The structure of spiral galaxies

L. S. Marochnik and A. A. Suchkov

Rostov-on-Don State University
Usp. Fiz. Nauk **112**, 275–308 (February 1974)

Chapter I (the Introduction) formulates the problem of the spiral structure. A survey of recent observational data on the structure of spiral galaxies is given in Chap. II. Data are submitted on the physical composition of the spiral arms, the distribution of the stellar and gaseous constituents of the galaxies, and the physical state of the interstellar medium. Chapter III deals with the wave theory of spiral structure. In the contemporary conception, the spiral arms are waves propagating in the galaxy at frequencies determined by some exciting mechanism (a generator), with their wave number determined by the dispersion properties of the galaxy. In this approach, the speed of rotation of the spiral pattern, which equals the frequency divided by the number of arms, does not depend on the distance to the center of the galaxy; this eliminates the problem of the twisting of the spiral arms by differential rotation. The conclusions of the wave theory are discussed, and a dispersion equation is given for the spiral structure of a galaxy. The problem of generation and maintenance of the sprial pattern is discussed. Numerical galaxy-modeling experiments are examined in Chap. IV. These experiments point to the formation of spiral structures of wave nature in disciform systems of gravitating particles. Critical remarks concerning the numerical experiments are offered. The Conclusion (Chap. V) discusses arguments for and against the wave theory. It is noted that this theory explains many of the observed facts and that many of its predictions have also been confirmed by observations.

CONTENTS

I. Introduction 85
II. Observational Data on the Structure of Spiral Galaxies 86
III. The Wave Theory of the Spiral Structure 89
IV. Numerical Experiments 97
V. Conclusion. Recent Data. Pros and Cons of the Wave Theory. Alternative. Hypotheses 98
Bibliography 100

I. INTRODUCTION

The mystery of the spiral structure of the giant galaxies has intrigued investigators for more than half a century. It is a subject of lively debate, inspiring clever hypotheses and sometimes even quite radical ideas affecting the most fundamental of the elements on which our physical picture of the universe is based. Back in 1928, the singular nature of the spiral-structure phenomenon moved Jeans to write that: "Each failure of an attempt to explain the origin of the spiral branches makes it more and more difficult to resist the suspicion that forces totally unknown to us are in operation in the spiral nebulae—forces that perhaps reflect new and unexpected metric properties of space. We return again and again to the notion that the centers of the nebulae are of the nature of "singular points." From these points, matter flows into our Universe from some other, totally extraneous space. To the inhabitant of our Universe, therefore, the singular points appear as points in which matter is continuously generated" [1a] (see [1b]).

These lines anticipated, as it were, the drama of ideas that has been unfolding in contemporary astrophysics with the discovery of a number of surprising phenomena and objects in the Universe.

It can be stated that the problem of the spiral structure was one of the first "dramatis personae" to appear on stage. Heisenberg, Weizsacker and Chandrasekhar, Fermi[2], and numerous other authors have attempted to solve it. But even in recent years, with our substantially advanced understanding of the processes that take place in stellar systems and with the discovery of possible mechanisms for the formation and existence of spirals, mechanisms that involve the physical properties of such systems, the situation is still, on the whole, unclear, and there is still room for unorthodox hypotheses.

In 1964, Vorontsov-Vel'yaminov arrived at the conclusion that the entire observed variety of galactic structural features cannot be explained on the basis of gravitational, electromagnetic, and other known interactions[3]. Arp[4] cites a series of observational data that he considers to favor the hypothesis in which the spirals originated as a result of gigantic explosions in the galactic nuclei. Hoyle does not reject the possibility that the spiral branches formed as a result of production of matter in the galactic nuclei, with subsequent outflow of this matter from the center. Nevertheless, the recent rapid development of theoretical research on the physical properties of stellar systems of the type of the observed spiral galaxies justifies optimism as to the possibility of explaining the spiral structure of the galaxies within the framework of known physical laws. It is found that the fundamental difficulties of theory are resolved if the possibility of collective excitations in stellar systems is taken into account. We refer here primarily to a property of rotating disk-shaped galaxies that consists in the possible existence of spiral mass-density waves in these galaxies.

According to the wave theory that is now being developed in many countries, the spiral branches of the galaxies are nothing else but density waves propagating across of the galactic disk, and the process of propagation is one of rotation of the wave as a solid (in contrast to the differential rotation of the matter in the galaxy).

Our object here is to review observational data in the spiral structure of galaxies and the present state of the theory of the spiral structure.

II. OBSERVATIONAL DATA ON THE STRUCTURE OF THE SPIRAL GALAXIES

1. Morphology and physical properties of the spiral galaxies. The "grand-design" brightness distribution patterns of most observable spiral galaxies have many features in common. This enables us to devise a kind of morphological classification that groups the galaxies on the basis of apparent (optical) criteria. The first classification of this kind was the well-known Hubble sequence[5], which has retained its central significance to this day (Fig. 1). This sequence divides all galaxies into three major types—types E, S, and SB. The first includes the spherical and elliptical galaxies, the second the so-called normal spirals, and the third spirals with a central bar. The second and third types are broken down into into three subtypes: Sa, Sb, Sc, and SBa, SBb, and SBc, which correspond to the transition from tight (closed) spirals to open spirals.

This simple classification is highly convenient in many respects. Its chief merit, in virtue of which it is still useful today, is apparently that it usually reflects not only the visible features of the galaxies, but also to a certain extent their physical properties.

The classification proposed by de Vaucouleurs[6] is widely used (Fig. 2). It is also based on visible (optical) attributes, and, as compared with Hubble's classification, takes more detailed account of differences in the spiral designs of the galaxies. In speaking of the visible features of global structure, which both Hubble and de Vaucouleurs classified as different types of spirals (and rings), it must be remembered that there are many galaxies with a regular grand-design structure that have little in common with the spiral galaxy. A considerable number of galaxies of this type that do not fit into the known morphological classifications are enumerated in[3].

2. Relation between the visible spiral design and the physical properties of the galaxies. In very many cases, it is possible to discern a dependence of the nature of the spiral design, specifically the tightness of the spiral branches (the basis of Hubble's classification) and their thickness, on the global mass distribution in the galaxy. Thus, Sa galaxies, which are characterized by tightly wound spirals, usually show a strong concentration of mass toward the center. Here the great bulk of the total mass is situated in the central region and forms what might be called a central body (bulge). The strong concentration of mass toward the centers of Sa galaxies is associated with a dense spherical subsystem of old stars (stars of late spectral classes), which have widely scattered velocities and small systematic (rotational) motion. The transition to more open and wider spirals, i.e., from Sa to Sb and Sc, is accompanied by a decrease in the significance of the bulge. Type Sc spirals have a much smaller concentration of mass toward the center. A flat subsystem of young stars (stars of early spectral classes) becomes a significant factor in their dynamics. Rotation is more important for this subsystem, while the scatter of the star velocities is small.

Van den Berg[7] recently drew attention to the following fact. As a rule, supergiant galaxies have an extensive, wide, and well-developed spiral structure. Normal galaxies usually have only rudimentary structure. And, finally, dwarf galaxies of the late type have no spiral structure at all. Thus, note is taken of a relation between the mass of the galaxy and the development of spiral structure.

3. Trailing and leading spirals. The question as to the direction in which the spiral branches are wound is a weighty one in the spiral-structure problem. Are the

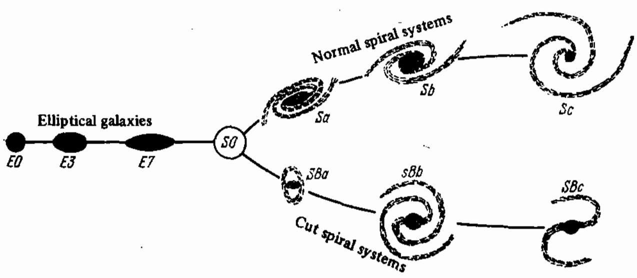

FIG. 1. Hubble's "tuning-fork" diagram.

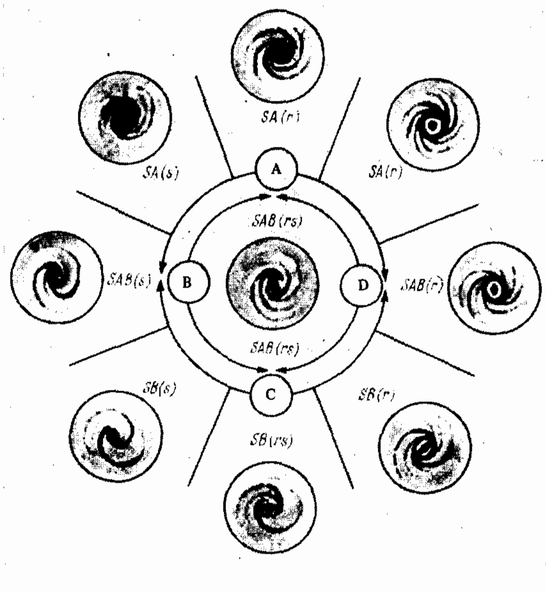

FIG. 2. Classification of galaxies according to de Vaucouleurs[6].

spiral branches trailing, i.e., do they open counter to the direction of rotation, or are they opening in the direction of rotation, i.e., leading? There is as yet no definite answer to this question. Nevertheless, the overwhelming majority of investigators are inclined to favor trailing branches. The state of this problem is discussed in detail in the review[6] (p. 376).

Vorontsov-Vel'yaminov's discovery of the so-called γ forms[3] is highly important for the problem at hand. Coexisting spiral branches with opposite directions of twist are really observed in several galaxies. On intersecting, these branches form a structure with a superficial resemblance to the Greek letter γ. Since the probability of coexistence in a galaxy of subsystems that rotate in opposite directions is very small, Vorontsov-Vel'yaminov concludes that the spirals in a galaxy can, in principle, be either leading or trailing. We do indeed observe coexistence in the γ forms. He arrived at the same result in an analysis of the structure of close pairs of spiral galaxies[8].

4. Physical composition of the spiral branches. The spiral structure of external galaxies is observed in the optical band (it has also been possible to detect spiral structure in the radio band in M 31, M 51, and NGC 4258[9]. Thus we see primarily a brightness distribution, which, generally speaking, does not reflect the over-all mass distribution. The spiral branches owe their high luminosity to the fact that the brightest objects in the galaxy are concentrated in them: giant stars of the early OB spectral classes and clusters of clouds of ionized hydrogen HII which have high luminosity in their emission lines.

Certain galaxies show a distinct spiral structure outlined by stars of later spectral class that belong to the old population of the galaxy's spherical and disk subsystems. This phenomenon was discovered by Zwicky[10] in the galaxy M51, which has broad spiral branches of yellow stars. A similar picture was later observed in certain other galaxies.

Analyzing the presently available data on the composition of the spiral branches, we may conclude the existence of four basic types of objects that indicate spiral structure: a) HII regions; b) belts of dust (clouds of absorbing dust); c) blue supergiants (with a certain admixture of red supergiants of the h-χ Persei type) and d) "nonblue" stars (see, for example, [11]. It also appears that pulsars concentrate in the spiral branches[12]. Morgan[11] called for particular attention to the fact that each of the four types of indicators determines the structure of the spiral branches in a different way. He observed that the HII regions determine arm segments that are narrow and noticeably irregular in shape. Arms defined by dust spots are also narrow, but are quite regular in the internal regions of the main body of the branch. The pattern defined by the dust spots often coincides with that obtained from the HII regions. Blue supergiants form spiral-like segments whose shape is less clearly defined than by objects a) and b). The "nonblue stars" in internal regions of the galaxy have the smoothest and most regular configurations. The local width of the branch is also different for the different objects: it increases from HII regions (narrowest branches) toward the "nonblue stars" (the broadest branches). These indicators often yield totally different pictures of the spiral structure in internal and external regions of the galaxy. Thus, there are two principal spiral branches of stars in the internal region of M 38; in the outer region, the structure is complex, with many arms, and is to a substantial degree independent of the other principal arms[11].

The substantially different spatial arrangement of the spirals of young (type a), b), and c)) and old (type d) objects would appear to be the most remarkable feature[10]. The corresponding branches are found to be shifted considerably with respect to one another in azimuth. This circumstance should be an important factor in the theory of the origin of the spiral branches, in attempts to explain the phenomenon of spiral structure.

Spatial noncoincidence of the indicators of spirals with different ages has also recently been detected in objects that do not differ so greatly in age. For example, it is reported in[13] that the younger Cepheids in the spiral branches of M 31 concentrate toward the interior of a spiral branch, and old ones toward the exterior. Thus, we observe a distinct stellar-age gradient across the arm. There have been references to the existence of the opposite gradient in M 33[14]. It is possible that a similar pattern obtains in our Galaxy[15,16].

But it also appears that the ages of the objects concentrated in the spirals (at least young ones) do not depend on distance from the center, i.e., there is no age gradient along the branch. This has been demonstrated by Markaryan for stellar associations in M 51 and M 101[17], as well in NGC 6946[18].

5. Distribution of neutral hydrogen H I. The spiral structure of our Galaxy. Neutral hydrogen H I, which is observed for the most part in the 21-cm radio line, is of special importance as an indicator of spiral structure. Our galaxy is transparent in this line, and the distribution of H I regions can be observed throughout most of its volume.

Westerhout[19] was the first to chart the neutral hydrogen distribution of the Galaxy in any detail. Oort et al.[20] interpreted the resulting distribution as representing the spiral-structure pattern of the Galaxy with tightly twisted spiral branches of the kind typical for type Sb galaxies. A refined picture was proposed by Kerr and Westerhout[21]. On the basis of recent observations, Kerr[22] obtained a more complete and accurate version of the H I distribution pattern (Fig. 3).

According to Kerr, H I forms spiral branches one of which passes through the part of the Galaxy outward of the sun, while the other two lie between the sun and the "three-kiloparsec arm." The twist angle of the branches is 7°.

A substantially different picture of the spirals was constructed by Weaver[24] (see also[25]). Here there is only a single principal spiral arm, which has a twist angle of about 12°.5 and passes near the sun. In the other regions, in which Kerr draws regular principal arms, Weaver finds only secondary local features, such as branchings, etc. (see Fig. 3). Thus, we now have two divergent pictures of the Galaxy's spiral structure. The difference between Kerr and Weaver arises out of the following state of affairs. With the increasing resolving power of radio telescopes, the finer structure of the radiohydrogen distribution has recently been established. It has proven to be extremely irregular; the hydrogen is concentrated in chaotically scattered and quite extensive but local formations. The global picture depends mater-

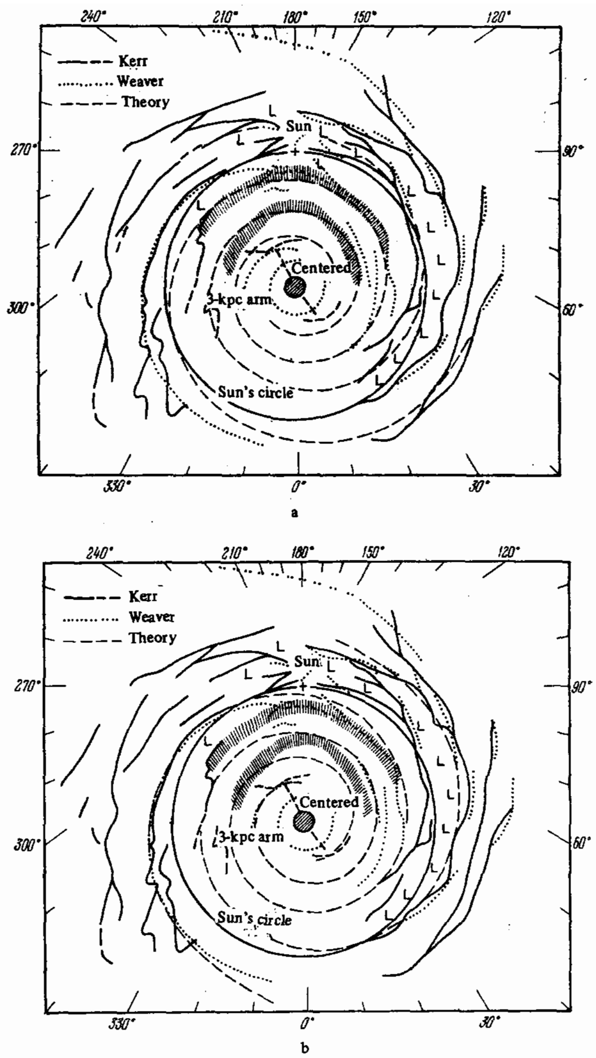

FIG. 3. Neutral hydrogen distribution charts according to Kerr (1969) and Weaver (1970) for the Galaxy[23] and the theoretical spiral structure[75]. The sun is situated at a distance of 10 kpc (plus sign). The dashed line indicates the theoretical design for mode k_0 with m = 2 and Ω_p = 23 km/sec-kpc for comparison with the charts of Kerr (a) and Weaver (b).

ially on how we join these local formations. Using different criteria for this joining, Kerr and Weaver obtained different pictures. On the whole, the strong irregularity of the HII regions that has recently been discerned has moved certain astronomers to favor the use of the term "spiral structural features" instead of "spiral arms"[23].

In regard to the optical spiral structure, most astronomers have now accepted the conclusion[23] that the concentration of young stars observed in the neighborhood of the sun, which form the so-called Orion arm, do not represent a principal spiral branch. On Kerr's chart, this optical formation has nothing in common with the H I distribution; it forms a rather large angle with the hydrogen spiral branch, and has a twist angle of ≈ 25°. According to Weaver[24], the region in which the Orion arm is localized coincides with a ramification, also at a twist angle of ≈ 25°, of neutral hydrogen from the main branch. It is Weaver's view that this deals once and for all with the long-disputed problem of the noncoincidence of the optical and radio patterns of the spiral branches in the vicinity of the sun.

The optical spiral structure of the Galaxy as recently constructed by Courtes et al.[26] now appears highly interesting. Here the optical H II regions serve as branch indicators. Observations of the radial velocities of 6000 H II regions were used to construct a picture that includes four spiral arms whose localizations agree with those of the Perseus, Cygnus, Sagitta-Carina, and Norma-Centaurus arms, which are known from H I regions. However, in disagreement with the 6–7° twist of the branches according to Kerr, the twist angle of the branches in the new pattern of the spiral structure is 20° (Fig. 4;[26b]).

6. State of the interstellar medium. Since our basic information on the spiral structure of the Galaxy is obtained from hydrogen data, it is important to know the physical state of the interstellar medium if we are to compare theory with observations. Study of processes in the interstellar medium, and of the large-scale motions of the interstellar gas in particular, may yield new clues as to the mechanism of formation and maintainance of the spiral design.

Three regions can now be distinguished in the interstellar medium: a hierarchy of gas concentrations (gas clouds), the intercloud medium, and the interarm medium[23]. Clark proposed in 1965 that the gas in the arms is in the form of cold, dense clouds and a rarefied hot intercloud gas that may be in pressure equilibrium with the clouds[27]. This has been confirmed by subsequent studies (see the review[28]). The approximate temperature range of the clouds is 20 to 160° K, while the intercloud medium has temperatures of the order of several thousand degrees Kelvin[23]. The interarm medium is extremely rarefied; its density is no more than one-tenth the average density in the arms, and its temperature is on the order of several thousand degrees (see, for example, [29]).

Large-scale motions of enormous gas complexes are observed in the plane of our Galaxy. These are the well-known "3-kiloparsec arm," which moves radially outward from the center at a velocity of 50 km/sec, and the "4-kiloparsec arm," which moves in the opposite direction, also away from the center, at a velocity of about 135 km/sec. These processes apparently testify to activity in the central regions of the Galaxy, which eject enormous masses of gas in explosive fashion. This model of the origin of these complexes was analyzed in[30]. There are indications that such explosive processes occur not only in the nuclei of galaxies, but also in the spiral branches (where, of course, they are less powerful). In certain galaxies, e.g., NGC 3486, numerous branchings that form a secondary spiral structure extend almost at right angles from the main branches. This can be linked to activity of the principal branches,

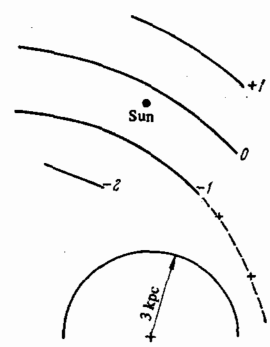

FIG. 4. Diagram of spiral arms in our Galaxy according to observations of H II regions [26b]. The distance to the sun is 8.2 kpc.

which results in the ejection of enormous masses of matter from the central body that are later set in rotation[31]. In our Galaxy, for example, Kepner[32] studied hydrogen clouds of low density that are situated at a considerable distance above the plane of the Galaxy (up to a kiloparsec). Oort associates these clouds with an explosion similar to that known to have occurred in the Small Magellanic Cloud[23], with the result that substantial masses of gas were ejected high above the galactic plane.

It appears that regular motion of neutral hydrogen coincides in our Galaxy with the motion of H II regions. It follows from a study by Mezger et al. (see[35]) that there is no systematic difference greater than a few km/sec between the velocities of H I and H II regions. In all probability, H I and H II regions and young stars generally move together[33]. Among other things, this circumstance can serve as an argument in favor of coincidence of the optical and radio pictures of the spirals in the Galaxy.

Various details of the spiral-structure pattern described above for external galaxies and our own were discussed in greater detail in the reviews[34-41].

III. THE WAVE THEORY OF THE SPIRAL STRUCTURE

The chief problem of spiral-structure theory is to explain the following points: a) why some galaxies have a clearly pronounced spiral design that extends over the entire disk of the galaxy, and b) why this design exists over many revolutions of the galaxy despite the destructive effects of differential rotation. The problem was elegantly formulated by Oort: "In systems with strong differential rotation, such as is observed in all unbarred spiral galaxies, spiral features are perfectly natural. Any structural inhomogeneity will probably be twisted to form part of a spiral. But it is not this phenomenon with which we should be concerned. We must explain the spiral structure that encompasses the entire galaxy, from its nucleus to its outermost reaches, and consists of two arms that emerge from two diametrally opposite points. Although this structure is often hopelessly clotted and irregular, the general pattern of the large-scale phenomenon can still be distinguished in many galaxies"[42].

The first part of the problem (a) is now often referred to as the problem of the existence of the global structure (grand design), while the second part (b) is called the "survival" (persistence) problem.

There are now serious grounds for the idea that the visible spirals in normal galaxies[1)] are spiral density waves, which can exist in such systems, and that their propagation is a process of rotation as a solid about the center of the galaxy at a certain fixed phase velocity. The density-wave hypothesis makes it possible to resolve the basic difficulty in the problem of spiral structure, which is that any nonaxisymmetric conglomerate of matter is quickly set in differential rotation, dissipating in the system after a few revolutions. But if the spiral branches are waves, and not conglomerates of matter, conditions exist for the rotation of such waves as solid bodies despite the general differential rotation of the system; thus, the wave theory answers both of the questions posed above.

The wave theory of the spiral structure was developed over many years by Lindblad, to whom we owe the very idea of spiral density waves in galaxies[44,45]. However, his work was not widely recognized. The principal reason for this seems to have been that his theory, which was constructed in terms of orbits and made use of numerous assumptions and approximations, is difficult to follow in its details. But what was perhaps more important for astronomers was the fact that Lindblad insisted on leading spirals, which was contrary to the prevailing opinion. A new stage in spiral-structure theory is associated with the work of Lin and Shu[46], in which the spiral waves were treated as collective motions in a self-gravitating disk that could be described by the equations of the continuous medium. A more adequate analysis was given later on the basis of a collisionless kinetic equation[47], and Lin et al. gave a complete formulation of the conceptual and quantitative aspects of the theory in[48]. The wave theory was elaborated concurrently on the basis of Lin's ideas by a number of other authors (see below, and also the reviews[23,38-41,49-54]). Many aspects of the wave pattern of the spiral structure have now been clarified as a result.

1. The model of the galaxy. In first approximation, the spiral density waves can be regarded as small perturbations against an equilibrium axisymmetric background distribution of the galactic mass. The properties of the waves are entirely determined by the equilibrium parameters of the system, and for this reason it is first necessary to construct an equilibrium model of the galaxy.

Generally speaking, spiral galaxies consist of a number of subsystems with different kinematic properties and different types of mass distribution.

The oldest stars form a spherical subsystem. Stars of this subsystem exhibit wide velocity dispersion, but the speed at which the subsystem itself rotates is not very high. The mass distribution is almost spherically symmetrical, and there is a strong concentration of mass toward the center. The youngest stars and the interstellar gas and dust form the plane subsystem. It rotates at high velocity, but the objects that form it have a small velocity dispersion. When we speak of the rotation of a galaxy, we usually have in mind the rotation of precisely this subsystem. In the plane subsystem, mass is concentrated in a thin disk, with little concentration toward the center, i.e., density varies little along the radius of the disk. It is important to note once again that it is the objects of this subsystem that trace the spirals in the galaxy.

Between the spherical and plane subsystems, there is a whole series of subsystems with intermediate properties. Five subsystems are recognized in our Galaxy[55,56]. The mass of the plane subsystem is $7 \cdot 10^9 M_\odot$, and the mass of the slowly rotating old subsystems—spherical, disk, and interval—is almost an order larger at $63 \cdot 10^9 M_\odot$. This makes it clear that the rotation of the plane subsystem is entirely determined by the subsystems of old stars, i.e., its equilibrium is not self-consistent.

On the whole, the spiral galaxy is an extremely complex system, and the construction of a self-consistent solution for it, one that would state the relationships among all of the equilibrium parameters of its subsystems (density, velocity of rotation, star velocity distribution function, etc.) is an extremely complex problem that has not been solved to this day. This circumstance forces an appeal to observational data in the choice of an equilibrium model. A self-consistent solution has been obtained in certain studies (see, for example,[57,58]) for an elementary model of an equilibrium galaxy regarded as a rapidly rotating thin disk of stars. However, the inadequacies of this model make it impossible to put these results to practical use in explaining processes in real spiral galaxies.

The distribution of the total mass of the galaxy's subsystems can be obtained from its observed rotation curve (for example, the model of Schmidt[59] for our Galaxy). But the difficulty is encountered when it comes to finding the mass distributions of the various subsystems.

In the equilibrium model, it is necessary to know, in addition to the mass distribution, the form of the velocity distribution function and the dependence of the distribution parameters on distance to the center of the galaxy. Observations indicate that the distribution in our Galaxy in the vicinity of the sun is close to the Schwarzschild distribution (an anisotropic Maxwell distribution)[60,61]. Investigation of certain possible relaxation mechanisms also leads to this distribution (for example,[62–64]). It can be supposed that this distribution is realized in most of the galaxy, although this is not certain. The velocity dispersion is often determined as a function of galactocentric distance on the basis of the hypothesis of marginal stability[2)] of the star disk[65], according to which

$$c_r = 0.085\, \lambda_T\, \varkappa(r), \tag{1}$$

where c_r is the radial dispersion,

$$\varkappa(r) = 2\Omega(r)\sqrt{1 + \frac{r}{2\Omega}\frac{d\Omega}{dr}}$$

is the epicyclic frequency,

$$\lambda_T = \frac{4\pi G\sigma(r)}{\varkappa^2(r)} = \frac{2\pi}{k_T}; \tag{2}$$

G is the gravitational constant, and $\sigma(r)$ is the surface density.

A thin rapidly rotating disk is taken as a model of the galaxy in many papers for analysis of the perturbations. The following arguments, among others, can be cited in favor of this model. Firstly, the spirality of the galaxy is actually conveyed basically by objects of the plane subsystem. Secondly, it can be expected that stars of the spherical subsystem, which have large peculiar velocities, will make a relatively small contribution to the self-field of the wave and will have low density contrast. These circumstances enable us to confine ourselves in first approximation to analysis of a thin-disk model[3)].

Two subsystems of stars corresponding to Populations I and II in the galaxy have been distinguished explicitly in some papers[66–70]. Having wide velocity dispersion, Population II stars can influence only the "subtle properties" of the waves (instability, etc.), without influencing their dynamics. But it is they that determine the rotation of the thin disk of Population I (for further details, see Sec. 6 of Chapter III).

2. Spiral waves in disk galaxies. In most cases, the spiral branches of galaxies represent only small deviations of the density from a density distribution that is axisymmetric in the mean. Accordingly, the gravitational field associated with the branches is a small deviation from a gravitational field that is axisymmetric in the mean. Therefore the spiral structure can be regarded in the first approximation of the theory as a small perturbation of the equilibrium axisymmetric state of the system, and the situation will be described by linear equations. The following statements of the problem then become possible[51,71]:

a) Find the vibrational modes of the self-consistent system.

b) Find the self-consistent solution with initial conditions.

c) Find the response of the system to a predetermined spiral potential.[4)]

In the mode problem, it is necessary to show whether and under which conditions spiral density and potential perturbations are solutions of equations describing the self-consistent behavior of small perturbations in the system. Let us turn our attention to this problem, which is the one of greatest interest in spiral-structure theory.

Let f_1 and φ_1 be small perturbations of the distribution function and the potential, respectively. The perturbed distribution function and the perturbed gravitational potential will be written in the forms

$$\tilde{f} = f_0(r, z, v, v_\theta, v_z) + f_1(r, z, \theta, v_r, v_\theta, v_z, t), \quad |f_1| \ll |f_0|, \tag{3}$$

$$\tilde{\varphi} = \varphi_0(r, z) + \varphi_1(r, \theta, z, t), \quad |\varphi_1| \ll |\varphi_0|. \tag{4}$$

Let us first consider infinitesimally thin disk models. Substituting (3) and (4) into the Liouville and Poisson equations and confining ourselves to linear terms, we obtain for f_1 and φ_1

$$\frac{\partial f_1}{\partial t} + v_r\frac{\partial f_1}{\partial r} + \left(\frac{v_\theta}{r} + \Omega\right)\frac{\partial f_1}{\partial\theta} + \frac{\partial\varphi_1}{\partial r}\frac{\partial f_0}{\partial v_r} + \left(\frac{v_\theta}{r} + 2\Omega\right)v_\theta\frac{\partial f_1}{\partial v_r} + \frac{1}{r}\frac{\partial\varphi_1}{\partial\theta}\frac{\partial f_0}{\partial v_\theta} + v_r\left(\frac{v_\theta}{r} - \frac{\varkappa^2}{2\Omega}\right)\frac{\partial f_1}{\partial v_\theta} = 0, \tag{5}$$

$$\frac{\partial^2\varphi_1}{\partial r^2} + \frac{1}{r}\frac{\partial\varphi_1}{\partial r} + \frac{1}{r^2}\frac{\partial^2\varphi_1}{\partial\theta^2} + \frac{\partial^2\varphi_1}{\partial z^2} = -4\pi G\delta(z)\int f_1\, dv_r\, dv_\theta. \tag{6}$$

It was recognized in the derivation of (5) that centrifugal force is cancelled by gravitation at equilibrium, i.e., $\partial\varphi/\partial r + \Omega^2 r = 0$.

In the general case, it is not possible to obtain a solution of Eqs. (5) and (6) in analytic form. It was shown in[46–48] that such a solution can be obtained comparatively simply by considering "tightly wound" spiral waves described by functions of the type

$$A_1(r, \theta, t) = A^0(r)\exp\{i[\omega t - m\theta + \Phi(r)]\}, \tag{7}$$

where $A^0(r)$ is the amplitude, which varies slowly with r, $\Phi(r)$ is the radial phase, which varies slowly (monotonically) and is a large quantity, ω is a certain constant, and m is an integer. The line of constant values of the functions A_1 is given approximately by the equation

$$m(\theta - \theta_0) = \Phi(r) + \Phi(r_0), \tag{8}$$

which represents a spiral design with m branches. The spiral design rotates at an angular velocity

$$\Omega_p = \frac{\omega_r}{m}, \tag{9}$$

where ω_r is the real part of ω. The radial wave number is $|k(r)|$, where

$$k(r) = \frac{d\Phi}{dr}. \tag{10}$$

Putting $\Omega(r) > 0$ and $m > 0$, we find that the spirals will be trailing if $k(r) < 0$ and leading if $k(r) > 0$.

The large value of $\Phi(r)$ enables us to employ an asymptotic analysis corresponding to the WKB approximation. Substituting f_1 and φ_1 in the forms (7) into Eqs. (5) and (6), we obtain the following dispersion equation as an existence condition for spiral-wave solutions in the first WKB approximation:

$$1=\frac{4\pi G\sigma_0(r)\,e^{-x}\,|k(r)|}{x\kappa^2}\sum_{s=1}^{\infty}\frac{s^2 I_s(x)}{s^2-\nu^2}, \tag{11}$$

where $x = k^2c_r^2/\kappa^2$, $\nu = (\omega - m\Omega)/\kappa$, and $I_s(x)$ are Bessel functions of imaginary argument. It is assumed in the derivation of (11) that the velocity distribution is Schwarzschildian with a dispersion ratio $c_r/c_\theta = 2\Omega/\kappa$, and that the peculiar velocities of the stars are small compared to their circular velocities. Equation (11) was obtained in[48,66]. Also given in[66] is an equation that is valid for an arbitrary distribution function.

Another notation of Eq. (11) is found to be more convenient[48]:

$$\frac{k_T}{|k|}(1-\nu^2)=F_\nu(x), \tag{12}$$

where

$$F_\nu(x)=\frac{1-\nu^2}{x}\left[1-\frac{\nu\pi}{2\pi\sin\nu\pi}\int_{-\pi}^{\pi}e^{-x(1+\cos s)}\cos\nu s\,ds\right]. \tag{13}$$

The function $F_\nu(x)$ is tabulated in[48]. Eqs. (11) and (12) are valid for

$$1-\nu^2>0. \tag{14}$$

The points at which $\nu = \pm 1$ (Lindblad resonance points) require special examination. Equations (11) and (12) cease to be valid in the neighborhood of these points (for details see Sec. 5).

The waves described by Eqs. (11) and (12) exhibit the following properties[48].

a) They occupy a region of the galactic disk in which the condition

$$\Omega-\frac{\kappa}{m}<\Omega_p<\Omega+\frac{\kappa}{m} \tag{15}$$

is satisfied. This region will obviously be more extensive the smaller m. Leaving aside the cases with m = 0 and 1, it follows from (15) that a two-arm structure occupies the largest area of the galactic disk. For the most typical of the observed rotation curves, the two-arm pattern occupies the entire area of the galaxy in which the spiral structure can be traced[48,72]. But for all $m > 2$, the region of existence of the wave is found to be very small for all allowed values of Ω_p. This property of spiral waves may explain why most galaxies are two-armed.

b) Figure 5[73] shows a dispersion curve giving the dimensionless wavelength λ/λ_T as a function of the dimensionless frequency $|\nu|$. It was plotted on the assumption that c_r equals the minimum value determined by the Toomre criterion. Its value is determined from Eq. (12) with the condition $\nu^2 = 0$. The range $\nu^2 > 0$ corresponds to neutral waves, and $\nu^2 < 0$ to instability. But even if $c_r \lesssim c_{r\,min}$, the Jeans critical length, which separates stable from unstable wavelengths, is found to be so large that the Jeans instability is apparently of no significance in spiral-structure theory.

c) It follows from (11) and (12) that trailing and leading spirals have an equal right to exist. Thus, preference may not be given to either type of spiral in the approximation under discussion here.

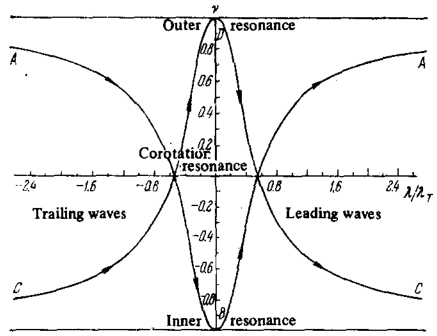

FIG. 5. Dispersion curve of disk marginal stability[73].

Let us consider the role of the subsystem of stars with high peculiar velocities and slow rotation[66–69]. This subsystem can be taken into account by including the density of the stars under consideration, $\int f_2 dv_r dv_\theta$, in the Poisson equation (6) and adding to the equation system (6), (7) the equation for the perturbed distribution function of the stars of subsystem II:

$$\frac{\partial f_2}{\partial t}+v_r\frac{\partial f_2}{\partial r}+\frac{v_\theta}{r}\frac{\partial f_2}{\partial\theta}+\frac{\partial\varphi_1}{\partial r}\frac{\partial f_2^0}{\partial v_r}+\left(\frac{\partial\varphi_0}{\partial r}+\frac{v_\theta^2}{r}\right)\frac{\partial f_2}{\partial v_r}+\frac{1}{r}\frac{\partial\varphi_1}{\partial\theta}\frac{\partial f_2^0}{\partial v_\theta}-\frac{v_r v_\theta}{r}\frac{\partial f_2}{\partial v_\theta}=0. \tag{16}$$

In the WKB approximation, this results in the appearance of an additional term in the right-hand side of (42):

$$\frac{\sigma_2^0}{k}\int_{-\infty}^{\infty}\frac{(\partial f_2^0/\partial v_r)\,dv_r}{(\omega/k)-v_r}, \tag{17}$$

where σ_2^0 is the unperturbed density of subsystem II and, in contrast to (16), the function f_2^0 is normalized to unity and integrated over v_θ.

Let f_2^0 be a Maxwell distribution. The dispersion equation can then be written

$$1=\frac{2\pi G\sigma_1^0}{|k|\,c_r^2}\left[\frac{F_\nu(x)}{1-\nu^2}x+\zeta\int_{-\infty}^{\infty}\frac{f_2^0\,dv_r}{1-(\omega/kv_r)}\right], \tag{18}$$

where

$$\zeta=\frac{\sigma_2^0}{c_2^2}\Big/\frac{\sigma_1^0}{c_r^2}. \tag{19}$$

If $\zeta \ll 1$, solutions of (18) can be sought by expanding in a series in powers of ζ. In the zeroth approximation, we shall obviously simply obtain Eq. (12). This result justifies consideration of only the rotating disk of stars if $c_r^2/c_2^2 \ll 1$ with comparable σ_1^0 and σ_2^0. However, it is evident that, in contrast to[46–48], it is then not the total surface density, but only the density of subsystem I itself that should figure in the dispersion relation.

With a given m, Eqs. (11) and (12) give the relation between ω and k. Two statements of the problem can be considered: a) for a given ω, find k from (11) and (12), and b) having given k, find the corresponding frequency ω. The first formulation corresponds to a physical situation in which a "generator" in the system excites oscillations at a fixed frequency ω in a certain region; propagating, these oscillations form a wave field with wave frequency ω and a wavelength determined by the dispersion properties of the system. If the medium is inhomogeneous, the wavelength will be a function of the coordinates and will be determined by the dispersion equation.

It is the possibility of this situation that offers an explanation for the existence of the global spiral structure and its "persistence" in the presence of differential rotation. In fact, let us assign the frequency with an arbitrary number in the interval (15) and determine the corresponding function $k(\omega, r)$ from (12). The phase velocity of the wave along the radius will then naturally depend on r: $\omega/k(r) = v_f(r)$. But it is important for us that the azimuthal phase velocity $\Omega_p = \omega/m$, i.e., that the angular velocity of rotation of the spiral design does not depend on r. Thus, spiral waves can rotate as solid bodies in systems with differential rotation.

In the above scheme, the velocity of rotation Ω_p of the wave (or the frequency ω) is not defined within the framework of the theory itself, and is a free parameter. Since the pattern of the spiral design correlates with other properties of the galaxies, it is natural to suppose that this frequency is set by a "generator" whose properties are determined by the general physical characteristics of the galaxy (see below).

We can use the equation to construct a picture of the galactic spiral design by determining the function k(r) for a fixed Ω_p. This program was carried out for our Galaxy in[48,74,75] and for the galaxies M 33, M 51, and M 81 in[72]. Here Ω_p was so chosen as to produce the spiral design in closest agreement with that observed.

Equations (11) and (12) determine two functions k(r) for each value of ω, one of which increases without limit at $\nu = -1$ (Lindblad inner resonance), while the other vanishes. The situation is reversed at the outer resonance, i.e., at $\nu = 1$ (Fig. 6). With the behavior near the inner resonance in mind, we shall call the former function the shortwave mode k_∞ and the latter the long wave mode k_0.

If Schmidt's 1965 model is used as the model of the galaxy, agreement can be obtained between the theoretical and observed designs only for the mode k_∞[48] (see Fig. 6). It is to this mode that we shall have reference everywhere in the exposition that follows, except for Sec. 6 of Chapter III, where we shall present alternative considerations that use the k_0 mode in construction of the theoretical spiral design.

3. Comparison of theory with observations. A comparison was drawn in[48] between results obtained on the basis of the wave theory and observational data for the Galaxy. The comparison was based on four points. It considered 1) the distribution of neutral hydrogen (H I regions) in the Galaxy, 2) the systematic motion of the gas, 3) the distribution of young stars and other optical objects, and 4) the migration of fairly young stars.

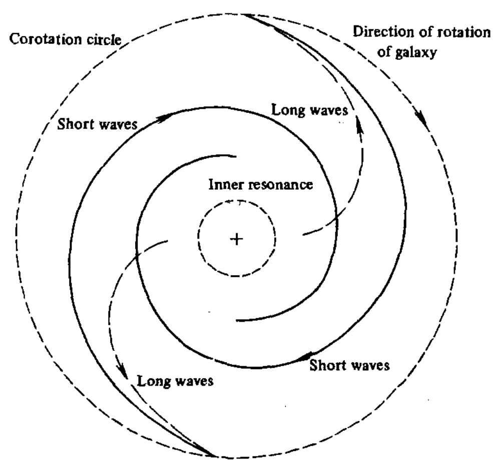

FIG. 6. Schematic representation of longwave and shortwave modes [54] for the Schmidt (1965) model of the Galaxy.

a) Distribution of neutral hydrogen. It was found using Schmidt's model of the Galaxy (Tables I and II in[48]) that the best agreement with the radiohydrogen distribution pattern[20–22] is obtained with Ω_p = 11–13 km/sec · kiloparsec. The inner Lindblad resonance point $\nu = -1$ then lies at a distance of ≈4 kiloparsecs from the center of the Galaxy. It was assumed in[48] that the spiral terminates in a ring according to the theory; this ring is associated with the "4-kiloparsec arm" in the Galaxy. But the situation here is more complex[74], and this conclusion is not obvious. On the whole, the model constructed for the spiral structure yields spiral arms whose spatial arrangement and number agree with Kerr[22]. The construction of the theoretical spiral-structure scheme is set forth in detail in[76].

b) Systematic motion of gas. The rotation curve of the Galaxy exhibits wavelike irregularities when it is constructed from the profiles of the 21-cm radio line on the assumption of purely circular motion of the gas (see, for example,[77]). Analysis of this effect shows that the waviness is only apparent, since the motion of the gas is not in fact purely circular: the gas has additional systematic motions with a velocity of the order of 10 km/sec. The systematic gas motion may owe its existence to the spiral wave. In fact, matter contracts as it flows across an arm and changes velocity in accordance with the continuity equation. To explain a velocity on the order of 10 km/sec, it is sufficient to have a spiral-wave potential amplitude of about 5% of the axisymmetric field. Here the density amplitude in the wave is about 10% of the average density (in the neighborhood of the sun), and the contributions of the gas and the stars are approximately equal. The theoretical profile of the 21-cm line and the apparent rotation curve are found to agree closely with observations if a value of 11.5 km/sec-kpc is taken for the rotational velocity of the spiral design.

The spiral-structure model[48,76,78,79] yields theoretical profiles of the 21-cm line and a theoretical "apparent" rotation curve that agree with observations[48,76,79–81].

c) Distribution of young stars. It is known that young stars of spectral classes O and B form a distinct spiral design. They are evidently born at points of maximum gas concentration, and since gas concentrates in the spiral arms, it is here that the birth of the stars should occur. Moving at the velocity of the general rotation of the matter, the newborn stars should be displaced with respect to the spiral wave design, thus migrating out of the arm with the course of time. At a velocity Ω_p = 13 km/sec-kpc, the displacement of the stars and the gaseous arm will be about 1.2 kpc after 10 000 000 years in the neighborhood of the sun, where $\Omega(R_\odot)$ = 25 km/sec-kpc. But because of the small inclination angle of the arm, the radial displacement will amount to only a tenth of this value, and would be difficult to detect. As a result, young optical objects in the Galaxy are disposed within its gaseous branches[48].

d) Migration of stars. A series of papers by Stromgren et al. developed a method for determining the ages of young stars accurate to 10–15%. Having determined the age of a star and having its proper motion from observations, the point of its birth can be established. Tables compiled in[82] make it easy to plot the orbits of young stars from the present back to their times of birth. The region occupied by the birthplaces of the stars would be expected to form a spiral arm that existed there at the time of star formation. Strömgren's first results[83,84] showed that the birthplaces of the stars fall into a spiral-arm region provided it is assumed that the arm rotates at a velocity different from the velocity of rotation of the galaxy. Yuan[85,86] calculated the pattern of stellar motion in detail backward in time. Taking account of the effects of the gravitational field of the spiral arm on the motion of a star, he showed that when this field has a value equal to 5% of the axisymmetric average Galactic field, the region of birthplace localization forms segments of spiral branches. According to[86], the present branches could have occupied these regions at the birth of stars if they rotate at a velocity Ω_p = 13 km/sec · kiloparsec[48,85].

In addition to the above four points, the wave theory explains qualitatively a number of other effects in galaxies, which offer a way to test it by quantitative comparison. These include:

e) The deviation of the vertex (i.e., the major axis) of the velocity ellipsoid.

f) The correlation between the ages and spatial arrangement of young objects.

It is known that the major axis of the velocity ellipsoid for young stars in the neighborhood of the sun deviates appreciably in direction from the line to the center of the Galaxy, contrary to the theory of the stationary axisymmetric Galaxy (see, for example, [60]). Among other possible causes, this deviation may be an effect of the field of the spiral wave on the velocity distribution[68,87-89].

The greater the ages of young stars, the farther do they recede from the point of maximal gas concentration at which they were born. Thus a star-age gradient should exist across an arm, its magnitude and sign depending on $\Omega(r) - \Omega_p$. Determining the magnitude of this gradient for the galaxy M 33, Dixon[14] obtained an estimate of $r\Omega(r) - r\Omega_p$: it was found approximately equal to -12 km/sec at a distance of 3.4 kpc. If this result is correct and if the spiral design rotates as a solid, then $\Omega_p = \Omega$ at r = 2.8 kpc and $\Omega_p < \Omega$ at $r < 2.8$ kpc in that galaxy.

The spiral structures of three galaxies—M 33, M 51, and M 81—were calculated on the basis of the density-wave theory in the recent paper[72]. The known rotation curves of M 33[84,86,90], M 51[91,92], and M 81[93] were used to construct equilibrium models for these galaxies by the method developed in[57,58,63], i.e., the mass distributions and velocity dispersions of the stars were found. In accordance with Lin's hypothesis as to the mechanism of spiral-wave generation in galaxies[94] (see also below), the velocity of rotation Ω_p of the spiral design was put equal to the velocity $\Omega(r)$ at the point r corresponding to the most remote regions of ionized-hydrogen (H II) clouds. It was found to be 16 km/sec · kiloparsec, 33 km/sec-kpc, and 21.5 km/sec-kpc for M 33, M 51, and M 81, respectively. Calculation of the spiral designs with these values of Ω_p for the galaxies constructed produced a picture that agrees with their observed spiral structure.

The velocity of rotation Ω_p of the spiral pattern can be determined directly by measuring the ages of various objects in the cross section of an arm, i.e., by determining the age gradient across the arm. Such an investigation was carried out by Courtes and Dubout-Crillon[95] for M 33. On the basis of their age calculations for objects in its branches, these authors concluded that Ω_p lies in the interval between -24 km/sec-kpc and $+25$ km/sec-kpc.

The problem of star formation in the spiral branches is an important aspect of the wave theory. In their early papers, Lin and Shu[46] advanced a hypothesis according to which stars are formed at the front of a density wave, i.e., at points of elevated density of the interstellar gas, and thus duplicate the form of the spiral arm. Quantitative calculations were made by Fujimoto[96] and Roberts[97]. Roberts considered the formation of stars in a shock wave that appears in the interstellar gas in the presence of a spiral density wave moving through the stellar component of the galactic disk. According to[98,99], a global (covering practically the entire disk) two-armed shock wave can exist in disk galaxies similar to our own and serve as a triggering mechanism for the collapse of gas clouds, which results in star formation. Here it is found that young stars in a H II region should be situated on the inner edge of a gaseous H I spiral branch, as is generally confirmed by observations (see Chap. II).

4. Difficulties of the density-wave theory. As we see from the above, the wave theory of spiral structure has produced a number of convincing arguments in its favor. However, there have been weak spots in the form in which the theory has existed up to the present time.

a) Behavior of the spiral-wave packet. The most fundamental difficulties arise in connection with a result obtained by Toomre[74]. He drew attention to the fact that the evolution of a given spiral pattern must be regarded in a realistic scheme as the evolution of a packet of spiral waves, which is determined by its group velocity c_g.

For the neighborhood of the sun, c_g is found to be ≈ 10 km/sec (κ/k_T = 65 km/sec, $\partial|\nu|/\partial|k^*|$ = 1.5). At this velocity, a packet covers a distance of 10 kpc in 10^9 years, i.e., approximately 4 galactic years (in the neighborhood of the sun). Generally speaking, c_g is a function of r, tending to zero as r tends toward the inner resonance (for a trailing spiral). Nevertheless, an exact calculation indicates that about 10^9 years are required to cover the distance from 12 kpc to 5 kpc (it will be recalled that the inner resonance in the spiral design of Lin et al. lies at about 4 kiloparsecs). During a time on the order of 4 revolutions of the Galaxy, therefore, the packet will be drawn toward the inner resonance, ultimately occupying a very small region. Accordingly, the spiral design will cease to exist as a global phenomenon occupying the entire galactic disk.

As the packet moves toward the inner resonance r_L its energy density and momentum will increase, i.e., all of the energy of the packet is concentrated toward the region in the vicinity of r_L. The density amplitude accordingly increases without limit, but, as has been shown in a recent paper by Mark[100], this is accompanied by absorption of the shortwave mode, so that the packet dissipates in accordance with Toomre's hypothesis[74].

In addition to the contraction of the wave packet toward the resonance, Toomre[74] considers it possible for the spiral wave to be distorted as a result of a slow $\omega(t)$ dependence. Contopoulos[53] also arrived at this conclusion from an analysis of the evolution of an initial perturbation.

b) The antispiral theorem. Linden-Bell and Ostriker[101] showed that spiral waves are not natural vibration modes in a rotating axisymmetric disk of gas when dissipation is left out of account. The natural modes in such a system form a design of the "wagon-wheel" type. Why this occurs is obvious: there are two conjugate spiral modes in this system for a given real value of ω, and they eventually result in a formation of a "wagon wheel."

The difficulties associated with the antispiral theorem are evidently not fundamental. They obviously result from the extremely simple symmetry of the problem, and it is sufficient to consider a more realistic system (for example, to include dissipation, magnetic fields, currents, etc.) to disturb this symmetry[50]. Finally, initial conditions introduce the necessary asymmetry[50]. It is argued in[102] that the antispiral theorem reflects the reversibility of the equations of the motion in time. Allowance for dissipation, which distinguishes the "time arrow," is cited as the factor that makes the reversal of time impossible and, consequently, violates the antispiral theorem. As we know, the "time arrow" is also distinguished by the initial conditions, so that the latter may be sufficient to circumvent the above difficulties.

Before turning to a discussion of the difficulties that arise due to the quick disappearance of the spiral-wave packet, let us take note of the following. Several authors endorse the view that the behavior of the wave near the Lindblad resonance is determined in the linear theory by the conditions of "matching" of the solutions in the WKB approximation far from the "turning points" and in the neighborhood of these "turning points." The possibility of reflection of the wave is considered here along with absorption (see, for example[102]). However, as we have already noted, the Lindblad resonance point cannot be a reflection point for the shortwave solution of interest to us, since $k_\infty = \infty$ at that point. The quantization rules are not satisfied for this reason, and, consequently, the Ω_p spectrum is continuous. In contrast to the "turning points," near which the conditions for validity of the WKB cease to apply, these conditions, are to the contrary, satisfied with increasing exactness near Lindblad resonance points, so that in this respect there is *no* need to speak of a special investigation of the solution in the neighborhood of resonances. The result obtained by Mark[100]—finiteness of k_∞ at the point $r = r_L$—does not affect this conclusion.

5. The problem of generation and persistence of the spiral design. Because of the serious theoretical difficulties encountered in the picture of the spiral structure according to Lin et al. that is under discussion here, what it requires most at the present time is clarification of the question as to the excitation and persistence of the spiral design. There are as yet only certain general considerations in respect to the generation and persistence mechanism. Let us cite some of them.

a) Gravitational fragmentation of gas. Lin[94] assumed that gravitational instability (Jeans instability) in the outer regions of a galaxy may constitute such a mechanism. His argument is as follows. According to Toomre[103], a disk is locally Jeans-unstable if the dispersion of the star velocities or the turbulent motion of the gas is below a critical level. There are few stars in the outer regions of a galaxy, and turbulent motion can be dissipated rapidly. This creates conditions for the instability considered by Toomre. Its development results in the formation of structural inhomogeneities, which, rotating at the local circular velocity, are stretched out into spiral fragments by the general differential rotation. The dynamics of such inhomogeneities was analyzed in detail in[104].

Perturbations that have appeared generate spiral waves in the disk. The velocity of rotation of the wave equals the velocity of the disk in the region of instability, i.e., this region is a region of corotation. Thus, the frequency of the wave is determined by the velocity of the galaxy in its outer regions; it is the structural inhomogeneities in rotation here that form the "generator" that determines the frequency.

More strictly speaking, it is necessary to consider the generation not of a specific mode, but of a wave packet that moves toward the inner resonance. It is not immediately obvious that wave systems generated in the outer region will produce a quasistationary structure over the entire disk. Firstly, the energy of perturbations arising at random in the outer parts of the disk is limited. Moreover, it is, generally speaking, dissipated as it propagates.

According to Lin[94], the following may offer a way out of these difficulties. Between the inner resonance (or the center of the galaxy if there is no resonance) and the corotation point (i.e., the regions in which $\Omega_p = \Omega$), he suggests the possibility of a feedback that permits pumping of energy from the resonance (center), where it accumulates, to the "generator." This is accomplished as follows. If there is no inner resonance, a wave converging toward the center has a high enough amplitude to cause distortion of the central part of the galaxy in the form of a "short bar." This configuration rotates at a velocity Ω_p corresponding to the velocity of the perturbing wave. The perturbed gravitational field associated with the "bar" propagates outward on "long waves" (short waves can propagate outward only in the case of a leading spiral). The effect of this wave will be felt most strongly at the corotation point, where the velocity of rotation Ω_p of the perturbed field coincides with the velocity of rotation Ω of the matter. According to Lin, therefore, at least partial transfer of energy occurs in a galaxy from its internal region to its external region (to the "generator"), and this makes the persistence of a relatively stationary structure possible.

If a sufficiently sharp Lindblad resonance exists, the wave cannot penetrate to the center of the system. As was shown by Contopoulos[51], the orbits of stars in the field of a spiral wave near resonance occupy a certain oval. This oval-shaped structure plays the same role as the bar in the former case, i.e., the wave is reflected through it from the resonance.

A number of serious objections can be raised to the qualitative pattern of generation and persistence of the spiral design proposed by Lin. Among other things, it would appear more probable that an instability at the edge of the galaxy would result in the formation of chaotically scattered clumps of matter that would generate a whole spectrum of waves. Superposition of these waves should therefore produce a chaotic pattern. There is doubt as to the possibility of reflection of a wave from the inner part of the galaxy. Near this region, the spiral turns into a ring ($\lambda \to 0$) and the field of the wave possesses rotational symmetry, so that bar distortion of the central region would appear to be impossible. According to[100], k_∞ is finite at $r = r_L$, but absorption rather than reflection of the wave occurs. And even if reflection does occur, a "short" leading wave should appear in addition to the "long" trailing wave, and waves of both types should then always be superposed. Without going into greater detail, we note that Lin's hypothesis appears, on the whole, to be quite artificial.

b) Close encounters with neighboring galaxies. Toomre[74,105] considered the possibility of formation of a spiral structure as a result of oscillations excited in a galaxy during a close encounter with a neighboring galaxy. In our Galaxy, for example, this process might have been associated with the passage of the Large Magellanic Cloud, which was at perigalacticon 20–25 kpc away about 5×10^8 years ago. The well-known galaxy M 51 and its companion, along with certain others, also draw attention to this possibility. A numerical experiment was designed in[106] to study the response of a model of a given galaxy in the form of a system of material points moving along closed orbits in the field of a large central body to a close passage of a massive body in the form of another galaxy. A developing spiral structure emerges clearly in the body of the first galaxy. And even though the model used was highly inadequate to the real system, the result is impressive.

Nevertheless, the enormous number of spiral galaxies without companions makes it appear improbable that this mechanism could be by any means universal.

c) Two-stream instabilities. In principle, a certain local instability of a globally Jeans-stable disk of stars could preserve a spiral wave design (ultrastability[74]). Lin and Shu[47] spoke of local instability of spiral modes arising in the second WKB approximation. However, a detailed analysis showed that this effect does not occur[74].

A local instability of the two-stream type resulting from the presence of a weakly rotating subsystem of stars in real galaxies was examined as a spiral-wave generating mechanism in[51,66,69]. The increment of this instability is found to be approximately equal to

$$\gamma(r) = \nu(r)^{-1} \cdot 2\pi^3 G^2 \sigma_1^0 \sigma_2^0 F_\nu(x) \left.\frac{\partial f_2^0(v_r)}{\partial v_r}\right|_{v_r = \omega/k(r)}. \quad (20)$$

We see from (20) that $\gamma(r)$ formally increases without limit as $\nu(r) \to 0$, so that the frequency of the wave being amplified satisfies the corotation criterion. The result is a situation in which a wave that has somehow been excited is thereafter preserved by this instability. As we have already noted, only an insignificant amount of energy is necessary in the corotation region to obtain a large wave effect. Thus the instability of (20), acting only in the corotation region, is not necessarily strong. But still the time of its development should be of the order of the lifetime of the packet, since otherwise it will will not be capable of preserving the pattern.

It was shown in[67,107] that spiral waves in the Galaxy are found to be Jeans-unstable when the spherical subsystem is taken into account. The significance of the various subsystems in the spiral-wave instability has also been discussed in[108].

d) Generation of waves by resonances. Linden-Bell made a qualitative analysis of the possibility of generation and maintainance of spiral structure by resonant processes unfolding in a corotation region in which "particle" resonance occurs (the epicenters of the "particles"—star and wave—move at the same velocity[109]). However, the detailed analysis in[110] showed that the qualitative arguments cited are incorrect. The role of the inner Lindblad resonance in the generation and maintenance of the spiral pattern was investigated in the same paper. The idea consists in the following. The equilibrium state of a disk in differential rotation is a state with an energy minimum at a given angular momentum distribution. Linden-Bell and Kalnajs[110] showed that this minimum can be lowered even further if nonaxisymmetric perturbations that transfer momentum outward exist. It is found that trailing spiral waves are the only possible perturbations of this type. As expressed by the authors, these results determine the "cause" of the existence of the spirals and their "purpose," which is to "help" the galaxy in its dynamic revolution. The actual question as to the specific mechanism of wave generation at the resonance remained open, but the results obtained are nevertheless of great importance for the entire wave conception. They enable us to view the problem from a somewhat different angle, namely from the standpoint of the general dynamic evolution of the galaxy. Specifically, though, the location of the wave generator is an extremely important question in the wave theory[75]. Indications that it should be at the inner resonance correlate very well with the results of[75] (see also Sec. 6 below), from which it follows that the spiral structure of our Galaxy must be generated in its central region.

e) Gradient instabilities. Systems with velocity-dispersion and density gradients along their radii may be unstable. The hypothesis has been advanced that the instabilities may be manifested in the formation of a spiral structure (see, for example, [111]).

Another variety of gradient instability was considered in[112]. In a differentially rotating self-gravitating infinitesimally thin disk consisting of magnetized interstellar gas and stars, spiral density waves are unstable with an increment proportional to $H_r H_\theta r d/\Omega/dr$. This instability is related to the presence of the magnetic field and the differential nature of the rotation.

f) Shear instability. Harrison[113a] sees a possibility of spiral-structure generation in the instability of oblique waves in a system with intersecting flows. He considered a simple model: plane oblique waves in a uniformly rotating system with radial flow. The wave amplitude increases in such a system. However, the excessive simplification of the situation examined here excludes direct application of the results to galaxies.

We offer the following general remarks concerning local instabilities.

In most cases, they cannot act as generators of waves that form a regular structure. In fact, at each given distance from the center of the galaxy, an instability excites a wave with its "own" local frequency, equal to the angular velocity of the rotation at that distance. In sum, these waves will produce a pattern with virtually no resemblance to a spiral design. At best, it will give rise to turbulence. But there may also be a situation similar to local (in velocity space) instability of a Maxwellian velocity distribution in a collisionless plasma. As we know, this distribution can, in the collisionless case, be represented as a system of mutually penetrating streams. Generally speaking, the streams are unstable, but the distribution on the whole is stable[113b].

To bring out a single frequency, the instability must develop only at some specific distance from the center. For Lin, for example, this occurs in the outer regions of the galaxy, where gravitational condensation is possible.

To summarize the above, it may be concluded that the fundamental question of the theory of spiral galactic structure remains open within the framework of the theory of Lin et al.: the question as to the generation and persistence of the spiral design. The situation will perhaps become more definite as a result of study of the behavior of a wave at the singular points of the galaxy: at the center, at the margin, and at the Lindblad resonance points. The phenomena that unfold here may be essentially nonlinear (see Sec. 7).

It appears more likely that the solution of the problem as a whole will be found in the scheme set forth below, which is, in a certain sense, an alternative to the scheme of Lin et al.

6. The k_0 mode and the spiral design of the Galaxy. The difficulties that arise in the scheme expounded by Lin et al. are absent in an alternative scheme of the wave theory that was proposed in[75]. This paper was based on the idea that instead of the total surface density $\sigma(r)$ of the Galaxy, only the surface density $\sigma_1(r)$ of the plane subsystem should be substituted into the dispersion equation, since the parameter ζ (19) is very small in our Galaxy, on the order of 10^{-2} to 10^{-3}.

Substituting σ_1 for $\sigma(r)$ changes the situation radically. The density of the plane subsystem varies weakly with distance from the Galactic center. It was therefore assumed in[75] that σ_1 = const. Then the mode k_0 rather than k_∞ is found to agree with the observed picture. Figures 3a and 3b present a comparison of the theoretically calculated spiral design with the neutral hydrogen distributions according to Weaver[24] and Kerr[22] for k_0. Very good agreement with observations is obtained with Ω_p = 23 km/sec-kpc and $\sigma_1 = 40 M_\odot/\text{nc}^2$. The inner and outer Lindblad resonances then take positions at approximately 2 kpc and 14 kpc, respectively. It is interesting that in the Weaver pattern, a ring of neutral hydrogen is situated at the distance of 2 kiloparsecs from the center and may therefore be associated with the presence of the inner resonance in this region.

According to the Schmidt model, the total density $\sigma = 40 M_\odot/\text{nc}^2$ corresponds to r = 14–15 kpc. It would seem reasonable to assume that the total density is practically the same as the density of the plane subsystem at this distance from the center of the Galaxy. Thus, the values found for Ω_p and σ_1 appear natural.

The group velocity of a packet of trailing waves of k_0 modes is positive, i.e., directed outward from the center of the Galaxy, unlike the case of k_∞. If, therefore, some constantly running "generator" exists in the central region, it could "chase" spiral waves across the galactic disk, and the problem of the "persistence" of the spiral design would not exist. A slowly rotating (by comparison with the disk) bar of old stars, such as is often seen in the central regions of galaxies, was proposed in[75] as such a generator. Obviously, the angular phase velocity Ω_p of the spiral waves must coincide with the angular velocity of the bar, which is thus found to be on the order of 23 km/sec-kpc for our Galaxy. It follows from the rotation curve that Ω = 93 km/sec-kpc at a distance of 2 kpc from the center, so that $\Omega/\Omega_p = 4$, i.e., it is of the order of the angular-velocity ratio of the plane and spherical subsystems. This is an argument in favor of the hypothesis of the bar of old stars. Otherwise, it appears that waves could also be generated by active processes in the galactic nuclei. This problem requires special treatment.

We note several important points in connection with the concept set forth above:

a) Disturbance of the regular spiral structure is observed at distances on the order of 14–15 kpc from the center (i.e., on the rim of the galactic disk) in Kerr and Weaver's pattern. Within the framework of the scheme considered here, this may be a consequence of a packet of k_0 modes behaving on its approach to the outer resonance ($\nu \to 1$) in the same way as a packet of k_∞ modes approaching the inner resonance (see Fig. 6). The amplitude and energy increase and nonlinear effects come into play and could probably result in turbulization and disintegration of the structure.

b) If the wave generator is in fact a rotating bar, it would then evidently be possible to understand, at least qualitatively, the origin of a "spectrum" of different transitional forms among the spiral galaxies, for example the basic type in de Vaucouleurs' classification in Fig. 2 (for greater detail see[75]).

There is a difficulty in this conception, arising out of the fact that agreement with the observed patterns of Kerr[22] and Weaver[24] is obtained at Ω_p = 23 km/sec-kpc, while Stromgren's estimate of the angular phase velocity of the waves[83] gives Ω_p = 20 km/sec-kpc, and it is even smaller according to Yuan[85] (see Sec. 3). Improvement of the numerical data may be necessary here.

7. Resonance effects of the spiral galaxies.

a) Lindblad resonance. Certain features connected with the behavior of a wave near a Lindblad resonance were examined by Contopoulos[114].

Two types of stable resonant periodic orbits were found near resonance. In response to the growing field of the spiral wave, increasing numbers of stars transfer from originally epicyclic orbits to new resonant periodic orbits. The result is a redistribution of density from the original axisymmetric distribution to one that has nearly quadrupolar symmetry. In Contopoulos' opinion, transfer of stars to resonant orbits can result in nonlinear limitation of the wave-amplitude increase near resonance. In[115], the response of the system near the inner Lindblad resonance to a given increasing spiral field was studied in the linear approximation. It was found that:

1) In all cases, the response in angle has a positive phase shift, i.e., a density maximum precedes a potential minimum in the direction of rotation.

2) This shift is small in the outer parts of the galaxy, and, generally speaking, increases inward. Near resonance, the shift reaches 90° in the case of trailing waves and rapidly growing leading waves, or 270° for slowly growing leading waves. The angle phase shift covers a larger region the larger the star-velocity dispersion.

3) The response to a trailing growing potential is a trailing density wave. The response to a weakly growing leading potential is a strongly trailing wave. And only the response to a rapidly growing leading potential yields a leading wave.

These results may indicate that the properties of a system near a Lindblad resonance identify trailing spiral density waves as the only kind possible in any situation, if we exclude the improbable case of a rapidly growing potential.

b) "Particle" resonance. Barbanis[116] considered resonant effects near a particle resonance r_p, i.e., in the corotation region. Under the action of a spiral field, the stars transfer here to orbits of a new type such that two density maxima appear at two diametrically opposite points of the disk. This results in disintegration of the spiral branches in the region of the corotational distance. Thus, nonlinear effects also make it impossible for a self-consistent spiral wave to exist near the region of corotation.

8. Nonlinear density waves. Berry and Vandervoort found solutions in the WKB approximation that describe stationary nonlinear density waves[70,117]. The system model was a differentially rotating disk described by hydrodynamic equations. In addition to its own gravitational field, the disk is also in an external gravitational field. This model corresponds to the presence of two subsystems in galaxies: subsystem I, which consists of gas and stars with a small velocity dispersion, and subsystem II, which is formed by stars with high peculiar velocities. Such a model was studied in[66-69] in the linear approximation.

The existence conditions of nonlinear stationary spiral waves require that a galaxy consist of distinct subsystems I and II. Here it is necessary that the ratio of the density of subsystem I to the total density of the system be a small quantity on the order of the small parameter in the WKB approximation. Hence it follows at once that the larger the relative density of subsystem II, the more tightly will the spirals be wound. Incidentally, the same conclusion also follows from the linear theory: the characteristic radial distance between branches equals $\lambda_c = 4\pi G\sigma^*/\omega^2$, where σ^* is the effective density of the subsystems participating in the wave[48].

For small amplitudes, the solutions of Berry and Vandervoort yield results that agree with those of[48].

Since the superposition principle is not satisfied for nonlinear waves, the antispiral theorem is not valid for them. But the results do not imply that one of the types of spiral structure is to be preferred.

The role of resonances for nonlinear stationary waves has not been established, and the ultimate fate of these waves, like that of the linear waves, is therefore unclear.

9. The nonasymptotic approximation in the wave theory of the spiral structure. If we do not restrict ourselves to the WKB approximation, the behavior of small perturbations in a disk of stars is described by the equation (see, for example,[102])

$$r\sigma(r) = \int_0^\infty K_{m\omega}(r, a)\,\sigma(a)\,da, \qquad (21)$$

where

$$K_{m\omega} = -4\pi G m^* \iint_{\Pi_0>0} \frac{dE_0\,dJ}{\Pi_0(r, E_0, J)} \left\{ \frac{\partial f_0}{\partial E_0} H_m(r, a) - \frac{r(\partial f_0/\partial E_0) + m(\partial f_0/\partial J)}{2\sin(\omega\tau_{12} - m\theta_{12})} \int_{-\tau_{12}}^{\tau_{12}} H_m(r^*(\tau), a)\cos[\omega\tau - m\theta^*(\tau)]\,d\tau \right\}, \qquad (22)$$

E_0 and J are the energy and momentum integrals, $f_0(E_0, J)$ is the unperturbed distribution function, m^* is the mass of a star, τ_{12} is the time of motion of a star between the farthest and nearest points on the perturbed orbit, θ_{12} is the azimuth angle corresponding to passage of the star between these points, and

$$\Pi_0(r, E_0, J) = \{2[E_0 - \varphi_0(r)] - (J^2/r^2)\}^{1/2}. \qquad (23)$$

The linear homogeneous integral equation (21) with kernel (22) gives a complete statement of the problem of the normal modes of a disk of stars. The eigenfunctions of this equation, which are to be found, are $r\sigma(r)$, and the ω are its eigenvalues.

Kalnajs[118] investigated spiral modes of arbitrary tightness in a galaxy in the epicyclic-orbit approximation by numerical integration of Eq. (21). The specific galactic model considered was M 31. The value obtained for the velocity of the spiral design was Ω_p = 30 km/sec-kpc, i.e., the rotation is very rapid compared to[72]. The wave proves to be strongly unstable: its amplitude increases by a factor of e in 10^9 years. The spiral tightness of the density wave is found to be different for subsystems with different velocity dispersions. A bar-type design forms in the central region of the galaxy. But the response of the gas to this design is a tightly wound wave with high density contrast, an effect of the presence of resonances in the system.

Thus, according to Kalnajs, the density wave in galaxies of the type of our own and M 31 is an open spiral, but the response to this wave from the subsystem of bright stars and gas, which has low density and small dispersion, produces a tightly wound wave. It is the latter that we detect as the spiral structure.

Because of its greater mathematical difficulty, Kalnajs' theory has not been developed as far as the theory of "tightly wound" spirals, and it is difficult to judge its adequacy to the real situation at the present time.

IV. NUMERICAL EXPERIMENTS

Numerical experiments yield highly interesting information on the evolution of the galaxy and the development of structure in it. A numerical solution of the problem of N bodies interacting in accordance with Newton's law was carried out in[119-122]. As many as N = 200 000 interacting points could be handled at moderate computing accuracy. The basic purpose of these calculations was to trace the evolution of plane disk-shaped (two-dimensional) galaxies. The computing method did not include close-range interactions of the points (collisions), since the models analyzed simulated collisionless star systems.

Miller and Prendergast[119] showed that a spiral structure forms quite quickly in an originally axisymmetric rotating disk consisting of gravitating points. But a serious difficulty was brought to light also in this paper: the system "heated up" very quickly, and the peculiar velocities were found to be equal to the rotational velocity. Ultimately, the system is held at equilibrium not by the rotation, but by the "pressure" of the gas of stars. Spiral structure is not observed in galaxies with such large velocity dispersions.

To eliminate the effects that stem from rapid heating, Miller, Prendergast, and Quirk[121] introduced two subsystems. The first was continuously artificially cooled (the velocity dispersion decreased while energy and momentum remained constant) and was called the "gas." The second subsystem, the "stars," was left to itself and therefore heated up rapidly. At time zero the entire system consisted of "gas." Then "star formation" was invoked, with the "coefficient of star formation," which defined at each step of the integration the number of points turning from "gas" into "stars," having a strong influence on the early evolution of the system. The problem is obviously a very good illustration of the evolution of real galaxies. It enables us to ascertain the relative importance of the galaxy's subsystems in the formation of its spiral structure, since all details of the system—rotation, average velocity, epicyclic frequencies, etc.—are computed in accordance with a standard scheme.

The computed evolution of such a system results in a spiral structure having much in common with that which we observed. In particular, the spirals are sharply defined by the "gas," while the "stars" evidence only insignificant structure.

It is remarkable that the "numerical" spiral structure is consistent with the conceptions of the wave theory. Thus, the individual points move across the spiral design, i.e., the latter is a wave; the design rotates more slowly than the "gas;" the structure extends from the inner Lindblad resonance to the corotation point, where the system ends.

Special experiments showed that the "stars" determine the basic (background) gravitational field in which the "gas," which has an asymmetric potential determined by spirality, moves. But the "gas" itself does not play a significant role in preservation of the spiral design. Self-gravitation is essential in the persistence of spiral structure. Thus, two important facts emerge:

1) a well-defined spiral structure can and should exist in galaxies having a subsystem with a small velocity dispersion;

2) preservation of the spiral structure is due to a subsystem of stars with large dispersion.

Vandervoort stressed the striking similarity between the conditions of the formation and existence of spiral structure in[121] and the existence conditions of nonlinear stationary spiral density waves[58].

Miller and Quirk discussed experiments in which it was found that spirals have a bar-type structure near their centers, but a short central bar apparently does not result in spirals. On the other hand, a general barlike distortion of a galaxy gives rise to a spiral wave[53].

Another group of experiments was performed by Hohl and Hockny[120, 122]. A cold disk rotating as a solid and having a density $\sigma(r) = \sigma(0)\ (1 - r^2/R_0^2)^{1/2}$, where R_0 is the boundary of the disk, was considered in[122]. The numerical experiment fully confirmed Toomre's results[65]. That is to say, it was found that such a disk is gravitationally unstable and, as a result, breaks up into three to five smaller systems during one revolution. A velocity dispersion equal to 27% of the velocity at the rim of the disk stabilizes it against breakup into local condensations.

But the disk remains unstable with respect to shape distortions, i.e., it assumes the form of a bar after approximately two revolutions. The same result was obtained for a disk with a Gaussian density distribution. It was found that only a high concentration of mass toward the center, of the exponential type, can suppress the instability with respect to shape distortion. Thus, yet another important condition for the possible existence of spiral structure in a disk galaxy was found: the presence of a strong mass concentration toward the center.

Schmidt's model of our Galaxy satisfies this condition. An experiment in[122] showed that a well-defined spiral structure is developed in such a system (Fig. 7). At a disk mass equal to 10% of the total system mass and an initial radius corresponding to 15 kiloparsecs, the structure persists for 8.5 rotations of the galaxy if the period of rotation at a distance of 10 kpc is taken as the time unit. The computation terminates at time 8.5. But purely kinematic considerations indicate that the spiral arms would be twisted by differential rotation within a time shorter than 8.5 rotations. This indicates that the spiral design obtained is of wave nature.

Rapid "heating" of the system takes place in[120, 122] as in[119, 121]. The cause of this effect is still unclear; it has been suggested that it is due not to real physical causes, but to the computing method[53]. This calls for caution in comparing the results with observations.

V. CONCLUSION. RECENT DATA. PROS AND CONS OF THE WAVE THEORY. ALTERNATIVE HYPOTHESES

The question as to how well the wave theory is confirmed by observational data and in particular by the most recent radioastronomical data, has recently become a topic of lively discussion (see below). The most important observed effects are associated with the interstellar gas—with the passage of the spiral density wave across it (see Sec. 3 of Chap. III). This led to a series of studies (beginning with one by Roberts[97]) of the response of the gas to a spiral gravitational-potential wave and of the corresponding physical processes in the gas (see, for example,[123, 124]). New effects were predicted theoretically: narrow regions of compressed gas at the inner edges of the spiral branches, age gradients of young stars and other objects across the arm, etc., effects that have been confirmed (in any event qualitatively) by a whole series of observations. Furthermore, new light has been cast on the old "classical" problems of the physics of the interstellar gas and stellar cosmogony. The problem of the origin of the cloud structure of the interstellar medium has now been under discussion for more than 20 years and remains unresolved. Greenberg[125] suggested that the clouds are produced in a compression wave in the gas that is induced by the spiral wave. This phenomenon was analyzed in detail in[126, 127], whose results suggest that such a process—a phase transition in interstellar gas compressed by a shock wave—is a real possibility. Incidentally, recent studies[128] support the view according to which the clouds have short lifetimes (~ 10^7 years) and hence must be generated with some intensity[129]. Perhaps the only sufficiently convincing scheme of their regeneration now available is given by the wave theory. The spiral-wave notion also opened new possibilities for solution of fundamental problem of astrophysics—that of the origin of the stars. Here, together with the triggering mechanism[97], the possibility of production of stars in phase transitions caused by the spiral wave in the interstellar medium is discussed[126, 129].

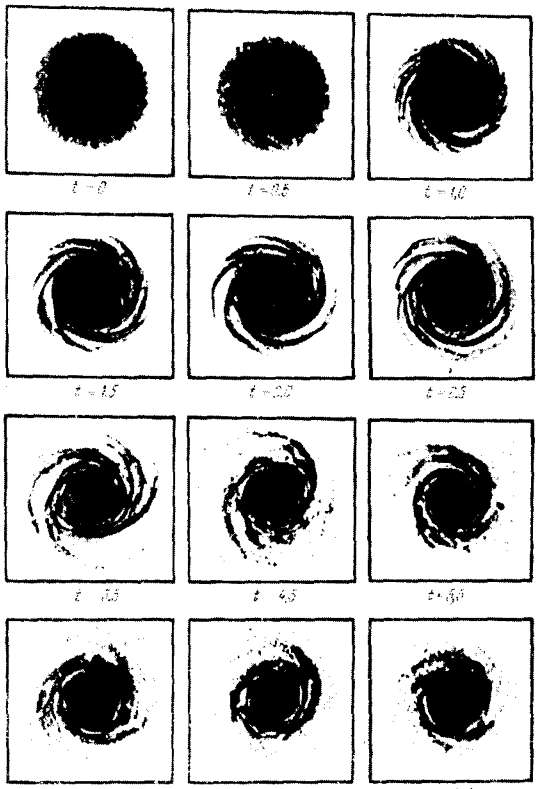

FIG. 7. Results of Hohl's numerical experiment[122]: evolution of an initially cold disk of 200 000 stars. Time is given in units of the period at a distance of 10 kpc; the disk radius is 15 kiloparsecs, and the disk is in equilibrium at t = 0.

From a program in which theory was compared with observations, the Lin-Shu-Roberts group arrived at the conclusion that all predictions of the theory are in good agreement with the observational data. The paper[90], which demonstrated the existence of a shock wave at the inner edges of the spirals and an associated enhancement of synchrotron radiation (by a factor of 100) in the galaxy M 51, is a strong argument in favor of the wave theory. A similar result was obtained for M 101 in[130]. Arguments have also been advanced for the existence of shock waves in barred spirals[131] and have been confirmed by various observations[131-133].

At the same time, Piddington[133,134] (see also[135]) recently came out with some serious criticism of the wave theory. Reevaluating the observational data that are regarded as confirmation of the wave theory, he arrived at the conclusion that they not only do not agree with it, but as much as refute it. We note at once that Piddington's most serious objections actually pertain not to the wave theory as a whole, but only to the specific theoretical models considered by the Lin-Shu group for the spiral structures of our Galaxy, M 51, M 33, M 101, and M 31. These objections are as follows: 1) In our Galaxy, the shock wave should amplify the synchrotron radiation in a unit volume by a factor of ~80. On the other hand, the observations of[136] give an intensity change by only a factor of three at a distance of several kiloparsecs. 2) The same applies to M 101[137]. For M 51, the synchrotron radiation should be amplified by a factor of ~270 instead of the observed hundredfold enhancement. 3) A global compression wave in the gas, and the amplification of the magnetic field in it, should produce a change in the cosmic-ray flux by a factor of about 8 to 10 in our Galaxy over 10^8 years. In fact, the flux changes by no more than a factor of two[138]. 4) The width of the gas-compression region in M 31 is 1/4 of the interarm distance[9], i.e., it is not a narrow band. (We note Weaver's observation[139] that, as a characteristic observed feature of many spiral galaxies, the dark dusty material concentrates at the inner edges of the spiral branches; this is demonstrated for our Galaxy in[140]. Dust clouds are usually closely associated with gas clouds.) 5) Separation of objects by ages is observed in the corotation theory, but should not be according to the theory.

Despite the fact that Piddington considers the above objections to be arguments against "wave conceptions" in general, it is our opinion that they can be treated only as objections to the specific theoretical spiral-structure models that were computed by Lin-Shu-Roberts and their colleagues. As was pointed out in[75], a more adequate theoretical model of the Galaxy's spiral structure results in a smaller difference between the rotational velocities of the wave and the gas. In this case, the theory predicts a lesser degree of gas compression and a wider gaseous arm[124]. The amplification of the synchrotron radiation will therefore prove to be smaller than in the Lin-Shu-Roberts models, and the region of gas compression will be wider. The corotation region will then occur at a different distance from the center of the Galaxy. We may therefore hope that these models will enable us to find an explanation for the facts cited by Piddington. Incidentally, Wester observed (see[141]) the HI concentration maximum in our Galaxy (and that in M 31) at a distance of 14–15 kpc from its center. And it is precisely at this distance that the model of[75,142] predicts a similar effect due to the presence of the outer Lindblad resonance in this region. The same applies for M 31[142]. Thus, we have yet another piece of evidence in favor of the indicated model[71,142].

Piddington sees a serious argument against the wave theory in the gas flow at velocities of 20–30 km/sec toward the plane of the Galaxy that has been observed at high and middle latitudes (see, for example,[143]). This is because the wave field cannot produce motions at velocities larger than ~ 2 km/sec. At the same time, Piddington believes that this phenomenon is naturally explained by his magnetohydrodynamic theory of the spiral structure[133,144].

However, it was shown in[145] that the fall of H I clouds toward the plane can occur as a result of convective-thermal instability developed as the spiral wave passes through the gas. The fall velocity of the clouds is determined by the total gravitational attraction along the z coordinate and reaches precisely the observed values of ~ 20–30 km/sec. Piddington states further that gas and dust do not concentrate at all in regular spiral branches, but form a chaotic or annular structure in all of the observed cases. This objection is disputed by Bok[141] among others.

The situation with the So galaxies, which resemble the S galaxies in all of their properties, but in which no spiral structure is observed (this can, incidentally, be explained as due to a whole series of causes), is intriguing. This fact is regarded in[134] as an argument against the wave theory. A remark of Oort[146] is interesting in this connection: a stepped rather than a smooth decrease in brightness from the center toward the rim of the disk is observed in certain So galaxies. According to Oort[146], this is most probably a consequence of the existence of spiral branches. Finally, multiple arms (NGC 2841, NGC 488, and others), branching of arms (NGC 1232 and others), and secondary features of the spiral structure should not be regarded as evidence against the wave theory[134]. The two-armed models constructed by Lin's group for specific galaxies do not by any means exhaust all of the variants that the wave theory can produce. Nor does this theory exclude multi-armed configurations or secondary features (see, for example,[124]) etc.

The enormous variety of structural features encountered among the various galaxies continues to stimulate the flow of new hypotheses and theories. Arp[4,147] (see also[148]) cites various observational data that can hardly be reconciled with the wave theory. He considers the spiral structure of a whole series of galaxies to be beyond doubt of explosive (eruptive) origin. Discussion of this possibility has flourished strongly since the paper of Van der Kruit, Oort, and Mathewson[9C] made its appearance. These authors constructed a radio map of the galaxy NGC 4258. It was found that this galaxy has two strong gaseous spiral branches that are shifted almost 90° from the two main spirals, which are visible in the optical band. The gas in the arms is moving outward at considerable velocity. This indicates that it was ejected from the nucleus of the galaxy comparatively recently. Accordingly, a hypothesis was advanced in[9C] under which the mechanism of spiral-structure formation consists of periodic (with a period of 500 million years) eruptions of gas with a mass on the order of $10^8 M_\odot$ from the Galactic nucleus, with each eruption renewing the spiral structure. This hypothesis also has its difficulties, which have been discussed of late in[9C,148,149]. In NGC 4258, for example, the mass of the eruptive spiral branches is one one-hundredth of the mass of the optical branches. Thus the former could hardly serve here even as a mechanism for generation of a wave spiral structure comparable in strength to the observed optical structure.

Along with the eruptive hypotheses of the origin of the spiral branches, "magnetic" hypotheses have continued to appear during the last few years (see, for example, the aforementioned work of Piddington[133,134], the paper by Pismis[150], and others).

The wave theory has offered a simple and orderly explanation of the nature of galactic spiral structure and has solved a fundamental problem with which astronomers were faced for decades: how can a global spiral structure persist over a long time in systems with differential rotation? This theory has been accepted by the overwhelming majority of astronomers and has made them confident that they will eventually understand why the galaxies "have a right" to spiral structure. It remains only to find out why they "should" have it.

The authors thank S. B. Pikel'ner and L. M. Ozernoĭ, who read the manuscript and offered a number of valuable comments.

[1)]We are not concerned in this paper with the problem of barred spirals (Sb), which requires a separate review of its own (see, for example, [37-43]).

[2)]Marginal stability means that the dispersion of the star velocities in each region of the disk is equal to a certain minimum value at which the disk is locally stable with respect to gravitational condensation.

[3)]There have been many papers in which cylinders were used as models of the galaxy. They are not discussed here.

[4)]Studies of this type are not dealt with in the present review. The corresponding references can be found in the reviews [51,52].

1. a) J. Jeans, Astronomy and Cosmogony, Cambridge, Cambridge Univ. Press, 1928, p. 352; b) V. L. Ginzburg, Usp. Fiz. Nauk **103**, 3 (1971) [Sov. Phys.-Uspekhi 14, 1 (1971)].
2. W. Heisenberg and C. F. Weizsäcker, Zs. Phys. **125**, 290 (1948); S. Chandrasekhar and E. Fermi, Astrophys. J. **118**, 113 (1953).
3. B. A. Vorontsov-Vel'yaminov, Astron. Zh. **41**, 814 (1964) [Sov. Astron.-AJ 8, 649 (1965)].
4. H. Arp, Sky and Telescope **38**, 385 (1969).
5. E. Hubble, The Realm of the Nebulae, New Haven, Yale Univ. Press, 1936.
6. G. de Vaucouleurs, in: The Structure of Stellar Systems (Russ. transl.), IL, 1962, p. 351.
7. S. Van den Berg, IAU Symposium No. 38, Dordrecht, D. Reidel, 1970, p. 478.
8. B. A. Vorontsov-Vel'yaminov, Astron. Tsirkulyar No. 457 (1968).
9. a) G. G. Pooley, Mon. Not. RAS **144**, 101 (1969); b) D. S. Mathewson, P. S. Van der Kruit, and W. N. Brown, Astron. and Astrophys. **17**, 468 (1972); c) P. S. Van der Kruit, J. H. Oort, and D. S. Mathewson, ibid. **21**, 169.
10. F. Zwicky, Morphological Astronomy, B., Springer-Verlag, 1957.
11. W. W. Morgan, see[7], p. 9.
12. P. J. Mills, ibid., p. 178.
13. Yu. N. Efremov, Astron. Tsirkulyar No. 639 (1971).
14. M. E. Dixon, Astrophys. J. **164**, 411 (1971).
15. L. F. Smith, Mon. Not. RAS **141**, 317 (1968).
16. W. Becker, Zs. Astrophys. **57**, 117 (1963).
17. B. Markaryan, Soobshch. Byurakanskoĭ Obs. **26**, 3 (1959).
18. É. E. Khachikyan and K. L. Saakyan, see[7], p. 87.
19. G. Westerhout, Bull. Astron. Inst. Neth. **14**, 215 (1958).
20. J. H. Oort, F. J. Kerr and G. Westerhout, Mon. Not. RAS **118**, 319 (1958).
21. F. J. Kerr and G. Westerhout, Stars and Stellar Systems, v. 5, Chicago—London, Chicago Univ. Press, 1965, p. 167.
22. F. J. Kerr, Austr. J. Phys., Astrophys. Suppl. No. 9 (1969).
23. S. C. Simonson, Astron. and Astrophys. **9**, 163 (1970).
24. H. Weaver, see[7], p. 126
25. R. D. Davies, Nature **237**, 88 (1972).
26. a) G. Courtès, J. Georgelin, J. P. Monnet, J. M. Monnet and A. Pourcelot, Interstellar Ionized Hydrogen, N.Y., Benjamin, 1968; b) G. Courtès, J. Georgelin, J. P. Monnet and J. M. Monnet, see[7], p. 209.
27. B. J. Clark, Astrophys. J. **142**, 1398 (1965).
28. U. Mebold, Beitr. Radioastron. **1**, 97 (1969).
29. G. Westerhout, see[7], p. 124.
30. P. C. Van der Kruit, Astron. and Astrophys. **13**, 405 (1971).
31. K. Rudnicki, see[7], p. 435.
32. M. Kepner, Astron. and Astrophys. **5**, 444 (1970).
33. F. J. Kerr, see[7], p. 95.
34. B. Bok, American Scientist **55**, 375 (1967).
35. B. T. Lynds, Sky and Telescope **33**, 343; **34**, 18 (1967).
36. T. L. Page, ibid., p. 372.

[37]S. B. Pikel'ner, Ann. Rev. Astron. and Astrophys. 6, 165 (1968).
[38]J. H. Oort, see[7], p. 1.
[39]J. H. Oort, ibid., p. 474.
[40]B. J. Bok, ibid., p. 457.
[41]B. J. Bok, Sky and Telescope 39, 21 (1970).
[42]J. H. Oort, Interstellar Matter in Galaxies, Ed. L. Woltjer, N.Y., Benjamin, 1962, p. 279.
[43]K. C. Freeman, see[7], p. 351.
[44]B. Lindblad, Stokholm Obs. Ann. 13, No. 10 (1941); 14, No. 1 (1942).
[45]B. Lindblad, ibid. 15, No. 4 (1948); 16, No. 1 (1950); 22, No. 5 (1963).
[46]C. C. Lin and F. Shu, Astrophys. J. 140, 646 (1964).
[47]C. C. Lin and F. Shu, Proc. Nat. Ac. Sci. US, 55, 229 (1966).
[48]C. C. Lin, C. Yuan and F. Shu, Astrophys. J. 155, 721 (1969).
[49]C. C. Lin, Ann. Rev. Astron. and Astrophys. 5, 453 (1971).
[50]K. H. Prendergast, see[7], p. 303.
[51]G. Contopoulos, see[7], p. 303.
[52]G. Contopoulos, Structure and Dynamics of the Galactic System, Univ. Thessaloniki, 1970.
[53]G. Contopoulos, Earth and Extraterr. Sci. 1, 185 (1971).
[54]R. Wielen, Astron. Rechen-Inst. Heidelberg Mitt., Ser. A, Nr. 47 (1971).
[55]J. H. Oort, in: "Stellar Population", Vatican Obs., 1958, p. 415.
[56]S. A. Kaplan and S. B. Pikel'ner, Mezhzvezdnaya Sreda [The Interstellar Medium], Fizmatgiz, 1963.
[57]P. Vandervoort, Astrophys. J. 147, 91 (1967).
[58]P. Vandervoort, ibid. 161, 67 (1970).
[59]M. Schmidt, in:[21], p. 513.
[60]K. F. Ogorodinkov, Dinamika Zvezdnykh Sistem [Dynamics of Stellar Systems], Fizmatgiz, 1963.
[61]R. B. Shatsova, Plankovskoe Raspredelenie Zvezd v Okrestnostyakh Solntsa [Planckian Distribution of Stars in the Neighborhood of the Sun], Rostov State University Press, 1965.
[62]D. Linden-Bell, Mon. Not. RAS 136, 101 (1967).
[63]F. Shu, Astrophy. J. 158, 505 (1969).
[64]L. S. Marochnik, Astrofizika 5, 487 (1969); Astron. Zh. 47, 46 (1970) [Sov. Astron.-AJ 14, 36 (1970)].
[65]A. Toomre, Astrophys. J. 139, 1217 (1964).
[66]L. S. Marochnik and A. A. Suchkov, Astron. Zh. 46, 319 (1969) [Sov. Astron.-AJ 13, 252 (1969)].
[67]L. S. Marochnik and A. A. Suchkov, ibid., p. 524 [411].
[68]L. S. Marochnik and A. A. Suchkov, Astrophys. and Space Sci. 4, 319 (1969).
[69]L. S. Marochnik and A. A. Suchkov, Byull. In-ta Astrofiz. Akad. Nauk TadzhSSR, No. 58, 3 (1971).
[70]P. Vandervoort, Astrophys. J. 166, 57 (1971).
[71]C. C. Lin, J. SIAM Appl. Math. 14, 876 (1966).
[72]F. Shu, R. Stachnik and J. Yost, Astrophys. J. 166, 465 (1971).
[73]F. Shu, ibid. 160, 99 (1970).
[74]A. Toomre, ibid. 158, 899 (1969).
[75]L. S. Marochnik, Yu. N. Michurov and A. A. Suchkov, Astrophys. and Space Sci. 17, 1456 (1972).
[76]C. Yuan, Astrophys. J. 158, 871 (1969).
[77]W. W. Shane and G. P. Bieger-Smith, Bull. Astron. Inst. Neth. 18, 263 (1966).
[78]C. C. Lin and F. Shu, IAU Symposium No. 31, Dordrecht, D. Reidel, 1967, p. 313.
[79]C. Yuan, see[7], p. 391.
[80]W. B. Burton and W. W. Shane, see[7], p. 397.
[81]W. B. Burton, Astron. and Astrophys. 10, 76 (1971).
[82]G. Contopoulos and B. Strömgren, Tables of Plane Galactic Orbits, N.Y., Inst. Space Studies, 1965.
[83]P. Strömgren, see[78], p. 323.
[84]J. C. Brandt, Astrophys. J. 131, 293 (1965).
[85]C. Yuan, ibid. 158, 889 (1969).
[86]G. Carranza, G. Courtès, Y. Georgelin, and G. Monnet, C. R. Ac. Sci. B264, 191 (1967).
[87]M. Mayor, see[7], p. 433.
[88]M. Mayor, Astron. and Astrophys. 6, 60 (1970).
[89]C. Yuan, Bull. Am. Astron. Soc. 2, 358 (1970).
[90]L. Volders, Bull. Astron. Inst. Neth. 14, 323 (1959).
[91]E. M. Burbidge and G. R. Burbidge, Astrophys. J. 140, 1445 (1964).
[92]G. Carranza, R. Grillon and G. Monnet, Astron. and Astrophys. 1, 479 (1969).
[93]G. Münch, Publ. Astron. Soc. Pac. 71, 101 (1959).
[94]C. C. Lin, see[7], p. 377.
[95]G. Courtes and R. Dubout-Grillon, Astron. and Astrophys. 11, 468 (1971).
[96]M. Fujimoto, IAU, Symposium No. 29, Dordrecht, D. Reidel, 1966, p. 453.
[97]W. W. Roberts, Astrophys. J. 158, 123 (1969).
[98]W. W. Roberts, see[7], p. 415.
[99]W. W. Roberts and C. Yuan, Astrophys. J. 161, 887 (1970).
[100]J. Mark, Proc. Nat. Ac. Sci. US 68, 2095 (1971).
[101]D. Lynden-Bell and J. D. Ostriker, Mon. Not. RAS 136, 273 (1967).
[102]F. Shu, Astrophys. J. 160, 89 (1970).
[103]A. Toomre, ibid. 138, 385 (1963).
[104]P. Goldreich and D. Linden-Bell, Mon. Not. RAS 130, 97, 125 (1965); W. H. Julian and A. Toomre, Astrophys. J. 146, 810 (1966).
[105]A. Toomre, see[7], p. 334.
[106]T. M. Eneev, R. A. Sunyaev and N. N. Koslov, Astron. and Astrophys. 22, 41 (1973).
[107]A. A. Suchov, Astron. Zh. 47, 1187 (1970) [Sov. Astron.-AJ 14, 947 (1971)].
[108]V. L. Polyachenko, V. S. Synakh and A. M. Fridman, ibid. 48, 1174 (1971) [15, 934 (1972)].
[109]D. Linden-Bell, see[7], p. 331.
[110]D. Linden-Bell and A. J. Kalnajs, Mon. Not. RAS 157, 1 (1972).
[111]G. S. Bisnovatyĭ-Kogan and A. B. Mikhaĭlovskiĭ, Astron. Zh. 50, 312 (1973) [Sov. Astron.-AJ 17, (1973)].
[112]L. S. Marochnik and N. G. Ptitzina, Astrophys. and Space Sci. 7, 437 (1970); N. G. Ptitzina, Astrophys. Lett. 5, 279 (1970).
[113]a) E. R. Harrison, Mon. Not. RAS 150, 443 (1970); b) T. Stix, Theory of Plasma Waves (Russ. transl.), Atomizdat, 1966.
[114]G. Contopoulos, Astrophys. J. 160, 113 (1970).
[115]G. Contopoulos, ibid. 163, 181 (1971).
[116]B. Barbanis, see[7], p. 343.
[117]C. L. Berry and P. O. Vandervoort, ibid., p. 336.
[118]A. J. Kalnajs, ibid., p. 318.
[119]R. H. Miller and K. H. Prendergast, Astrophys. J. 151, 699 (1968).
[120]F. Hohl and R. W. Hockny, J. Comput. Phys. 4, 306 (1969).
[121]R. H. Miller, K. H. Prendergast and W. J. Quirk, see[7], p. 365.
[122]F. Hohl, ibid., p. 367.
[123]S. B. Pikel'ner, Astron. Zh. 47, 752 (1970) [Sov. Astron.-AJ 14, 602 (1971)]; W. W. Roberts, Astrophys. J. 173, 1 (1972); F. Shu, V. Milione, W. Gebel, C. Yuan, D. W. Goldsmith and W. W. Roberts, ibid.,

p. 557; A. Kalnajs, Astrophys. Lett. 11, 41 (1972); W. W. Roberts and F. Shu, ibid. 12, 49; K. Rohlfs, A Astron. and Astrophys. 17, 246 (1972); M. Tosa, Publ. Astron. Soc. Japan 25, 191 (1973).
[124] F. Shu and V. Milione, On the Nonlinear Development of Gaseous Density Waves, Astrophys. J. (1973) (in press).
[125] J. M. Greenberg, in: Kosmicheskaya Gazodinamika [Cosmic Gasdynamics (Russ. transl.)], Mir, 1972, pp. 102, 103.
[126] P. Biermann, R. Kippenhahn, W. Tscharnuter, and H. Yorke, Astron. and Astrophys. 19, 113 (1972).
[127] D. W. Goldsmith, Astrophys. J. 161, 41 (1970); S. B. Pikel'ner, Comm. Astrophys. and Space Sci. 4, 129 (1972).
[128] W. J. Quirk, Astrophys. J. 180, 25 (1973); P. Meszaros, ibid. 381, 397.
[129] E. Parker, in:[125], p. 102.
[130] D. H. Rogstad, Astron. and Astrophys. 13, 108 (1971).
[131] W. W. Roberts, et al., Bull. Am. Astron. Soc. 3, 369 (1971).
[132] S. A. Kaplan, S. B. Pikel'ner, Ann. Rev. Astron. and Astrophys. (1974) (in press).
[133] J. H. Piddington, Mon. Not. RAS 162, 73 (1973).
[134] J. H. Piddington, Astrophys. J. 179, 755 (1973).
[135] M. C. Wright, P. J. Warner and J. E. Baldwin, Mon. Not. RAS 155, 337 (1972); J. E. Baldwin, Observatory 92, 2 (1972).
[136] J. E. Baldwin, see in[78].
[137] D. H. Rogstad and G. S. Shostak, Astron. and Astrophys. 13, 99 (1971).
[138] P. Meyer, Ann. Rev. Astron. and Astrophys. 7, 1 (1969).
[139] H. Weaver, in:[125], p. 35.
[140] B. Linds, IAU Symposium No. 39, Dordrecht, D. Reidel, 1970.
[141] B. J. Bok, American Scientist 60, 708 (1972).
[142] Yu. N. Mishurov, in: Dinamika galaktik i zvezdnykh skopleniĭ (Dynamics of Galaxies and Star Clusters), Nauka, 1973, p. 25.
[143] H. Weaver, in: Problems of Galactic Spiral Structures, Tucson, Steward Obs., Univ. of Arizona, 1972, p. 31.
[144] J. H. Piddington, Cosmic Electrodyn. 3, 129 (1972).
[145] Yu. Shchekinov and A. A. Suchkov, Astrofizika (1974).
[146] H. H. Oort, Mitt. Astron. Gesellschaft, Nr. 32, 15 (1973).
[147] H. Arp, Atlas of Peculiar Galaxies, Astrophys. J. Suppl. 14 (1966).
[148] Yu. N. Efremov, Zemlya i Vselennaya, No. 4, 54 (1971).
[149] N. Banicelli, O. Havnes, J. Hemphil and E. Bolviken, Astrophys. Lett. 12, 37 (1972).
[150] P. Pismis, see[7], p. 452.

Translated by R. W. Bowers