On the Spiral Patterns in Disk Galaxies

C. C. LIN
Massachusetts Institute of Technology
Cambridge, Massachusetts

1. Introduction

The problem I shall discuss today concerns the behavior of a galaxy of stars (Fig. 1) like the Milky Way we live in. In particular, I shall consider the mechanism through which the regular spiral pattern is created and maintained. I shall report on the work which was done on this problem by several of us at Cambridge, Mass. (Kalnajs, 1965; Lin, 1965 a,b; Lin and Shu, 1964, 1966; Toomre, 1964). There were other related investigations which I shall not be able to cover in this review.

Students in Applied Mechanics may derive satisfaction from the fact that the general concepts and mathematical methods developed and used in their own subject have found application in these extra-terrestial objects. Indeed, the galaxy of stars, which number about 100 billion, can be shown to behave like a gigantic plasmoidal medium, extending over a distance of the order of 100 thousand light years. We can even show that the mechanisms for the creation and the maintenance of the spiral patterns have their counterparts in plasma physics under laboratory conditions. One might wonder whether one could build an experimental model of our galaxy by using plasmas. Such attempts would not only enhance our understanding of the galactic problem, but might also help with the construction of devices for plasma containment.

To establish a feeling for the physical processes in the galaxy, let us first examine the orders of magnitude of the quantities involved. Consider first the stars. In the following table (Table I), we compare a star with its planets, such as the solar system, with a single hydrogen atom, and compare the galaxy of stars with the gas.

TABLE I

	Gas: hydrogen atom	Galaxy: solar system
1.	Bohr radius of hydrogen atom	Radius of earth orbit (1 A.U.)
	0.5×10^{-8} cm.	0.5×10^{-8} kpc.
2.	Unit of length of measurement	
	1 cm.	1 kpc. = 3×10^{21} cm.
3.	Number density	
	$n = 10^{8}\ (\text{cm})^{-3}$	$0.1\ (\text{pc})^{-3} = 10^{8}\ (\text{kpc})^{-3}$
4.	Mean free path $\lambda = (n\pi\sigma^2)^{-1}$	
	3×10^{7} cm.	3×10^{7} kpc.

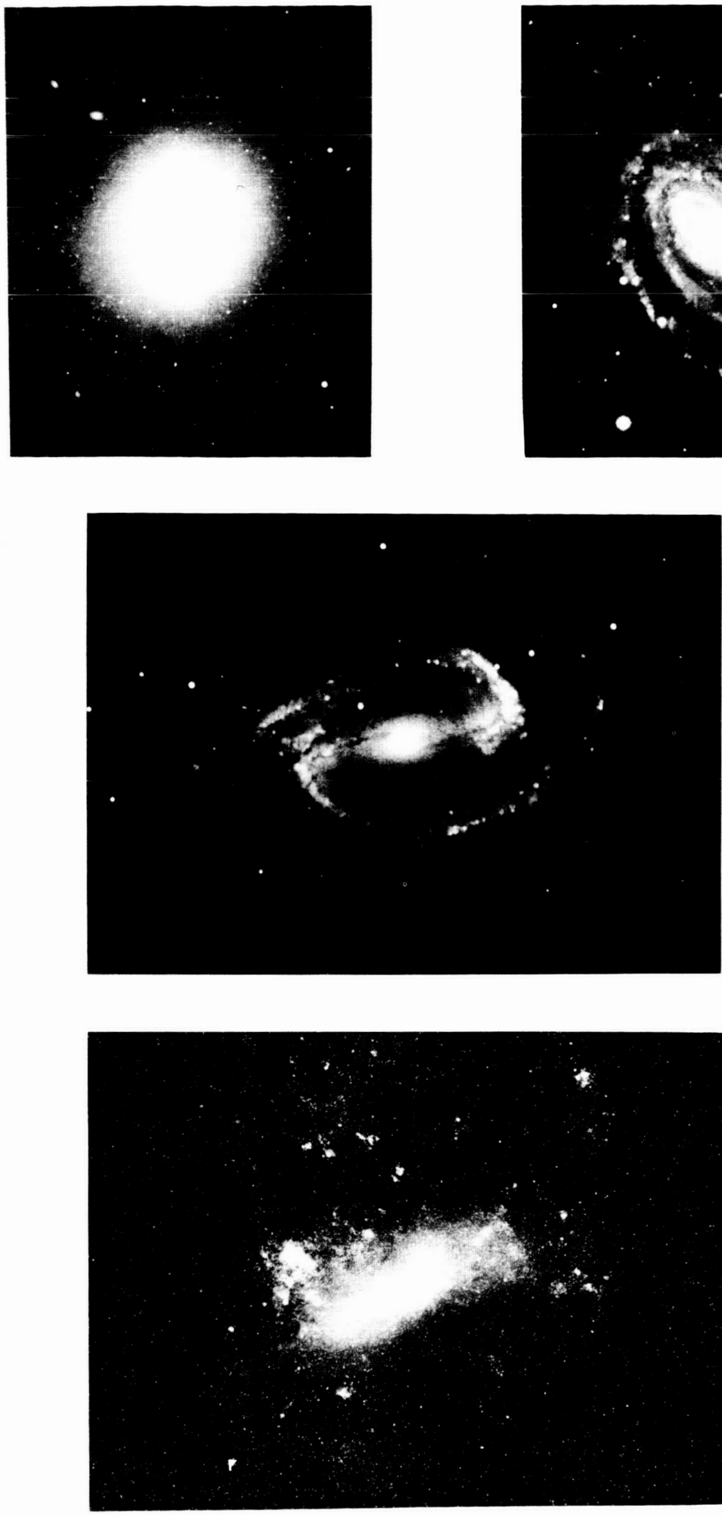


FIGURE 1. EXAMPLES OF GALAXIES OF VARIOUS TYPES

The mean free path of 3×10^{7} kpc is much larger than the galactic system, even much larger than the size of the cluster of galaxies. It is somewhat of an over-estimation, since we have taken the size of the "molecule" to be the diameter of the orbit of the earth. If we should adopt the orbit of pluto to be an indication of the size of the solar system, the mean free path would be reduced by a factor $(40)^{2}$; and the resultant value would be 2×10^{4} kpc, which is still extremely large.

Other relevant numerical relationships are as follows:

1 A.U. subtends an angle of 1 sec at a distance of 1 parsec (pc)

1 km/sec x 10^{6} years = 1 pc

20 km/sec x 10^{6} years = 2×10^{4} kpc

age of universe $\cong 10^{10}$ years

From these numerical values, we conclude that the stellar system is a "highly rarefied gas, held together by the mutual gravitational attraction of the molecules." It may therefore be described as a <u>gravitational plasma</u>.

There is also interstellar gas (consisting mostly of hydrogen, possibly all in atomic form); and there is an interstellar magnetic field. The gas is highly rarefied according to terrestial standards (1 atom per cubic centimeter being a typical value), but it may be regarded as a continuous medium, since the mean free path is still very much smaller than the linear dimensions involved. The gas is only very slightly ionized (perhaps to the extent of one part in 10^{3} or 10^{4}), yet it may be regarded as an infinitely conducting medium, again because of large linear scales involved. Thus, the gas and the magnetic field are governed by the usual laws of hydromagnetics.

The basic dynamical equations governing the behavior of the stars, the gas, and the magnetic field are summarized elsewhere and will not be reproduced here. (See, for example, Lin, 1965a, Section 2). Information about other components in a galaxy and about galaxies in general may be found in Appendix I of the same article.

2. Statement of the problem

Although not all galaxies are disk-shaped like our own, these galaxies comprise about 70% of the total galactic population. With the exception of a small class called SO galaxies, all the disk-shaped galaxies have a rather prominent spiral structure. (Imagine the surprise of the first astronomer who found these regular spiral-like objects in distant space!) The mechanism for the formation of the spiral patterns has been the subject for intensive investigation for the past few decades. There is little doubt, from the observational data available, that these magnificent manifestations are associated with the gas and the young stars being born in them. But could the old stars also play an important role in the formation of the spiral structure? Could the gas?

A flat galactic system must be in rapid rotation; otherwise, the gravitational field would have pulled the system together. It is well-known from observations that the galaxies are generally not in uniform rotation. Indeed, in our own galaxy, the mean velocity of the stars in rotational motion remains nearly constant at 250 km sec, over a wide range of distances from the galactic center. Thus the angular velocity increases as one moves toward the center. Individual stars also have their own peculiar velocities. Thus, to construct a theory of the spiral structure, one must bear in mind the following components of a galaxy:

(a) The stars — with their gravitational forces, circular velocity, and velocity dispersion.

(b) The interstellar gas — with its gravitational field and pressure.

(c) The magnetic field — which exerts its influence through the highly conducting interstellar gas.

A complete theory should take all these components and forces into account, and put their relative importance into perspective.

One of the important features of the spiral pattern is the following. There is a correlation of the spacing for spiral galaxies of the various types as determined by other physical characteristics, as shown in the following table:

	Sa→ Sb→ Sc.
Nuclear concentration	decreasing
Gas content	increasing
Arm spacing	increasing
Total mass	decreasing

Now, one would naturally attempt to associate a spiral arm with a given body of matter. However, in a differentially rotating system, one would expect the pattern to change: the spacing in one given galaxy would increase or decrease rather rapidly in the course of time, which would be contrary to observational evidence. Furthermore, there appears to be a grand design over the whole disk. The situation is very well described by Oort (1962):

> "In systems with strong differential rotation, such as is found in all non-barred spirals, spiral features are quite natural. Every structural irregularity is likely to be drawn out into a part of a spiral. But this is not the phenomenon we must consider. We must consider a spiral structure extending over the whole galaxy, from the nucleus to its outermost part, and consisting of two arms starting from diametrically opposite points. Although this structure is often hopelessly irregular and broken up, the general form of the large-scale phenomenon can be recognized in many nebulae."

All this leads to the rather natural conclusion that the spiral pattern is essentially a wave pattern rotating around the galactic center. If one can establish the existence of wave patterns, the appearance of a grand design would be easy to explain. The superposition of several such patterns would create something less perfect and comparatively transitory, such as that observed in a typical galaxy. But at least a typical scale would then persist. Such a theory has, however, to pass certain elementary kinematical tests. For one thing, the stars at various locations are moving at different angular velocities whereas the pattern must essentially rotate at a given angular velocity. Thus, some of the young stars must eventually drift out of the gas concentration where they are supposed to be born. It is however known that the most brilliant stars of the O and B types invariably stay close to the gas con-concentration. This dilemma can be avoided by noting that these stars have an age of less than 10 million years and that detailed calculations can be made to show that the separation produced during such a short time is always small, within the range of observations.

A similar kinematical test has to be passed by the observed motion of the gas. We shall not go into the details here. Suffice it to say that the pattern concept meets with no essential difficulties.

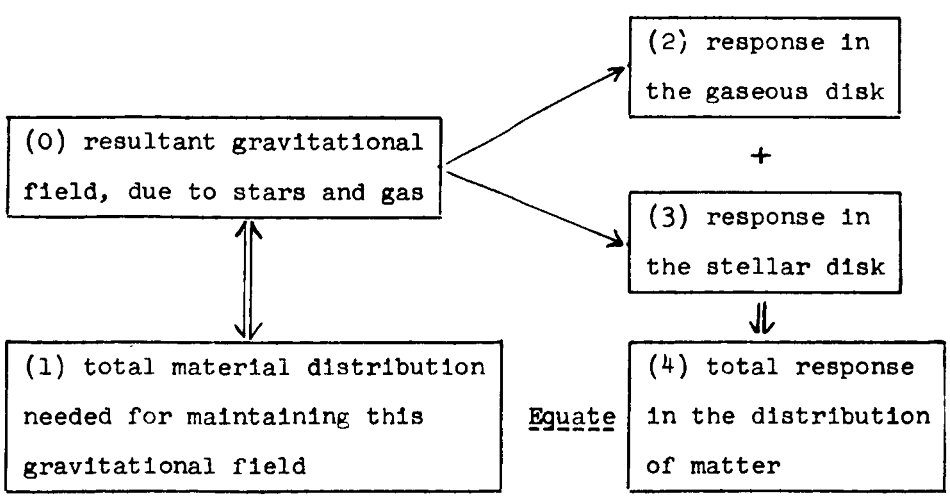
For a detailed discussion of both kinematical tests, see Lin, 1965b.

3. Dynamical processes

We now turn to the consideration of the mechanism for the creation and maintenance of the spiral pattern. After an estimation of the relative importance of the various components mentioned above, we assert the following hypothesis:

QSSS Hypothesis. The total stellar population, which has various degrees of velocity dispersion, forms a quasi-stationary spiral structure in space. This is primarily due to the effect of gravitational instability as limited by velocity dispersion (and secondarily to the influence of the gas and the magnetic field). The extent of density variation in the spiral pattern may be only a small fraction of the symmetrical mean density distribution.

To demonstrate the hypothesis, we adopt a method of self-consistent gravitational field. Suppose such a wave pattern were maintained by gravitational forces, then there must be an associated gravitational field with a spiral pattern. We shall start with this field and carry out the analysis as indicated in the following diagram:



The resultant gravitational field (0) must be associated, according to Poisson's equation with a certain distribution of matter (1), which may consist partly of gas and partly of stars. The distribution of gas (2) may be calculated in terms of the resultant gravitational field without any further reference to the distribution of the stars. Similarly, the distribution of the stars (3) may be calculated without any further reference to the distribution of the gas. The sum of these two distributions yields a total distribution of matter (4) that must be identical with the density distribution (1) which is needed to give rise to the field. The last condition is the equation to be solved for the unspecified functions and parameters that occur in the resultant gravitational field (0) initially assumed.

The analysis has so far been carried out only for the linear theory. Althouth the non-linear effects may be expected to tend to prefer trailing patterns, such a preference will be seen to occur already in the linear theory.

To carry out the analysis, we adopt the natural cylindrical coordinate system (ϖ, θ, z) such that the galactic disk is in the plane $z = 0$ with its center at the origin. In the linear theory, the gravitational potential (Item (0) above) may be assumed to be given by a superposition of spiral modes, and the response to these individual modes may be treated separately. Let the potential of each of these modes be given by

$$(3.1) \quad \mathcal{V}_1(\varpi, \theta, t) = A(\varpi) \exp\left\{ i \left[\omega t - m\theta + \Phi(\varpi) \right] \right\}$$

where $A(\varpi)$ and $\Phi(\varpi)$ are real functions of the axial distance ϖ, ω is a complex parameter, and m is a positive integer. The function $A(\varpi)$ is slowly varying with ϖ, whereas the function $\Phi(\varpi)$ is of the form $\varepsilon^{-1}\phi(\varpi)$, where $\phi(\varpi)$ is slowly varying, and ε is a small parameter of the order of the angle of inclination i of the spiral arms. Indeed, the function (3.1) clearly has a spiral structure described by the family of curves

$$(3.2) \quad m(\theta - \theta_0) = \Phi(\varpi) - \Phi(\varpi_0),$$

which has an angle of inclination i given by

$$(3.3) \quad \tan i = m(k\varpi)^{-1} \quad \text{where} \quad k = \Phi'(\varpi).$$

Thus, a natural approach is to adopt an asymptotic solution based on a rapidly varying phase angle for all the functions of the general form (3.1).

4. Density waves

The detailed analysis may be found elsewhere (cf. Lin 1965a, Lin and Shu, 1966). We shall give here only the final results that relate the characteristics of the density wave to those of the basic distribution function.

Suppose that the basic distribution has a surface mass density given by $\sigma_*(\varpi)$ for the stars and $\sigma_0(\varpi)$ for the gas. There is an associated symmetrical gravitational field pointing toward the galactic center. The circular velocity that is needed to balance this field will be denoted by $\varpi\Omega(\varpi)$. Corresponding to this circular velocity, there is an epicyclic frequency κ defined by

$$\kappa^2 = (2\Omega)^2 \left\{ 1 + \frac{\varpi}{2\Omega} \frac{d\Omega}{d\varpi} \right\}, \tag{4.1}$$

i.e. relative to an observer moving with the circular velocity, a class of stars will be seen to be moving in epi-cycles with an angular velocity κ with the observer as the guiding center. The peculiar velocity of the stars at (ϖ, θ) relative to the above-mentioned observer, will be denoted by (c_ϖ, c_θ). It is known from both observations and theoretical considerations that these should satisfy the Schwarzschild elliptical distribution as a first approximation. The Schwarzschild distribution is similar to a two-dimensional Maxwellian distribution, but with the mean square values $\langle c_\varpi^2 \rangle$ and $\langle c_\theta^2 \rangle$ not equal to each other. Indeed, it can be shown that

$$\langle c_\varpi^2 \rangle : \langle c_\theta^2 \rangle = (2\Omega)^2 : \kappa^2. \tag{4.2}$$

The response of such a stellar disk and of a similar gaseous disk to the gravitational potential (3.1) is found to yield a density distribution that is appropriate to maintain the original field, thus resulting in self-sustained density waves. Indeed, to the order of the initial approximation in the small parameter ε, these waves are found to be neutral. When the theory is carried out to the order of ε, it will be seen that trailing patterns, with $k(\varpi) = \Phi'(\varpi) < 0$, are preferred. It is generally believed that indeed only trailing patterns have been observed. It may then be expected, by analogy with the theory of hydrodynamic stability that eventually the pattern will settle down to a small but finite amplitude that is proportional to the square root of the amplification rate.

The radial wave number of the neutral wave (whether trailing or leading) is given by

$$|k(\varpi)| = \frac{\kappa^2 (1 - \nu^2)}{2\pi G [\sigma_0 + \sigma_* \mathcal{F}_\nu(x)]}, \tag{4.3}$$

where G is the gravitational constant,

$$\nu = (\omega - m\Omega)/\kappa, \tag{4.4}$$

$$x = k^2 \langle c_\varpi^2 \rangle / \kappa^2, \tag{4.5}$$

and $\mathcal{F}_\nu(x)$ is the <u>reduction factor</u> defined by

$$\mathcal{F}_\nu(x) = \frac{1 - \nu^2}{x} \left\{ 1 - \frac{\nu\pi}{\sin \nu\pi} \mathcal{G}_\nu(x) \right\}, \tag{4.6}$$

where

$$\mathcal{G}_\nu(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-x(1+\cos s)} \cos \nu s \, ds. \tag{4.7}$$

The significance of $\mathcal{F}_\nu(x)$ is seen as follows.

Equation (4.3) is derived by assuming that the gas molecules have negligible kinetic motion. The stars are however assumed to have a velocity dispersion measured by the parameter defined by (4.5). This velocity dispersion naturally tends to smooth out any uneven distributions, especially those at smaller scales (higher values of $|k|$). Accordingly, their gravitational effect is <u>reduced</u> by a factor $\mathcal{F}_\nu(x)$, whose value decreases with increasing values of x (cf. Fig. 2).

Similarly, there should be a reduction

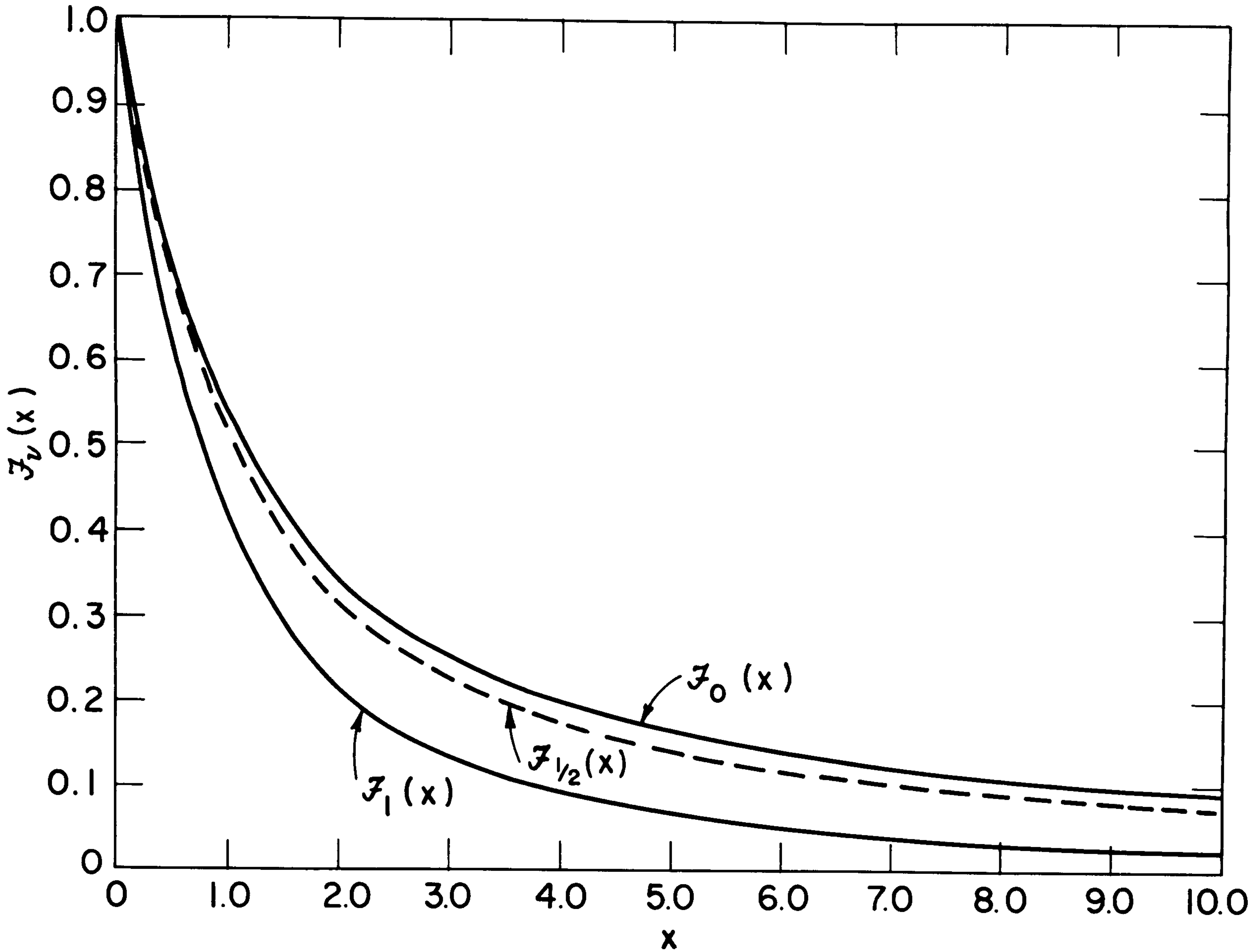


FIGURE 2. THE REDUCTION FACTOR

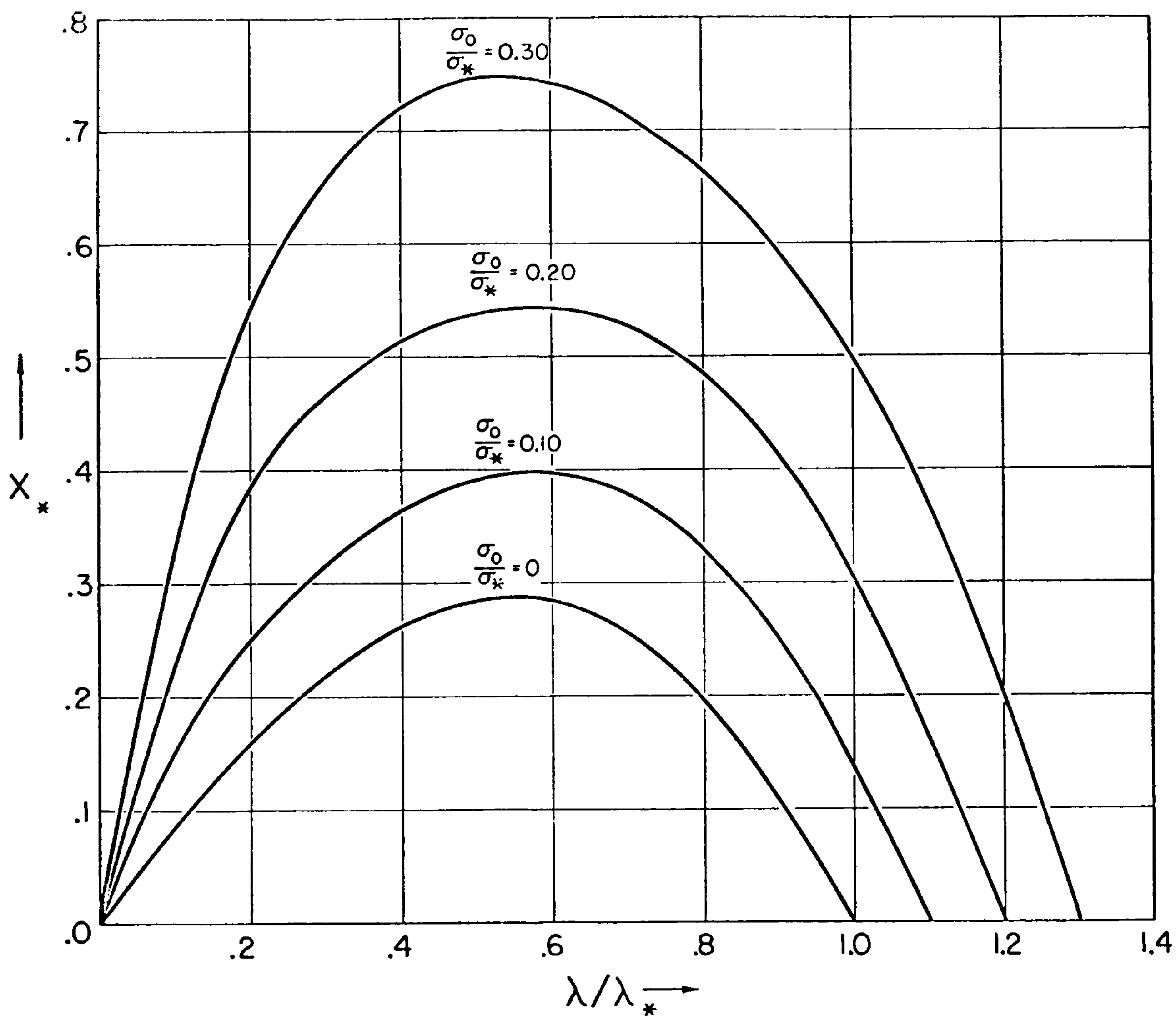


FIGURE 3. CURVES SHOWING THE MARGIN OF STABILITY UNDER VARIOUS CONDITIONS

factor applied to the gaseous density $\sigma_0(\varpi)$ when the effect of pressure or turbulent motion is included.

5. Application to a galactic model

(a) Eq. (4.3) shows the possible existence of neutral waves over a range of frequencies and scales. It also includes the unstable disturbances with $\nu^2 < 0$. The margin of stability is given by $\nu^2 = 0$. For a given value of σ_0/σ_*, this yields a relation between $x = k^2\langle c_\varpi^2\rangle/\kappa^2$ and $|k|/k_*$, where

$$k_* = \kappa^2/2\pi G\sigma_0 . \tag{5.1}$$

This relationship is shown in Fig. 3, which was calculated with the inclusion of a turbulent velocity in the gas, with a mean square value equal to one-tenth of that of the stellar dispersion. The ordinate $x_* = x(k_*^2/k^2)$ is a direct measurement of the dispersion velocity. The curves in Fig. 3 show that the dispersion velocity must reach a certain level before the system can be free from strong instabilities in the nature of a gravitational collapse. On the other hand, if the dispersion velocity is initially insufficient, this gravitational instability would be the mechanism through which dispersion velocities can be increased. (Toomre, 1963). Thus, the existing value of the dispersion speed might be expected to be given by the maximum point of the curve for each value of σ_0/σ_*. From this condition, we may then obtain an estimate of the distribution of $\langle c_\varpi^2\rangle$ as a function of the radial distance.

(b) Once $\langle c_\varpi^2\rangle$ is given, Eq. (4.3) gives, for each chosen value of m, and each chosen value of ω, a radial wave length. Observationally, we know that two-armed spiral patterns ($m = 2$) are always observed. The reason for this is rather subtle. It is not the exclusion of other patterns, but rather that two-armed patterns are most prominent. From (4.3) it is seen that the condition

$$\kappa^2 - (\omega - m\Omega)^2 > 0 \tag{5.2}$$

must be fulfilled. We shall refer to the range of values of ϖ for which (5.2) is satisfied as the <u>principal part</u> of the spiral pattern. Some spiral features might still be present outside this principal part, because the material there must be driven by the density waves. It turns out that with the usual distributions of $\Omega(\varpi)$ and $\kappa(\varpi)$, the principal part of the spiral pattern has <u>an appreciable extent only in the case</u> $m = 2$. Thus, in our own galaxy, if one adopts the values $m = 2$, and $\Omega_p = \omega/m$ = 20 km/sec per kpc, the principal part extends between 2 kpc and 12 kpc, which includes most of the galactic disk.

(c) What then determines the scale of the final observed features? There are several reasons for small scale features to become prominent, when there is a superposition of modes of all scales. Firstly, one has to consider the components of interest to us, namely, gas and young stars with low velocity dispersion. These play a larger role in the <u>total</u> gravitational field when the scale is small. Thus, if we have comparable gravitational fields at various scales, the small scale features would show up more prominently.

Secondly, when one makes an observation, the large scale features represent a gradual change, and are therefore less noticeable. Thus, a mode on a scale extending over the whole disk ($m = 2$) would presumably only cause the disk to take on a slightly elliptic shape without obscuring the small scale features. (This is in line with the familiar Gibbs phenomenon in Fourier series).

Thirdly, when finite amplitude calculations are made, it might turn out that the small scale features are preferred.

The above arguments would show that the smaller the scale, the more prominent the feature should appear. The observed 2 kpc spacing must then be caused by an actual diminution of the amplitude of the waves at still smaller scales. The reason is not too difficult to find. The effective thickness of the stellar disk is about 600 pc. This causes a substantial reduction of the relevant gravitational effect even at 2 kpc. This particular limiting effect increases rapidly with decreasing scale, and it is likely that density waves at smaller scales can hardly be maintained.

Finally, there is the question of a prominent pattern discussed as item (b) above. If one adopts the value $\Omega_p = 20$ km/sec per kpc, one gets an arm spacing of about 2 kpc throughout the galactic pattern. If a very different value of Ω_p is adopted, the extent of the pattern would be substantially reduced while the scale is changed.

(d) Implications on observations

One of the most significant implications of the gravitational theory is the modification suggested in the reduction of data obtained from 21 cm. radio-frequency observations of the gaseous motion. In particular, the method used to locate the gas depends on the assumption of exactly circular motion. According to the present theory, this is not correct, and significant modifications may therefore result in the data reduction.

We have also checked the theory against observations of gas motion in the solar vicinity, which present some puzzling features if one does not assume the existence of a spiral pattern for the distribution of gaseous velocity. These difficulties are immediately cleared up in the present theory; but they are too detailed an issue to be discussed here.

6. Concluding remarks

(a) The magnificent spiral patterns in disk-like galaxies can be satisfactorily explained in terms of density waves. These are the collective modes of the stellar system analogous to plasma waves. Indeed, if one expands the reduction factor $\mathcal{F}_\nu(x)$ in terms of a Mittag-Leffler series in the complex variable x, one obtains from (4.3) a dispersion relationship familiar in plasma physics, with only one minor difference caused essentially by differences in geometry. This serves to remind us that the study of all natural phenomena would usually aid one another in ways which may not be suspected beforehand, even though it might become quite obvious a posteriori.

(b) There is still a great deal of work to be done on the problem of galactic patterns if we wish to bring the theory to the usual standards of sophistication and exactness in Applied Mechanics. Even at the present stage of the development of the theory, we are in a better position to tackle the problems at a smaller scale: the problem of star formation - and the subsequent problem of the birth of the planets. Looking toward larger scales, we have the problems of the distribution of galaxies in clusters, and the distribution of the clusters in the universe. These are fascinating problems which will remain a challenge to us for a long time to come.

Acknowledgment

This work was partially supported by the National Science Foundation and the National Aeronautics and Space Administration.

Bibliography

1. Kalnajs, A. J. (1965) "The stability of highly flattened galaxies"; Ph.D. Thesis, Department of Astronomy, Harvard University.

2. Lin, C. C. (1965a) "On the mathematical theory of a galaxy of stars"; lecture at SIAM Symposium in honor of Richard Courant; to appear in SIAM Journal.

3. Lin, C. C. (1965b) "Stellar dynamical theory of normal spirals"; lecture at AMS Summer Seminar at Cornell University, 1965; to appear in the Proceedings.

4. Lin, C. C. and Shu, F. H. (1964) "On the spiral structure of disk galaxies" Ap. J., vol. 140, pp. 646-655.

5. Lin, C. C. and Shu, F. H. (1966) "On the spiral structure of disk galaxies II. Outline of a theory of density waves" Proc. Nat. Acad. Sci., February 1966 issue.

6. Oort, J. H. (1962) "Interstellar matter in galaxies," ed. L. Woltjer (New York, W. A. Benjamin, Inc.) pp. 234-244.

7. Toomre, A. (1964) "On the gravitational stability of a disk of stars" Ap. J., vol. 139, pp. 1217-1238.