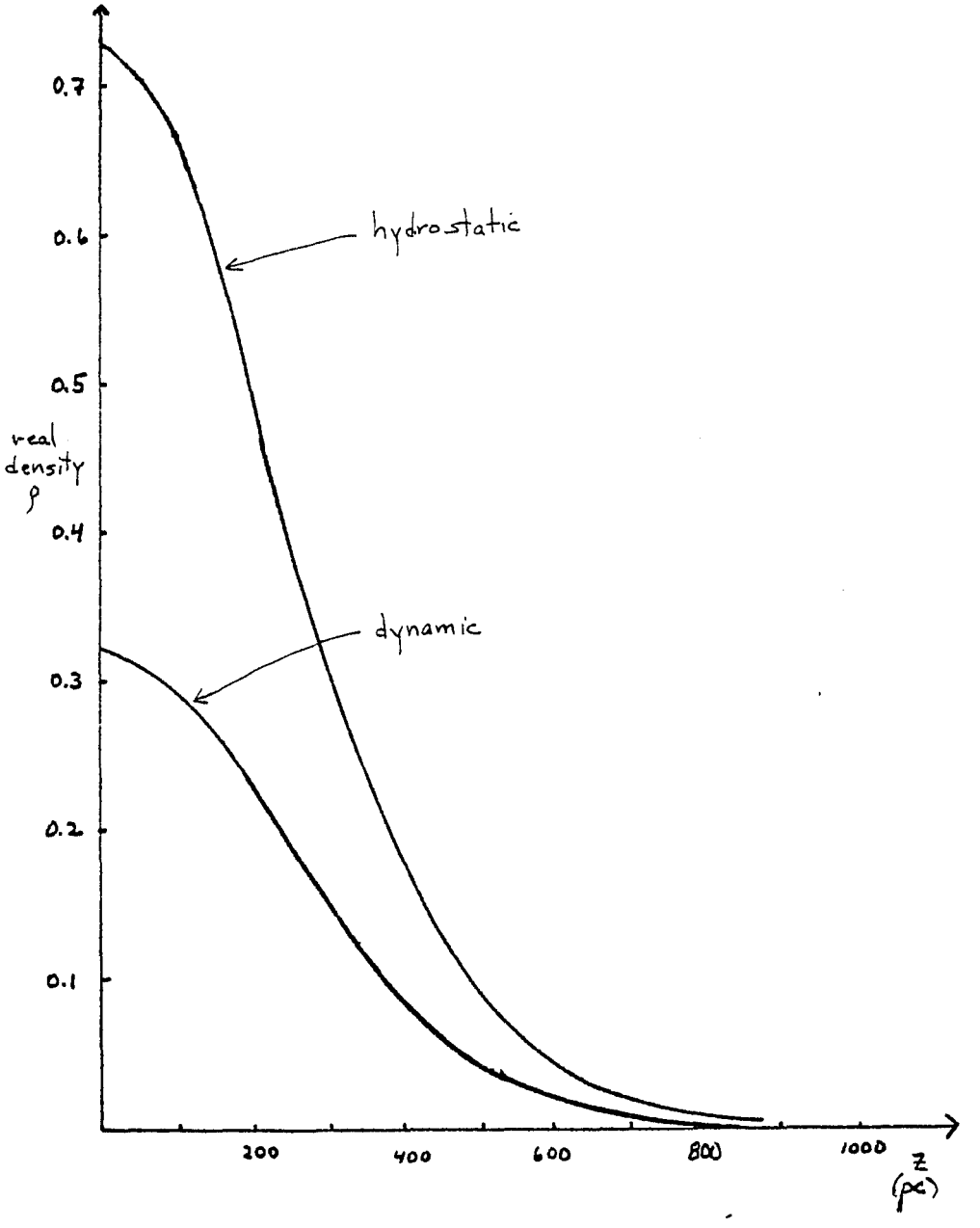


FIGURE 61: Hydrostatic and dynamic solutions for Model A. The dynamic distribution is for the flow shown in figure 53, just after the shock.

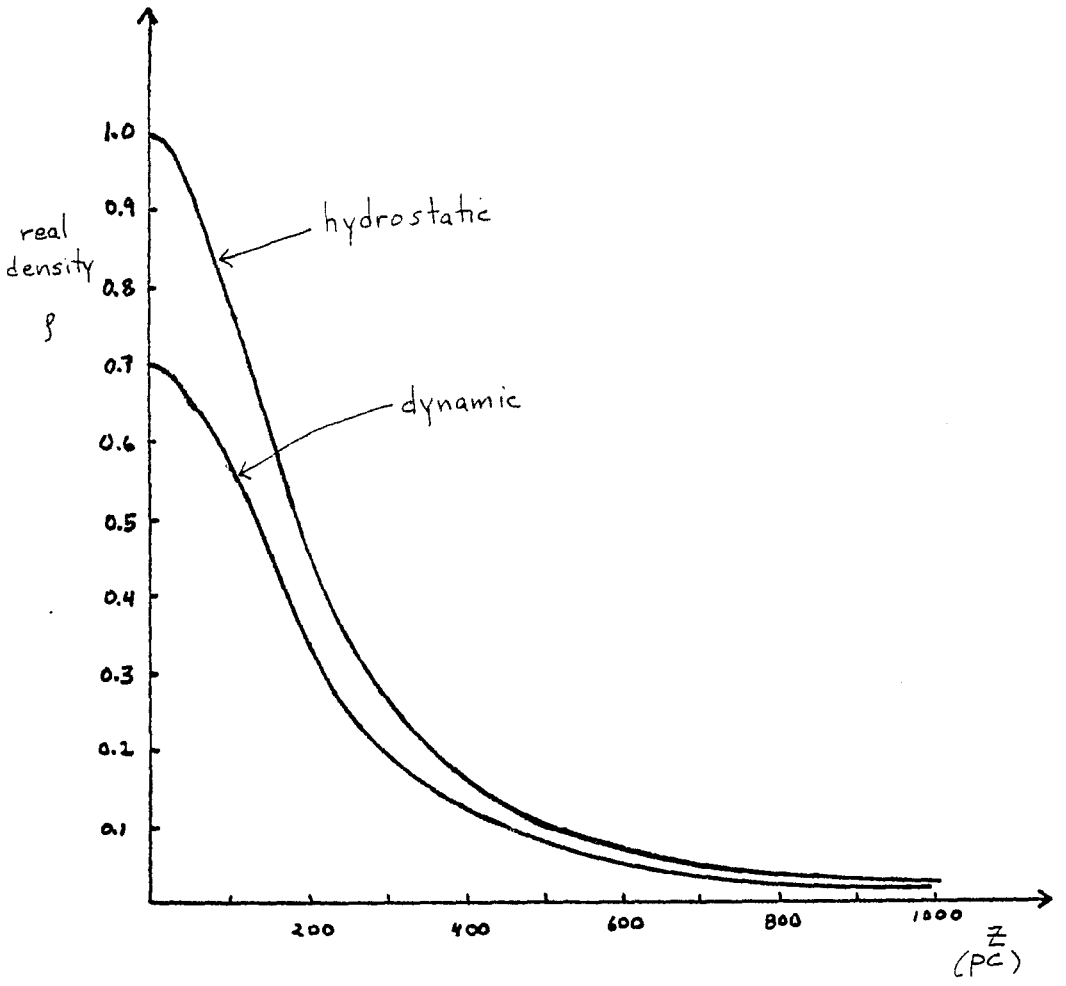


FIGURE 62: Hydrostatic and dynamic solutions for Model B. The dynamic distribution is for the flow shown in figure 56, just after the shock.

tic model. Therefore, neither has pushed up much extra gas to large z distances from the plane (the z-velocities of less than 2 km/sec are too small).

But, although the gas is not pushed up to higher latitudes by a shock wave near the galactic plane, the shock itself extends to large distances from the plane, compressing the gas already there to observable densities. The density contrast across the shock decreases only by about 25% for Model A and 15% for Model B over the first 600 pc away from the plane. Thus, although there is not much extra gas at high z, the gas compression due to the shock there will make it observable.

The shocks in these calculations produce density contrasts in the plane that are relatively weak (3.3:1) compared to the ratios of at least 5:1 found by Roberts (1969). Since we are interested only in the z structure of the flow and have omitted the Coriolis force, the effective force felt by the gas in our treatment is weakened (see section 3.3[b]), and thus the shock in the plane is weakened. However, this should not affect the fractional decrease in shock strength with z.

At the position of the shock in Roberts (1969) calculations, both the Coriolis force and the spiral force are negative, thus reinforcing each other, and increasing the effective force felt by the gas. Since an increase in the spiral field strength F extends the shock front to larger z in Model B (see figures 56a and 58), the solution to the equations including the Coriolis

force should make it even easier to compress the gas at high latitudes.

APPENDIX A

RANKINE-HUGONIOT RELATIONS ACROSS A SHOCK

Assuming conservation of mass and momentum, we derive two conditions that must be satisfied across a discontinuity (shock) in the density, velocity and pressure of a one-dimensional gas flow.

We apply the general conservation principles to a column of gas. The column covers, at time t, an interval $a_1(t) \leq x \leq a_2(t)$, where $a_1(t)$ and $a_2(t)$ are the positions of the particles that form the ends of the column. For this column we have

conservation of mass

$$\frac{d}{dt}\int_{a_1(t)}^{a_2(t)} \rho\, dx = 0 \tag{A.1}$$

and conservation of momentum

$$\frac{d}{dt}\int_{a_1(t)}^{a_2(t)} \rho u\, dx = P(a_1,t) - P(a_2,t) \tag{A.2}$$

If we assume that ρ and u are continuous and differentiable in the entire column, we can integrate to get the usual equations of motion of a gas. But if, instead, we assume a point of discontinuity within the column at $x = \xi(t)$, moving with velocity $\dot{\xi}(t) = U(t)$, we can derive the jump conditions. We assume that the

time derivatives of the density and velocity remain bounded.

Both equations (A.1) and (A.2) have the same form for the integral:

$$I = \int_{a_1(t)}^{a_2(t)} K(x,t)\, dx \tag{A.3}$$

where the integrand K is discontinuous at $x = \xi$. In order to differentiate, we break the integral into two parts:

$$\frac{dI}{dt} = \frac{d}{dt}\int_{a_1(t)}^{\xi(t)} K(x,t)\, dx + \frac{d}{dt}\int_{\xi(t)}^{a_2(t)} K(x,t)\, dx$$

$$= \int_{a_1(t)}^{a_2(t)} \frac{\partial K(x,t)}{\partial t}\, dx + K_1 \dot{\xi}(t) - K(a_1,t)\, u(a_1,t)$$

$$+ K(a_2,t)\, u(a_2,t) - K_2 \dot{\xi}(t) . \tag{A.4}$$

The quantities K_1 and K_2 are the limits of K(x,t) as x approaches ξ from the sides $x < \xi$ and $x > \xi$, respectively.

Now we perform a limiting process, letting the length of the column of gas approach zero. Since we had assumed the time derivatives of the variables to be finite, the integral in equation (A.4) approaches zero. Since $K(a_1,t) \longrightarrow K_1$ and $K(a_2,t) \longrightarrow K_2$, we get

$$\lim_{a_2 - a_1 \longrightarrow 0} \frac{dI}{dt} = K_2(u_2 - U) - K_1(u_1 - U) \tag{A.5}$$

If we define

$$v_1 = u_1 - U \tag{A.6a}$$

$$v_2 = u_2 - U \tag{A.6b}$$

to be the flow velocities relative to the shock velocity, we can immediately write down the jump conditions:

$$\rho_2 v_2 = \rho_1 v_1 = J \tag{A.7a}$$

$$(\rho_2 u_2) v_2 - (\rho_1 u_1) v_1 = P_1 - P_2 \quad . \tag{A.7b}$$

Thus the mass flux J is conserved across the shock, even though neither density nor velocity is conserved.

We can rewrite equation (A.7b) in a form that only involves the relative velocities v_1 and v_2:

$$P_1 + \rho_1 v_1^2 = P_2 + \rho_2 v_2^2 \quad . \tag{A.8}$$

All calculations done in this thesis are in a frame in which the shock is stationary, so $v_1 = u_1$ and $v_2 = u_2$. In addition, for the equation of state used in this thesis, $P = a^2 \rho$, we combine equations (A.7a) and (A.8) to get

$$u_1 u_2 = v_1 v_2 = a^2 \tag{A.9}$$

Thus a supersonic flow changes to subsonic after the discontinuity.

The jump conditions across two-dimensional shock fronts could be derived in the same manner. However, observed from a suitable moving coordinate system, a two-dimensional shock front is

always equivalent to a stationary one-dimensional shock front (Courant and Friedrichs 1948): the velocity component parallel to the shock front is continuous and the one-dimensional jump conditions hold for the velocity component normal to the discontinuity line.

APPENDIX B

ONE-DIMENSIONAL FINITE-DIFFERENCE EQUATIONS

(a) The Atmosphere Problem (Chapter 5)

The differential equations can be written in matrix form as

$$\frac{\partial V}{\partial t} + \frac{\partial G}{\partial z} = S \tag{B.1}$$

where there is no x-dependence in any of the two-dimensional vectors V, G and S. We derive the finite-difference equations from the full equations (4.10), (4.11) and (4.12) by letting $F \equiv 0$ and dropping all x-dependence in the other vectors.

We define forward, backward and centered spatial differences analogous to equations (4.6), (4.7) and (4.8) as

$$\Delta^{+} f(z) = f(z + \Delta z) - f(z) \tag{B.2a}$$

$$\Delta^{-} f(z) = f(z) - f(z - \Delta z) \tag{B.2b}$$

$$\Delta^{c} f(z) = \frac{1}{2}[f(z + \Delta z) - f(z - \Delta z)] \quad . \tag{B.2c}$$

We also define an average quatity

$$f_{av}(z) = \frac{1}{2}[f(z) + f(z + \Delta z)] \quad . \tag{B.3}$$

The first step calculates the intermediate values V^{*} from

$$V^{*} = V^{n}_{av} - \frac{\Delta t}{\Delta z} \Delta^{+} G^{n} + \Delta t \, S^{n}_{av} \quad . \tag{B.4}$$

These values are used to update G and S to G^{*} and S^{*}.

The second step calculates V^{n+1} using

$$V^{n+1} = V^n - \frac{\Delta t}{2\Delta z}(\Delta^c G^n + \Delta^- G^*) + \frac{\Delta t}{2}(S^n + S^*_{av}) + Q^n \qquad (B.5)$$

The artificial viscosity Q^n is given by the one-dimensional analogy to equation (4.12):

$$Q^n = \frac{B}{2} \frac{\Delta t}{\Delta z}(|\Delta^+ w| \Delta^+ V^n - |\Delta^- w| \Delta^- V^n) \qquad (B.6)$$

where w is the velocity.

(b) The Spiral Potential Problem (Chapter 6)

There is no z-dependence, and the differential equations can be written as

$$\frac{\partial V}{\partial t} + \frac{\partial F}{\partial X} = S . \qquad (B.7)$$

We go again to equations (4.10), (4.11) and (4.12) and set $G \equiv 0$ and ignore all z dependence in V, F and S. The forward, backward and centered differences are

$$\Delta^+ f(x) = f(x + \Delta x) - f(x) \qquad (B.8a)$$

$$\Delta^- f(x) = f(x) - f(x - \Delta x) \qquad (B.8b)$$

$$\Delta^c f(x) = \frac{1}{2}[f(x + \Delta x) - f(x - \Delta x)] \quad , \qquad (B.8c)$$

and the average is

$$f_{av}(x) = \frac{1}{2}[f(x) + f(x + \Delta x)] \quad . \tag{B.9}$$

The first step calculates V^*,

$$V^* = V^n_{av} - \frac{\Delta t}{\Delta x} \Delta^+ F^n + \Delta t \, S^n_{av} \quad , \tag{B.10}$$

and then uses V^* to update F and S to F^* and S^*.

The second step calculates V^{n+1}:

$$V^{n+1} = V^n - \frac{\Delta t}{2\Delta x}(\Delta^c F^n + \Delta^- F^*)$$

$$+ \frac{\Delta t}{2}(S^n + S^*_{av}) + Q^n \quad . \tag{B.11}$$

The artificial viscosity is given by

$$Q^n = \frac{B}{2} \frac{\Delta t}{\Delta x}(|\Delta^+ u| \, \Delta^+ V^n - |\Delta^- u| \, \Delta^- V^n) \tag{B.12}$$

where u is the velocity.

APPENDIX C

FLOWS IN A ONE-DIMENSIONAL CONVERGING-DIVERGING NOZZLE

We consider only flows that are subsonic at the entry section (the upstream boundary) and exit into a receiver in which we can fix the pressure. Since $P = a_o^2 \rho$, we freely replace pressure with density or vice versa, for better physical understanding.

Case 1: Subsonic Throughout

The flow, entering the nozzle at a pressure P_1 and a subsonic velocity u_1, speeds up in the converging section (figure 45). It reaches a maximum velocity, still subsonic, at the throat, and slows down until it exhausts at pressure P_2. Since the nozzle is symmetric about the throat, the subsonic flow must also be symmetric. That means $P_1 = P_2$. If we specify $P_2 \neq P_1$, there is a transient period of readjustment, and the new steady flow has $P_1 = P_2$, and therefore $u_2 = u_1$ (since $J = \rho_1 u_1$ = constant). The downstream conditions, therefore, determine some of the upstream conditions.

Case 2: Transonic Flow Without a Shock

The gas speeds up in the entry section and reaches the sonic velocity at the throat. The flow may now pass continuously into the supersonic range, or pass back into the subsonic range. Which of the two situations prevails depends on the pressures specified at the upstream and downstream boundaries. If the subsonic solution prevails, we again get a symmetric flow, characterized by

an exit pressure $\overline{P}$. If, however, the gas becomes supersonic past the throat, it continues to speed up.

There are now two flow regimes. The subsonic flow can no longer receive any information from the downstream boundary, because no signals can travel upstream through the supersonic gas. Thus, no matter what happens to the supersonic gas past the throat, the flow up to the throat remains the same for a given upstream pressure.

The supersonic flow continues to speed up at a rate dependent only on the geometry of the nozzle; the exit velocity and the exit pressure P^* are thus determined by the shape of the nozzle. If the real exit pressure P_2 happens to equal P^* we have what is called ideal flow in the nozzle: subsonic flow in the converging part, a smooth transition to supersonic flow at the throat, and a smooth exit flow.

Case 3: Transonic Flow With a Shock

If the downstream pressure P_2 is not equal to P^*, the expanding supersonic flow must adjust to this exit pressure. If the exit pressure P_2 is a little lower than $\overline{P}$ described above (the pressure for a subsonic flow that just reaches the sonic point), then the flow must exit subsonically and the gas forms a shock soon after it becomes supersonic. The gas is compressed and slowed down by the irreversible processes within the shock to subsonic speed with increased pressure. The new velocity obtained from the jump conditions across the shock (see appendix B) continues to decrease;

the pressure continues to rise. The position and strength of the shock are automatically adjusted so that the pressure at the downstream exit is P_2.

If the specified exit pressure is lowered even more, the location of the shock moves downstream toward the exit. When the shock reaches the exit, the flow can no longer adjust inside the nozzle, and must adjust outside, via oblique discontinuities that cannot be treated with a one-dimensional analysis.

BIBLIOGRAPHY

Burstein, S. Z. 1964, AIAA Journal, 2, 2111.

----------. 1967, J. of Comp. Phys., 1, 198.

Burton, W. B. 1970, Astr. and Ap. Suppl., 2, 261.

Courant, R., Friedrichs, K. O., and Lewy, H. 1928, Math. Ann., 100, 32.

Courant, R., and Friedrichs, K. O. 1948, Supersonic Flow and Shock Waves (New York-London: Interscience).

Courant, R., and Hilbert, D. 1953, Methods of Mathematical Physics (New York: Interscience).

Dalgarno, A., and McCray, R. A. 1972, Ann. Rev. Astr. and Ap., 10, 375.

Falgarone, E., and Lequeux, J. 1973, Astr. and Ap., 25, 253.

Field, G. B., Goldsmith, D. W., and Habing, H. J. 1969, Ap. J., 155, L 149.

Field, G. B. 1974a, Lectures given at the 1974 les Houches Summer School in Theoretical Physics.

----------. 1974b, "Hot Gas In and Between Galaxies" in W. R. S. Garton (Ed.), I.A.U. Colloquium No. 27 on UV and x-ray Spectroscopy of Astrophysical Plasmas (Boston: Reidel).

Ganz, T., and Serra, R. 1974, AIAA Journal, 12, 263.

Gorenstein, P., and Tucker, W. H. 1972, Ap. J., 176, 333.

Habing, H. J., and Goldsmith, D. W. 1971, Ap. J., 166, 525.

Henderson, A. P. 1972, published by Astronomy Program, University of Maryland.

Jenkins, E. B., and Meloy, D. A. 1974, Ap. J. (Letters), 193, L 121.

Kepner, M. 1970, Astr. and Ap., 5, 444.

Kilkenny 1973, Ph.D. thesis, St. Andrews.

Landau, L. D., and Lifshitz, E. M. 1959, Fluid Mechanics (London: Pergamon Press).

Lax, P., and Wendroff, B. 1960, Comm. Pure Appl. Math., 8, 217.

----------. 1964, Comm. Pure Appl. Math., 17, 381.

Liebovitch, L. S. 1978, Ph.D. thesis, Harvard University.

Lin, C. C., and Shu, F. H. 1964, Ap. J., 140, 646.

Lin, C. C., Yuan, C., and Shu, F. H. 1969, Ap. J., 155, 721.

Matthewson, D. S., van der Kruit, P. C., and Brouw, W. N. 1972, Astr. and Ap., 17, 468.

Moretti, G. 1969, Physics of Fluids Suppl., II, 13.

Potter, D. 1973, Computational Physics (London: Wiley).

Roache, P. J. 1972, Computational Fluid Dynamics (New Mexico: Hermosa Publishers).

Richtmyer, R. D., and Morton, K. W. 1967, Difference Methods for Initial Value Problems (New York: Interscience).

Roberts, W. W. 1969, Ap. J., 158, 123.

----------. 1972, Ap. J., 173, 259.

Roberts, W. W., and Yuan, C. 1970, Ap. J., 161, 877.

Roberts, W. W., Roberts, M. S., and Shu, F. H. 1975, Ap. J., 196, 381.

Sanders, R. H., and Huntley, J. M. 1976, Ap. J., 209, 53.

Schwarz, J., McCray, R., and Stein, R. 1972, Ap. J., 175, 673.

Serra, R. A. 1972, AIAA Journal, 10, 603.

Shalloway, A. M., Mauzy, R., and Greenhalgh, J. 1972, NRAO Electronics Division Internal Report No. 125.

Shapiro, A. H. 1953, The Dynamics and Thermodynamics of Compressible Fluid Flow, Vol. I (New York: Ronald Press).

----------. 1954, op. cit., Vol. II (New York: Ronald Press).

Shu, F. H. 1968, Ph.D. thesis, Harvard University.

Shu, F. H., Milione, V., Gebel, W., Yuan, C., Goldsmith, D. W., and Roberts, W. W. 1972, Ap. J., 173, 557.

Shu, F. H., Milione, V., and Roberts, W. W. 1973, Ap. J., 183, 819.

Silk, J. 1973, Ann. Rev. Astr. and Ap., 11, 269.

Spitzer, L., and Scott, E. H. 1969, Ap. J., 158, 161.

Tosa, M. 1973, Publ. Astr. Soc. Japan, 25, 191.

Tubbs, A. D. 1979, Ph.D. thesis, University of Pittsburgh.

Vandervoort, P. O. 1970a, Ap. J., 161, 67.

----------. 1970b, Ap. J., 161, 87.

----------. 1971, Ap. J., 166, 37.

von Mises, R. 1958, Mathematical Theory of Compressible Flows (New York: Academic Press).

Weaver, H., and Williams, D. R. W. 1974, Astr. and Ap. Suppl., 17, 1.

Westerhout, G. 1973, Maryland - Green Bank 21-cm Survey, 3rd ed.

Wielen, R. 1974, Publ. Astr. Soc. Pac., 86, 341.

Woodward, P. 1973, Ph.D. thesis, U. Cal. Berkeley.

----------. 1975, Ap. J., 195, 61.

York, D. G. 1974, Ap. J. (Letters), 193, L 127.

Yuan, C. 1969, Ap. J., 158, 889.

Yuan, C., and Wallace, L. 1973, Ap. J., 185, 453.