THE ASTROPHYSICAL JOURNAL, 338:78–103, 1989 March 1

MODAL APPROACH TO THE MORPHOLOGY OF SPIRAL GALAXIES. I. BASIC STRUCTURE AND ASTROPHYSICAL VIABILITY

G. BERTIN
Scuola Normale Superiore, Pisa, Italy
AND
C. C. LIN, S. A. LOWE, AND R. P. THURSTANS
Massachusetts Institute of Technology
Received 1988 July 5; accepted 1988 August 13

ABSTRACT

The principal objective of the investigations described in the present paper is the demonstration of the viability of the modal approach to global spiral structures in galaxies of all Hubble morphological types. This is done through (1) the *identification of the appropriate basic states* of galaxy models from the dynamical point of view, and (2) the demonstration of their *compatibility with observations* from the physical point of view. The modal approach is preferred to a direct evolutionary approach because it is believed that the observed spiral structure in the majority of galaxies is associated with the *late* states of an evolutionary process and that the pitch angle of the spiral arms of a galaxy can be used as a criterion for determining its Hubble type.

From a dynamical point of view, it is shown that barred spiral modes are likely to occur for relatively large disk masses. Specifically, the ratio R_A for *active* disk/total mass (see § III) within four exponential scale lengths for such spiral galaxies should be on the order of 30% or larger. There is in general only a single important unstable mode; thus, a regular quasi-stationary barlike structure may be expected. Normal spirals occur for lower active disk masses, especially those with lower pitch angles (Sa and Sb galaxies). From the physical point of view, this lower active disk mass is associated with the three-dimensional distribution of the stellar component, including the relatively large nuclear bulge.

A long-standing challenge in the case of normal spirals of small pitch angle is resolved. A new perception of the basic structure of these galaxies is proposed. For such normal spirals, the gaseous component plays an essential role, both in the excitation of the unstable spiral modes and in the stable maintenance of the final spiral structure. Indeed, dynamical studies indicate that a condition of marginal instability must be realized at large galactocentric distances and that a large part of the active galactic disk, with the possible exception of the central part, must be relatively "cool." Physically, this is realistic because of the three-dimensional distribution of stellar mass, especially in the central regions. The length scale of the active mass distribution is often found to be about twice as large as the exponential length scale of the optical disk. For the type of basic states considered, we expect in many cases two-armed spiral structures extending over a few exponential length scales, often with multiple winding. In particular, when the number of important spiral modes is small, the spiral structure is expected to have a highly regular grand design and to evolve in time in a quasi-stationary manner.

Subject headings: galaxies: internal motions — galaxies: structure — stars: stellar dynamics

I. INTRODUCTION

Hubble's classification scheme (Hubble 1926; de Vaucouleurs 1959; Sandage 1961) is based on a number of observational criteria such as the gas content, the size of nuclear bulge, the resolution of the spiral arms, and the pitch angle of the spiral arms. One of the principal challenges in the study of galaxies is to try to understand the physical basis of these criteria and their interrelationship (see, e.g., Whitmore 1984). For example, one would wish to know how normal spirals differ from barred spirals, and how the (small) pitch angle of an Sa galaxy is related to the (large) size of the nuclear bulge. Specifically one may pose the following questions: How can we construct a galactic model that simulates a galaxy or classes of galaxies of a given morphology? How can we account for the transitions observed among the various morphological types? How much can we infer on the basic state of a galaxy from its observed spiral structure? Why are there galaxies like NGC 6951 (Sandage 1961, p. 46) that exhibit characteristics of both SBb(s) and Sb systems? Why are regular spiral structures generally two armed?

In the context of the density wave theory, it is natural to attempt to interpret the various Hubble types by associating the *large-scale* structure of each of them with a global mode of oscillation, or a superposition and interaction of such modes, in an appropriate galaxy model. Here we assume that, in the galaxies we observe now, we are dealing with the "late" stages of an evolving dynamical system, where the modal description is usually preferable over nonmodal descriptions. If this modal approach is to be successful as a general perception, it is necessary to demonstrate the possibility to do so for *all* principal Hubble morphological types. Indeed, the main purpose of this study is to address the issue of the *identification of the basic state* that is to generate and to support spiral structure of a given shape. Besides possessing internal dynamical consistency, such a basic state must be realistic, so that astrophysical applications are justified. This is not a simple matter, as we shall see presently; in particular, the *physical* nature of the basic state of normal spirals poses a major challenge (see §§ III, IV).

From the dynamical point of view, it is essential that *the*

viability of the modal approach should be established; for it provides a clear perception and a basis for the discussion of a number of important issues such as the resolution of the winding dilemma in morphological classification and the nature of the hypothesis of quasi-stationary spiral structure (QSSS hypothesis). Although the latter is essentially a working hypothesis based on empirical grounds, the *dynamical support for its plausibility*, and indeed the proper limit to its applicability, can be best provided in the modal perception. Indeed, it would be very difficult to provide a theoretical basis for the QSSS hypothesis unless the modal approach is viable. The discussion of the regularity and the evolution of the spiral structure, can also be best carried out in the modal context. In a previous publication (Bertin and Lin 1987) we have reviewed several aspects of the issues related to the adoption of the QSSS hypothesis to describe grand design spiral galaxies. In that article we have also emphasized the role of *self-regulation* (see also Bertin and Romeo 1988) as a crucial process in the identification of the relevant basic states. In the following, self-regulation will be described mostly in § III, and a discussion of the morphology and the evolution of spiral structure will be given in § V.

These modal studies, in addition to providing a unified framework for the classification of spiral galaxies and for the structure of the corresponding basic states, can also provide a first step for the construction of detailed models for given galaxies (such as the Milky Way or M81).

Our study will be presented in two parts. In Paper I, by first illustrating the results of a modal survey of a family of basic states, we offer a unified approach to the dynamical basis of the morphological types in the Hubble diagram. The focus will be on the identification of the relevant basic states and on their physical justification. We also discuss other issues of direct astrophysical interest, such as the amount of mass in the dark halo. Paper II, as a separate self-contained analysis, gives a unified account of the relevant dynamical mechanisms at the basis of the processes of the excitation and maintenance of spiral modes in terms of waves governed by a simple cubic dispersion relation. Of course, the two articles are intimately related. On the one hand, much of the guidance to our numerical survey of Paper I derives from the analytical methods developed in Paper II. On the other hand, much of the confidence in the asymptotic methods and in the resulting physical interpretation proposed in Paper II is based on the concrete examples produced by the numerical survey of Paper I.

We shall now proceed to describe the issues addressed in Paper I.

a) Modal Approach to Spiral Structure: A Description of the Adopted Method of Investigation

Since global modes are the intrinsic characteristics of a given system, in order to cover *all* Hubble types in a coherent perspective (e.g., to determine whether it is true that normal spirals are associated with low active disk mass and barred spirals with high active mass), we have considered a family of basic states so that all of them could be covered through the *natural* variation of certain general parameters such as the ratio of disk mass to total gravitational mass. An atlas of modes is provided for hundreds of models (Thurstans 1987) together with their dynamical characteristics. Technically, this is made possible by devising a method of automated extrapolation in locating eigenvalues in order to overcome the difficulties that often occur with the calculation of modes of low growth rate. The equations used are presented in Paper II.

A modal investigation of this kind is quite complex and, as we shall demonstrate, involves a number of important physical issues. Therefore, we have adopted the following strategy. We begin our study with an *extensive exploratory survey* of basic states by following the properties of *one representative two-armed spiral mode* and the way these properties change while several parameters characterizing the basic equilibria are changed.

Already at this stage, a good fraction of the models that are considered can be discarded on the basis of the growth rate of the mode. In fact, we should focus on cases where the mode is characterized by *moderate growth rate* ($\gamma P \lesssim 1$; see text for definitions). These are the modes that can be stabilized through nonlinear mechanisms. Basic states subject to violent instabilities are expected to evolve rapidly, and observed galaxies are presumably beyond such a stage of rapid evolution (also through mechanisms of self-regulation).

At this point we are still only in the middle of the investigation, because two important physical issues have yet to be faced. First of all, we should check that the basic states finally adopted do have physical justification. In particular, we should carefully examine the nature of the profiles of the parameters that govern the distribution of mass and random motions. Some of these considerations can be made and indeed were made in advance, that is, in the initial choice of basic states for the exploratory survey. In fact, our survey gives reasonable *barred* morphologies in very reasonable basic states. However, the need for a further examination of the nature of the basic state may become apparent, should the survey indicate special difficulties. This was indeed the case for spiral morphologies of the *normal* type, especially those with a small pitch angle. Some of these difficulties may be intrinsic to the family of models adopted. Given the importance of this issue (which is at the core of Paper I, §§ III–IV), we shall elaborate further on it in the Introduction (§ I*b*).

Finally, after the appropriate set of basic states is identified and assured of its physical justification, the characteristics of *all* the global modes for each of these basic states must be studied in order to determine the expected morphological type, the appearance of the spiral structure, and its evolution in time. In particular, *three-armed* modes must be studied to see whether, as we expect, they may be easily weakened or eliminated by Landau damping at the inner Lindblad resonance. It should also be determined whether and how the morphological type of the spiral structure for a given basic state will change over a long period of time.

This discussion clearly shows why preliminary results published earlier (Berlin *et al.* 1977; Haass, Bertin, and Lin 1982) represent only an incomplete step in the direction just described. The major limitation of those studies, besides the somewhat small number of investigated cases, is that too little attention was given to the roles of gas and self-regulation, and that moderate instabilities were not considered. An adequate discussion of the choice of the basic state was also not provided.

b) A Major Physical Issue: How Do We Explain Tightly Wound Normal Spirals?

There is a long-standing difficulty in identifying models appropriate to normal spirals, especially those with small pitch angles. In the literature, failure to find two-armed normal

spirals modes with small pitch angles has been reported by several authors. As we shall see as a result of the exploratory survey of § II, the cause of these difficulties may be traced to the type of basic states adopted (see detailed discussion in § II*f*).

Tightly wound spiral modes with moderate growth rate are found only in *cool* disks. In addition, their morphology and radial extent are realistic (see § IV) only if the *mobile* part of the disk mass in the modal calculation differs significantly from the exponential profile suggested by *photometric* studies. The issue of the *cool* disk is often recognized as a crucial point of dispute (see, e.g., van Albada and Sancisi 1986, p. 455; see also Bertin and Romeo 1988). The term *cool* is used here to denote a condition close to that of *marginal* stability with respect to local axisymmetric disturbances. The resolution of the difficulties found in our exploratory survey of § II depends on the recognition of the following physical characteristics of the galactic disk in the construction of our models. (A quantitative description will be given in the main text later, especially in § III).

First, the galaxy is composed of at least *two* components: stars and gas. With reference to the stellar component, it is necessary to recognize the finite thickness of the galactic disk and the presence of a nuclear bulge which merges smoothly with the disk, and of an invisible halo. The *three-dimensional* configuration of mass distribution significantly reduces the Jeans instability of the stellar component, and the *active* disk mass to be used in the theoretical model with an infinitesimally thin disk. This enhances the importance of the gaseous component relative to the stellar component. With reference to the gaseous component, we note that the disk is *thinner* and the dispersive velocity is *lower*, and consequently it is much more responsive to the gravitational field. Furthermore, the dispersive velocity in the gas is *dissipative*, since it is turbulent. There is thus a process of *self-regulation* that may keep the disk *cool* when there is a moderate amount of gas. In the end, the conclusion is reached that the gaseous component plays essential roles both in the excitation and in the maintenance of normal spirals, even in those of types Sa and Sb where the total gaseous component is *small*. The crucial point is then the determination of the amount of gas needed to ensure that the outer disk may be maintained in a condition of marginal stability with respect to axisymmetric disturbances. We found that it is likely that there is in general sufficient gas to do so, especially in view of the presence of molecular gas. The reduction of the effective mass density of the thicker stellar disk, often by as much as a factor of 2, is another essential element.

The mechanism of dynamical instability of such normal spirals differs significantly from that in barred spirals, where the disk thickness and the gaseous component play less significant roles. Nevertheless, the mechanism of instability for normal spiral modes is still *gravitational* in nature, in contrast to other mechanisms that may be proposed (see, e.g., discussion after the paper by Rubin 1987, p. 64).

c) *Summary of the Main Results*

1. The main conclusion of the exploratory survey is that *all* morphological types can be represented through a systematic change of parameters of the relevant basic states. The results show surprising simplicity. If one considers the ratio R_M of disk mass to total mass within four scale lengths of the exponential disk adopted in our model, all morphological types may be obtained for a mass ratio below 0.5. Normal spirals are typically obtained for *active* disk mass ratio R_A of the order of 0.1–0.3; barred configurations show up for higher active disk masses. (See §§ III, IV for the precise defintions of active disk mass.) There is thus a one-dimensional sequence of morphological types with increasing disk mass, provided that the condition of moderate instability is imposed.

2. A closer examination reveals that the physical basis for the *normal* spiral modes obtained in the exploratory survey appears unsatisfactory. The roles of gas and three-dimensional distribution of mass must be properly incorporated. Without such an inclusion, the corotation radius of the spiral pattern is generally only a little larger than the luminous exponential length scale h_*. Within a realistic model it is of the order of *three* length scales so that the corotation zone lies in a region where the gaseous mass is sufficiently plentiful to play a significant dynamical role. (See §§ III, IV for the detailed discussions.) In general, the primary part of the spiral pattern is found to extend over the range $h_* < r < 3h_*$. This seems to be in general agreement with observations. The disk is cool roughly for $r > 2h_*$. In barred spirals, the role of the gaseous component is less crucial.

3. It turns out that, within the framework just described, the results originally obtained in the exploratory survey can be reinterpreted by suitable rescaling so that they can still be useful as a crude first approximation, with the size of the spiral pattern roughly 3 times the size of the luminous exponential disk.

4. A final assessment of the results obtained and a proper consideration of the role of inner Lindblad resonance indicate that we have established the general viability of the modal approach to cover all Hubble types. A number of implications on regularity, morphology, and evolution of spiral structure are discussed (§ V). Several open questions for future research are identified, where explicit inclusion of nonlinearities and full two-component calculations appear to be needed (§ VI). Most of these call for an improved *theoretical* and *observational* study of the interstellar medium in galaxies.

5. The present work is mostly concerned with the properties of *classes* of spiral galaxies, rather than with individual objects. However, we think that substantial progress is now made and that we can offer now a good basis for a modal study of specific galaxies. In particular, it would be desirable to reexamine the model of M81 constructed by Visser and Haass (see Haass 1982) in the present perspective, since that study was made only with the asymptotic theory for tightly wound spirals, and not all the relevant factors, such as the role of the gaseous component, were taken into account in the determination of the basic state of the model.

II. EXPLORATORY SURVEY OF GLOBAL MODES

We shall now carry out the program outlined in § I*c*, beginning with an extensive exploratory survey of a family of basic states. In this exploratory survey, our aim is to get a general idea of how the characteristics of a representative unstable mode vary with the change of the parameters in the basic state adopted. We shall therefore give only a brief description of the justification of the type of basic state adopted in our survey but defer the detailed discussion to later sections (§§ III, IV). It is expected that the survey would exhibit linear sequences of basic states that support spiral modes of the normal type and of the barred type. It is also expected, however, that not all of the models will have a sound physical justification, especially those subject to tightly wound spiral modes. Emerging diffi-

culties will be discussed together with those experienced by other investigators (§§ II*e–f*). In order to overcome these difficulties, we are led to a thorough examination of the physical basis of the models adopted and to their modification (§§ III, IV).

a) Simple One-Component Dynamical Models

As a model of a flat galaxy we consider an infinitesimally thin disk of *active mass* in the presence of a "fixed" or *inactive* component. For simplicity in describing the thin disk, a *one-component fluid model* of zero thickness is adopted. (See Paper II for the relevant mathematical description.) Therefore in the following, but also in any calculation of this kind, special attention must be paid to the *choice of the basic state* and to the evaluation of the relevant results, in view of the complex nature of the actual physical system that is being represented.

A real galaxy contains a number of components (stars, gas, etc.), distributed in a *three-dimensional* geometry. Only a part of these materials may be said to lie in the galactic disk. It is therefore not easy to construct a single-component model to mimic the essential dynamical characteristics of the galactic disk. The simple family of basic states described in § II*b* is indeed constructed in view of these physical issues. However, the proper physical justification will be best discussed (§ III) *after* we evaluate the results obtained from an extensive modal investigation of such a family of galaxy models. We shall find that the three-dimensional distribution of mass makes a crucial difference between the dynamics of open spiral modes and barred modes, on the one hand, and that of tightly wound normal spirals, on the other. The effect of the three-dimensional distribution of matter is clearly larger in the latter case. This discussion will allow us to proceed to identify the most appropriate basic states corresponding to the observed spiral morphologies (§§ III–V).

b) A Simple Family of Basic States

The family of basic states adopted has a simple rotation curve suggested by observations (see, e.g., van Albada and Sancisi 1986; Rubin 1987). It has an exponential disk, but with the active disk mass resulting from the modification through a correction for the three-dimensional distribution of mass. The linear velocity of rotation $V(r)$, the active surface density $\sigma(r)$, and the Q-parameter for our models are given by the following formulae. (See Paper II for a systematic listing of definitions.) The active disk density has two contributions: the first term represents the stellar mass; the second term σ_g represents the gaseous mass (in this section, we shall set for simplicity $\sigma_g = 0$):

$$V = V_\infty \frac{r}{r_\Omega}\left[1 + \left(\frac{r_2}{r_\Omega}\right)^2\right]^{-1/2},$$
$$\sigma = \sigma_0 e^{-r/hf} + \sigma_g, \quad \sigma_0 = (1 + \Delta)\sigma_{00}, \qquad (2.1)$$
$$Q = Q_{\rm OD}[1 + qe^{-(r/r_Q)^2}],$$

where r_Ω, h, r_Q are the scale length parameters, and V_∞, $(1 + \Delta)\sigma_{00}$, $Q_{\rm OD}$, and q are the magnitude parameters (OD refers to outer disk). The reduction factor f is introduced essentially to account for the presence of a bulge and the effect of finite thickness. It is taken to be of the form

$$f = 1 - f_0(r/r_{\rm cut}) + \tfrac{1}{6} f_0(r/r_{\rm cut}) \exp(r/h), \qquad (2.2)$$

where $f_0(r/r_{\rm cut}) = [1 + 4(r/r_{\rm cut})][1 - (r/r_{\rm cut})]^4$ for $r \le r_{\rm cut}$ and $f_0 \equiv 0$ for $r > r_{\rm cut}$. At this stage, the adopted *analytic* forms for Q and f are just a convenient smooth choice. The basic state with $r_{\rm cut} = (\frac{1}{2})h$ is illustrated in Figure 1. The effect of the reduction factor is shown in Figure 2, where σ is plotted for the cases $r_{\rm cut}/h = \frac{1}{2}$, 1, 2, 3, 4, 5, 6. Note that in all the cases shown, the reduction operates mostly in the *inner disk*. The reference scale length is h, which is fixed to be 4, and the reference velocity is V_∞, which is taken to be 140. If lengths are measured in kiloparsecs and velocities in kilometers per second, then the disk density turns out to be expressed in $M_\odot$ per square parsec. ($\pi G = 13.54$ in these units.)

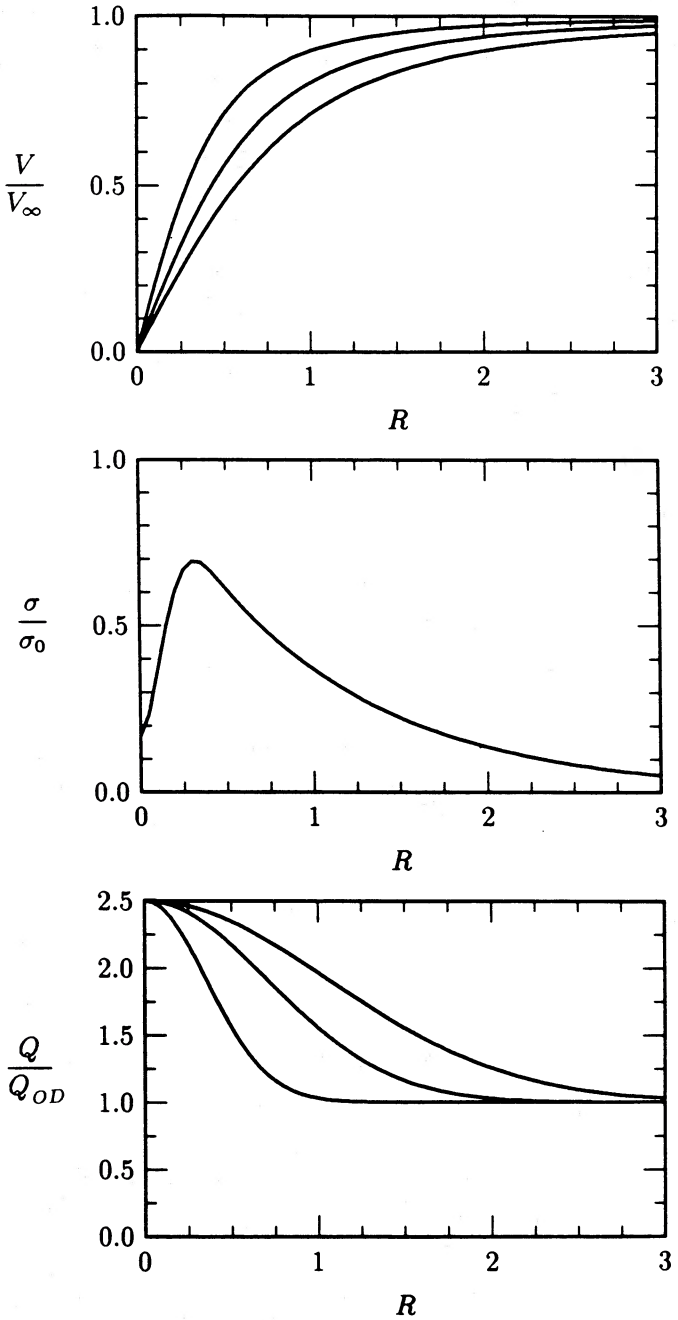

FIG. 1.—Properties of the basic states in the present survey. Quantities are plotted as a function of $R = r/h$. *Top*: rotation curves for $r_\Omega/h = 1/2$, 3/4, 1. *Middle*: surface mass density for $r_{\rm cut} = (1/2)h$. *Bottom*: Q-profiles for $r_Q/h = 1/2, 1, 3/2$.

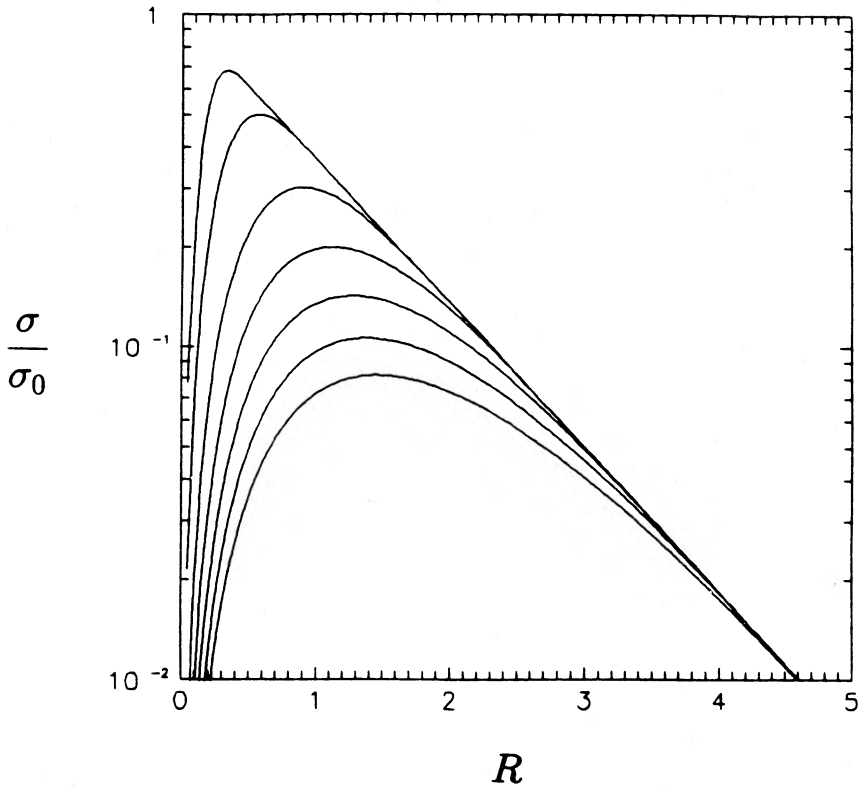

FIG. 2.—Effect of reducing the mass density from the basic exponential disk. The density σ is shown as a function of $R = r/h$ for $r_{cut}/h = \frac{1}{2}$, 1, 2, 3, 4, 5, 6 (from top to bottom).

The rotation curve defined in equation (2.1) is partly supported by an inactive component which we may associate with a spheroidal bulge halo. Part of the inactive mass could also be in the form of a thick disk. In this paper we do not discuss the relative "composition" of the inactive mass since it does not change the stability analysis. If, for simplicity, we refer to the model with $r_\Omega = 2, f = 1$ for the case where the inactive component is taken to be spherically distributed, then the disk to total mass ratio R_M, at the location $r = 4h$, is approximately given by

$$R_M \approx 0.38(1 + \Delta) . \tag{2.3}$$

For the survey presented in this section, since we are taking $f \simeq 1$, the whole disk mass is regarded as active ($R_A \simeq R_M$), and a decrease of Δ implies a "transfer" of mass into the spherical distribution. In some cases, this may even introduce a mass concentration in the central regions that tends to simulate a nuclear bulge. Observationally, there is a corresponding uncertainty in the determination of disk mass since the mass to luminosity ratio M/L is usually thought to be known only within a factor of 2. Referring to this basic state, the so-called "maximum disk" solution (see van Albada and Sancisi 1986) corresponds to $\Delta \sim 0.25$, i.e. $R_M \approx 0.5$. For cutout models with sizable values of r_{cut}, the maximum value of Δ can change from this number.

The family of models described above is characterized essentially by six dimensionless parameters (q, Q_{OD}, Δ; r_Ω/h, r_Q/h, r_{cut}/h). A routine survey with 10 values chosen for each parameter would lead to 10^6 cases, which would be a nightmare to attempt to analyze. As it turns out, it is possible to reduce the number of cases needed in a significant survey down to 10^2, without missing any essential morphological type. For completeness, especially for a better understanding of the dynamical mechanisms (see Paper II), $\sim 10^3$ basic models were actually considered (Thurstans 1987). On the other hand, for application to morphological studies, there is a natural division of models into subgroups within which law and order prevail. Thus, the number of representative cases can be actually reduced to ~ 12.

The present survey begins with a three-parameter survey in which the changing parameters are essentially Δ, Q_{OD}, and (r_Q/h). These parameters are expected to be especially important for the determination of stability characteristics. In future work, we shall deal with the variation of the other dimensionless parameters, particularly q and r_Ω/h. In the following $q = 1.5$ and $r_\Omega = h/2$ (unless specified otherwise). Variation of the parameter (r_{cut}/h) is partly considered here (see following definition of the B-surveys and the discussion of tightly wound normal spiral modes). Because of these limitations, we shall refer to the models adopted in the present survey as the "rudimentary models."

c) *Modal Survey of the Family of Basic States: A Set of Two-Parameter Subsurveys*

By following the properties of the fastest rotating two-armed spiral model we have systematically studied the family of basic states introduced in § II*b*. As explained above, this is only a preliminary survey to provide a general perspective.

We shall describe a three-parameter survey based on the changes of the basic state through changes in the three parameters r_Q, Q_{OD}, and Δ. Three types of *two-parameter subsurveys* are considered: type A, surveys with r_Q specified; type B,

surveys with Q_{OD} specified; type C, surveys with Δ specified. The details of these surveys may be found in Thurstans (1987). Here we focus on the properties of the following four major subsurveys, all with $r_{cut} = r_Q$:

$$\begin{aligned} &\text{Survey } A_1: \quad r_Q = h/2 . \\ &\text{Survey } A_2: \quad r_Q = h . \\ &\text{Survey } B_1: \quad Q_{OD} = 1 . \\ &\text{Survey } B_2: \quad Q_{OD} = 1.2 . \end{aligned} \tag{2.4}$$

For each case we have a systematic diagnostics in terms of six charts: (*a*) modal shape (perturbed density contour), (*b*) α-spectrum, (*c*) propagation diagram, (*d*) density eigenfunction (real and imaginary part), (*e*) stability characteristics, and (*f*) conditions at the corotation circle. Items (*b*), (*c*), and (*f*) are mostly of theoretical interest and are discussed in Paper II, although we may note that the α-spectrum has also been used as diagnostics of observed galaxies (see Iye *et al.* 1982). (The full set of data available from the various surveys can be found in Thurstans 1987.) In § V below we show a sample of these data that are more relevant for astrophysical applications. Here we highlight surveys A_1 and B_1 by giving modal shapes in Figures 3 and 4. Note the coherence and gradual transitions realized in the surveys. In general, *barlike spirals appear at higher disk masses*, and *normal spirals appear at lower disk masses*. The bar modes found here are the *long* bars (see Bertin 1983*a* for an example of such a mode with the propagation diagram); some of them extend directly to the corotation circle. Short bars were found by Haass, Bertin, and Lin (1982) through a lower-

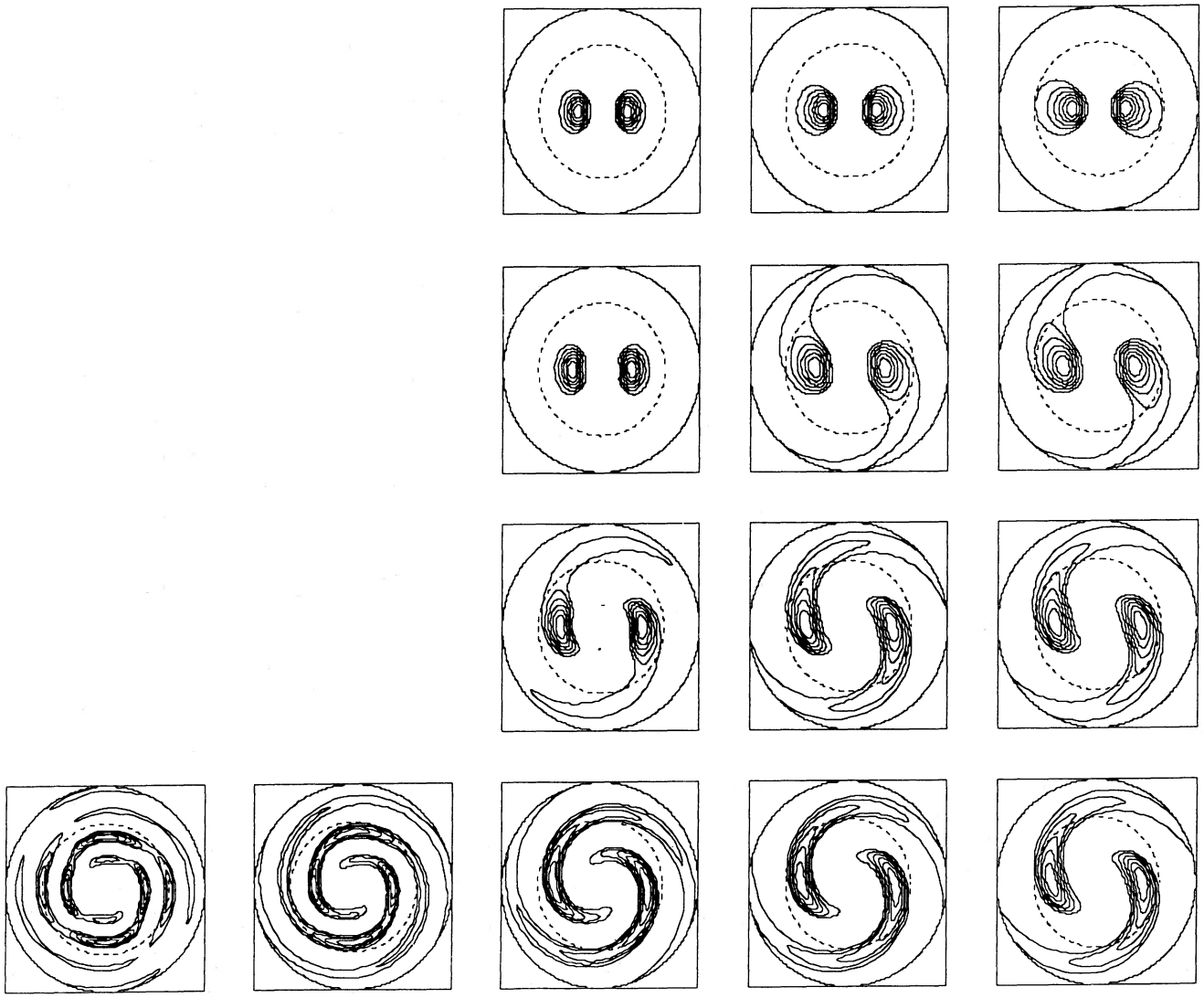

FIG. 3.—Modal shapes for the A_1 survey. Here, and in the following diagrams of modal shapes, perturbed density contours are given in arbitrary units in steps of one-seventh of the peak value, from the ($\frac{1}{7}$) contour upward; dotted circles identify corotation, solid circles correspond to $\nu = 0.75$. This survey is characterized by $r_Q = 2$. Top frames have $Q_{OD} = 1.5$, second row $Q_{OD} = 1.3$, third row $Q_{OD} = 1.1$, bottom frames have $Q_{OD} = 1$. From left to right Δ changes in the following way: -50, -35, -15, $+5$, $+25$(%). The six frames of the right low triangular corner are associated with large growth rate ($\gamma P > 1$).

ing of the wave barrier in the interior part (sometimes called "the Q-barrier"). Further discussion of these morphological types will be given in § V.

d) Overall Perception of Instability Characteristics

As noted in § I*a*, strongly unstable modes cannot be present in observed galaxies; yet for certain parameter regimes, excitation processes may be very powerful. How can such regimes be avoided? Obviously, the answer lies in either (i) high velocity-dispersion for a given value of the disk mass, or (ii) low disk mass for a given velocity dispersion.

The stability characteristics of the four major subsurveys are presented in Figure 5. These results conform to, and actually were to some extent anticipated by, the studies of a single

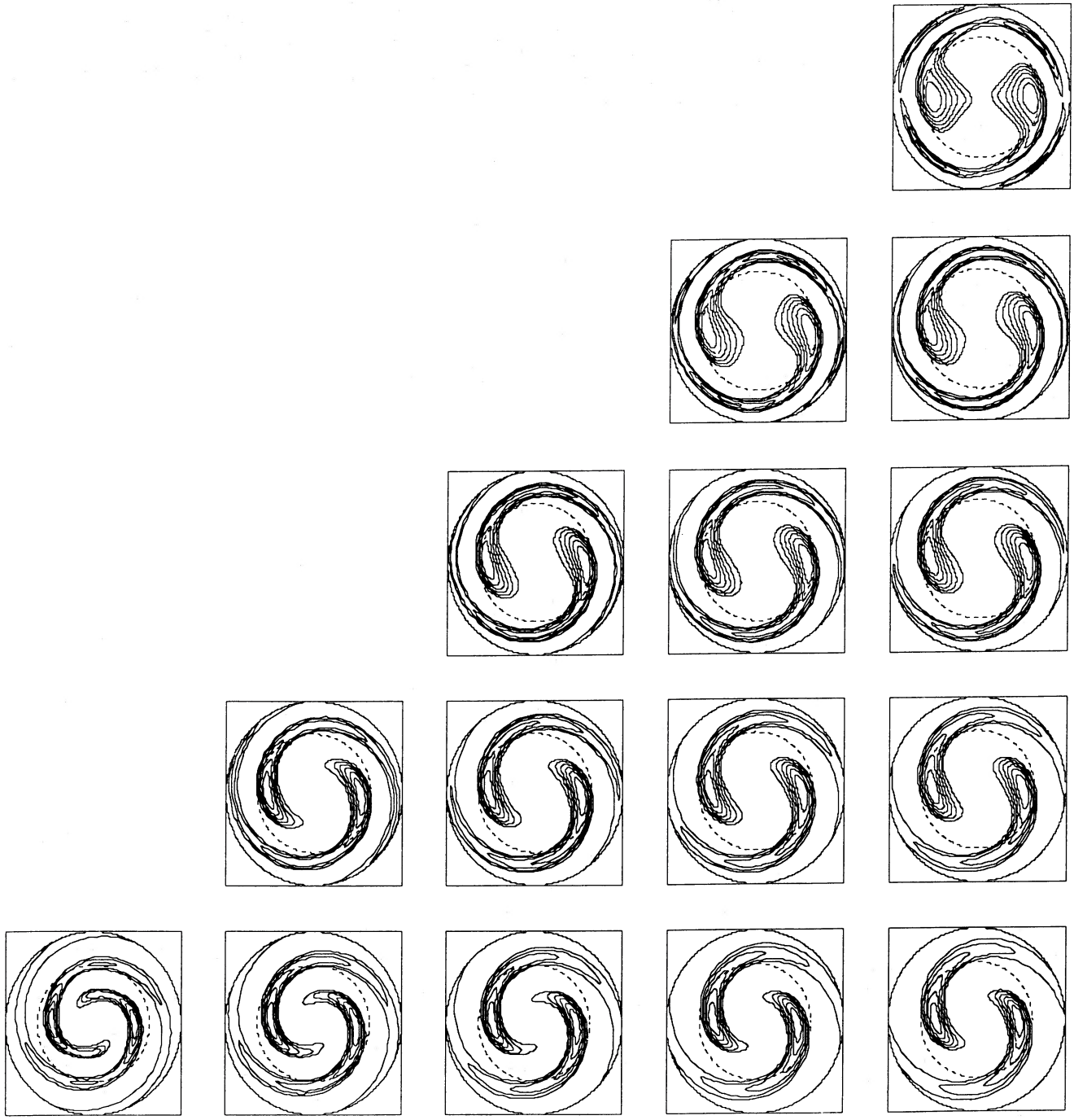

FIG. 4.—Modal shapes for the B_1 survey. This survey is characterized by $Q_{OD} = 1$. From the top downward r_Q decreases from 6 to 2 in equal steps. From left to right Δ increases from -25% to $+15\%$ in equal steps. The frames on the diagonal boundary on the upper left are associated with unstable modes with moderate growth ($\gamma P \lesssim 1$). The others are violently unstable ($\gamma P > 1$).

simple local dispersion relation, which will be discussed in Paper II. Here we shall only present the results obtained; the dynamical mechanisms supporting the modes found and the relation of their morphology to the dispersion relationship will be described in Paper II, through the use of diagrams in a plane of two parameters (J, Q) defined there.

From the A_1 survey ($r_Q = h/2$), we see a sharp change of stability characteristics between the condition $Q_{OD} > 1.1$ and the condition $Q_{OD} < 1.1$. For the A_2 survey ($r_Q = 4$), we see that the separation of the two domains occurs at $Q_{OD} = 1$. The disk in the case $Q_{OD} \approx 1$ is already quite massive and has a quite open spiral inside of the corotation circle. There is a more tightly wound spiral outside of the corotation circle, as is typical in baried spirals. In the B_2 survey ($Q_{OD} = 1.2$), instability is in general weak. We find SB0 (and SBa) galaxies simulated. In survey $B_1(Q_{OD} = 1)$, we see that the conditions of moderate or low instability may occur over a wide range of disk masses, provided that there is an associated change of the

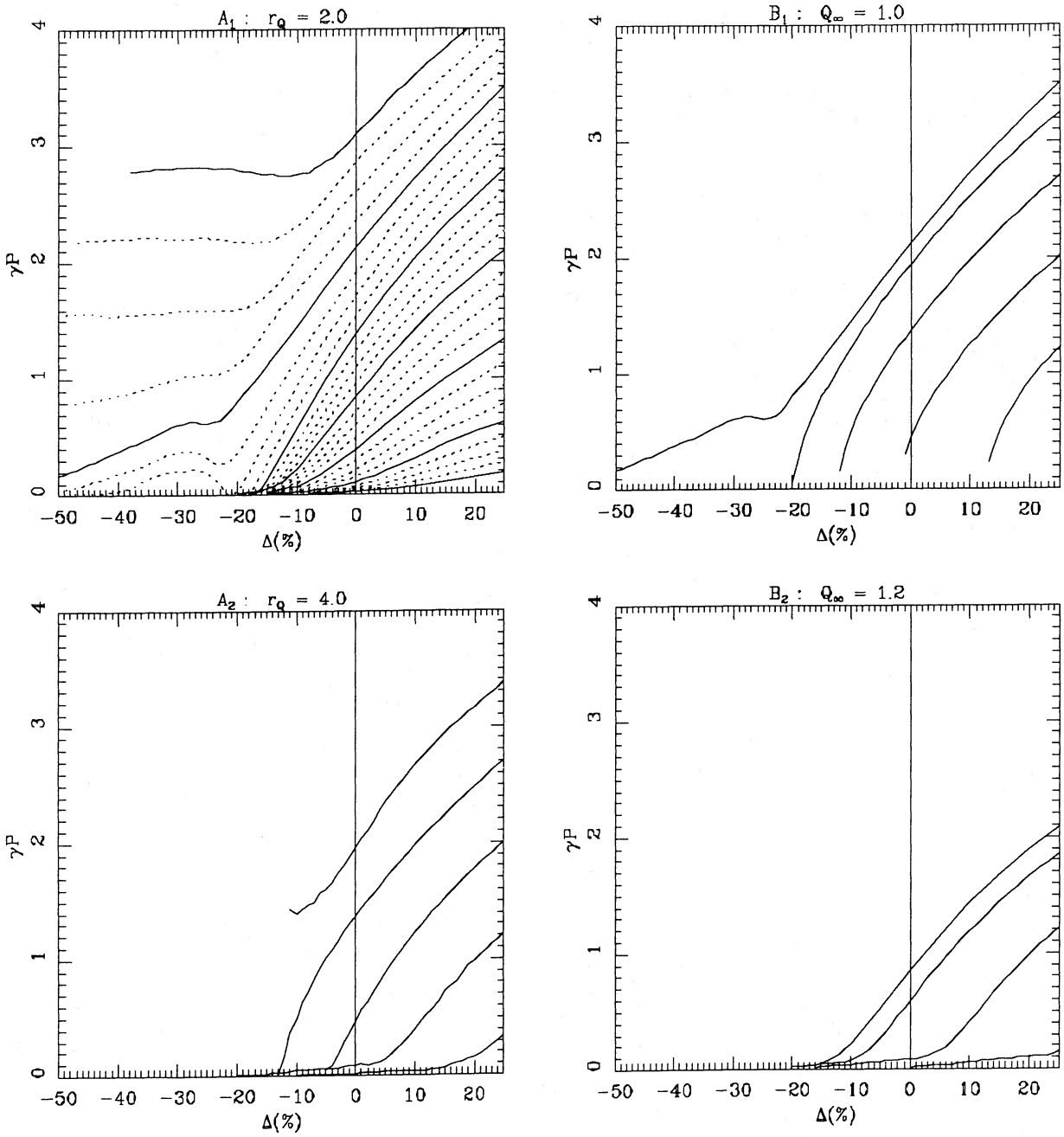

FIG. 5.—Stability diagrams for various surveys. These diagrams give a synthetic view of the relevant regimes of instability and of the change of character of the mechanisms of instability when an equilibrium configuration is tracked along various paths in parameter space. *Top left:* A_1 survey, $r_Q = 2$; tracks at constant Q_{OD}, solid lines in steps of 0.1; lowest track at $Q_{OD} = 1.5$, uppermost track at $Q_{OD} = 0.9$. *Bottom left:* A_2 survey, $r_Q = 4$; tracks at constant Q_{OD} in steps of 0.1 from $Q_{OD} = 0.9$ (*top*) to $Q_{OD} = 1.3$ (*bottom*). *Top right:* B_1 survey, $Q_{OD} = 1$; tracks at constant r_Q in steps of 1.0 from $r_Q = 6$ (*bottom*) to $r_Q = 2$ (*top*). *Bottom right:* B_2 survey, $Q_{OD} = 1.2$; tracks at constant r_Q in steps of 1.0 from $r_Q = 2$ (*top*). The character of the stability diagrams can also be recovered by using a simple ordinary differential equation as suggested by the asymptotic theory of various regimes (see Paper II).

scale length r_Q. With high values of r_Q, the *overall* value of the Q-parameter is high, so that $Q > Q_{OD}$ at the relevant corotation circle, and we find moderate instability even at fairly high disk masses.

The A survey indicates the existence of two types of instability characteristics, depending on whether $Q_{OD} > 1.1$ or $Q_{OD} < 1.1$. (This is why survey B was done in the manner given above.) It turns out that cases with $Q_{OD} > 1.1$ correspond to open spirals or barred spirals, whereas $Q_{OD} < 1.1$ corresponds to normal spirals. Through an examination of the data for the cases $Q_{OD} \cong 1$–1.05, in the surveys A_1 and A_2, we can thus see that *all barred spirals are stabilized if* $\Delta < -15\%$; i.e., if $R_M \lesssim 0.3$. This same conclusion is obtained from the other surveys too. However, it should also be noted that Figure 5 clearly shows that there is still room for *instability with respect to (normal) spiral disturbances at much lower values of* R_M; significant instability exists even for $R_M \approx \frac{1}{5}$. In surveys A_2 and B_2, instabilities are not found at such low disk masses. To be more precise, our survey shows that, *if barlike modes are to be present*, R_M must exceed the lower limit 0.3 (at $\Delta = -20\%$). This implies an upper bound of 2.3 for the ratio M_H/M_D of halo-to-disk mass at four exponential scale lengths. For normal spirals, however, this upper bound of M_H/M_D may be as high as 4.

Therefore we recognize that some limits on important quantities such as R_M can be set on the basis of instability arguments, but only if the *morphology* of the spiral mode is properly taken into account. This contrasts with the approach taken by Athanassoula, Bosma, and Papaioannou (1987).

Note that all the bar modes discussed here are *long* bars, i.e., the bar has a length comparable to the diameter of the corotation circle. Under different circumstances, e.g., those expected to favor *short* bars like the case of the SBc(s) spiral NGC 7741 (which is close to an Sc spiral), we argue that there could be more halo mass (e.g. M_H/M_D close to 3). Questions of this kind give another reason for pursuing our survey further, beyond the limitations of the present three-parameter investigation.

Finally, we note that morphological properties of the relevant unstable modes, such as the length of the bar, as a fraction of the diameter of the corotation circle, are very important for the study of the reaction of the gaseous component. In our work, we can provide quite specific information about this and other dynamical characteristics of the mode for a given basic state (see following discussions in §§ IV, V).

e) *Merits and Limitations of the Survey*

As outlined in § I*a*, a survey of this kind has a few limitations that must be removed when astrophysical applications are considered. In particular, we should select models for which the relevant modes have *moderate growth rate.* Then we should consider in detail the physical justification of the basic states thus selected, and we should make sure that other modes (especially those with multiple arms) do not change the picture suggested by the representative two-armed mode that we have been tracking. As we have stated at the end of § I*a*, a solution to the issue of other competing modes is to be found in the role of the inner Lindblad resonance, which can inhibit three-armed modes even when these are expected to be strongly unstable on the basis of the overreflection mechanism at corotation (see Paper II); this point will be addressed in detail later in § IV*c*.

Here we draw the attention of the reader to the issue of the physical justification of the basic state. In particular, while *barred morphologies* with moderate growth rate are associated with a broad class of galaxy disks with relatively high mass and occur with shapes and properties that look consistent with those of many SB's, the *tightly wound normal* spiral modes obtained in this rudimentary survey are found for models which are not satisfactory. First of all, in order for tightly wound spiral modes with moderate growth rate to emerge from the survey, the models must be characterized by a *cool* disk ($Q_{OD} = 1$) in a very *narrow* "temperature" range. In particular, Q should be "tuned" to be very close to unity over a large fraction of the galaxy disk (e.g. for the survey A_1 we must take $r/r_Q = 2$ at $r = h$ so that $Q = 1.03$). Such a fine tuning is likely to be realized in actual galaxies only when specific physical mechanisms concur to this goal (i.e. self-regulation; see § III*b*). We should stress that the major physical problem here is not so much the "coldness" of the disk (after all, we can get a relatively "warm" disk with $Q = 1.2$ just by increasing by 15% the value of the velocity dispersions used in some cases of the A_1 survey), but rather the narrowness of the "temperature" range. Therefore we should be ready to explain how the system can maintain a given value of the velocity dispersion, even when most dynamical mechanisms are likely to induce a (secular) increase of random motions. The second reason for concern for tightly wound spiral modes, emerging from this rudimentary survey, is that they seem to be too *small* in linear size, while observational evidence indicates that regular spiral arms often extend over several exponential length scales in the whole galactic disk.

Identifying these difficulties in clear terms is also among the merits of the survey; we shall see in the next two sections how the role of gas and a more realistic description of the *active* mass disk can resolve these difficulties. In addition, there are other obvious merits in the extensive survey that has been performed. First, by means of diagrams of the kind shown in Figure 5 we have produced an enormous data base for dynamical studies (see Paper II). Second, smooth transitions among various morphologies are clearly demonstrated and the basic physical regimes are thus identified. Third, a large fraction of models, mostly those leading to barred morphologies but also some of the cases displaying normal spiral structure, appear to have a reasonable physical justification and a good chance of applicability. Finally, as we shall make clear in the following sections, most of the survey *can* actually be used for astrophysical applications, even for tightly wound spirals, *provided a proper reinterpretation and physical rescaling of the model are made* (see § IV*b*).

f) *Other Modal Studies*

Before we proceed to present the resolution of the difficulties in our own case, let us digress to note that other authors did not find adequate basic states for supporting normal spiral modes.

In a pure stellar disk, the disk mass density approaches zero at large distances but the stellar velocity dispersion may not. To be sure, the gravitational instability of a low-mass disk may not lead to $Q > 1$, but if the velocity dispersion in the stars should be increased by any means, there is no way for it to be reduced. Thus, the Q-parameter can become very large at large radii. This is indeed true in the distribution

$$Q(r) = Q_0[1 + (r/2a)^2]^{1/2} \tag{2.5}$$

adopted by Aoki, Noguchi, and Iye (1979). Note that another contrast with our Q-distribution (eq. [21]) is the *absence* of a *peak in* $Q(r)$ *near the center.* Thus choices as in equation (2.5)

ignore the presence of a nuclear bulge and a disk with high velocity dispersion near the center.

The sharp contrast between equations (2.1) and (2.5) results in normal modes with completely different characteristics. As pointed out by Haass (1982), the low value of Q at $r = 0$ and the rise of Q outside lead to a resonant cavity for which the wave pattern tends to show a peak density near the center in the form of a short bar. Tightly wound spirals and long bars were therefore not found by Aoki *et al.* Most of the modes found by these authors also have excessive growth rates, since a full mass disk is used.

Haass also pointed out that the Q-distribution simulated with the device of the "softened gravity" model used by Toomre (1981) and Erickson (1974) has characteristics similar to that used by Aoki *et al.* if one identifies the velocity dispersion (or acoustic velocity) with the product of the epicyclic frequency and the "softening length." Unlike Aoki *et al.*, Toomre (1981) studied higher modes at full mass or uniformly reduced mass and obtained the "edge mode" (D-mode) as the dominant modes for reduced mass. The growth rate of this mode remains very high, even at reduced disk masses; its shape is a very open spiral. Tightly wound normal spirals are not found to be dominant for these models of basic states.

Another situation was reported by Haass (1983) with models subject to too many modes with high growth rate, especially three-armed and four-armed modes. In realistic models, inner Lindblad resonance is expected to limit considerably the number of modes (see § IV*c*) and lighter disks would be less unstable.

Therefore, we should be aware that certain properties of modes and instabilities may *just reflect the special properties of the choice of the basic state*, and therefore need not be of physical concern. As an example we may cite the singular model used by Zang (1976) which lacks a nuclear bulge and systems with exaggerated gradients which may lead to the edge mode just mentioned. Before rushing to general conclusions, the physical justification of such basic states should be carefully examined.

The results of Athanassoula and Sellwood (1986; see Fig. 2 of their paper), who found that there is modal instability only if the disk/total mass ratio is approximately one-third or more, are apparently consistent with our survey, to the extent that they are *restricted* to open bar modes. Strictly speaking, the basic distribution functions for their basic states are quite different from ours, and any obvious comparison would not be expected. The fact that there is this degree of agreement with respect to *barlike* instability is somewhat surprising. (There is perhaps underlying some general dynamical principle, in the nature of that suggested by Ostriker and Peebles 1973). On the other hand, normal spiral modes have not been found in their model, for dynamical reasons explained elsewhere in this paper. Basically these are traceable to the failure to simulate the role of the gaseous component and to acknowledge the three-dimensional distribution of the stellar mass.

One should not think that the difficulties that are described in the construction of tightly wound normal spirals are special to the modal approach. Other scenarios are found to face even more serious difficulties, and they have not yet produced *quantitative* results for suitable application to observational data.

We shall not pursue further the study of the type of physical objects represented by models described in this subsection, but return to the type of models described in § II*b* and show how the difficulties described in § II*e* can be resolved.

III. PHYSICAL BASIS OF MODELS WITH COOL OUTER DISKS: THE ROLES OF THE GAS COMPONENT AND OF THE THREE-DIMENSIONAL DISTRIBUTION OF MATTER

In view of the difficulties identified in our exploratory survey we shall carefully review the physical basis of the modeling process. We shall first describe the essential physical nature of the issues to be examined (§ III*a*), to be followed by a discussion (§ III*b*) of the roles of the gas, and the specific process of self-regulation that ensures a *cool* disk in the outer parts ($r > 2h_*$). This analysis will provide us with the tools to respond to the specific difficulties raised above in the case of tightly wound two-armed spirals. The issue of three-armed spirals and other competing models will be discussed later (§ IV*c*).

a) Dynamical Models versus Real Galaxies

We begin by considering a number of essential points to be noted in the construction of simple models for galaxies.

i) *Stellar Dynamics*

In a first approximation, the thin disk would be best represented as a collisionless stellar system governed by the equations of *stellar dynamics*. Thus, in the simpler and more flexible fluid model, we may have to invoke a proper interpretation in the stellar context in order to avoid any unrealistic features. One important feature to be kept in mind is the role of the inner Lindblad resonance for the weakening or elimination of spiral modes with more than two arms. In general, an indiscriminate use of a fluid model can be misleading, since resonances, pressure anisotropy, and other subtle issues can be improperly overlooked. Some of these important effects can be easily included by a judicious use of *appropriate boundary conditions* in the modal calculations (see discussion by Lin and Bertin 1981).

ii) *Active Mass*

For dynamical studies we must keep in mind that the luminous bulge is essentially *inactive* and that there is also an inactive "halo" mass. Both contribute to the support of the observed rotation curve. Indeed, even a sizable fraction of the *observed disk mass* is also to be considered *inactive* because of the finite thickness of the disk. Thus the disk surface density σ of the *basic state* that will be perturbed in the *linear stability* analysis is in general *not* the same as the projected density of the observed disk. Some *inactive* mass would be counted as disk mass by observers and should *not* be counted as *active* disk mass by dynamicists studying the problem of stability.

To be more specific on the role of thickness, Shu (1971, especially p. 322) finds that in the solar vicinity the responsiveness of the gas component is larger by a factor of ~ 6 and that the reduction factor for the stellar component is ~ 0.57. Values for the reduction factor close to 0.5 are easily derived (see Vandervoort 1970; Yue 1982) when the vertical structure of the disk is taken to be of the kind reported by Bahcall and Casertano (1984) for NGC 891, NGC 4565, and NGC 5907. In addition, one should note that, for a disk with *constant* thickness z_0 and with approximately constant pitch angle of spiral structure, the reduction factor is smaller (i.e. less mass is active) at small radii, because the ratio z_0/r is larger.

Another point to keep in mind is that observations suggest that the disk thickness z_0 is constant only in the cases where the bulge contribution is small (see van der Kruit and Searle 1982). On the other hand, little is known about the structure

and, indeed, the very existence of the disk *inside* the bulge. It may well be that when a bulge is observed, it actually *replaces* the disk. If this is the case, the active mass inside the bulge is likely to be reduced to insignificant values (cf. Fig. 7 in Lin and Lau 1979).

A direct consequence of this discussion is that the disk of *active* mass, to be used in dynamical studies, can easily be lighter than the *observed* disk by a factor of 2; this reduction may have dramatic implications for the dynamics of the disk.

Even by adopting the maximum disk models that are compatible with a given rotation curve (see van Albada and Sancisi 1986), one may turn out to find relatively light disks for the dynamical context. For example, a model of NGC 3198 very close to the maximum disk solution (Begeman 1987; Casertano 1988, private communication) characterized by $R_M \approx 0.65$ would have $J \cong 0.83, 0.5, 0.45$ at $r = 2h, 3h, 4h$, respectively, if zero thickness is assumed; thus a very reasonable thickness correction can bring even the maximum disk model into the domain of normal spiral structure (see discussion of the J, Q diagram in Paper II). In addition, we note that the maximum disk solutions need not be realized as a general rule, since there is growing evidence for lighter disks in spiral galaxies (see, e.g., van der Kruit 1988; Kuijken and Gilmore 1988).

iii) *Scales*

Related to the above discussions is the problem of identifying the scales of the *dynamical* model with the scales of the *actual* system. For example, as the simplest first approximation for the mass distribution in observed disks, one can take an exponential law with the scale length h_* of the luminous disk. The mass distribution corresponds in sizable part to *inactive* mass. Thus it is reasonable to start out with equation (2.1) with an exponential disk with scale length $h = h_*$, but it should be emphasized that, especially since gas is more abundant outside, the *mass scale length* $h_{\rm eff}$ *of the active disk* is expected to be significantly larger than h_* and *the distribution of active mass is, in general, not exponential.* In addition, in the presence of a bulge the active mass in the center is highly reduced and often negligible (see specific choices in § IV*a*). Thus, as we often stressed, even though cutting the active disk mass by a factor $1 + \Delta$ evenly at all locations may often be the simplest way of reducing the disk mass in dynamical studies, distributing the mass reduction by cutting more at the center and less in the outer regions is the better choice for astrophysical applications. Indeed, this choice can make the whole difference in the results of dynamical studies of spiral structure.

iv) *Dispersion Speed*

Another crucial step is the proper choice of the equivalent acoustic profile $a(r)$ in the construction of a one-component fluid model. It is best discussed by focusing on the *equivalent Q-parameter.* (See following § III*b* for its proper definition.) Again we must stress that $a(r)$ does *not* represent the velocity dispersion of the stars, nor the turbulent speed of the gas component. Instead, in the one-component model, it is an *equivalent quantity* to be eventually chosen so as to represent the process of self-regulation in the more complex *multiple-component/multiple-thickness real system.* In turn, results from stability investigations of simple one-component systems should always be interpreted in view of the physical processes occurring in actual galaxy disks. Results of N-body simulations, for example, if taken at face value and not interpreted properly, can be completely misleading. In fact, N-body experiments naturally *overestimate the amount of heating* in the disk. On the one hand, their fluctuation level is usually larger than in the real system because of the small number of particles (see Lin and Bertin 1985); in addition, the role of the gaseous component is not included, and thus the important mechanism of self-regulation is ignored.

In deciding on the general structure of the Q-profile, one should keep in mind that in the central regions of a galaxy both the absence of cold gas and the transition in geometry from the disk to the nuclear bulge make it natural for the Q-parameter to *exceed* unity. Thus profiles of the kind chosen in § II*b* (see also Lau, Lin, and Mark 1976; Bertin *et al.* 1977) are not arbitrary functions. They actually represent a well-defined *physical choice* which, in our opinion, is the best representation of the actual astrophysical system. The only case where observations have provided a profile for the stellar dispersion speed is that of the Milky Way (Lewis and Freeman 1988); these data indicate the presence of a relatively cool disk with properties that are consistent with the picture that we have adopted.

b) Roles of the Gaseous Component: Self-Regulation

Extensive observational data are available on the amount of neutral hydrogen present in external galaxies (see, for example, Wevers 1984; Wevers, van der Kruit, and Allen 1986). It is clear from the relative distributions of gaseous and stellar masses present that the mass of the gaseous component becomes gradually more important as we move to the outer parts of the galactic disk. In these outer parts, the process of gravitational self-regulation discussed by Bertin and Romeo (1988) leads to the establishment of a condition of marginal instability, which bears directly on the parameter $Q_{\rm OD}$ to be used in the dynamical models.

With the perception of the choice of models discussed in § II*b*, let us consider the more specific determination of the parameters in the models through the dynamical process of self-regulation.

As already indicated (§ I), a major feature of the present investigation is to focus on *moderately unstable modes* and therefore on systems where spiral structure is self-excited but where violently unstable modes are not present. *In particular, at low values of disk mass why should the outer galactic disk be characterized by a value of Q so so close to unity?* In general, the question is why real systems should actually conform to the conditions of moderate growth, as we shall describe more precisely in Paper II. Here we focus our attention on the specific issue of cases of low-disk masses.

The physical process by which this situation is realized we call *self-regulation.* This process is expected to be *more effective* where the parameter region of moderate growth is *narrower*, as is the case of low disk mass systems. The key ingredient of this process is gas (and, to some extent, the presence of low dispersion stars).

i) *Shocks*

One role of the gaseous component is to provide dissipation and therefore a saturation mechanism at low amplitudes for the growing spiral modes (see Kalnajs 1972; Roberts and Shu 1972; Shu 1985; Lubow, Balbus, and Cowie 1986; Lubow 1986). Shocks in the gas trace the smooth underlying spiral field by sharp optical features. Thus one role of the gas is to *regulate* the otherwise exponential growth of spiral modes.

ii) *Excitation*

A second crucial role of the gas, also well known (see Lin and Shu 1966; Lynden-Bell 1967; Ostriker 1985), has been

recently further clarified by the analysis of Bertin and Romeo 1988 (see especially their Figs. 5 and 9). The existence of a gas disk in real galaxies *reduces* the value of the *equivalent* Q-parameter in the model.

For convenience of reference, we give a brief description of the definition of the equivalent Q-parameter defined by Bertin and Romeo (1988). To characterize the relationship between the gaseous component and the stellar component, we introduce two ratios: a density ratio α and a temperature ratio β (i.e., the ratio of the square of the velocity dispersion in the gas to that in the stars). For any combination of α and β, it is possible to determine a condition of marginal instability with respect to axisymmetrical disturbances. For the condition of marginal stability, the usual Q-parameter for the *stellar* component is then called $\bar{Q}(\alpha, \beta)$. (See Fig. 5, Bertin and Romeo). The equivalent Q-parameter for the combined disk is then defined by

$$Q_{\rm eq} = Q_*/\bar{Q}(\alpha, \beta) < Q_* , \tag{3.1}$$

where Q_* is the actual Q-parameter defined for the stellar component alone. Then, as described with a simple model by Bertin and Romeo (1988), a process of self-regulation can be easily ensured (see § III*d* of their paper). It can be traced to a proper balance of the following physical processes: (i) the cooling of the interstellar medium by turbulent dissipation; (ii) the conversion of gas into stars; (iii) the heating of the interstellar medium by the young stars; (iv) the dynamical increase in the dispersion speed of the stars. It is found that the system can indeed *self-regulate* because if the *equivalent* Q exceeds marginal stability, then the cooling process (i) rapidly reduces the value of the *equivalent* Q, and if too much cooling occurs, then dynamical instability, via (iii) and (iv), brings back the system to higher values of the *equivalent* Q-parameter. Note that the presence of gas is crucial, since it is subject to fast cooling processes in contrast to the stellar component which is subject to perennial heating only.

Note that this discussion is likely to hold even when the gaseous density is locally only one-third or even less than that of the stellar component. In fact, for each individual component, the value of the Q-parameter may be quite high, since the reference mass density in each case is lower than the total. Thus, it can be as high as 1.7 for the stellar component and 2.4 for the gaseous component, even though the effective value for the combined system is found to be close to unity. However, the calculations of Bertin and Romeo (1988; see Figs. 3, 4, 5 of their paper) also show that if there is too little gas (e.g. $\alpha \lesssim 0.1$), the process of self-regulation may not always be effective. In that case the turbulent velocity in the gas may be so low compared to the stellar velocity dispersion that the two components may no longer be coupled to each other.

As a result, it is of primary importance to estimate the values and distribution of $\alpha(r)$ in specific objects. We have checked in a number of cases that the role of gas is definitely significant. In particular, we refer to the data reported by Bertin and Romeo (1988, their Fig. 7) for reasonable models of NGC 4565 and NGC 5907. Similar values for α can be derived for the Sb spiral NGC 4258 (Wevers 1984) which has $\alpha \approx 0.17$ at $r = 3h_*$. For the maximum disk model of NGC 3198 mentioned above (Begeman 1987) one finds $\alpha \approx 0.11$, 0.31 at $r = 3h_*$, $4h_*$, respectively. However, all these numbers *heavily underestimate* the values of α to be used, since the gas density is taken to be that of *atomic hydrogen* and the stellar disk is taken to have *zero thickness*. Most likely the above values of α should be multiplied by a factor of 4 because of (i) the presence of molecular hydrogen (see also discussion by Allen 1987), (ii) cosmological helium, (iii) thickness corrections for the stellar disk. Furthermore, there is yet *no compelling* evidence for the maximum disk solution, so that the numbers for NGC 3198 might be increased further.

To be sure, the above galaxies are Sb's and Sc's. We have not been able to find convincing data for Sa objects except for one case with a small amount of atomic hydrogen by Warmels (1988). However, the above discussion makes it clear that one should not be misled by the small numbers that are often given for the *total* amount of atomic hydrogen in disk galaxies. For example, the ratio of the total mass of atomic hydrogen to total mass of the galaxy within $4h_*$ for NGC 3198 is just 2.8%. Still the local value of the *density* ratio α at $3h_*$ may well be close to 50%. (Recall that $e^{-3} = 0.05$, and thus the stellar mass density is a very small fraction of that at the center.) For Sa spirals, there is in general a larger nuclear bulge, and hence the corotation circle of the mode is expected to lie further out in the galactic disk in a region where gas is relatively plentiful (see § IV).

As long as the two components are effectively coupled, the process of self-regulation maintains the Q-parameter very close to unity since the stability characteristics are *extremely* sensitive to a change of Q. A little calculation shows that instability is very high when Q is 0.95. In the survey by Thurstans (1987) modes with $Q_{\rm OD} = 0.9$ are included. These show very high growth rates.

The discussion of this subsection has indicated the conditions under which a dynamical model with $Q_{\rm OD} = 1$ is physically justified.

iii) *The Scales of the Active Mass and of the Q-Profile*

A third role of the gaseous component is its impact on the large-scale structure of the modes through its influence on the *mass* distribution in the *active* disk (§ III*a*[iii]). In this perception, the outer part of a galactic disk is essentially gaseous, with $Q \approx 1$. This gaseous disk extends inward to where the stellar component begins to appear, where the condition $Q \approx 1$ continues to hold. The Q-parameter then rises gradually inward. The *scale* of the active disk is therefore *longer* than the *scale* of the Q-distribution which is determined largely by the stellar disk. Clearly, this role of gas is more important if the active disk mass is smaller, i.e., when the pitch angles are smaller, since the thickness correction is more important for shorter wavelengths. But even for open spirals, there is a significant reduction.

In our modal survey of § II, tightly wound normal spirals were found only for the case $r_Q \approx h/2$, not for the cases $r_Q \approx h$ ($r_Q = 4$, 5, 6); only barred spiral modes are found in those cases. As mentioned above, in the case $r_Q = h/2$, and if h is essentially identified with the scale length h_* of the stellar disk, the disk is still cool where the surface density is still considerable, and is most likely to be pure stellar. Indeed, the above discussions suggest that it is only reasonable to choose $r_Q \approx h_*$, (e.g., $3 \leq r_Q \leq 6$ for $h = h_* = 4$), since the stellar disk should be an integral whole with a single scale for both mass and velocity dispersion. In that case, the type of survey scheme described above *yields only barred spirals*. (See survey A_2.) The resolution of this difficulty is to be found in the inclusion of the mass of the gaseous component and the proper three-dimensional distribution of matter (see § IV) in modal calculations.

In conclusion, let us repeat that the explicit inclusion of the gaseous component may be expected to produce significant differences only in those cases where the stellar disk has a low active surface density, since the density of the gaseous disk is limited. Thus, in models of barred galaxies with high disk mass and high velocity dispersion, the explicit inclusion of the gaseous component is not expected to play a highly significant *active* role in the excitation of the spiral mode. The *passive* role of the gas would, of course, be important for the configuration of dust lanes and H II regions.

IV. NORMAL SPIRALS WITH SMALL PITCH ANGLE

The study of § IV gives us the essential tools that are necessary in order to resolve the difficulties for tightly wound normal spiral modes that have emerged from the exploratory survey of models presented in § II. In this section, by means of explicit examples and quantitative data we show that, by properly incorporating the roles of the gaseous component and of the three-dimensional distribution of matter, we get models where the corotation circle is found to occur at ~ 3 times the length scale of the exponential disk. At such a large distance from the center, gas is sufficiently plentiful to control the equivalent stability parameter Q and thus to provide for the excitation of the spiral mode.

a) Simple Models for Early-Type Normal Spirals

Normal spiral galaxies of the early type (Sa's and Sb's) possess a sizable bulge component. In the modeling process, the presence of a bulge has three major effects: (i) a modification of the active density profile as measured by the parameter J, (ii) an impact on the profile of the stability parameter Q, and (iii) a possible modification of the shape of the rotation curve. For the present purposes, we have chosen to minimize the introduction of new parameters. Therefore we have tried to operate within the family of models presented in § II (see eq. [2.1]) by suitable parameter variations. In practice, item (iii) will only be marginally considered in the following, item (i) will be incorporated by a variation of $r_{\rm cut}$, and item (ii) by variation of r_Q.

To be more specific about the choice of the active mass density profile, since little is known of the disk structure inside the bulge region and sizable departures from the exponential law are often observed (see Kent 1984), we have two natural options: either the exponential disk is actually *replaced* by the bulge in the central region, or it *coexists* with the latter, although dominated by the bulge. In the following, we shall adopt the former picture that allows us to leave the shape of the rotation curve substantially unchanged and to act on very few parameters, essentially $r_{\rm cut}$. The second option should also be explored, but we postpone a more complete analysis of the role of the bulge to future studies, when we deal with specific objects with a well-defined set of observational constraints, such as M81. Thus we are going to consider sizable reduction and large values of $r_{\rm cut}$ over the basic exponential disk studied in § II. Roughly speaking, for a given choice of the reduction factor f, one should interpret f as simulating the effect of *thickness* as far as $f \gtrsim \frac{1}{2}$ and *replacement by the bulge* where $f < \frac{1}{2}$. Notice that the parameter $r_{\rm cut}$ is not representative of the scale of the bulge. Even when $r_{\rm cut}$ largely exceeds h (see Fig. 2), the transition region (from bulge to disk) defined as the location where $f \approx \frac{1}{2}$ tends to occur at r in the range h to $2h$. Within the picture that the bulge actually replaces the disk, the smooth transition implied by our choice for the f-function is natural and sensible. However, the *analytic* form for f is just a convenient choice that may be revised if necessary and if specific observational constraints are considered.

The roles of gas are incorporated by "turning on" the σ_g-contribution to the active mass density of equation (2.1) in our *equivalent* one-component disk and by imposing the self-regulation prescription proposed by Bertin and Romeo (1988) that justifies the use of a cool disk with an *effective* Q close to unity (i.e. $Q_{\rm OD} = 1$; see § III). For simplicity, we refer to models with constant σ_g (see Fig. 6) and with uniform velocity dispersion associated with the gas component. This choice for the gas distribution may be justified if the corotation radii of the modes obtained are of the order of $3h_*$, since the gas typically extends out to $4h_*$ and beyond (see Wevers 1984; Wevers, van der Kruit, and Allen 1986, esp. Fig. 7; van Albada and Sancisi 1986) where the stellar density is negligible. As we shall see, the main part of the mode obtained extends approximately over the part of the disk where $h_* < r < 3h_*$.

In our models with $Q_{\rm OD} = 1$ the *cool outer disk* may be identified with the region outside $r = 2r_Q$, since $Q(2r_Q) \approx 1 + 0.02q$. Note that, if we take $r_Q = h_*$, the (total) stellar density in the disk at $r = 2r_Q$ is only 14% of the extrapolated central density of the exponential disk, i.e. $0.14\sigma_0$. Thus if we take $\sigma_g = 0.02\sigma_0$, the local density ratio $\alpha = \sigma_g/\sigma_*$ at $r = 2h_*$ is already quite large ($\approx 15\%$). This number can be increased further when the reduction factor f is taken into account (see also discussion of observational data in § III). Larger values for r_Q also enhance the significance of the gas component, since the dynamically active region is moved further outward.

A detailed description of various models and modes within the framework outlined above has been given by Lowe (1988). Here we report a few significant results. A sequence of models has been surveyed, characterized by $r_\Omega = 0.625h_*$, $Q_{\rm OD} = 1$, $q = 1.5$, $r_Q = h_*$, $r_{\rm cut} = 6h_*$, variable Δ and variable σ_g. In Figure 7*a* the two lowest (i.e. fastest rotating) two-armed modes for the case $\Delta = 7\%$, $\sigma_g = 0.02\sigma_0$ are illustrated. The first mode has the corotation radius at $r_{\rm co} = 2.5h_*$ and is characterized by moderate growth rate $\gamma P = 0.48$; the second mode extends further out, with $r_{\rm co} = 2.9h_*$ and has also a moderate growth rate, with $\gamma P = 0.30$. The values of the J-parameter at corotation are $J_{\rm co} = 0.45$ and 0.44, respectively. In the cool outer disk that has been assumed, the value of the gas to star density ratio is sufficiently high. We note that a superposition of these two modes gives rise to a quasi-stationary, regular, two-armed structure that persists for more than one rotation period, because there is only a slight difference in the angular speeds and in the pitch angles between the two modes. No other important modes are found in this model. The reasons for the absence of other competing modes will be addressed in § IV*c*.

Probably the most satisfactory case, although it has been obtained for a somewhat high overall profile of the α parameter, is illustrated in Figure 8. This could serve as the prototype for models of grand design spirals of the Sb type. Here the model is characterized by $r_\Omega = 0.375h_*$, $Q_{\rm OD} = 1$, $q = 2$, $r_Q = 4h_*$, $r_{\rm cut} = 6h_*$, $\Delta = 7\%$, $\sigma_g = 0.02\sigma_0$. This model is subject to one dominant mode (shown in Fig. 8) characterized by $r_{\rm co} = 2.5h_*$ and $\gamma P = 0.39$; the value of the J parameter at corotation is $J_{\rm co} = 0.47$. For this model, all other modes are irrelevant since they either are of much lower growth rate (the second two-armed mode has $\gamma P = 0.03$) or they are expected to be suppressed by inner Lindblad resonance. We note that at $r = 2h_*$ this model would be consistent with the corrections for

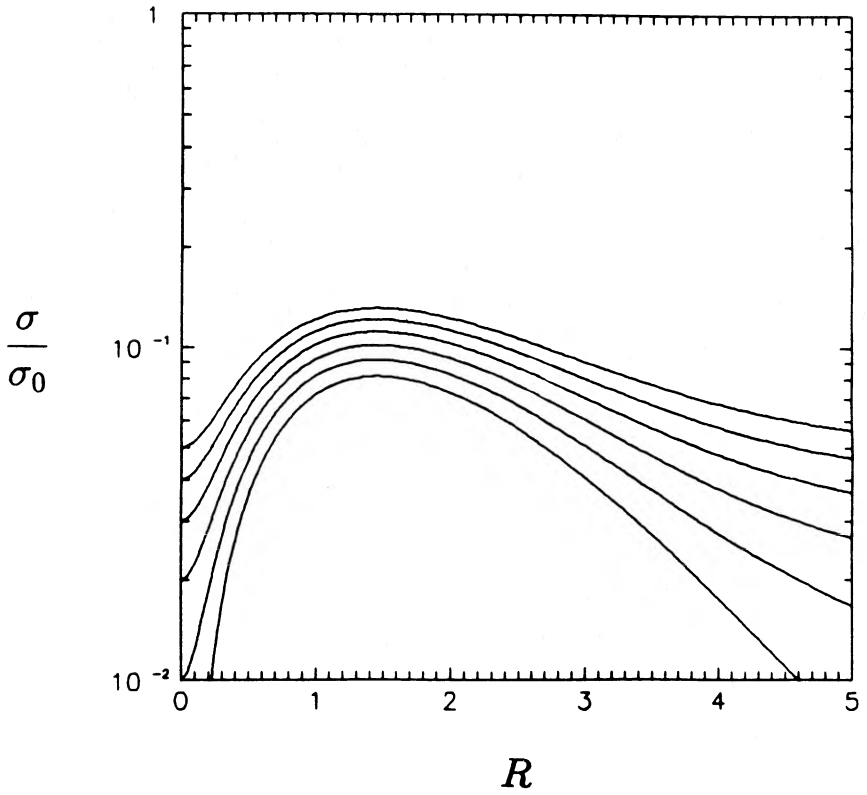

FIG. 6.—The effect of the addition of a flat density distribution (gas component) on the activity density profile for the case $r_{cut} = 6h_*$, with σ_g ranging from 0 (*bottom*) to $0.05\sigma_0$ in equal steps.

thickness, gas content, and stellar dispersion speed that are estimated for the Milky Way in the solar vicinity. This model also serves to illustrate an important point regarding the Q distribution: in the presence of a cool gas layer, instability is possible in a composite stellar/gaseous disk even when the stellar component is stable. In the paper by Bertin and Romeo (1988), the relationship between the mass densities, dispersive speeds, and stability parameters that connect a two-component disk with an equivalent one-component model have been worked out. By applying their prescription to our case, we obtained the data shown in Table 1 and Figure 9. Note that the Q-profile employed in the mathematical model used in the stability analysis is the *effective* stability parameter for the equivalent one-component system, and that Q_* is derived from it through the prescription mentioned above. As can be seen in the table, the composite equivalent system is at marginal stability ($Q = 1$), even though the stellar disk is quite stable. Indeed, Q_* is quite high ($Q_* > 1.6$) and rises in the outer disk.

TABLE 1

A TWO-COMPONENT MODEL

R	f	α	J	Q	Q_*	β
1.5.....	0.37	0.24	0.36	1.21	1.80	0.26
2.0.....	0.54	0.27	0.45	1.04	1.60	0.24
2.5.....	0.69	0.35	0.47	1.00	1.71	0.22
3.0.....	0.81	0.49	0.45	1.00	1.99	0.22
3.5.....	0.90	0.73	0.41	1.00	2.47	0.24
4.0.....	0.96	1.14	0.37	1.00	3.28	0.25

b) Reinterpretation and Rescaling of the Original Survey of Models

An examination of the active density profiles for the models constructed in § IV*a* shows that the mass distribution is actually reasonably well approximated, in the dynamically active region, by a declining exponential having a scale length $r_\sigma = h_{\rm eff}$ *considerably longer* than the scale h_* of the assumed stellar distribution (see Fig. 10). This conforms to the general arguments outlined in § III*a*(iii), but now we can be more specific, since for the relevant models we find $h_{\rm eff} \approx 2.5h_*$, and $r_Q \approx h_*$. This suggests the possibility of recovering many of the results obtained in our basic exploratory survey of § II, provided a proper *reinterpretation* is given. In particular, since $h_{\rm eff}$ is the scale pertinent to the computational model, the exponential scale length h used in the exploratory survey of § II *could be interpreted as* $h_{\rm eff}$, *not* h_*. Therefore, those models that were subject to modes with corotation radius $r_{co} \gtrsim h$ could be reinterpreted to have $r_{co} \gtrsim h_{\rm eff} \cong 2.5h_*$, a value more in line with observations. Thus, results of the exploratory survey, initially rejectable as lacking physical justification, can be applicable provided proper rescaling is made.

In order to quantify and to confirm this important point, a calculation was carried out using an exponential disk to mimic one of the models discussed in § IV*a*. The model considered had $\sigma_g = 0$, $r_\Omega = 0.625h_*$, $h = 2.5h_*$, $r_{\rm cut} = 3h_*$, $r_Q = h_*$, $Q_{\rm OD} = 1$, $q = 1.5$, and $\Delta = -75\%$. The pattern speeds and growth rates for the two lowest modes in this exponential model, which are shown in Figure 7*b*, differ insignificantly from those of the corresponding modes shown in Figure 7*a* in the model described in § IV*a*. As can be seen by comparing the density contours, the patterns are also remarkably similar. The

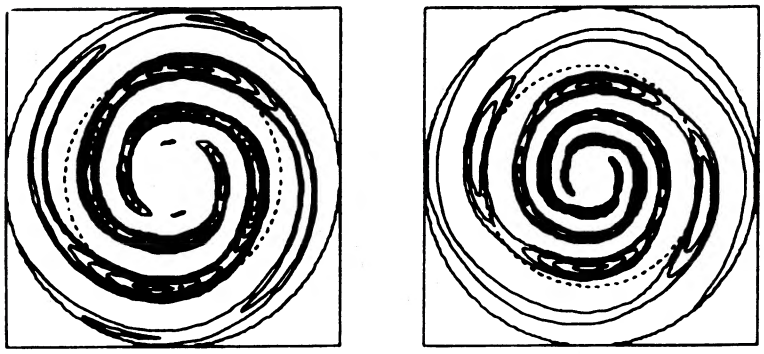

FIG. 7a

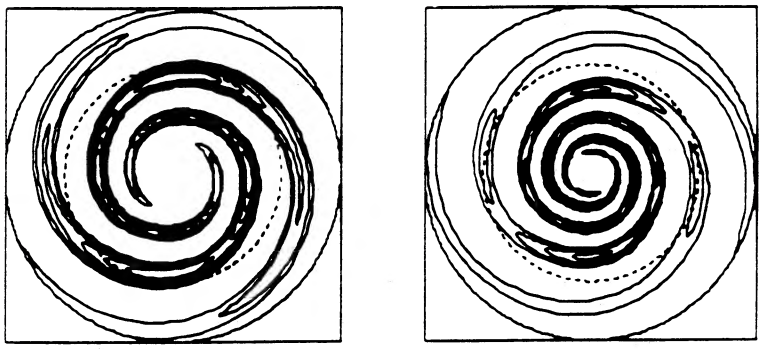

FIG. 7b

FIG. 7.—Examples of normal spiral modes with small pitch angle and with large corotation radii. (*a*) The two dominant modes are displayed for a model characterized by $r_\Omega = 0.625h_*$, $\Delta = 7\%$, $r_{cut} = 6h_*$, $\sigma_g = 0.02\sigma_0$, $r_Q = h_*$, $q = 1.5$, and $Q_{OD} = 1$. The corotation circles (*dotted circles*) are at $r_{co}/h_* = 2.5$ and 2.9, respectively. (*b*) The two dominant modes are displayed for a model characterized by $r_\Omega = 0.625h_*$, $h = 2.5h_*$, $\Delta = -75\%$, $r_{cut} = 3h_*$, $\sigma_g = 0$, $r_Q = h_*$, $q = 1.5$, and $Q_{OD} = 1$. The eigenvalues of these two modes are essentially identical to those of (*a*). This shows that many of the normal spiral modes of the survey of § II can be recuperated for astrophysical applications, provided proper rescaling of the basic state is made.

simultaneous existence of two (or more) similar modes would lead to a spiral structure whose corotation circle is hard to ascertain empirically (see Binney and Tremaine 1988, p. 391), but there is clearly no difficulty to the determination of an angular velocity sufficiently accurate for an approximate description of the spiral structure as a whole (values of $\nu(r)$ not much changed between the two modes).

The above discussion shows that the introduction of the density contribution for the gas component (eq. [4.1]) does not invalidate the earlier modal calculations of the survey of § II for a cool galactic disk, but a different interpretation has to be given to the models for normal spirals. The length scale of the mass distribution $\sigma(r)$ should refer to that of the total *active* mass distribution. The length scale h_* of the observed stellar disk itself does not appear in the final model, but its value is approximately given by r_Q, the length scale of the Q-distribution. The corotation circles of the spiral modes are then found to lie in the gas-rich region.

c) *Three-armed Modes and Higher Two-armed Modes*

In order for regular structure to be a common feature in observed galaxies, there must be modes in which only one or two modes are dominant; indeed, such modes must be fairly natural. We have examined the cases of (1) higher two-armed modes (those that are slower rotating and thus have larger corotation radii), and (2) three-armed modes. In both cases, the inner Lindblad resonance plays a crucial role in suppressing these competing instabilities. Specifically, by examining the propagation diagrams we can determine whether the short trailing wave propagating inward from corotation reaches ILR, in which case Landau damping will absorb the density wave, thus suppressing the mode. For several cases we have carried out such an examination, and we have found that all the three-armed modes and all but one or two of the two-armed modes are suppressed in this way.

That this is a natural occurrence is best exhibited by considering the dispersion relationship. As shown in Bertin, Lin, and Lowe (1984) and Paper II, the local dispersion relationship can be reduced via a similarity transformation to a form in which the dimensionless Doppler-shifted frequency ν does not appear. Relevant to the present discussion is the relation

$$Q_0 = Q(1 - \nu^2)^{1/2} ,$$

where Q_0 is the similarity variable. In the low disk mass regime, wave propagation is permitted if Q_0 is smaller than ~ 1. Thus, values of $|\nu|$ approaching 1 *encourage* wave propagation by decreasing Q_0. Two-armed modes having larger corotation radii are, therefore, more vulnerable to ILR because of the central rise for the quantity Q occurs where ν is closer to -1 (i.e., farther from corotation where $\nu = 0$). For three-armed instabilities ν changes more rapidly with r than is the case for two-armed modes, so three-armed modes also tend to reach ILR easily.

d) *Concluding Remarks*

We have examined normal spiral modes from a global point of view, beginning with an emphasis on the role played by the nuclear bulge. On the other hand, the excitation of spiral struc-

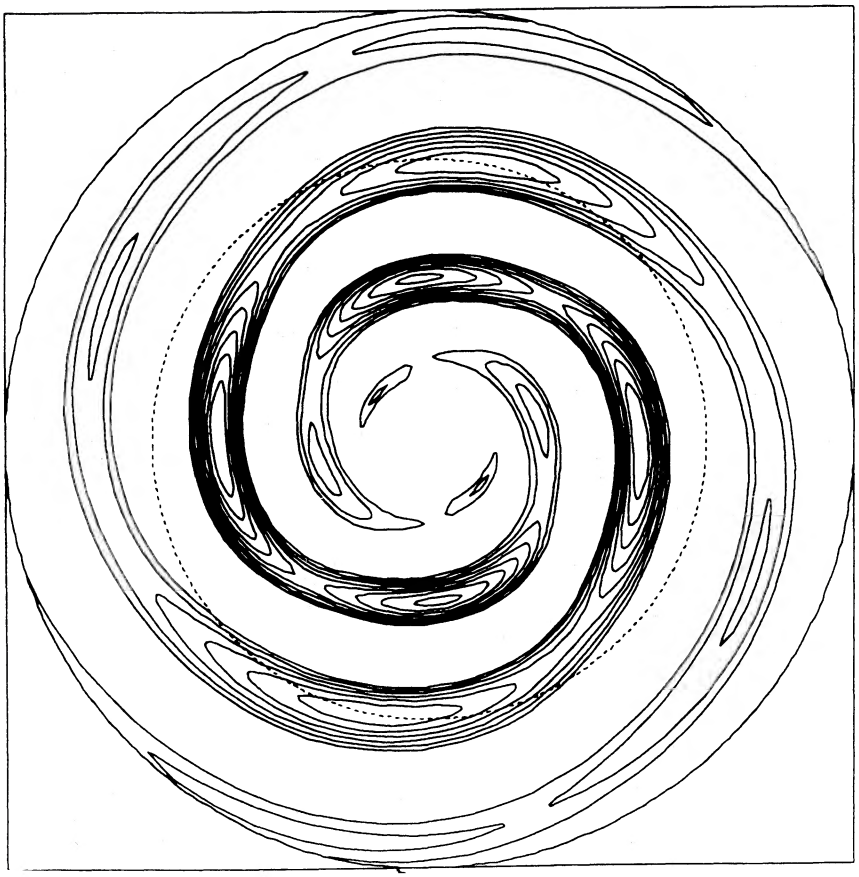

FIG. 8.—Multiwinding spiral mode. This particular mode is outside the main surveys presented in § II. The basic parameters are $Q_{OD} = 1$, $r_Q = h_*$, $\Delta = 7\%$, and $r_{cut} = 6h_*$. This model, with a major change in density distribution, represents a case where the active disk essentially coincides with a thin Population I layer. The interesting properties of this mode are its multiple winding and its radial extent (the dotted circle is at $r_{co} \sim 2.9h_*$).

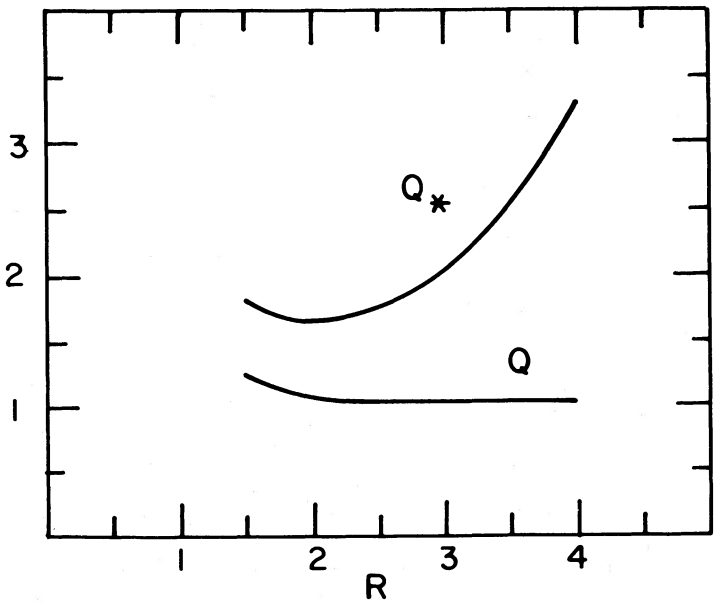

FIG. 9.—Behavior of the effective Q-profile and of the corresponding parameter Q_*, defined for the stellar component alone, for the model subject to the spiral mode displayed in Fig. 8. The stellar disk may appear to be very hot even when the disk is effectively cool.

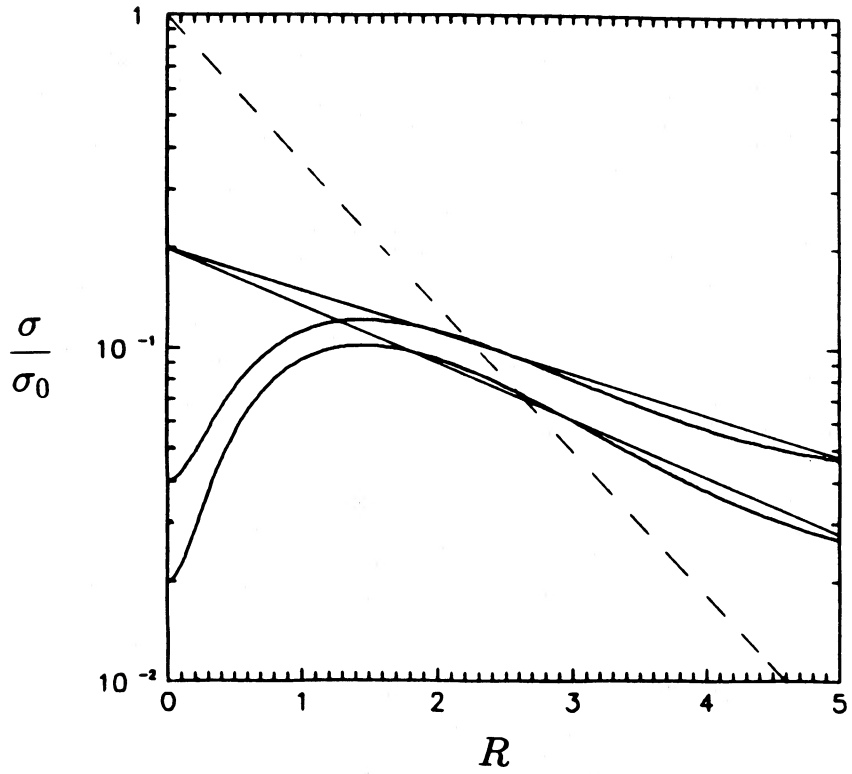

FIG. 10.—Active disk mass density profiles compared to the basic exponential disk (*dashed line*). The two curves correspond to $r_{cut} = 6h_*$ for $\sigma_g = 0.02\sigma_0$ (*lower*) and $\sigma_g = 0.04\sigma_0$ (*upper*). Straight solid lines represent exponential distributions with length scale $2.5h_*$ (*lower*) and $3.5h_*$ (*upper*). This figure illustrates how the scale of the active disk, because of the role of gas and geometry, is expected to be significantly larger than the scale of the optical disk.

ture clearly originates in the region where gas is relatively plentiful. Indeed, if one first focuses attention on this outer region, one can always determine a reference location in a galaxy where the ratio $\alpha = \sigma_g/\sigma_*$ is ~ 0.4. As we move inward from this location by approximately one scale length h_*, we have $\alpha \approx 15\%$, and we would expect the Q-parameter to begin to rise and to form a Q-barrier. The wave generated in the cool region around the reference location will be turned back to complete a feedback cycle and form a mode. However, it is now essential to emphasize the three-dimensional distribution of matter, since a lowered mass in the central region is necessary (see survey B_1 and § IV*b*) for the formation of a normal spiral mode. A central point at issue is really this: what fraction of the normal spirals observed contains enough gas, in appropriate amounts in the various parts of the galactic disk, so that spiral modes may be excited through gravitational instability? The resolution of the winding dilemma—i.e., the usefulness of the pitch angle as one of the criteria for Hubble classification—would suggest that there is enough gas in a *statistical majority* of such galaxies. An inspection of the gas content in many specific galaxies has indeed shown that this is likely to be the case.

The solar vicinity deserves special attention since it is the only location where the mass density can be independently determined without using the uncertain values of mass to luminosity ratio. Recently Kuijken and Gilmore (1988) found that the total surface density of the disk mass might be as low as 44 $M_\odot$ (pc)$^{-2}$. While we are not in a position to evaluate such analyses, we should examine its potential implications on our theory if there is such a large reduction in surface density. It turned out that our general conclusions remain unchanged.

For the reduction in density indicated, the value of α may be higher than 0.5. If we refer to Figure 5*b* of Bertin and Romeo, we see that this is more than sufficient to cool a galactic disk with $Q_* = 2$. Thus, even though this reduction of mass density would raise the value of Q_* from that given by Lewis and Freeman (1988), we are still dealing with a cool galactic disk, in terms of the effective Q. The relatively high density in the gaseous component also implies a longer scale for the total mass distribution, as can be seen from Figure 10. This provides an even stronger basis for the excitation of unstable spiral modes with large corotation radii.

V. MORPHOLOGY, REGULARITY, EVOLUTION

The main theme of our study has been the identification of basic states that support spiral structure of a given shape. A large effort has been devoted to the issue of the astrophysical justification of the models used. In turn, it seems that *all* the relevant morphologies have been covered by our survey. Here below we illustrate how these results can be made to correspond to the Hubble diagram. Section VI will focus on some important issues that have to be addressed in the future.

In making comparisons with observations, we should keep in mind that in this paper we have been dealing with "typical" galaxies, as representative of a statistical majority of cases. In another perspective, we wish to give primary attention to grand design spirals (see Elmegreen and Elmegreen 1982 for the division of galaxies into 12 categories on the basis of regularity); flocculent galaxies may result in the same modal perception if there are several unstable modes or if there is irregularity in the distribution of the interstellar medium.

a) Correspondence with the Hubble Morphologies

A coherent picture of all the major morphological types encountered in our modal survey has been exhibited in Figures 3 and 4 (§ II). We recognize prototypes of SB0 spirals, SB(s) spirals, and S-spirals (normal spirals), with modes of transitional types in between these prototypes.

For application to observed galaxies in the Hubble morphological classification, we have in general to impose the condition of moderate instability (see (§ I*a*). We envisage that if the basic state has only modes of moderate instability ($\gamma P \lesssim 1$), these modes may be brought to general equilibrium through the damping effect of the gaseous component (see § III). The more gas there is in the galaxy, the larger the value of γP may be allowed. On the other hand, if a basic state has highly unstable modes, these modes are expected to grow rapidly and possibly lead to a change of the basic state itself into a more stable configuration. In the new basic state, the final mode observed will again be one of moderate growth which is brought to saturation through the damping mechanism associated with the gaseous component. Once a state of moderate growth is reached, the long-term evolution induced by spiral modes should be slow (see also Bertin 1983*b*).

i) *Prototypes of Modes*

In Figure 11, we show representative examples of all three prototypes of modes with moderate instability ($\gamma P \lesssim 1$), together with one normal spiral mode of very high growth rate. Clearly, the mode of high growth rate will evolve. Indeed, in such cases there are expected to be other modes with considerable rates of growth (see Haass 1983). We may visualize the following scenario for the evolutionary process.

The rotation curve and the distribution of disk mass should not be significantly modified during the evolutionary process, but the distribution of velocity dispersion is likely to be influenced, thus leading to a change of $Q(r)$. If the gaseous component plays an active dynamical role so that $Q_{\rm OD}$ remains at unity (see § III*b*), the high-growth Sc spiral will evolve into a morphology like the SB(s) spiral shown in the same figure. On the other hand, if the gaseous component does not play an active dynamical role, $Q_{\rm OD}$ may rise, and the final configuration would be more like an SB0 spiral. (If r_Q does not change, it will be precisely the SB0 spiral shown in the upper left of Fig. 11. If r_Q increases, the mode will be an SB0, but somewhat different from the one shown.) The dynamical properties of the modes shown are described in Paper II.

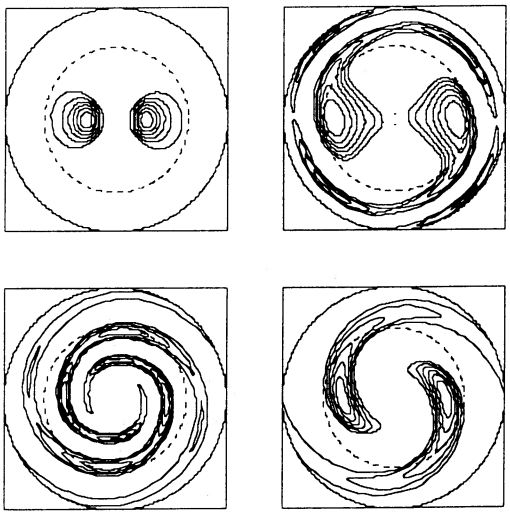

FIG. 11.—Mode prototypes. Four key morphological types are compared: SB0, SB(s), and S, all with moderate growth; a violently unstable S mode at the low right corner. The dynamical properties of these modes are discussed in Paper II.

ii) *Normal Spiral Modes*

Normal spiral modes appear to be excited only if the gaseous component plays an important role; their corotation circles are expected to occur in the gas-rich region where $r \approx 3h_*$. All the normal spiral modes exhibited in our exploratory survey of § II have only one turn up to the corotation circle; furthermore, the size of the spiral structure, as measured by the radius of the corotation circle, is less than two scale lengths of the exponential disk in all the normal modes calculated. The fact that galaxies are observed with spiral arms with a 540° turn indicates that for these cases the dynamically *active* disk is not the exponential disk itself. However, normal spiral modes of the exploratory survey can be recuperated by proper rescaling of the relevant basic states (see § IV).

iii) *Spiral Sequences and Transitions*

From the S-spirals and the SB(s) spirals in surveys A_1 and B_1 one may identify *one continuous sequence of spiral modes of moderate instability* as the disk mass is increased. Indeed, these may all be regarded as members of survey B_1, where $Q_{\rm OD} = 1$ and the two parameters Δ and r_Q are varied. The condition $\gamma P \lesssim 1$ selects such a sequence from the two-parameter field of modes. This is illustrated in the left frame of Figure 12. The zone along the boundary between the unstable regime and the stable regime corresponds to the conditions along this sequence.

The modal shapes and dynamical characteristics of this sequence are shown in Figure 13*a*. The parameter ranges are

$$-0.25 \leq \Delta \leq 0.15 ,$$

$$2 \leq r_Q \leq 6 .$$

There is clearly a continuous change of all the dynamical characteristics exhibited. In contrast, in the right frame of Figure 12, we exhibit conditions for barred structures SB0 and SBa; the latter morphology is expected when the gaseous component plays a passive role (see § II). At low disk masses, we expect no spiral structure (S0 galaxies).

Two important characteristics of transitions among normal spirals should be mentioned:

1. *Spiral structure in low-mass disks.*—In Figure 13*b*, we give further details for normal spirals for a slightly different set of basic states. These have a slightly modified distribution of surface density from the exponential disk. The three-dimensional distribution of mass in the disk is taken into account. In this set of modes, we note that there is a noticeable change of the propagation diagram between the cases $\Delta = -0.30$ and $\Delta = -0.20$, but there is no abrupt change of modal shape.

2. *The subdivisions Sa, Sb, and Sc.*—The primary physical parameters that mark out these subdivisions are the bulge size and the gas content. This is consistent with the dynamical characteristics of the spiral modes obtained. In addition, when more gas is present, the *active* disk mass is higher. At the same time, the *damping* mechanisms are stronger, thus allowing for a higher value of γP of the mode. Indeed, modes with higher γP

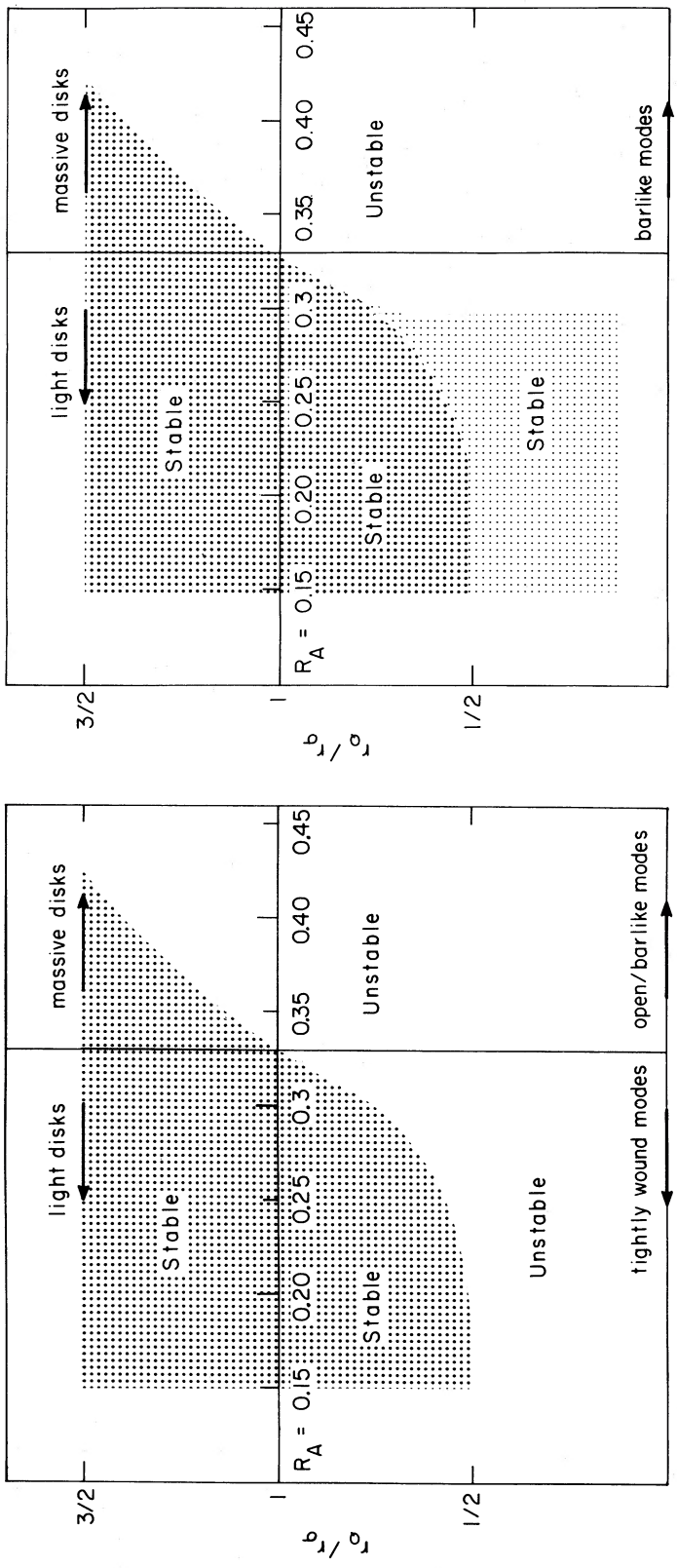

FIG. 12.—*Left:* Stability characteristics of global spiral modes in galactic disks in which gas plays an active role in excitation. Modes of moderate instability, corresponding to conditions along the zone dividing unstable and stable regimes, take on a sequence of shapes of tightly wound normal spirals, open spirals and barred spirals, as shown in Figs. 13*a* and 13*b*. *Right:* similar diagram for cases in which gas plays only a passive role. The modal shapes are SB0 and SBa spirals. S0 galaxies lack gas and have little mass attributable to the active disk. This figure summarizes the stability properties of modes found in the present survey of modes. The dynamical basis of this behavior will be examined in Paper II, especially Figs. 3 and 7.

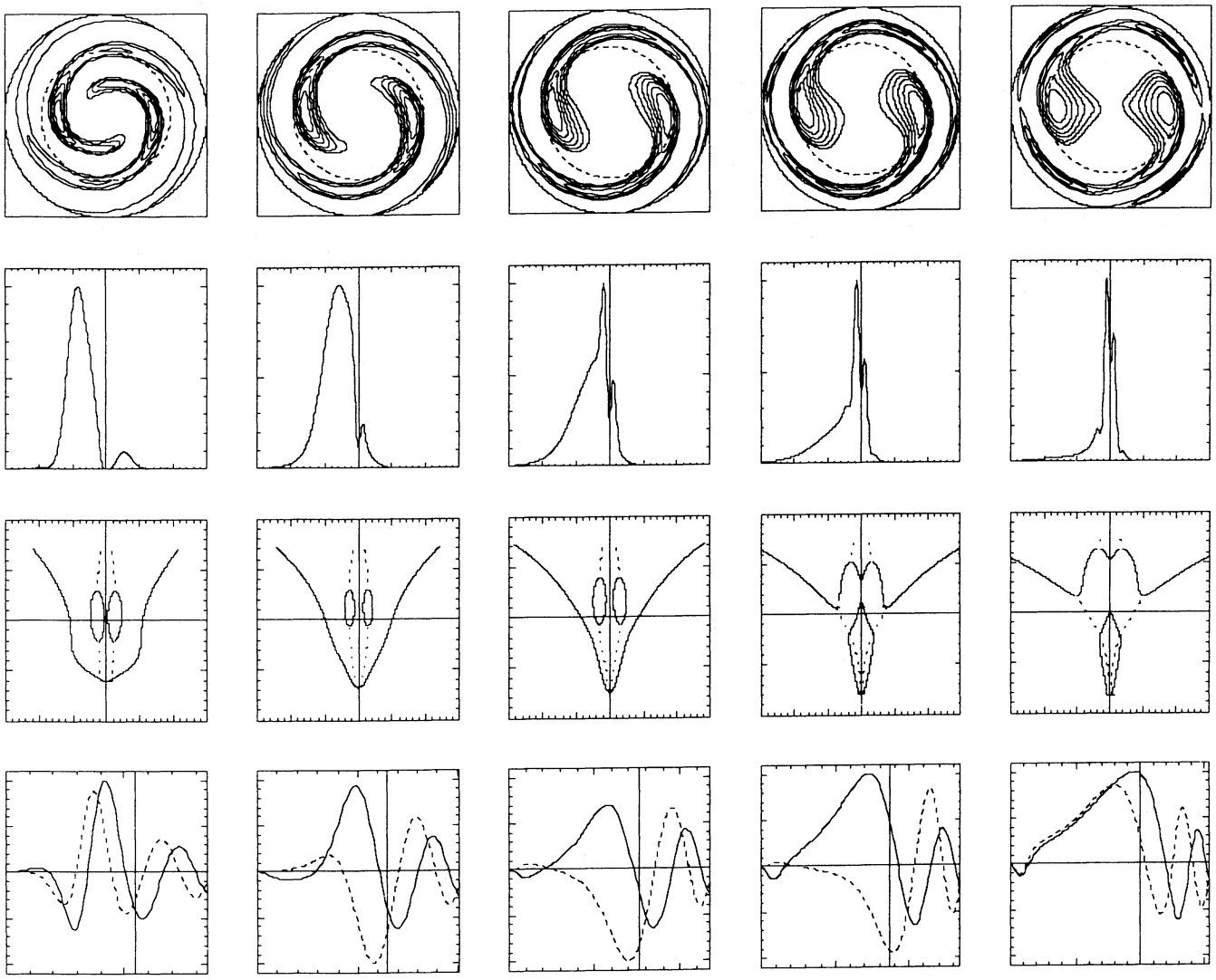

FIG. 13*a*.—The B_1 spiral sequence. Modes are selected from the B_1 survey (see § II) on the basis of the moderate growth criterion. For each mode we give model shape, α-spectrum, propagation diagram, eigenfunction. The α-spectra (*second row*) are given in arbitrary units; the horizontal axis has α/m from −15 to +15. The real and imaginary part of the mode eigenfunction (*fourth row*) are given as a function r; the vertical line indicates the location of the corotation radius. The propagation diagrams (*third row*) have ν on the vertical axis from −1 (*bottom*) to +1 (*top*); on the horizontal axis μ runs from −15 to +15. A key feature of this interesting spiral sequence is that outside corotation the mode is within the low-J regime of normal spiral structure, while inside it is in the high-J open wave regime. A complete definition and discussion of the available diagnostic tools that are used is presented in Paper II.

TABLE 2

DATA FOR FIGURE 13*a*

PARAMETER	$r_Q = 2.0$ $100\Delta = -25.0$	3.0 −15.0	4.0 −5.0	5.0 5.0	6.0 15.0
J_{co}	0.553	0.590	0.596	0.577	0.538
Q_{co}	1.00	1.01	1.04	1.07	1.10
Ω_p	24.23	20.62	17.83	15.62	13.81
γP	0.612	0.807	1.005	0.912	0.521
r_{co}/h	1.35	1.62	1.90	2.18	2.49

NOTE.—Each column corresponds to a column of Fig. 13*a*.

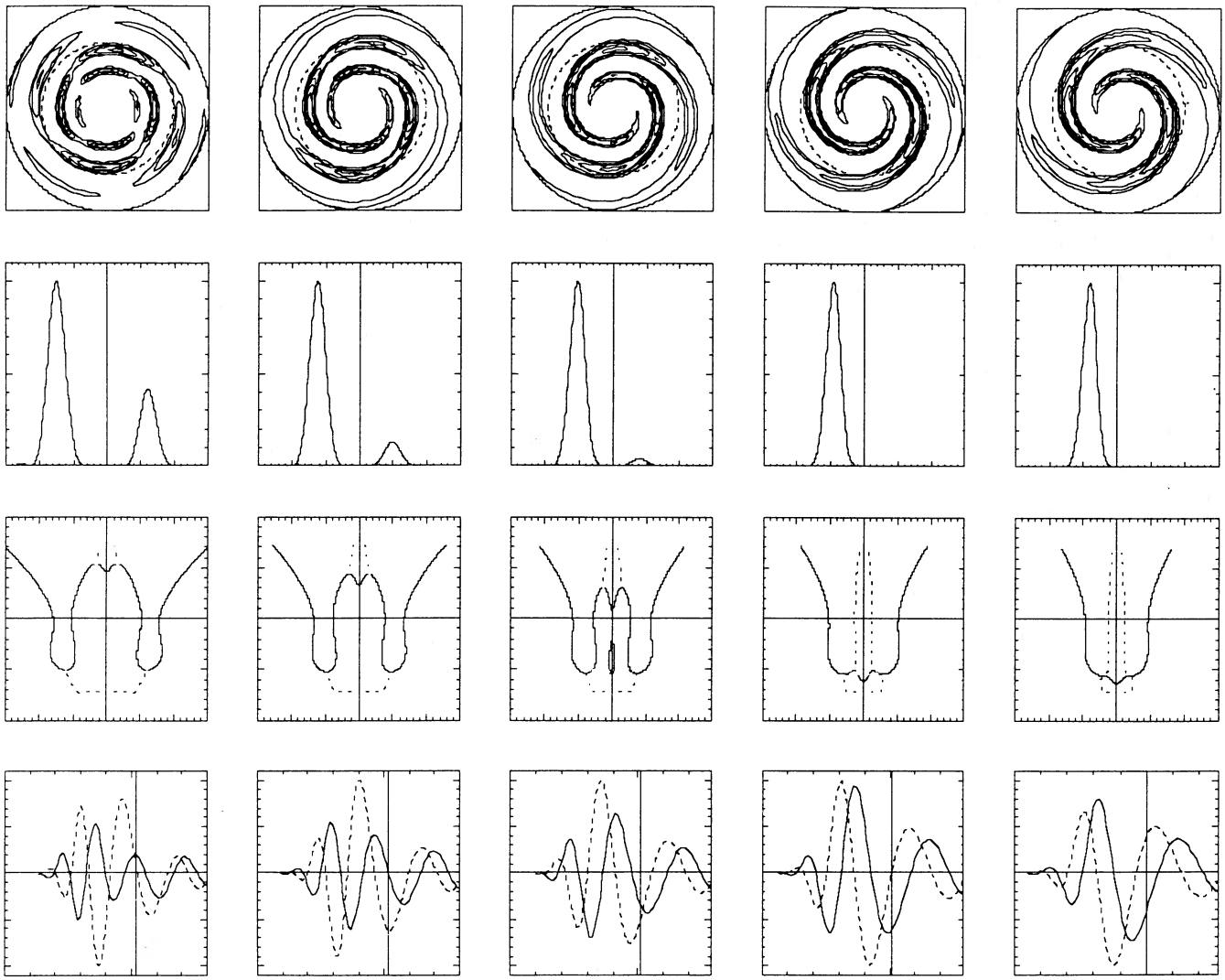

FIG. 13*b*.—The normal spiral sequence. This is a kind of continuation (in the low-density domain) of the spiral sequence of Fig. 13*a*, and it is presented in the same format. Two cases of relatively high J_{co} (high surface density) and high growth are included to better describe the continuous transition in the maintenance mechanisms. We should note that even if the propagation diagrams display abrupt, topological transitions, the actual mechanisms are expected to follow a relatively smooth change of character. Note that the long wave branches are expected to participate even when they become complex (dotted lines in the propagation diagrams).

TABLE 3

DATA FOR FIGURE 13*b*

PARAMETER	$Q_{OD} = 1.00$ $100\Delta = -50.0$	1.00 -40.0	1.00 -30.0	1.00 -20.0	1.00 -10.0
J_{co}	0.379	0.454	0.530	0.605	0.680
Q_{co}	1.00	1.00	1.00	1.00	1.00
Ω_p	26.05	26.09	26.20	26.42	26.81
γP	0.437	0.533	0.684	0.884	1.153
r_{co}/h	1.29	1.29	1.28	1.27	1.25

NOTE.—Each column corresponds to a column of Fig. 13*b*.

occur with higher disk mass, and they also exhibit a more open morphology. They are thus open Sc spirals.

iv) *Small Bars*

Modes with a "small" bar, mostly of a transition S/Sb type, were considered by Haass, Bertin, and Lin (1982) by keeping $Q_{OD} = 1$ and by lowering the central Q-barrier. This corresponds to lowering the q-parameter introduced in § II. The bar is called "small" because it affects only a small fraction of the disk inside the corotation circle. Note that, in the absence of a Q-barrier, one should pay special attention to the rotation curve, which should rise gently in such a way that ILR does not occur. This situation might be realized in some gas-rich, bulge-free S/SBc galaxies of the (s) type. These should be cases (possibly like M51 or NGC 7741) where the cool gas disk extends into the central regions. In order to have a better appreciation of the issues and of the possible morphologies involved, further explorations of cases with changing q are desired. Preliminary study with a model with relatively large r_Ω ($= h_*$), small r_Q ($= h_*/2$), and lower q ($= 0.2$) indeed leads to a spiral pattern which resembles that of M51.

In general, we find that the modal approach is a viable basis for the Hubble classification scheme. For two-armed spiral modes, we find a coherent total picture of the Hubble types as a one-dimensional sequence extending from tightly wound spirals to open spirals to barred spirals. This sequence is in the line of increasing ratio R_A of active disk mass to total mass derived within four length scales of the exponential disk mass distribution.

b) Regularity and Evolution

We would like to stress that the viability of the modal approach that has been demonstrated in this article goes well beyond the category of *grand design* spirals. In fact, grand design spirals are only a fraction of the total. The modal approach best applies to the set of galaxies where spiral structure occurs on a *large scale*, even if not so regular (such as M101). Indeed, in our *linear* modal analysis it is not easy to find basic states that support only one dominant mode. Most likely, because of the role of Landau damping at inner Lindblad resonance, the number of important modes is reduced to a small value, like two or three. But indeed there are isolated galaxies which appear to have a single *dominant* two-armed mode as shown by the empirical analysis of the spiral structure and rotation curve of the galaxy NGC 2885 (Roelfsma and Allen 1985), and so stated by these authors.

The importance of the modal approach is that it associates the morphological characteristics of the spiral structure to the *intrinsic* properties of the basic states. No need for external excitation is found (although processes like tidal interactions and gas infall could be in principle analyzed in terms of modes) if some modes are unstable. In addition, a superposition of these modes is likely to produce morphologies within the Hubble class identified by each of the modes present.

Therefore modes are a useful description of the morphologies observed at present. How should we expect spiral structure to evolve? The modal approach advocates that *the Hubble type* is expected to change, if at all, only on a very long time scale, thus resolving the winding dilemma. However, *the regularity* of spiral structure may well change on a shorter time scale. Most likely grand design and regular spirals are indeed a *long-lasting phase* of the system with a very small number of modes involved. However, the full basis of this assertion in the modal perception is to be found in the *nonlinear theory* (see detailed comments in § VI) both because of the role of gas and because of the nonlinear mode interaction that is anticipated. This does not mean that we have to wait for the full nonlinear theory before drawing any conclusion or making any comparison with observations. The encouraging correspondence between basic states, modes, and Hubble morphologies is an indication that linear modes are likely to provide a successful framework for *observed patterns*, much like in the theory and experiments of rotating fluids. This latter comment brings us back to the early empirical statement of the QSSS hypothesis, with the substantial progress added now by the identification of the proper basic state, or class of basic states, corresponding to a given morphology.

Therefore we can see the modal approach as a coherent framework which can be used to develop a number of *quantitative* models (see § VI). We note that, to our knowledge, no other approach has been tested for consistency at any comparable level of quantitative details, nor would it be available for so many quantitative predictions.

We should add that our confidence in the detailed dynamical mechanisms that are at the basis of the excitation and maintenance of spiral modes, which underlie the spiral morphologies surveyed in this paper, is based on the use of simple and powerful "propagation diagrams," which incorporate the role of excitation, inhomogeneities, feedback, and boundary conditions. This theoretical tool will be developed in Paper II.

The hypothesis of quasi-stationary structure (QSSS hypothesis) for regular grand designs has sometimes been misinterpreted as one of essentially stationary structure, i.e., that the spiral structure remains stationary over *many* orbital periods (see, for example, Binney and Tremaine 1988, pp. 337, 384). Naturally, such a restrictive interpretation can be expected to hold only in rare circumstances, e.g., when the system has only a single unstable mode, while the QSSS hypothesis is expected to hold over a much wider set of circumstances, contrary to the description given by Binney and Tremaine (p. 398). Over a time span of many orbital periods, the spiral pattern generally evolves in a quasi-periodic manner *without* change of morphological type.

c) Dynamical Instability, Morphology, and Dark Matter

After the pioneering work of Ostriker and Peebles (1973), attempts have been made recently to constrain the amount of dark matter in disk galaxies on the basis of studies of the stability of disks with respect to spiral modes (e.g., see Athanassoula, Bosma, and Papaioannou 1987; van Albada and Sancisi 1986). From the present article it is clear that we agree that some dynamical constraints on the amount of dark matter can be made, but only on the basis of the *observed morphology, not just on general stability grounds.* (In fact, the determination of the ratio of disk mass to total mass is part of the goals of our dynamical approach to the classification of spiral galaxies.) General stability arguments are thought to be insufficient since "stability" is automatically taken care of by the process of self-regulation at *any* ratio of disk mass (see Fig. 12). Still we do not believe that *simple* and *strong* constraints can be easily set, especially in the case of Sa's and Sb's. The main physical reason for this difficulty is that different physical ratios are involved in dynamical studies (e.g., the amount of gaseous mass to that of the active stellar mass) and in observational discussions of mass ratios. In general terms, a model that includes a spherical component, and a disk of zero thickness is

oversimplified for the purpose at hand. The reader is referred to our previous discussions in § III for other cautionary remarks on the subtle issue of the choice of the basic state in dynamical studies.

In conclusion, we note that for *any* value of the disk mass (e.g., even the case of a thin light disk in the presence of a large amount of bulge-halo matter) there is in general room for spiral instabilities of a gravitational nature. However, certain *morphologies* (such as open or barlike structures) are found only at certain values of the disk mass (as measured by the parameter J; see Paper II). Thus it is on the basis of the observed morphologies that we can try to put dynamical constraints on galactic structure. But in doing this we should be very careful, in view of the limitations of our one-component disk models.

A more promising approach in setting dynamical constraints on the amount of luminous matter seems to be a more direct correlation between observed photometry and kinematics of spiral structure (see again van Albada and Sancisi 1986) as an extension of the work by Visser (1977).

VI. OPEN ISSUES FOR FUTURE RESEARCH

The present study offers a unified approach to the morphology of spiral galaxies. A number of issues have been addressed in quantitative details. In Paper II we shall make a thorough examination of various dynamical questions. However, already from the physical discussion provided here, one clear limitation appears to occur in our study. In fact, we have often referred to the multiple-component nature of galaxy disks (see § III), but deliberately limited our discussion to an *equivalent* one-component model for simplicity of dynamical perception. This, of course, is just a first important step, but eventually global structures and modes should be investigated in a true multiple-component system (see Lubow 1986; Bertin and Romeo 1988). In addition, the process of self-regulation involves nonlinear aspects that require new analyses.

We believe that a natural direction for progress will involve not only new theoretical and numerical efforts, but especially a renewed interaction with present and future observations. The key missing item, we may suggest, is a clear perception of the physics of the interstellar medium and its interaction with the stellar disk. In a sense, after the long effort in studying the dynamics of large-scale spiral structure of the last two decades, we are back to the original motivation of the theory, i.e., the search for a framework for the study of the activity in the interstellar medium, especially the process of star formation, that is observed in coherent spiral arms. We note indeed that a similar point can be made from new observations of spiral structure (Lindblad and Jörsäter 1987; Allen 1988) that show notable discrepancies between the observed shock structure and the simple picture originally put forward by density wave theorists (Roberts 1969). Only a continual interaction between better and more specific observations and construction of more realistic dynamical models can generate progress in our understanding of the structure of galaxy disks.

We shall now conclude our article by bringing up certain specific topics where future progress may be made. The motivation for many of these issues is best indicated by the example of Figure 14. By superposition of a bar mode onto its basic axisymmetric state, we obtain a model that closely resembles some SB0 galaxies such as NGC 2859 (see Sandage 1961, especially the description of $SB0_2$ galaxies on p. 22). However, in constructing this model a finite amplitude for the mode was used, and this is beyond the present linear theory of modes. In addition, in the presence of some amount of gas the same dynamical structure of Figure 14 might originate a model that could resemble the SBa galaxy NGC 2217 (see Sandage 1961).

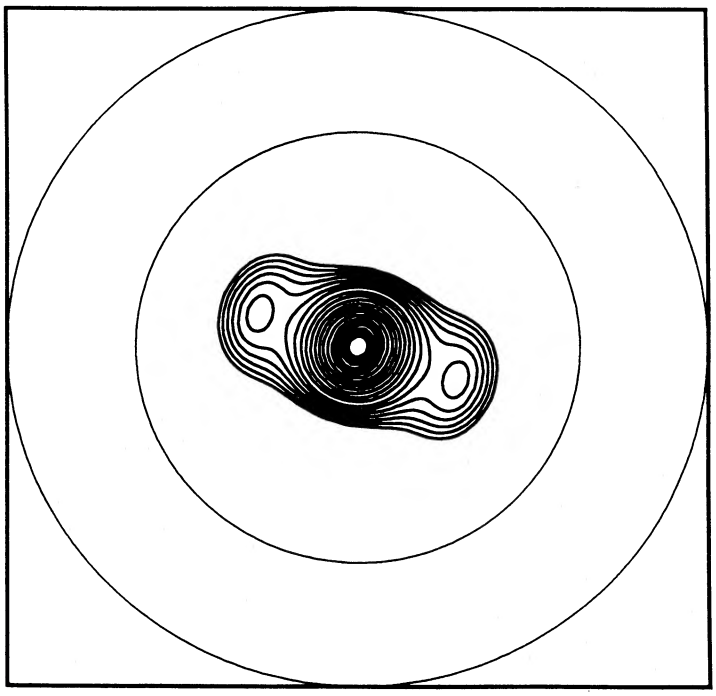

FIG. 14.—Model of a barred galaxy of SB0 type. The model is obtained by superposing a mode of the type shown in the upper left diagram of Fig. 11 onto its basic axisymmetric mass model.

a) Theoretical and Observational Studies of Multicomponent Systems

For the study of certain specific classes of morphologies, a *one-component* description of the galactic disk gives only an incomplete indication of the physical processes. On the other hand, in many cases the gas component may play an essentially passive role, even though giving rise to prominent observable features. Although these phenomena also involve nonlinear mechanisms, they are technically somewhat easier to analyze and to compare with observations. Some examples are described below.

i) *Bar Driving in Two-Component Disks*

A bar or an oval distortion of the mass distribution in the central regions can drive spiral structure in a disk. To this situation we referred to sBB spiral structure in an earlier publication (Lin and Bertin 1981). We recall that for this case the optical appearance does not coincide with the distribution of mass.

Dynamically we may conceive a two-component system for the disk, with the young stars and the cold gas forming the dynamical Population I and the older stars forming the dynamical Population II. The Population I subsystem is relatively cold, thin, and light, while the Population II is relatively hot, thick, and massive. Therefore the former component tends to support tight density waves and the latter subsystem more open structures (see Fig. 1 in Bertin and Romeo 1988).

Consider now the scenario where an oval distortion is present in the Population II component. Rather than an arbitrary distortion, as many authors have considered (see Sanders and Huntley 1976; Roberts, Huntley, and van Albada 1979), we may refer to an internal bar mode. This approach could possibly identify which are the relevant bar structures for a given basic state; for example, we should discard as unrealistic those broad oval distortions that extend too far out, since the present modal studies indicate that such structure are not naturally supported. Once a reasonable bar is taken to affect the hot massive disk, its gravitational field could excite short trailing waves in the Population I thin disk (see Feldman and Lin 1973; Lin and Lau 1975; Goldreich and Tremaine 1979; Yuan 1984; Cheng 1987). These trailing waves do *not* have to form a mode since, although propagating, they are continually replenished by an "external" source. When the bar is relatively strong, the reaction of the light Population I component could then follow essentially the predictions of some hydrodynamical codes, i.e., to give rise to offset shocks inside the bar, and develop tighter spiral arms mostly beyond the corotation circle. This picture just completes the scenario of SB0 modes in the presence of gas and some SB(s) modes (e.g., NGC 1398 or NGC 1300). This seems to be a viable picture, and essentially repeats and develops the concept of sBB spiral structure briefly mentioned at the beginning of this discussion.

However, at present it is hard to draw a clear boundary between the case of an almost "passive" Population I and that of a more "active" gas disk. In other words, a quantitative implementation of the shock structure associated with the various prototypes of modes illustrated in § V is not yet available and should be the focus of future research in this field.

ii) *Formation of Rings*

Besides the preceding discussion of bar-driven spiral structures, a two-component model is essential for the study of the formation of rings. The phenomenology of rings has attracted great attention recently (see Buta 1986, and references quoted therein) and certainly goes beyond the simple application of one-component systems, as illustrated in § V.

Outer rings most likely trace the location of OLR for the relevant mode involved. Inner rings have often been claimed to trace the location of ILR, but this seems to imply a relatively transient nature of the spiral structure that is observed. One possibility is to envisage a system where a light thin Population I disk extends into the bulge region. In this case spiral structure can develop on two different scales, inside and outside the bulge region. The situation may represent galaxies with "dual" spiral structure, such as NGC 5364 (see § VI*a*[iii] and Lin and Bertin 1981).

The ring structure that wraps around the tip of the bar in NGC 1398 (Sandage 1961, p. 47) is typical of SB(r) galaxies. Indeed, this is one important gap in our studies of morphology; i.e., what distinguishes SB(r) galaxies from SB(s) galaxies (see Sandage 1961, p. 24). The process for the formation of the ring may be intimately related to the nonlinear response of the gas to the mode calculated in our basic survey. Possibly an examination of the nonlinear response of gas to the strong bar modes found in subsurvey A, would yield a good model for NGC 1398.

iii) *Galaxies with Dual Characteristics*

There are galaxies like NGC 1097 and NGC 6951 (Sandage 1961, p. 46) which have dual characteristics of Sb galaxies and SBb galaxies (straight dust lanes). One may speculate whether this might be associated with a two-component system with the stellar component showing both a normal Sb structure on the outside and a barlike structure near the central regions. The gaseous component would then show straight dust lanes like those in NGC 1300. Such a complex modal shape is more likely to occur in a two-disk dynamical system since it has an additional degree of freedom.

b) Nonlinear Behavior in Large-Scale Structures

Most of the issues raised in § VI*a* clearly involve nonlinear mechanisms. We comment, in the following, on other nonlinear processes, of a more general nature, which need not depend on the multiple-component structure of galaxy disks.

i) *Multiple-armed Spiral Structure and Flocculent Galaxies*

So far we have been concerned mostly with equilibria that support a few discrete unstable spiral modes. When ILR is less efficient in damping some of the modes, we could have equilibria that support higher m-modes and a number of two-armed spirals. (Note that high m-modes are likely to be *open* even in low-mass disks. Thus flocculent spiral structure with small pitch angle is likely to be related to the presence of many two-armed spiral modes.) If the total number of growing modes is still relatively low, their superposition would initiate multiple spiral structures and some irregular features such as spurs (see Haass, Bertin, and Lin 1982, Fig. 1). When the total number of modes is larger, flocculent spiral structure would be generated. Notice that for those galaxies that do not have a grand design, as is the case of some multiple-armed and flocculent spirals, mechanisms that produce transient spiral features without a grand design can well explain most of the observed features. Here we might refer to stochastic processes of star formation in the presence of differential rotation (see Gerola and Seiden 1978) and to clump-induced spiral wakes (Julian and Toomre 1966). In view of the winding dilemma, the most

likely scenario is the random generation and regeneration of waves through Joans instability from the *gas-rich* outer region of the galactic disk and the propagation of these waves into the interior. Since these waves have limited variation in angular velocities, their pitch angle at any given interior location would remain roughly the same, consistent with the Hubble classification scheme (see Bertin and Lin 1987 for comments on nonstationary scenarios and references to earlier papers).

In the modal context, nonlinearities offer two interesting alternatives that should be tested by future investigations of the possible relevant dynamical conditions. One possibility is that when many modes are present, resonance overlapping across the disk could rapidly lead to chaotic behavior and thus "explain" flocculent spiral structure. On the other hand, it is also possible that nonlinearities under certain conditions would lead to "locking" phenomena which would produce slower and more coherent "vacillations" of spiral structure around preferred configurations, even in those cases where linear superposition of modes indicates a somewhat rapid beating process.

The resolution of these issues appears to be a far distant goal. A more limited result that could be achieved in a nearer future could be the shock structure and small-scale instabilities in the gaseous component in the presence of two linearly superposed spiral modes.

ii) *Smooth-armed Spirals and S0 Galaxies*

For those disks that are gas deficient, such as S0 galaxies, especially if they are characterized by *thick* disks, there should in general be no unstable spiral modes of astrophysical interest. For the few cases, possibly characterized by *thin* stellar disks, where spiral modes can be excited (possibly by mechanisms other than those described in this paper), shock saturation at low amplitude by the gas would not be available, and the modes would be likely to grow to larger amplitudes and to evolve. A smooth structure could still be guaranteed by the relatively large stellar epicycles. These cases could be the dynamical interpretation of smooth armed spiral (see Strom and Strom 1978; Sandage 1983). To be sure, no complete calculation of these nonlinear stellar modes has yet been attempted.

In this respect, one natural concern is on the existence of mechanisms that could limit the growth of these modes and the related fast evolution of the disk. One possibility could be that the orbital response becomes incoherent at a certain amplitude level (see also the amplitude condition stated as $\eta < 1$ by Bertin, Coppi, and Taroni 1977). Some interesting results are being obtained by the extensive numerical iterative techniques of Contopoulos and Grosbol (1986), who try to reach, by purely stellar dynamical methods, a nonlinear quasi-stationary self-consistent spiral structure. We think that their efforts could take advantage of our modal studies, if their procedure is initialized with self-consistent linear global modes of the kind we obtain in our surveys.

c) *Application to the Construction of Models for Specific Galaxies*

An ultimate objective of modal studies is the construction of models for specific galaxies on the basis of all the available observational constraints. In the earlier literature (see Lin and Shu 1967; Lin, Yuan, and Shu 1969; Roberts, Roberts, and Shu 1975), the first step of the modeling process was to estimate the value of the relevant pattern frequency Ω_p empirically. For external galaxies, this was determined on the basis of the observed rotation curve and the location of the corotation zone associated with the "tip" of the optical spiral arms. The local dispersion relationship for the short *wave* branch was then used to calculate the spiral pattern, generally with the additional stipulation of $Q = 1$. Such a procedure was quite adequate for a first study, but there was really no attempt at calculating the amplitude of the wave along the arms nor at showing how such a spiral structure could be self-supporting, even though the need for a general feedback process was recognized.

More recently, a few attempts have been made from the *modal* point of view (Haass 1982; Visser and Haass work in preparation quoted by Haass 1982), but for this purpose modes were calculated only by means of the simplified ordinary differential equations described in Paper II (§ IV). The adopted method of investigation was *ad hoc* in that there was no general perception of the parameter regime for the general category of galaxies under study, and the physical justification of the choice of the relevant basic state (see § III) was essentially overlooked.

Now we have a better basis for the modeling of regular spiral structure in specific objects. In fact, the results of the present study provide a general guidance to the regime of the various dynamical parameters as expected from the observed morphology. As before, we may start out with the observed rotation curve and an estimated value of Ω_p. Then we may determine the general range of the length scale r_Q of the Q-distribution from the observed length scale of the optical (exponential) disk. The identification of the distribution of the relevant *active* surface density is likely to be the most difficult part of the modeling process, since it depends on the relative contributions of stars, gas, and dark matter and their three-dimensional distribution. Especially since the observational data will provide only incomplete information, we should be prepared to face an *iteration* process, starting out with reasonable distributions of the relevant parameters consistent with the data. Plans are under way to apply this procedure to specific galaxies such as M81.

During the preparation of this paper, the authors have benefited from discussions with a large number of observers and theoreticians in the study of spiral structure, too many to be individually acknowledged. We wish to acknowledge the partial financial support of the National Science Foundation of the USA and of the MPI and of the CNR of Italy.

REFERENCES

Allen, R. 1988, in *Proc. Symposium in Honor of C. C. Lin*, ed. D. J. Benney, F. H. Shu, and C. Yuan (Singapore: World Scientific), p. 299.
Aoki, S., Noguchi, M., and Iye, M. 1979, *Pub. Astr. Soc. Japan*, **31**, 737.
Athanassoula, E., Bosma, A., and Papaioannou, S. 1987, *Astr. Ap.*, **179**, 23.
Athanassoula, E., and Sellwood, J. A. 1986, *M.N.R.A.S.*, **221**, 213.
Bahcall, J., and Casertano, S. 1984, *Ap. J. (Letters)*, **284**, L35.
Begeman, K. 1987, Ph.D. thesis, University of Groningen.
Bertin, G. 1983*a*, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel), pp. 119, 174.
———. 1983*b*, *Astr. Ap.*, **127**, 145.
Bertin, G., Coppi, B., and Taroni, A. 1977, *Ap. J.*, **218**, 92.
Bertin, G., Lau, Y. Y., Lin, C. C., Mark, J. W.-K., and Sugiyama, L. 1977, *Proc. Nat. Acad. Sci.*, **74**, 4726.
Bertin, G., and Lin, C. C. 1987, in *Evolution of Galaxies*, ed. J. Palous, *Pub. Astr. Inst. Czechoslovakian Acad. Sci.*, **69** (4), 255–262.
Bertin, G., Lin, C. C., and Lowe, S. A. 1984, in *Plasma Astrophysics*, ed. J. Hunt and T. D. Guyenne (ESA SP-207), p. 115.
Bertin, G., Lin, C. C., Lowe, S. A., and Thurstans, R. P. 1989, *Ap. J.*, **338**, 104 (Paper II).
Bertin, G., and Romeo, A. B. 1988, *Astr. Ap.*, **195**, 105.

Binney, J., and Tremaine, S. 1988, *Galactic Dynamics* (Princeton: Princeton University Press).
Buta, R. 1986, *Ap. J. Suppl.*, **61**, 609.
Cheng, Y. 1987, Ph.D. thesis, City University of New York.
Contopoulos, G., and Grosbol, P. 1986, *Astr. Ap.*, **155**, 11.
de Vaucouleurs, G. 1959, in *Handbuch der Physik*, Vol. **53**, *Astrophysics IV, Stellar Systems*, ed. S. Flügge (Berlin: Springer), p. 275.
Elmegreen, D. M., and Elmegreen, B. G. 1982, *M.N.R.A.S.*, **201**, 1021.
Erickson, S. A. 1974, Ph.D. thesis, Massachusetts Institute of Technology.
Feldman, S., and Lin, C. C. 1973, *Studies Appl. Math.*, **52**, 1.
Gerola, A., and Seiden, P. E. 1978, *Ap. J.*, **223**, 129.
Goldreich, P., and Tremaine, S. 1979, *Ap. J.*, **233**, 857.
Haass, J. 1982, Ph.D. thesis, Massachusetts Institute of Technology.
———. 1983, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel), p. 121.
Haass, J., Bertin, G., and Lin, C. C. 1982, *Proc. Nat. Acad. Sci.*, **73**, 3908.
Hubble, E. 1926, *Ap. J.*, **64**, 321.
Iye, M., Okamura, S., Hamabe, M., and Watanabe, M. 1982, *Ap. J.*, **256**, 103.
Julian, W. H., and Toomre, A. 1966, *Ap. J.*, **146**, 810.
Kalnajs, A. J. 1972, *Ap. Letters*, **11**, 41.
Kent, S. M. 1984, *Ap. J. Suppl.*, **56**, 105.
Kormendy, J. 1985, in *IAU Symposium 106, The Milky Way Galaxy*, ed. H. van Woerden, R. J. Allen, and W. B. Burton (Dordrecht: Reidel), p. 541.
Kuijken, K. H., and Gilmore, G. 1988, in *The Mass of the Galaxy, Proc. Workshop of the Canadian Institute for Theoretical Astrophysics*, ed. Michel Fich.
Lau, Y. Y., Lin, C. C., and Mark, J. W.-K. 1976, *Proc. Nat. Acad. Sci.*, **73**, 1379.
Lewis, J., and Freeman, K. 1988, *A.J.*, submpitted.
Lin, C. C., and Bertin, G. 1984, *Adv. Appl. Mech.*, **24**, 155.
———. 1985, in *IAU Symposium 106, The Milky Way Galaxy*, ed. H. van Woerden, R. J. Allen, and W. B. Burton (Dordrecht: Reidel), p. 513.
Lin, C. C., and Lau, Y. Y. 1975, *SIAM J. Appl. Math.*, **29**, 352.
———. 1979, *Studies Appl. Math*, **60**, 111.
Lin, C. C., and Shu, F. H. 1964, *Ap. J.*, **140**, 646.
———. 1966, *Proc. Nat. Acad. Sci.*, **55**, 229.
———. 1967, in *IAU Symposium 31, Radio Astronomy and Galactic System*, ed. H. van Woerden (London: Academic), p. 313.
Lin, C. C., Yuan, C., and Shu, F. H. 1969, *Ap. J.*, **155**, 721.
Lindblad, P. O., and Jörsäter, S. 1987, in *Evolution of Galaxies*, ed. J. Palous, *Pub. Astr. Inst. Czechoslovakian Acad. Sci.*, **69** (4), 289.
Lowe, S. A. 1988, Ph.D. Thesis, Massachusetts Institute of Technology.
Lubow, S. H. 1986, *Ap. J. (Letters)*, **307**, L39.
Lubow, S. H., Balbus, S. A., and Cowie, L. L. 1986, *Ap. J.*, **309**, 496.
Lynden-Bell, D. 1967, in *Lectures Appl. Math.*, Vol. **9**, *Relativity Theory and Astrophysics, Galactic Structure*, ed. J. Ehlers (Providence: AMS), p. 131.
Ostriker, J. P. 1985, in *IAU Symposium 106, The Milky Way Galaxy*, ed. H. van Woerden, R. J. Allen, and W. B. Burton (Dordrecht: Reidel), p. 638.
Ostriker, J. P., and Peebles, P. J. E. 1973, *Ap. J.*, **186**, 467.
Roberts, W. W. 1969, *Ap. J.*, **158**, 123.
Roberts, W. W., Huntley, J. M., and van Albada, G. D. 1979, *Ap. J.*, **233**, 67.
Roberts, W. W., Roberts, M. S., and Shu, F. H. 1975, *Ap. J.*, **196**, 381.
Roberts, W. W., and Shu, F. H. 1972, *Ap. Letters*, **12**, 49.
Roelfsma, P. R., and Allen, R. J. 1985, *Astr. Ap.*, **146**, 213.
Rubin, V. C. 1987, in *IAU Symposium 117, Dark Matter in the Universe*, ed. J. Kormendy and G. R. Knapp (Dordrecht: Reidel), p. 51.
Sandage, A. 1961, *The Hubble Atlas of Galaxies* (Washington: Carnegie Institute).
———. 1983, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel), p. 367.
Sanders, R. H., and Huntley, J. M. 1976, *Ap. J.*, **209**, 53.
Shu, F. H. 1971, in *Astrophysics and General Relativity*, Vol. **2**, ed. M. Chrétien, S. Deser, and J. Goldstein (New York: Gordon & Breach), pp. 314–322.
———. 1985, in *IAU Symposium 106, The Milky Way Galaxy*, ed. H. van Woerden, R. J. Allen, and W. B. Burton (Dordrecht: Reidel), p. 530.
Strom, S. E., and Strom, K. M. 1978, in *IAU Symposium 77, Structure and Properties of Nearby Galaxies*, ed. E. M. Berkhuijsen and R. Wielebinski (Dordrecht: Reidel), p. 69.
Thurstans, R. P. 1987, Ph.D. thesis, Massachusetts Institute of Technology.
Toomre, A. 1981, in *The Structure and Evolution of Normal Galaxies*, ed. S. M. Fall and D. Lynden Bell (Cambridge: Cambridge University Press), p. 111.
van Albeda, T. S., and Sancisi, R. 1986, *Phil. Trans. Roy Soc. London, A*, **320**, 447.
van der Kruit, P. C. 1988, *Astr. Ap.*, **192**, 117.
van der Kruit, P. C., and Searle, L. 1982, *Astr. Ap.*, **110**, 61.
Vandervoort, P. O. 1970, *Ap. J.*, **161**, 87.
Visser, H. C. D. 1977, Ph.D. thesis, University of Groningen.
Warmels, R. H. 1988, *Astr. Ap. Suppl.*, **72**, 427.
Wevers, B. M. H. R. 1984, Ph.D. thesis, University of Groningen.
Wevers, B. M. H. R., van der Kruit, P. C., and Allen, R. J. 1986, *Astr. Ap. Suppl.*, **66**, 505.
Whitmore, B. 1984, *Ap. J.*, **278**, 61.
Yuan, C. 1984, *Ap. J.*, **281**, 600.
Yue, Z. Y. 1982, *Geophys. Ap. Fluid Dyn.*, **20**, 1.
Zang, T. A. 1976, Ph.D. thesis, Massachusetts Institute of Technology.

G. BERTIN: Scuola Normale Superiore, Pisa 56100 I, Italy

C. C. LIN, S. A. LOWE, and R. P. THURSTANS: Massachusetts Institute of Technology, Room 2-330, Cambridge, MA 02139

THE ASTROPHYSICAL JOURNAL, 338:104–120, 1989 March 1

MODAL APPROACH TO THE MORPHOLOGY OF SPIRAL GALAXIES. II. DYNAMICAL MECHANISMS

G. BERTIN
Scuola Normale Superiore, Pisa, Italy
AND
C. C. LIN, S. A. LOWE, AND R. P. THURSTANS
Massachusetts Institute of Technology
Received 1988 July 5; accepted 1988 August 13

ABSTRACT

A unified approach to spiral modes is carried out through the use of a cubic dispersion relationship developed earlier. This allows us to demonstrate the processes of maintenance and excitation of spiral modes of various morphologies separately for (*a*) normal spiral modes and (*b*) open spiral modes.

In order to understand the relevant physical mechanisms and the general roles of the dynamical parameters (specified by the radial distribution of angular velocity, *active* surface density, and *equivalent* dispersive speed) the waves are described in terms of sustained wave *trains*. This allows us to examine the various dynamical issues discussed in Paper I. Very good accuracy is found when the present analytical results are compared with exact numerical data.

The analytical approach makes it much easier to accomplish the process of *identification of basic states* from the observed spiral structure when a wide range of variation of the parameters of the basic state is involved. Our characterization of parameter regimes in terms of a few simple algebraic relations will be a useful *guide* for the construction of detailed quantitative models of specific galaxies and for general numerical surveys of classes of galaxies of given morphological types.

Subject headings: galaxies: internal motions — galaxies: structure — stars: stellar dynamics

I. INTRODUCTORY REMARKS

In a previous paper (Bertin *et al.* 1988, hereafter Paper I) we have provided a unified framework for the morphology classification of spiral galaxies by means of a modal survey of a family of galaxy models of astrophysical interest. Numerical surveys of this kind require a good knowledge of dynamical mechanisms. In fact, on many occasions, in the choice of the basic state and in the discussion of the relevant spiral modes, we have referred to various dynamical arguments. It is the purpose of this article to address in detail the many dynamical issues that have been raised. The main result of our analysis is a very simple unified approach (in terms of a cubic dispersion relation) that covers the processes of maintenance and excitation of spiral modes of various morphologies, in particular normal spiral modes and open barlike modes.

In order to understand the relevant physical mechanisms and the general role of the dynamical parameters (specified by the radial distribution of angular velocity, *active* surface density, and *equivalent* dispersive speed) we shall describe the modes in terms of sustained *trains* of waves which satisfy a certain dispersion relationship. As we shall see (§ III), such a description helps us to identify the parameter regimes that correspond to stable, moderately unstable, or violently unstable disks. In fact, this gave us valuable guidance in our numerical survey. In general, we can now choose the dynamical parameters so as to simulate galaxies of the normal type or of the barred type using the unifying framework of a *single simple dispersion relation.* In passing, we note that dynamical mechanisms and concepts described in this paper have some relevance to other astrophysical contexts, such as planetary rings and accretion disks and tori (see Papaloizou and Pringle 1984).

In the present paper special attention will be given to open or barlike modes (§ V) since the theory of normal spiral modes (briefly addressed in § IV) has already been described in the past, but even in this latter case we shall present some new results. In particular, we find that the growth rates calculated by the analytical theories, for both open spirals and tightly wound spirals, agree quite accurately with those obtained from global integration for a wide range of parameters. Before developing the relevant analytical theory, in the next section we address some important dynamical issues of general interest. This will complete some of the arguments presented in Paper I (e.g., see §§ I and VI of that paper) and will form the framework for the results of the present article.

II. SPIRAL MODES IN THE FLUID MODEL

The formulation of the theory of linear spiral modes is well known both in the *stellar* theory and in the *fluid* approach. In the latter, the phenomenon of spiral structure is approached by studying the stability against spiral perturbations of a state of axisymmetric equilibrium characterized by differential rotation $\Omega(r)$, equivalent acoustic speed $a(r)$, and active disk density $\sigma(r)$, according to the physical picture explained in Paper I. In recent years, efforts have been devoted toward the understanding of the nature of these modes. Generally speaking, after the rotation curve is specified, the characteristics of modal instability of a galactic disk have been found to be determined by the radial profiles of two parameters, J and Q (which are defined in terms of Ω, a, and σ in § III). Normal spiral modes are associated with lower values of the J parameter; moderately growing barred spiral modes, with higher values. The Q parameter is expected to be determined by a physical process of *self-regulation* (see Paper I). In many cases the difference in

the results obtained by various authors may indeed be traced to the differences in the distributions of these parameters in the basic states adopted.

The modal theory is self-contained. The maintenance of spiral modes may be visualized in terms of oppositely propagating wave trains that satisfy a certain dispersion relationship. The *excitation* mechanism for such modes can be seen as a process of overreflection near the corotation circle, in the form of a WASER (Wave Amplification by Stimulated Emission of Radiation) (see Mark 1976 for the regime of tightly wound normal spiral structure). Another reflection mechanism must be present near the central region of the galaxy, so that the wave trains are linked in a closed cycle or "feedback loop." In the following we shall discuss certain general issues that serve as a background to the analysis of the present paper.

a) Modes in the Fluid Model: An Integrodifferential Equation

Modal studies may be carried out with two distinct objectives in mind:

1) For the study of dynamical mechanisms;
2) For the simulation of galaxies of various morphology types.

The latter objective requires appropriate identification of models of the basic states of galaxies and was discussed in Paper I. Here we focus mostly on objective (1).

We consider elementary waves so that any perturbed quantity $q(r, \theta, t)$ is of the form

$$q = q_1(r) \exp(i\omega t - im\theta) \tag{2.1}$$

in cylindrical polar coordinates (r, θ, z); t is the time, m and ω are constants. Starting from the Euler and the continuity equations, we derive (cf. Lin and Lau 1979) the basic equation for perturbations:

$$L(h_1 + \psi_1) = -Ch_1 , \tag{2.2}$$

where

$$L = \frac{d^2}{dr^2} + A\frac{d}{dr} + B , \tag{2.3}$$

and

$$A = -\frac{1}{r}\frac{d \ln \mathscr{A}}{d \ln r} , \quad \mathscr{A} = \frac{\kappa^2(1 - \nu^2)}{\sigma r} , \tag{2.4}$$

$$B = -\frac{m^2}{r^2} - \frac{4m\Omega(r\nu')}{\kappa r^2(1 - \nu^2)} + \frac{2\Omega m}{r^2\kappa\nu}\frac{d \ln(\kappa^2/\sigma\Omega)}{d \ln r} , \tag{2.5}$$

$$C = -\frac{\kappa^2(1 - \nu^2)}{a^2} . \tag{2.6}$$

Here κ is the epicyclic frequency, and $\nu = (\omega - m\Omega)/\kappa$ is a dimensionless frequency. The quantity h_1 represents the enthalpy perturbation, which in our model is related to the density perturbation by

$$h_1 = \frac{a^2\sigma_1}{\sigma} . \tag{2.7}$$

The potential ψ_1 is related to the active disk density σ_1 through the Poisson equation, which can be written in the integral form

$$\psi_1(r) = -2\pi G\int_0^\infty K(r, r')\sigma_1(r')\,dr' . \tag{2.8}$$

The kernel $K(r, r')$ is well known; it exhibits a logarithmic singularity at $r = r'$.

Within the present fluid model, these equations are exact and have been adopted by Pannatoni (1979, 1983) in a flexible numerical code. Thus in the numerical studies made by using Pannatoni's code (or another version of the code devised by Haass 1982) the long-range gravity force is taken into account in its exact integral form. Therefore, in the following, we shall refer to results obtained by using the code mentioned above as the "exact results." Different models that have been developed by other researchers in order to calculate global modes (e.g., Erickson 1974; Bardeen 1975; Zang 1976; Kalnajs 1977; Aoki, Noguchi, and Iye 1979; Toomre 1981; Athanassoula and Sellwood 1986) have been referred to in Paper I. In this article we shall compare our exact results to certain simpler approximate results derived mostly analytically. For the convenience of such analytical studies, it is necessary to reduce the problem to a simple differential form. This requires a careful treatment of the Poisson equation, under a certain systematic procedure of asymptotic approximations (see Bertin and Mark 1979 and Appendix B).

In dealing with the computation of spiral modes one must pay attention to the *appropriate boundary conditions*, since these play a crucial role in the global stability problem. As a way to incorporate physical processes such as those described in § II*f*, we shall impose *a radiation boundary condition just inside* the outer Lindblad resonance (OLR) circle (see Lau, Lin, and Mark 1976; Pannatoni 1983). Other physical processes related to the turbulent cold gas component can favor the same choice of outer boundary condition. Thus OLR, which would be improperly handled by a fluid model, is actually avoided in our integration scheme. (If the inner Lindblad resonance happens to occur in the propagating region of our mode calculation, we will reject the calculated mode as a damped mode, according to the well-known results of stellar dynamics; § II*f*).

b) Diagnostics of Modes: Spectral Representation

In Paper I we have illustrated a three-parameter modal survey of a family of basic states. In the following, numerical results on "exact" modes will be compared with a theory based on the wave train description (§§ IV, V). For this purpose we notice that a useful representation of a mode which helps us to "diagnose" the underlying wave composition is the spectral representation. For a given value of m we can evaluate the α-spectrum

$$\sigma(\alpha) = \frac{1}{2\pi}\int_0^\infty \sigma'(r)e^{i\alpha \ln r}\frac{dr}{r} \tag{2.9}$$

and refer to the quantity $P(\alpha) = |\sigma(\alpha)|^2$. This logarithmic representation has been found to be convenient for a number of reasons (see Kalnajs 1965; Bertin and Mark 1979). For linear modes we can use arbitrary units for $P(\alpha)$.

The spectra calculated by means of equation (2.9) include all the information of the radial structure (of the modes) and therefore all the features associated with the inhomogeneous system. Thus they should *not* be confused with the spectral representation *localized* in the corotation zone that will be discussed in Appendix C.

c) Diagnostics of Modes in Terms of Wave Trains: Propagation Diagrams

One natural way to understand a spiral mode is to describe it in terms of propagating waves satisfying a local[1] dispersion

[1] In this paper, we also use the term "localized in the corotation zone" or simply "localized." Such a term refers to a more restrictive situation where a narrow corotation zone alone is being studied.

relationship at various points in an inhomogeneous disk. In § III we will discuss the dispersion relationship for spiral waves which is of the form

$$F(K, \nu^2; Q^2, J^2) = 0 . \quad (2.10)$$

The α-spectrum is only *one* important diagnostic tool. In order to analyze better the mode structure we have to resort to the *propagation diagram.* Once we fix the time frequency of a mode (i.e. its eigenvalue), the quantity ν is a simple function of r which depends only on the properties of the rotation curve of the equilibrium model. Therefore we can read the local dispersion relation (2.10) as a relation which, for any galactocentric radius r, fixes the possible value(s) of the local radial wavenumber. This results in a diagram in the (k, r)-plane which we call the propagation diagram. There we can easily include information on propagation regions, resonances, and reflection points. With an arrow we can record the direction of group propagation on each branch. A clear example of the usefulness of the representation is shown in the figure of the paper by Bertin (1983*a*) where the mechanisms for the two regimes of normal spiral structure and of open (barred) structure are well identified.

d) Modes in the Fluid Model: Reduction to Ordinary Differential Equations

This simple description of the mode in terms of propagating wave trains suggests the possibility of developing an analytic theory in terms of ordinary differential equations. In §§ IV, V, we shall give the full details of an analytic theory appropriate to each of the two regimes of normal spiral modes and of open (barred) modes. The theory is based on the dispersion relationship for short waves, i.e. large *total* wavenumber (§ III); it is simpler than the "exact" theory (§ II*a*), and yields similar numerical results (see Fig. 5) in a large number of cases tested. It can thus be used conveniently for the purpose of exploration in a survey of the type described in Paper I, and for checking numerical results obtained by other methods. The easy physical interpretation of the processes of maintenance and excitation is, of course, the strongest point in an analytic theory.

e) Excitation Processes

Because of the differences between the nature of the waves composing the normal spiral modes and the open (bar) modes, their growth rate is expected to depend differently on the dynamical parameters J and Q. This will be discussed in the following sections. Here we stress the fact that the basic excitation mechanism can always be defined as an overreflection process at the corotation circle. This occurs as a result of the conservation of wave action, when wave action is transferred from the region inside the corotation circle (where the wave energy density is negative) to the region outside (where the wave energy density is positive) (see also Bertin 1983*b*). Overreflection between short and long waves (i.e. for normal spiral structure, below the transition line) is found to be moderate for $Q \approx 1$. Overreflection between open waves (i.e. for larger values of J, above the transition line) is too high for $Q \approx 1$ and becomes moderate only for larger values of Q. Thus if one starts out with the condition $(J, Q) = (1, 1)$, the value of Q is expected to develop soon into higher values as the dispersive speed increases as a result of violent instability. The regime of moderate overreflection would be reached by a process of self-regulation (see Paper I).

A comment is required for a comparison with the studies of "swing" amplification mechanism (Goldreich and Lynden-Bell 1965; Julian and Toomre 1966; Goldreich and Tremaine 1978; Toomre 1981). It has been shown by Drury (1980) that this mechanism, usually discussed in a *homogeneous* model of the vicinity of the corotation zone where narrow wave *packets* swing from the leading to the trailing form, has a counterpart in the theory of steady wave trains. This point has been elaborated further by Lin and Thurstans (1984), who unified the discussion by means of a systematic spectral theory. Some of these arguments are summarized in Appendix C below. Here we would like to reiterate that in this respect there is no conflict between the "swing" mechanism and the present modal theory. Indeed, our study of overreflection of open waves above the transition line (see Bertin 1983*a*) describes the *same* physical process as that called "swing" in an alternative formalism. Thus it should be kept in mind that the "swing," as interaction between leading and trailing waves, is not necessarily powerful nor necessarily transient (see Lin and Thurstans 1984, p. 128).

f) Stellar Dynamics

The fluid model that we have described and that we will adopt in the present paper is a convenient tool of investigation if used with care and judgement. In fact, extensive studies of stellar dynamic models support the view that a fluid model is generally (qualitatively) adequate *except at the location of resonances.* Many important issues have been reviewed before (see, e.g., Bertin 1980; Lin and Bertin 1981). Here we just recall a few points that are most relevant to the studies in this article and in Paper I:

1. Stellar dynamic resonances can be studied by analogy with wave-particle interaction in ordinary plasma physics (see Landau 1946; Lynden-Bell and Kalnajs 1972; Bertin and Haass 1982). Note that heating processes for linear quasi-stationary modes operate at the resonance locations only. In contrast, violent instabilities, with large amplitudes and large growth rates, are expected to produce a distributed heating and therefore increase the overall velocity dispersion in the disk.

2. The absorption of waves at the inner Lindblad resonance can interrupt the relevant *feedback* wave cycle by preventing waves from reflecting back from the central to the outer regions. This is found to damp the spiral mode (see also the numerical results by Zang 1976). Thus the welcome role of inner Lindblad resonance is that of considerably reducing the number of unstable modes that can be supported by a galactic disk (see Bertin *et al.* 1977). The presence of a modest concentration of nuclear mass can easily lead to a model in which the higher $m (m \geq 3)$ modes and the higher n modes (lower frequency) are damped. In particular, by this argument, we have checked that the two-armed spiral mode used in our modal survey can indeed be taken as a good representative mode for the family of basic states considered (see Lowe 1988).

3. If the Lindblad resonances are too close (less than two epicyclic radii) to the corotation circle, *chaotic behavior* of stellar orbits is expected to occur (see Contopoulos 1983), and hence the standard analysis of physical processes is invalidated. In particular, coherent phenomena such as angular momentum transfer across the corotation zone (and therefore the analogous time-dependent process of swing amplification) are expected to be *inhibited.* Most cooperative disturbances with $m \geq 4$ are expected to suffer from this difficulty and thus to be unimportant, in contrast to the indication that might

derive from naive local analyses or global calculations that do not include this dynamical feature.

4. The (J, Q)-diagram (see Fig. 3) should be used by including the *global* stability properties of the disk. The parameter J *is proportional to* m, so that in principle the disk displays *different* levels of stability depending on the azimuthal structure of the mode. High-m modes, being characterized by higher values of J, would be expected to be more unstable. However, in our discussion we shall always focus, for astrophysical applications, on *the reference value of* $m = 2$. In fact, our view is that most galaxy disks of astrophysical interest have the $m = 3$ *modes inhibited by ILR*, consistent with the empirical fact that coherent three-armed structures are rarely observed. Note that *fast rotating* $m = 3$ modes that might be ILR–free would have *small corotation radii* (usually considerably smaller than the corotation radius of two-armed modes) in the region where the Q-profile is expected to be relatively high. Actually this could even provide part of the physical reason why the *Q-profile* is rising toward the central regions of the disk. Therefore for the models that we consider of physical interest the (J, Q)-diagram should be discussed on the basis of two-armed modes. Higher m-modes are likely to be unimportant. The $m = 1$ *modes* are expected to coexist but with less prominence, given their smaller J-values.

5. Before comparing the precise numbers for Q, as indicated by the (J, Q)-diagram, with observations, another feature to keep in mind is that a *kinetic* calculation is bound to identify an overall *higher "marginal curve"* on the high-J side just because of the physical role of pressure anisotropy in the stellar dynamic system.

g) *Discrete Spectrum of Modes*

The evolution from a given equilibrium configuration of a dynamical system is usually not too sensitive to the initial conditions, provided that they are "reasonable." However, if the system does contain a set of modes with a continuous spectrum, the situation may be different. A superposition of neutrally stable oscillations (part of a continuous spectrum) often leads to a solution that depends very much on the initial conditions and decays algebraically in time. Thus, even such a set of modes is *not* important when there are significantly unstable modes. In a previous paper (Lin and Bertin 1981), we have also explained why this issue is generally not important in the theory of spiral structure. Here we only notice that direct experience of spiral mode calculation in various galaxy models shows no evidence of realistic disks with a continuous spectrum of spiral modes (see also Zang 1976).

III. THE LOCAL DISPERSION RELATION IN AN INHOMOGENEOUS DISK

A local stability analysis of axisymmetric self-gravitating fluid disks (see statement of the problem in § II*a*) leads to the following dispersion relation:

$$\frac{Q^2}{4} = \frac{1}{K} - \frac{1 - \nu^2}{K^2 + J^2/(1 - \nu^2)}, \tag{3.1}$$

where

$$\nu = \frac{(\omega - m\Omega)}{\kappa}, \tag{3.2}$$

$$J = m\epsilon_0 \left(\frac{4\Omega}{\kappa}\right) \left|\frac{d \ln \Omega}{d \ln r}\right|^{1/2}, \tag{3.3}$$

$$Q = \frac{a\kappa}{\pi G\sigma}, \tag{3.4}$$

$$\epsilon_0 = \frac{\pi G\sigma}{r\kappa^2}, \tag{3.5}$$

$$K = 2kr\epsilon_0, \tag{3.6}$$

$$k^2 = k_r^2 + k_\theta^2 = k_r^2 + \frac{m^2}{r^2} = \frac{m^2}{r^2}(1 + \mu^2). \tag{3.7}$$

Here ν is the dimensionless frequency, Q (see Toomre 1964) and J are the relevant dimensionless stability parameters, ϵ_0 measures the (local) self-gravity, and K is the magnitude of the appropriate dimensionless total wavenumber. The radial wavenumber k_r appearing in equation (3.7) is defined by the usual WKB prescription, but it is allowed to vanish, if necessary. Indeed, k_r and $\mu = k_r r/m$ can take on both positive and negative values. In fact, equation (3.1) is derived under the ordering:

$$\epsilon_0^2 \ll 1, \tag{3.8}$$

and

$$K^2 = O(1). \tag{3.9}$$

Comments on the derivation of equation (3.1), which essentially repeats that given by Lau and Bertin (1978), are provided in Appendix A. The assumed ordering of parameters is required for a local treatment of the potential theory. The dispersion relation (3.1) essentially describes the Jeans stability mechanism in disk geometry in the presence of differential rotation.

For subsequent discussions it is convenient to introduce the following quantities: the pattern frequency Ω_p,

$$\Omega_p = \mathrm{Re}\,(\omega)/m, \tag{3.10}$$

the shear parameter s,

$$s = -\frac{d \ln \Omega}{d \ln r}, \tag{3.11}$$

and the quantity

$$\chi = \frac{J^2}{K^2} = \frac{\chi_0(s)}{1 + \mu^2}. \tag{3.12}$$

a) *The* (J, Q)*-Plane*

The dispersion relation (3.1) clearly identifies J and Q as the stability parameters of the problem. Therefore the properties of the dispersion relation are naturally discussed in a (J, Q)-plane. Actually a $(\ln J, \ln Q^2)$ representation turns out to be most convenient.

For the marginal case $\nu^2 = 0$ we can plot the K-contours, i.e. the contour curves of constant K. These are shown in Figure 1. The family of K-contours can be divided into three groups:

i) Short waves, with $2 < K$;
ii) Long waves, with $\frac{1}{2} < K < 2$;
iii) Open waves, with $K < \frac{1}{2}$.

The long waves can be further divided into two subgroups: those with $K > 9/8$ and those with $K < 9/8$.

The heavy contours in Figure 1*a* are the envelopes of the long wave contours. They delineate a lower left portion of the (J, Q)-plane, where the solution surface is *three-sheeted*. Within this sector, the dispersion relation possesses three real solu-

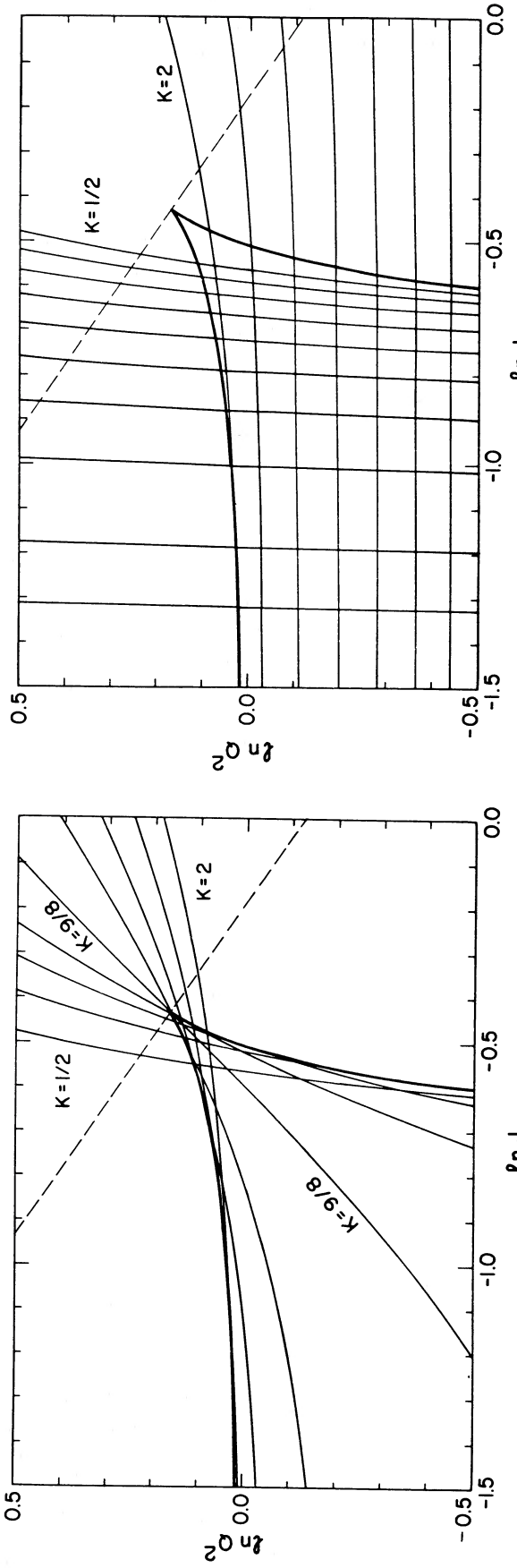

FIG. 1.—Contour curves of constant dimensionless wavenumber K in the (J, Q)-plane for the marginal case $\nu^2 = 0$. Dashed diagonal line represents the *transition line*. *Left*: curves for $\frac{1}{2} \leq K \leq 2$. These contours have an envelope which is shown as a thick line. Each contour changes sheets at its point of tangency with the envelope. *Right*: curves for $K \leq \frac{1}{2}$ and $K \geq 2$. The envelope from the left frame is reproduced here for reference. These contours do not change sheets.

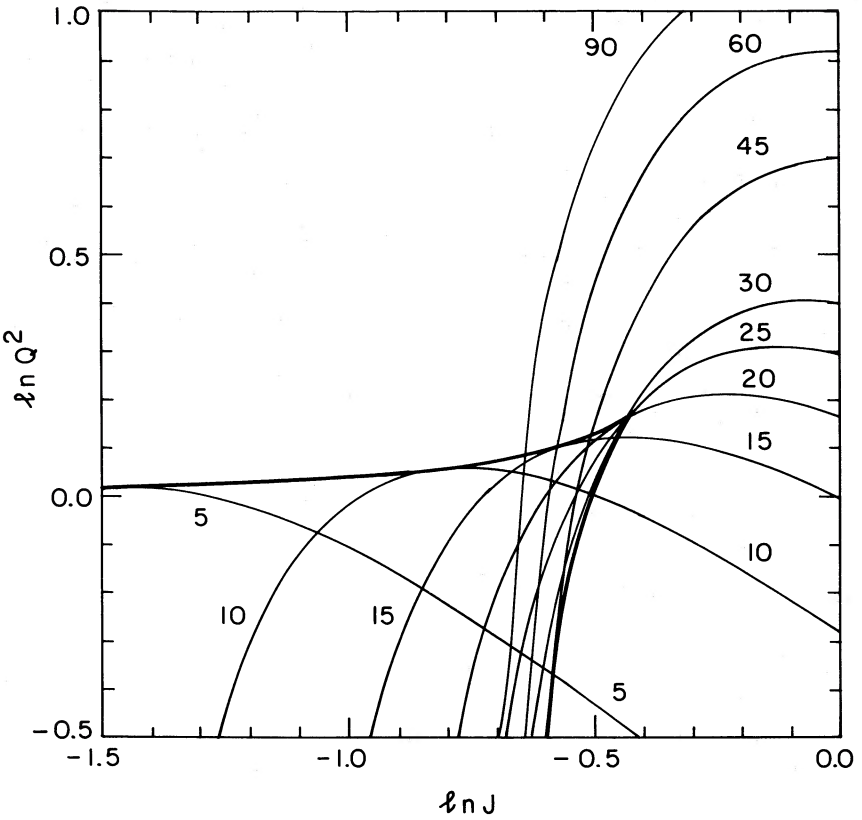

FIG. 2.—Curves of constant pitch angle $i = \cot^{-1} \mu$ in the (J, Q)-plane. These curves are derived from the cubic dispersion relation (3.1) for the conditions $s = 1$, $v = 0$. Curves are labeled by the value of i in degrees. For $v \neq 0$, refer to the similarity transformations described by Bertin, Lin, and Lowe (1984).

tions for K; outside, it has only one. For the wave corresponding to this latter case, the direction of group propagation is the same as that for the short wave. It is important to note that the three-sheeted region extends only as far as the cuspidal point P_c, where J and Q have the values:

$$J_c = \sqrt{27/64} \cong 0.6495 \; ; \quad Q_c = \sqrt{32/27} \cong 1.0887 \; . \tag{3.13}$$

This means that moderate to large values of J and/or Q will place us squarely in the *one-sheeted* part of the plane, where the feedback cycle must consist of one leading and one trailing wave. Note that in the Lau-Bertin (1978) plot of J-contours for $v^2 = 0$, this point P_c lies on the curve $J = J_c$ and corresponds to the location where the point of inflection has a horizontal tangent.

By examining the structure of the K-contours when v^2 is varied, it is found that the cuspidal point P_c moves in the (J, Q)-diagram and divides it into two parts by the *transition line*

$$\ln J + 3/2 \ln Q^2 = \ln (16\sqrt{2}/27) \; . \tag{3.14}$$

Above this line there is only *one* solution for *all* values of v^2 (provided that the condition $\mu^2 \geq 0$ is met; see § III*b*).

For convenience of reference, we also present in Figure 2 (prepared by Yue, private communication) the curves of constant pitch angle $i = \cot^{-1} \mu$ in the (J, Q)-plane for the conditions $s = 1$ and $v = 0$ (from eqs. [3.1] and [3.12]).

b) *The Geometrical Constraint*

Because of a constraint which is intrinsic to the disk geometry, and not present in the plane geometry, the total wavenumber must exceed a minimum value m/r (see eq. [3.7]). This imposes the following physical boundaries

$$K \geq 2m\epsilon_0 = \frac{J}{\sqrt{\chi_0(s)}} \; , \tag{3.15}$$

and

$$\chi \leq \chi_0(s) = \frac{4\Omega^2}{\kappa^2} \, s \; . \tag{3.16}$$

Therefore the geometrical constraint brings in the shear rate, through χ_0, as an independent parameter for the stability analysis. Note that, for low values of J, some of the solutions shown in Figure 1*b* (as nearly vertical lines) are invalidated by this constraint. Those solutions are naturally omitted from Figure 2.

c) *Regimes of Spiral Structure*

We can classify different regimes of spiral structure on the basis of the value of J around the corotation region. The terminology that will be adopted refers to the morphology of spiral modes along the region of moderate growth (see § III*d*).

1. The regime of *normal spiral structure* is characterized by $J \lesssim \frac{1}{2}$. The regime of *tightly wound* spiral arms is obtained from the dispersion relation (3.1) by taking the limit $J \to 0$. The regime of *finite inclination of spiral arms* can be investigated by considering the first-order corrections due to the presence of a small J.

2. The regime of *open spiral structure* is characterized by $J \gtrsim \frac{3}{4}$. In this regime moderate growth occurs only for very open structures ($\mu^2 < 1$).

3. A *transition* regime is found when $\frac{1}{2} \lesssim J \lesssim \frac{3}{4}$. In this case the maintenance and the morphology of spiral structure can be quite complicated, as indicated by the propagation diagrams of the relevant modes.

The first regime has been studied extensively in the literature and is well understood from the dynamical point of view. (Its physical basis is described in Paper I.) It will be briefly summarized in § IV. Most of the new results of the present paper refer to the second regime, which will be described in § V. The transition regime should be studied mostly numerically. Empirically (see survey of Paper I) the morphology of modes is found to change in a relatively smooth manner in the transition regime; here extrapolation from the regime of either normal or open spiral structure turns out *a posteriori* to be still of interest (see discussion at the end of §§ IV*b* and of V*c*).

d) The Strip of Moderate Growth

In the theory of normal spiral structure it has long been recognized that moderate growth occurs when the relevant wave cycle is all trailing (i.e. $\mu < 0$) and the short and long wave branches merge and interact with each other at the corotation region ($v^2 = 0$). Indeed, a "marginal" growth (corresponding to the case $Q = 1$ for tightly wound spirals) is found along the upper boundary of the envelope of Figure 1 where the long and the short waves coincide. This is easily demonstrated in the WASER formalism (see § IV).

In the regime of open spiral structure, overreflection operates because of the interaction between leading ($\mu > 0$) and trailing ($\mu < 0$) waves. Therefore, a moderate growth is expected when μ^2 is very small at the corotation region. Indeed the "marginal" growth condition is found by setting $\mu^2 = 0$ and $v^2 = 0$ in the dispersion relation (3.1). This is also easily shown in the WASER formalism (see § V). Thus the region of moderate growth in the regime of open spiral structure is identified by drawing the χ-contours in the (J, Q)-plane (see Bertin, Lin, and Lowe 1984), and by selecting the contour where $\chi = \chi_0(s)$ (see condition [3.16] imposed by the geometrical constraint). For a given value of s, the contour $\chi = \chi_0(s)$ is the "marginal" growth line, while the others $[\chi < \chi_0(s)]$ can be seen as contours of constant pitch angle ($\mu^2 = \text{constant}$; see Fig. 2).

A synthetic view of the regimes of moderate growth is given in Figure 3.

e) Wave Propagation

The general characteristics of wave propagation can be discussed by calculating, from the dispersion relation (3.1), the group velocity of radial wave propagation

$$c_g = -\frac{\partial \omega}{\partial k_r} \tag{3.17}$$

along the various wave branches. This is important when constructing the relevant wave cycles for the maintenance of spiral modes. Another point to keep in mind is that signals are absorbed at the Lindblad resonances ($v^2 = 1$; see § II*f*). Then it

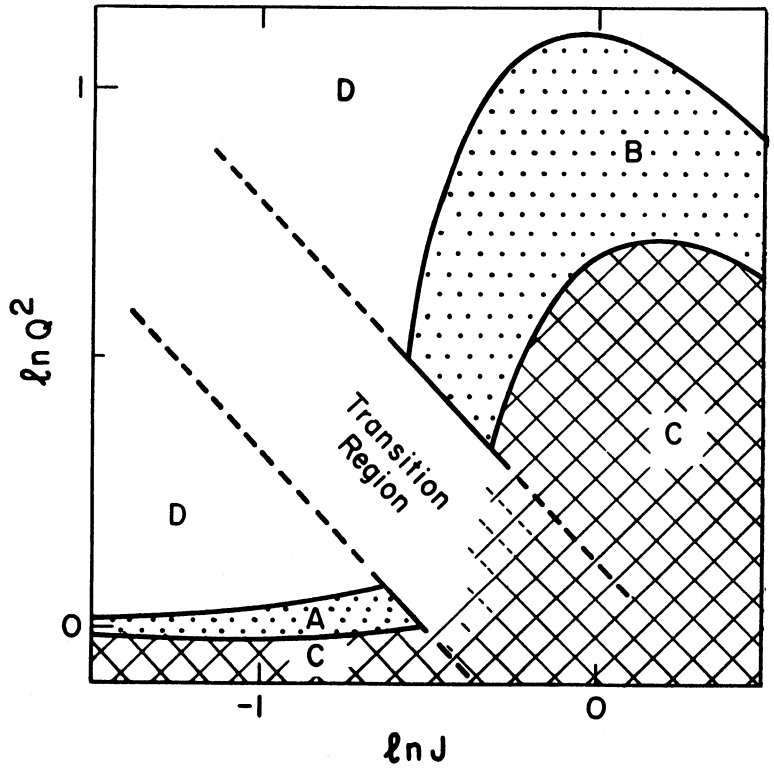

FIG. 3.—The (J, Q)-diagram. Regions of various amounts of overreflection are identified in the relevant parameter space. In regions A and B the overreflection factors correspond to moderate instability (A, tight spirals; B, open spirals). In region C the strong overreflection implies rapid growth of modes. The simple two-wave analysis (see §§ IV, V) does not apply in the transition region, but trends are found to hold when compared to numerical surveys. Boundaries of region B are here calculated for $I_r = \ln 2$ and $I_r = 2$ in the case $s = 1$ (see § V), but their precise definition is less important for the overall perception. The key feature of the diagram is that it gives the essential information for interpreting stability diagrams, such as those shown in Fig. 5 of Paper I. However, judicious use is required, since inhomogeneties and boundary conditions enter the global problem.

is useful to check which pitch angle range is suggested by the fluid dispersion relation in the vicinity of the Lindblad resonances. There we have $K \sim 4/Q^2$; i.e.,

$$\mu^2 \sim \left(\frac{4\sqrt{\chi_0}}{JQ^2}\right)^2 - 1 = \left(\frac{2}{m\epsilon_0 Q^2}\right)^2 - 1 . \quad (3.18)$$

Thus, in general, we expect a fairly small pitch angle ($\mu^2 \gg 1$) at the outer Lindblad resonance where $\epsilon_0 Q^2$ is expected to be small (the situation could change for high m-modes). In contrast, in the inner regions of a galactic model we may expect a larger value of ϵ_0 and of Q. Thus it will not be unusual to have waves refracted outward at an "inner turning point" before reaching the inner Lindblad resonance.

Note that a Lindblad resonance is reached by only one wave branch. From equations (3.17) and (3.1) it is found that only trailing waves approach Lindblad resonances. Note also that in the regime of tightly wound spiral waves ($J \to 0$, $k \sim k_r$) the long wave branch formally reaches Lindblad resonance with solution $k^2 \sim 0$. In the present context, this is a spurious solution which is removed in our analysis (eq. [3.1]) that includes the geometrical constraint (see § III*b*).

f) *Prototypes of Modes*

Due to inhomogeneities and the complicated structure of the transition region, many modes defy a simple analytical treatment. However, it is possible to identify a few prototypes of modes that can be easily understood in terms of our dispersion relation. These actually provide the key morphological structures of a survey like the one presented in Paper I.

Two of these prototypes were essentially described earlier by Bertin (1983*a*). In one case, the waves in the mode are uniformly in the regime of normal spiral structure (low Q and low J; see § IV). In the other case the mode is uniformly in the domain of open spiral structure (high Q and high J; see § V). These prototypes will be shown in Figure 6.

A third important prototype, which is naturally identified in surveys of the B-type of Paper I, corresponds to a situation where the region inside the corotation circle is essentially in the domain of open spiral structure and the region outside the corotation radii is in the regime of normal spiral structure. This leads to morphology of the SB(s) type, and will be further discussed in § VI.

The other simple alternative, viz. a regime of open spiral structure outside and normal spiral structure inside corotation, does not appear to occur naturally in the astrophysical context.

IV. THE REGIME OF NORMAL SPIRAL STRUCTURE

Investigations of tightly wound spiral density waves (see, e.g., Lin and Lau 1979; Bertin 1980, and references cited therein) have led to the conclusion that some realistic galaxy models (see Paper I) can support self-excited global *normal spiral modes*, which owe their maintenance to the presence of trailing waves with opposite propagation properties and are excited mostly as a result of a WASER (overrflection) mechanism (Mark 1976) at corotation. In the regime of finite inclination of spiral arms the presence of shear and self-gravity enhances the amplification of spiral waves, as described by the parameter J (Bertin and Mark 1978; Lau and Bertin 1978). In fact, tangential forces ease the transfer of energy to the outer regions, thus enhancing the overreflection process at corotation. As a result, sizable growth rates of normal modes can be found even in models characterized by values of Q slightly larger than unity, which is the marginal case for axisymmetric disturbances. Nevertheless, the same studies show that the growth rates of normal modes are still quite sensitive to the value of Q.

a) *A Turning Point Equation*

We rederive the earlier results on normal spirals in the present context of a cubic dispersion relationship. We consider the asymptotic ordering $J \sim \epsilon_0 \ll 1$ and $\mu^2 \gg 1$, so that $K^2 \sim K_r^2 \equiv (2r\epsilon_0 k_r)^2$. Then the dispersion relation (3.1) reduces to

$$K_r^2 \frac{Q^2}{4} \sim K_r - (1 - \nu^2) + \Delta , \quad (4.1)$$

where

$$\Delta = \frac{J^2}{K_r^2} \sim \frac{J^2 Q^4}{4} . \quad (4.2)$$

In estimating the value of the correction Δ we have assumed that K_r is close to its double root $K_r^0 \cong 2/Q^2$, where the short and the long wave branches merge. This algebraic dispersion relation is associated with the following turning point equation:

$$u'' + \frac{1}{4r^2\epsilon_0^2}(\delta K)^2 u = 0 , \quad (4.3)$$

where

$$(\delta K)^2 = \frac{4}{Q^2}\left(\nu^2 - 1 + \frac{1}{Q^2} + \Delta\right) . \quad (4.4)$$

Note that, in the WKBJ limit far from corotation, equation (4.3) corresponds to a wave solution characterized by

$$K_r = K_r^0 \pm (\delta K) . \quad (4.5)$$

Originally the theory developed in the regime of tightly wound spiral arms ($J = 0$) had been questioned by some authors, based on the argument that the important long-wave branch could disappear for spiral waves with finite inclination. The analysis that led to equation (4.3) showed that when J is relatively *small*, the theory of spiral modes is not changed qualitatively and that the relevant correction enhances the WASER process by easing the transfer of angular momentum across the corotation circle. This can be described by introducing the effective stability parameter

$$\frac{1}{Q_{\rm eff}^2} = \frac{1}{Q^2} + \frac{1}{4} J^2 Q^4 . \quad (4.6)$$

We wish to emphasize that equation (4.3), based on the fluid model, is supported by the results of the stellar dynamic theory (Bertin and Mark 1978).

The approximation used in deriving equation (4.3) is such that the "marginal" growth case is characterized by $Q_{\rm eff} = 1$, i.e. $\delta K = 0$ at $\nu^2 = 0$. In other words, within this approximation we are identifying the upper boundary of the envelope of figure 1 by means of the equation

$$\frac{1}{Q^2} + \frac{1}{4} J^2 Q^4 = 1 . \quad (4.7)$$

This simple relation is accurate only for small values of J. A more general turning point equation can be constructed in the

following way. We *define* K_r^0 as the exact double root of the dispersion relation (3.1), where the short and the long wave branches meet (i.e., as the value of K_r along the upper part of the envelope of Fig. 1). Then we *define* δK as the difference between K_r and K_r^0 (see eq. [4.5]). In this way we have implicitly formulated a turning point equation of the form (4.3) that can be easily handled by numerical investigation of equation (3.1). This equation has the advantage of possessing a "marginal" growth line in exact agreement with the upper part of the envelope of Figure 1.

b) Numerical Analysis

Numerical results on normal spiral modes, in regimes that are suitable for the simple treatment discussed in the previous subsection, have been presented in a number of papers, after the first announcement by Bertin *et al.* (1977) based on stellar dynamic equations, and the first results obtained using Pannatoni's code (Pannatoni and Lau 1979; Pannatoni 1983).

Figure 6 will show the properties of an exact normal spiral mode for a model in a sequence that was studied in Paper I. A simple calculation based on the integration of the ordinary differential equation (4.3) can reproduce its main characteristics very well. The α-spectrum for the normal spiral mode is also shown in Figure 6. On the trailing side, the long and the short waves are not detected separately, even though they are known to cooperate in the mode maintenance. As shown in the examples given in Figure 2 of the paper by Haass, Bertin, and Lin (1982), the tightly wound spirals appear to be dominated by a peak in the energy distribution in the trailing waves. A small feature on the leading part of the spectrum can also be present. This is interpreted as a "contamination" due to the possibility of a leading wave cycle. This issue of "contamination," which adds more continuity to the transition between normal tight and open spiral modes, has been studied in detail by Lin and Yue (1989).

The propagation diagram for the same normal spiral mode is given in another frame of Figure 6. Here it is very easy to identify the inner reflection point of equation (4.3), as determined by the properties of the equilibrium model in the bulge region. The relevant wave cycle, which is excited by conversion of long trailing waves into short trailing waves at the corotation circle, can also be recognized as well as the possibility of contamination with the "leading resonant cavity."

The role of long waves is best illustrated in Figure 4. Here we show the first two-armed modes for a model described in Paper I; the density contours of the modes has been shown in Figure 7*a* of that paper. The α-spectrum clearly reveals the presence of long waves and their relative importance in the mode maintenance.

Therefore we recognize that normal spiral modes with a small pitch angle may have a complex wave composition; on the other hand, the simple approximate theory based on an all-trailing wave cycle gives a fairly accurate prediction and description of the relevant global modes, so far as the pitch angle does not exceed 10^0.

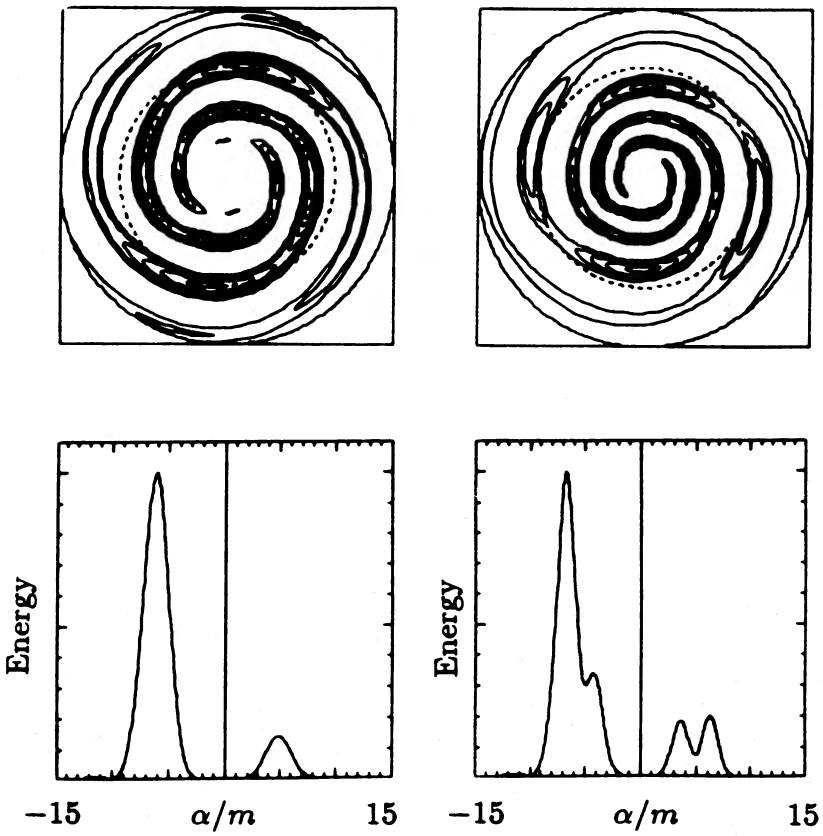

FIG. 4.—Examples of normal spiral modes with small pitch angle. The two dominant modes are displayed for a model discussed in Paper I (see Fig. 7*a*, Paper I). Their α-spectra (*bottom*) show a contamination by leading waves; note, however, the clear appearance of long waves in the second mode on the right.

V. THE REGIME OF OPEN SPIRAL STRUCTURE

In this section we restrict our attention to those conditions where the cubic dispersion relations (3.1) admits only *one* real root for K. This holds when the parameter J is fairly high ($J \gtrsim \frac{3}{4}$, for example).

a) A Turning Point Equation

Since there is only one real solution for K, there is only one real solution for μ^2 (see definition [3.7]). Thus we have

$$\mu^2 = \hat{\mu}^2(v^2; J, Q, s) , \tag{5.1}$$

where by $\hat{\mu}^2$ we denote the analytic relationship. Then we can construct the following associated differential equation:

$$\frac{d^2Y}{dx^2} + \left(\frac{m}{R}\right)^2 \hat{\mu}^2(v^2; J, Q, s)Y = 0 , \tag{5.2}$$

where $x = R \ln (r/R)$ and R is a reference radius. This ordinary differential equation leads to the dispersion relation (5.1) or (3.1). The physical meaning of Y can be left open for the moment. In fact, it is not important in the determination of the eigenvalues of the problem.

Equation (5.2) should be integrated with an appropriate boundary condition to yield eigenmodes. *A radiation* boundary condition should be applied to correspond to the propagation of waves from the corotation zone to infinity without reflection. The resulting growth can be interpreted in terms of a WASER process between leading ($\mu > 0$) and trailing ($\mu < 0$) waves.

In order for the WASER process to be characterized by moderate overreflection, the function $\hat{\mu}^2$ should be close to a double zero for $v^2 \sim 0$ (see following § V*b*). Therefore it is convenient to expand the dispersion relation in powers of μ^2 and to examine the special case where $\mu^2 < 1$. In this situation we may expect to obtain a good approximation by neglecting terms $O(\mu^4)$, as was effectively done earlier by Bertin (1983*a*).

For $\mu^2 \ll 1$, equation (3.1) can be approximated by

$$\frac{Q^2}{4} = \frac{\sqrt{\chi_0(s)}}{J} - \frac{(1 - v^2)\chi_0(s)}{J^2[1 - v^2 + \chi_0(s)]} + \left\{ -\frac{\sqrt{\chi_0(s)}}{2J} + \frac{(1 - v^2)^3\chi_0(s)}{J^2[1 - v^2 + \chi_0(s)]^2} \right\}\mu^2 , \tag{5.3}$$

where we have made use of definitions (3.6) and (3.12). Then in the vicinity of the corotation region ($v^2 \ll 1$) we can expand equation (5.3) further and find

$$\hat{\mu}^2 = \mu_1^2\left[v^2 + \frac{\mu_{co}^2}{\mu_1^2} + O(v^4) \right], \tag{5.4}$$

where μ_{co} and μ_1 are functions of Q, J, and $\chi_0(s)$. Thus in this case of small μ^2 we have derived a simple expression for the function $\hat{\mu}^2$ (see eq. [5.1]).

In Appendix B we study this regime of open spiral structure ($\mu^2 < 1$) by direct inspection of the Euler-Poisson equations, *without* the intermediate step of deriving equation (3.1). The resulting differential equation differs from equation (5.2) with the specification of equation (5.4) by the addition of a term in the derivative of the first order. Numerically, the difference is found to be unimportant (Lowe 1988). In addition, the analysis identifies Y as the enthalpy perturbation (multiplied by a slowly varying factor).

b) The Overreflection Factor

From the above expressions we can easily discuss the overreflection process at corotation. We consider a feedback cycle of leading and trailing waves. The leading wave approaching the corotation zone is reflected as a departing trailing wave with higher energy density (or flux of angular momentum), accompanied by another departing trailing wave traveling to large radii, were this is absorbed (radiation boundary condition). To calculate this overreflection factor we use the WASER formalism. If we neglect terms $O(v^4)$, then Eqs. (5.2) and (5.4) lead to the following formulae for overreflection:

$$\Gamma_r = \Gamma_f + 1 , \tag{5.5}$$

$$\ln \Gamma_f = \pi\mu_{co}^2 \Big/ \left(\frac{s\Omega}{\kappa}\mu_1\right) . \tag{5.6}$$

The first relation is just the equation for the conservation of wave action across the corotation circle. The quantity Γ_r is the factor of overreflection, i.e. the ratio of the flux of angular momentum in the reflected (trailing) wave to that in the leading wave approaching the corotation circle from the central regions of the galaxy.

Note that the formula (5.6) can be used to estimate the overreflection of open waves even when μ_{co}^2 is not small. For this purpose, one should apply the expansion (5.4) directly to the general relation (5.1) without considering the intermediate expansion (5.3). To be sure, when μ_{co}^2 is not small, equation (5.6) shows that high growth occurs. In the context of global modes this would also imply high growth rates, i.e. a large imaginary part of ν. These situations of high growth require a more careful treatment of the overreflection process, especially when applied to the estimate of the growth rates of the global modes. Therefore expressions such as equation (5.6) are best suited to describe regimes of moderate growth where the expansion (5.3) is a reasonable approximation.

In Figure 3 we had indeed shown contours of constant index of overreflection $I_r \equiv \ln \Gamma_r$ in the (J, Q)-plane, for the special case $s = 1$ bounding the region B of moderate growth. In the same figure, below the transition region, we also draw the analogous contours for the regime of normal spiral structure. For the regime of open spiral structure the contours generally resemble the χ-contours shown by Bertin, Lin, and Lowe (1984). The contour $I_r = \ln 2$ is the "marginal" growth line that would correspond to $Q_{\rm eff} = 1$, in the notation used in § IV. This curve is easily identified because it corresponds to $\mu_{co} = 0$ (see eq. [5.6]). From equation (5.3) we see that it is determined by the equation:

$$\frac{Q^2}{4} = \frac{\sqrt{\chi_0(s)}}{J} - \frac{\chi_0(s)}{J^2[1 + \chi_0(s)]} . \tag{5.7}$$

By varying s, equation (5.7) describes the "marginal" growth lines for open waves. These curves were shown earlier by Lin and Bertin (1985) (see also previous discussion in § III*d*). Note that *necessary* conditions for local instability, i.e. for having $\mu_{co}^2 > 0$, are

$$Q^2 < 1 + \chi_0 , \tag{5.8}$$

which identifies the maximum of the "marginal" growth curve, and

$$J > \frac{\sqrt{\chi_0}}{1 + \chi_0} , \tag{5.9}$$

which essentially identifies the transition to the regime of open spiral structure. The latter condition corresponds to Toomre's condition for the efficiency of the "swing amplifier" (Toomre 1981).

As an application of these results we may consider Toomre's example (Toomre 1981; Fig. 8, p. 125) of "swing" amplification. The situation that he studied is characterized by flat rotation curve ($s = 1$, $\chi_0 = 2$), relatively high dispersion speed ($Q = 1.5$), and moderately high active disk mass ($J = 2^{1/2}/2$). Note that both conditions (5.8) and (5.9) are satisfied. In fact, this example lies slightly below the "marginal" growth curve for open waves, and it is subject to moderate amplification. This result can be used to interpret Toomre's example and is obviously in contrast with the predictions obtained by extrapolating the regime of normal spiral structure. Indeed, the use of the dispersion relation for tightly wound spiral waves (eq. [4.1] with $\Delta = 0$) would have indicated a wide area of forbidden propagation around the corotation circle (as shown by the dotted circle in Toomre's Fig. 8), but such an application is not valid in this regime of open spiral structure.

A more interesting application of our analysis is the use of our equation (5.2) to calculate *global open modes*, as we shall discuss in § V*c*. This is possible because in our approach we can easily include the role of *inhomogeneities* in the galaxy model and therefore of feedback mechanisms that may operate in the inner regions. Mathematically, from the propagation diagram associated with equation (5.1) it is found that equation (5.2) can have a simple turning point at a radius not necessarily close to the galactic center. Thus the WASER process at corotation is supplemented by a feedback process that can maintain a global mode. Discrete self-excited global open modes are then obtained as a result of proper amplitude and phase matching at the turning points, as required by the appropriate boundary conditions of evanescent wave at the origin and of outgoing wave outside the corotation region. With the overreflection factor obtained, it is then easy to establish the approximate relationship

$$2\gamma\tau = I_r \tag{5.10}$$

where τ is the time of the feedback cycle, calculated on the basis of equation (3.17), properly modified to take into account the difference between the value of Γ_r for $\gamma = 0$ and that for $\gamma \neq 0$. We note that equation (5.10) provides an *estimator* for the growth rate γ of a mode in the following sense. For each assumed value of Ω_p we may calculate a value for γ, even when Ω_p is not an eigenvalue. Therefore we obtain an estimate of $\gamma(\Omega_p)$ for all *potential* values of Ω_p, even before calculating the physically admissible values of Ω_p.

c) *Numerical Analysis*

In order to consider a model suitable for applications of this regime of open spiral structure, we should take cases of relatively high J. Figure 6 shows the density contours of an exact mode for a model in a sequence that was studied in Paper I. The mode is very open. In the figure the dotted corotation circle gives an indication of the "size" of the bar, which is much longer than the "short" bar in some of the modes presented by Haass, Bertin, and Lin (1982). It is more similar to the simple example presented by Bertin (1983*a*). Modes of this type can be obtained by integrating directly the simple differential equation (5.2). The α-spectrum for the exact mode is also shown in Figure 6. The leading and the trailing waves are not detected as separate peaks, even though they are known to cooperate in the mode maintenance. The α-spectrum is thus quite different from that of a normal spiral mode (see the relevant frames in Fig. 6). The mode is maintained and amplified by the process of feedback and overreflection, as described in § V*b*.

For the simple ordinary differential equation (5.2), a synthetic representation of its accuracy and of its domain of application is given in Figure 5. For a number of modes, comparison is made between the value of the growth rate vertical axis predicted by the ordinary differential equation and the actual exact value determined by the integro-differential equation. Perfect agreement would have all points concentrated on the dashed line at 45°. Scatter around this line is found to be within acceptable limits. The detailed properties of the survey used to construct Figure 5 are given by Lowe (1988). Here we just mention that this diagram is a direct proof of the flexibility of the asymptotic theory of open modes, since many of the modes considered are not as "clean" as the prototype of Figure 6, especially when the transition region of Figure 3 is approached. Thus a certain degree of extrapolation is allowed from the conditions described in the present section, yet the numerical agreement is still very good.

d) *Various Types of Bar Modes*

In the previous Paper I we have addressed the astrophysical interpretation of open bar modes and we have related some of the present work to the concept of bar-driven spiral structure. At this point, we would like just to comment briefly on the dynamics of bar modes.

Much of the past work on "bar instabilities" actually refers to N-body simulations that recognized the possibilities of "violent instabilities" leading to more stable configurations characterized by a bar shape (e.g., see Miller, Prendergast, and Quirk 1970; Hohl 1971). In contrast, in the regime that we have considered, as in the example by Bertin (1983*a*), the bar mode has only a moderate growth rate. In this sense it may differ from some "violent" bar modes reported in the literature (which have presumably been calculated in different parameter regimes; note that the modes presented by Haass, Bertin, and Lin 1982 have a relatively small bar and high growth rates). On the other hand, some numerical simulations may just have started out in the high growth region C of the (J, Q)-diagram (see Fig. 3) at relatively low values of Q. Then the resulting violent instability (which initially is expected to have an open *spiral* appearance; see the high growth prototype of Fig. 6) may lead to a rapid increase of the dispersion of stellar velocities and eventually to a new state of equilibrium. Therefore the disk, in the absence of a gaslike component, could have evolved mostly along a vertical line in the (J, Q)-plane and reached the domain of moderate growth characterized by open bar morphology. However, we feel that interpretations of N-body experiments should be examined more carefully, in view of the many delicate issues involved (see Lin and Bertin 1985). It is to be regretted that Athanassoula and Sellwood (1986) did not present sufficient information between modal shape and growth rate in their identification of modes.

VI. CONCLUSION

In this article we have examined the physical properties of spiral modes calculated in a fluid model by means of a local dispersion relation. This equation is cubic in the magnitude of the total dimensionless wavenumber K and, for any value of

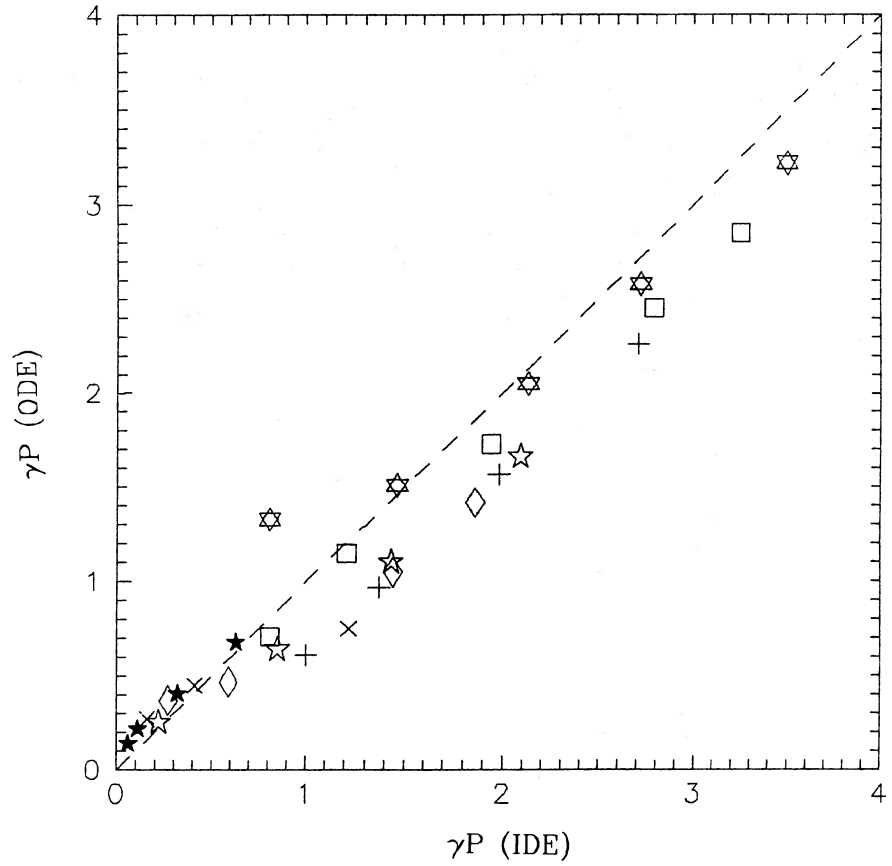

FIG. 5.—Test of the asymptotic theory. For a number of modes the predictions of the simple ordinary differential equation for open waves (see this § V) are compared with the exact results obtained from the integro-differential equation (see § II and Paper I). Various regimes are covered: for $r_Q = 2$, $Q_{OD} = 1$(✡), $Q_{OD} = 1.2$(☆), $Q_{OD} = 1.4$(★); for $r_Q = 3$, $Q_{OD} = 1$(□), $Q_{OD} = 1.2$(◇); for $r_Q = 4$, $Q_{OD} = 1$(+), $Q_{OD} = 1.2$(×). The general agreement is very satisfactory.

TABLE 1

MODE PROTOTYPES

SB0	SB(s)	S (low growth)	S (high growth)
$Q_\infty = 1.500$	$Q_\infty = 1.000$	$Q_\infty = 1.000$	$Q_\infty = 1.000$
$\Delta = 15\%$	$\Delta = 15\%$	$\Delta = -35\%$	$\Delta = 15\%$
$r_Q = 2.0$	$r_Q = 6.0$	$r_Q = 2.0$	$r_Q = 2.0$
$r_{cut} = 2.0$	$r_{cut} = 6.0$	$r_{cut} = 8.0$	$r_{cut} = 2.0$
$r_\Omega = 2.0$	$r_\Omega = 2.0$	$r_\Omega = 1.5$	$r_\Omega = 2.0$
$\Omega_P = 15.0$	$\Omega_P = 13.8$	$\Omega_P = 26.1$	$\Omega_P = 26.7$
$\gamma P = 0.10$	$\gamma P = 0.52$	$\gamma P = 0.60$	$\gamma P = 3.00$
$r_{co}/h = 2.27$	$r_{co}/h = 2.49$	$r_{co}/h = 1.29$	$r_{co}/h = 1.21$
$r_{olr}/h = 3.96$	$r_{olr}/h = 4.31$	$r_{olr}/h = 2.27$	$r_{olr}/h = 2.20$
$r_{ce}/h = 0.57$	$r_{ce}/h = 0.30$	$r_{ce}/h = 0.69$	$r_{ce}/h = 0.47$
$J_{co} = 0.604$	$J_{co} = 0.538$	$J_{co} = 0.492$	$J_{co} = 0.858$
$Q_{co} = 1.500$	$Q_{co} = 1.096$	$Q_{co} = 1.002$	$Q_{co} = 1.004$
$s_{co} = 0.954$	$s_{co} = 0.961$	$s_{co} = 0.922$	$s_{co} = 0.855$

NOTE.—The data in each column refer to those for one morphological type shown in the collection of four patterns in the upper left frame of Fig. 6. SB0 is the upper left pattern in this collection, SB(s) is the upper right pattern, S (low growth) is at lower left, and S (high growth) is at lower right.

the dimensionless frequency v, and hence may have three real solutions, but it admits only *one* real solution above a certain *transition line* in the (J, Q)-diagram (see Fig. 1). In general, on this basis we can identify *six* types of waves, or wave branches. They are the short, long, and open waves, and in each case we can distinguish trailing and leading waves. The analytical theory of normal spiral modes is based on short and long (trailing) waves. The theory of open barlike modes is based on open waves of leading and trailing forms. Note that open waves have the same propagation direction as short waves. It has been shown that the two different kinds of modes occur in separate domains in the space of dynamical parameters [see the (J, Q)-diagram, Fig. 3]. The two domains are connected by a "transition region" where a mixture of various wavecycles is expected to take place, and the corresponding modes have features that appear in both normal spiral structure and barred spiral structure.

Modes of general surveys such as the one described in Paper I can be studied, interpreted, and predicted in terms of these dynamical mechanisms. A summary of the basic prototypes of modes that correspond to various regimes identified in the present paper is given in Figure 6 (see also Table 1). Note that the open spiral mode shown in the lower right corner is supported by a leading-trailing wave cycle, but is not expected to

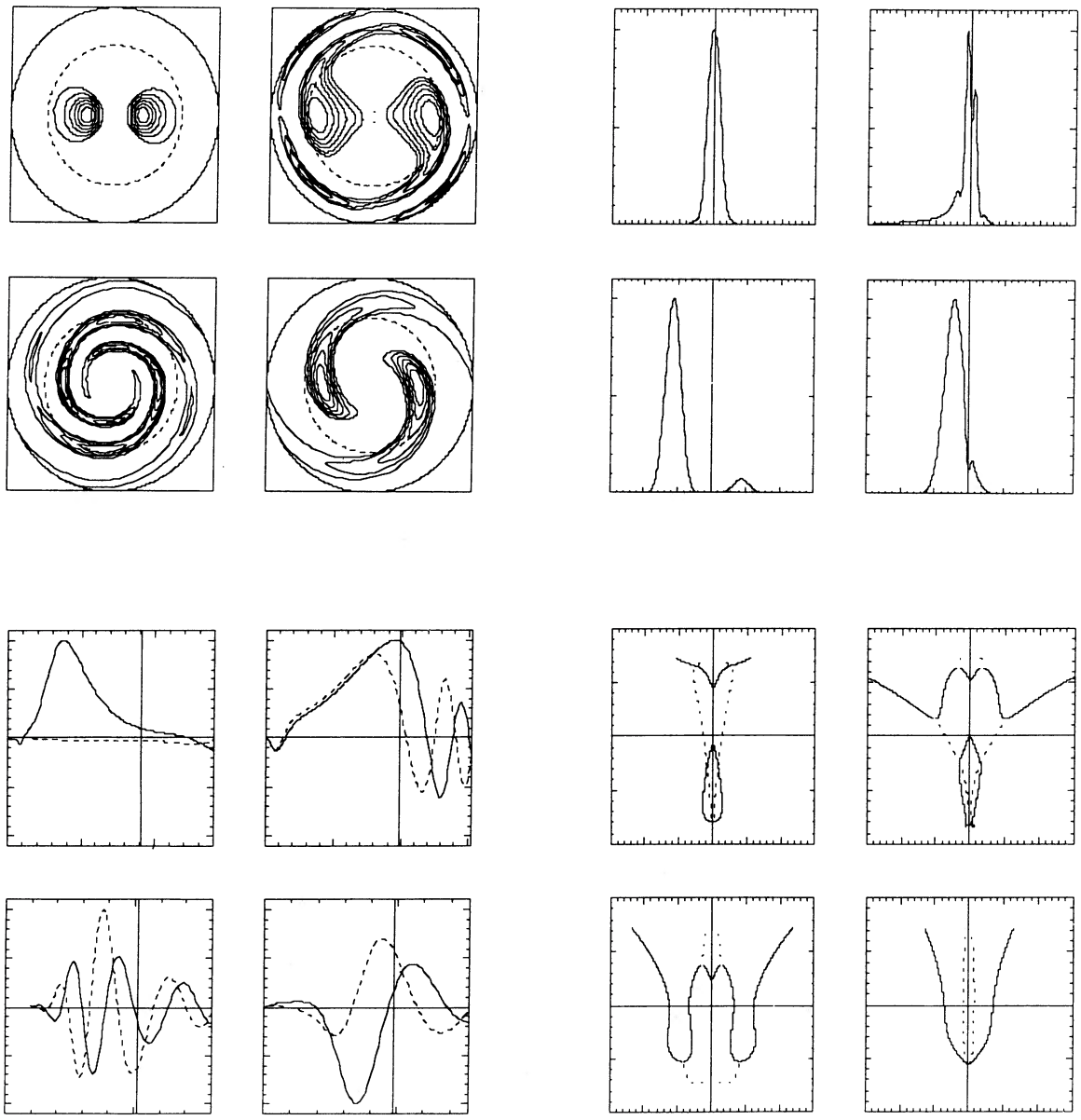

FIG. 6.—Mode prototypes: four key morphological types are compared (SBO, SB(s), and S, all with moderate growth; a violently unstable S mode at the low right corners). The model shapes are on the top left part of this composite figure. The α-spectra (*top right*) are given in arbitrary units; the horizontal axis has α/m from -15 to $+15$. The real and imaginary part of the mode eigenfunction (*bottom left*) are given as a function of r; the vertical line indicates the location of the corotation radius. The propagation diagrams (*bottom right*) have v on the vertical axis from -1 (*bottom*) to $+1$ (*top*); on the horizontal axis μ runs from -15 to $+15$. Numerical data for each of these prototypes are given in Table 1. Note that, even in the case where the mode is supported by a leading-trailing cycle, the waves are refracted outward by the bulge region and do not reach the center.

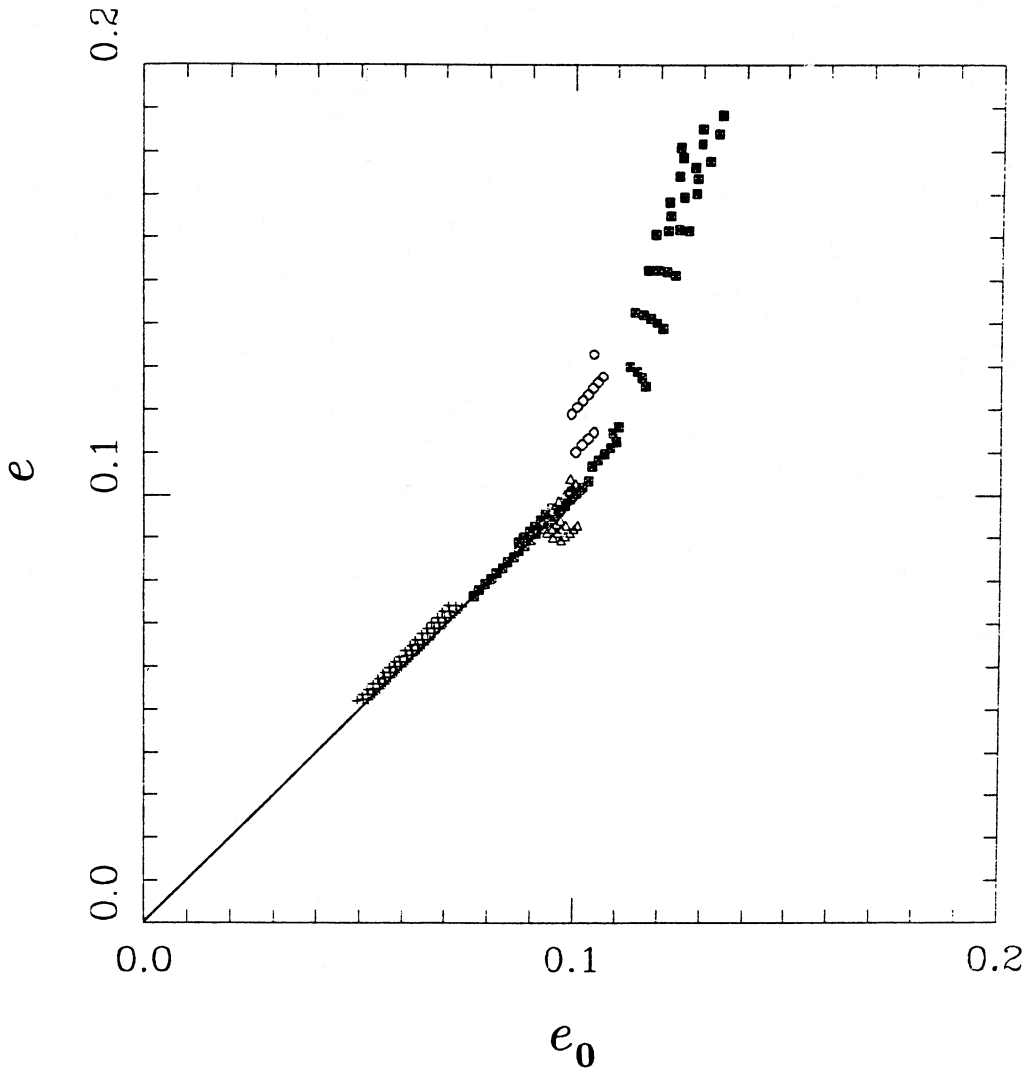

FIG. 7.—Sequence of models selected to be subject to modes with moderate growth rate ($\gamma P \approx 0.8$). The horizontal scale, $e_0 = (\frac{1}{4})2\pi G\sigma/r\Omega^2$, is a measure of the active surface density, the vertical scale, $e = c/r\kappa$, is a measure of the equivalent dispersion speed. Each symbol represents the conditions at the corotation circle for a model taken from the numerical survey of Paper I. The straight line identifies the condition $Q = 1$. This figure is the empirical counterpart of the (J, Q)-diagram discussed in Fig. 3.

be observed because it is violently unstable. The fact that extensive modal surveys such as that of Paper I can be categorized in terms of the simple analytical theory presented here is further illustrated in Figure 7. In this figure the models that are displayed are selected on the basis of the moderate growth rate criterion. The data indeed group themselves roughly in two straight lines with a rather sharp turn at $(e_0, e) = (0.1, 0.1)$. The lower part corresponds to region A in Figure 3 (*normal spirals*), the upper part to region B in Figure 3 (*open barred spirals*).

It is interesting to note that the same modes that are used as prototypes for illustration of the dynamical mechanisms in this paper are also the basic prototypes for the extensive survey of Paper I which has allowed us to draw a complete correspondence between modal shapes and Hubble morphological types of spiral galaxies.

APPENDIX A

ON THE DERIVATION OF THE LOCAL DISPERSION RELATION IN AN INHOMOGENEOUS DISK

The dispersion relation has been derived by following the procedure and the arguments given by Lau and Bertin (1978) under the ordering defined by equations (3.8) and (3.9). The present equation (3.1) generalizes their dispersion relation (see their equation [12]) in that here we have retained the $(1 - \nu^2)$-dependence of the B term (see their eq. [B9] and our eq. [2.5]). In turn, we find a $(1 - \nu^2)$-dependence in the denominator of the second term of the right-hand side of equation (3.1). Note that the original Lau-Bertin dispersion relation is obtained from the present equation (3.1) by replacing $J^2/(1 - \nu^2)$ with J^2.

The $(1 - v^2)$ factor had been omitted for simplicity, since Lau and Bertin (1978) focused on the cases where J^2 is small and on the marginal case ($v^2 = 0$). Obviously their dispersion relation coincides with equation (3.1) when $v^2 = 0$. In addition, results based on the study of the dispersion relation are not affected qualitatively by the $(1 - v^2)$-term. In fact, in recent announcements of results on open spiral structure (Bertin 1983*a*; Lin and Bertin 1985) we made use of the original Lau-Bertin relation. However, for quantitative applications, equation (3.1) should give more accurate results. This explains why, in a more complete investigation (starting with the paper by Bertin, Lin, and Lowe 1984), we have decided to make use of equation (3.1).

Here we list a few points where some quantitative differences between the present more general dispersion relation and the Lau-Bertin equation are found. (i) The expression for $Q_{\rm eff}$ (4.6) is slightly changed (compare the present eq. [4.4] with eq. [3] in Lau and Bertin 1978 or eq. [71] in Lin and Lau 1979). (ii) The slope of the transition line changes (compare eq. [3.14] with the condition stated in Lin and Bertin 1985, p. 520). However, the location of the cuspidal point P_c (see eq. [3.13] is unchanged. (iii) The overreflection factor (5.6) is slightly modified, but the marginal growth line for open waves (5.7) is unchanged. In particular, conditions (5.8) and (5.9) can be derived also from the original Lau-Bertin dispersion relation.

As noted by Hunter (1983) the Lau-Bertin relation is open to criticism because some "out-of-phase" terms were omitted. Here we do not add anything new to this point. We just recall that the terms mentioned above were calculated and listed by Lau and Bertin (1978) and refer the reader to the arguments given in that paper for their omission. The merits of equation (3.1) stand on its wide range of applicability in studying and in interpreting "exact" results and on its simplicity. These features would be lost in more complete formulations (for the regime of open spiral structure; see Appendix B).

A comment should also be made on the resonant term at corotation. Here we follow the arguments and the quantitative analysis by Bertin and Haass (1982), who showed that the stellar corotation resonance is weak and thus the fluid resonant term should be omitted. Their analysis strictly applies to the regime of normal spiral structure alone, but we expect their conclusion to hold qualitatively unchanged, even in the regime of open spiral structure. Note that in certain galaxy models the fluid resonant term is unimportant because the relevant gradient is small.

For those applications where the corotation zone is gas-dominated, and thus subject to a more "fluid" behavior, the role of the corotation resonance term should be reconsidered.

APPENDIX B

THE DIFFERENTIAL EQUATION FOR OPEN WAVES

A differential equation for open waves can be derived directly from the Euler-Poisson equations without the intermediate step of constructing the local dispersion relation (3.1). We start with equation (2.2) that we rewrite as

$$\mathscr{L}(h_1 + \psi_1) = \frac{(1 - v^2)}{\epsilon^2} h_1 , \tag{B1}$$

with

$$\mathscr{L} = \frac{d^2}{d\lambda^2} + H \frac{d}{d\lambda} + M . \tag{B2}$$

Here $\lambda = x/R$ is a dimensionless logarithmic radial coordinate (see eq. [5.2]), and H, M correspond to the functions A, B (see eqs. [2.4], [2.5]) of the operator L (eq. [2.3]). The epicyclic parameter ϵ is defined as $\epsilon = \epsilon_0 Q$. Then the Poisson equation reads

$$\left(\frac{\partial^2}{\partial\lambda^2} + R^2 e^{2\lambda} \frac{\partial^2}{\partial z^2} - m^2\right)\psi_1 = 4\pi G R^2 e^{2\lambda} \sigma_1 . \tag{B3}$$

The regime of open waves (which corresponds to $\mu^2 \ll 1$ of § V) is defined by the ordering:

$$m \gg 1 , \tag{B4}$$

$$\frac{1}{m} \frac{d}{d\lambda} \ll 1 , \tag{B5}$$

and

$$\epsilon \ll 1 . \tag{B6}$$

On the basis of equations (B4) and (B5), the Poisson equation (B3) admits the local solution

$$\psi_1 \cong -\left[1 + \frac{1}{2m^2}\left(\frac{d^2}{d\lambda^2} + \frac{d}{d\lambda}\right)\right]\left(\frac{2}{m\epsilon Q} h_1\right) . \tag{B7}$$

This can be used to eliminate ψ_1 from equation (B1) and obtain the following equation for h_1:

$$\left[-\frac{2}{m\epsilon Q}\left(1+\frac{1}{2m^2}M\right)+1\right]\frac{d^2h_1}{d\lambda^2}+H\left(1-\frac{2}{m\epsilon Q}\right)\frac{dh_1}{d\lambda}$$
$$+\left[M\left(1-\frac{2}{m\epsilon Q}\right)-\frac{(1-v^2)}{\epsilon^2}-H\frac{d}{d\lambda}\left(\frac{2}{m\epsilon Q}\right)-\frac{M}{m^2}\left(\frac{d^2}{d\lambda^2}+\frac{d}{d\lambda}\right)\left(\frac{2}{m\epsilon Q}\right)\right]h_1=0\,. \quad \text{(B8)}$$

Contributions due to higher order derivatives are found to be small. If (i) we drop the corotation resonance term in M (see eq. [2.5]), (ii) we drop the gradients of the quantity (ϵ, Q), and (iii) we omit the first-order derivative term in equation (B8), then the derived differential equation coincides with equation (5.2) for $\mu^2 \ll 1$, provided we identify Y with h_1. The resonant contribution (i) is understood in a satisfactory way (see also comment in Appendix A). The role of the effects due to terms related to (ii) and (iii) should be discussed by means of quantitative numerical analysis. This latter test has been performed by Lowe (1988), who found the differences with respect to equation (5.2) to be unimportant.

APPENDIX C

COMMENTS ON THE PROCESS OF OVERREFLECTION

In the present paper we have described the process of overreflection by means of the WASER formalism based on the use of a simple dispersion relation (see §§ IV, V). A different formulation which holds in a "corotation zone" has often been applied to study the relevant excitation mechanisms in a large class of problems ranging from plasma physics to meterology. This considers a *homogeneous* sheet with *uniform* shear and refers to *"swinging" wave packets* (for the problem of spiral structure, see Goldreich and Lynden-Bell 1965; Julian and Toomre 1966; Goldreich and Tremaine 1978; Toomre 1981; for the general hydrodynamic problem, see Marcus and Press 1977; Tung 1983, and references therein). Hunter (1983) remarked that "the question of how to fully reconcile this transient kind of instability with normal mode analysis remains." Drury (1980) had taken an important step in showing that the amplification resulting from the conversion of leading into trailing waves can indeed be described in the context of *steady wave trains.* A more complete discussion of these issues was provided by Lin and Thurstans (1984), who contrasted the behavior between wave trains and swinging wave packets and showed how such a diverse behavior can be unified in a *spectral approach.* Thus it is found that "swing amplification" has a counterpart in the context of steady wave trains. In particular, it is not necessarily transient, nor necessarily powerful. An important issue that is more easily resolved in the modal approach of the present paper, where *inhomogeneity* and the boundaries are recognized from the beginning, is the "joining" of solutions near the corotation zone with those outside such a zone in the general field. In terms of swinging wave packets in the homogeneous uniform shear sheet this issue remains as yet unresolved.

The *spectral approach* considered by Lin and Thurstans (1984) refers to the Fourier representation:

$$\Psi(x, y, t)=\sum_{m=-\infty}^{\infty}\int_{-\infty}^{\infty}\Psi_m(\xi, t)e^{i(\xi x-\eta y)}\,d\xi\,, \quad \text{(C1)}$$

where $x = R \ln (r/R)$ and $y = R\theta$ are Cartesian coordinates localized at the corotation circle and $\eta = m/R$ is the azimuthal wavenumber. The general solution for disturbances in the homogeneous uniform shear model can be written as

$$\Psi_m(\xi, t)=\sum_{i=1}^{2} f_i^{(m)}(\xi+\bar{c}t)\hat{\Psi}_i^{(m)}(\xi)\,, \quad \text{(C2)}$$

where $\bar{c} = s\Omega\eta$, f_i are two *arbitrary* functions, and $\hat{\Psi}_i^{(m)}(\xi)$ are two linearly independent solutions of the *time-independent* "oscillator equation" (see Lin and Thurstans 1984, eq. [7]). By a proper choice of the arbitrary functions one can describe quasi-steady wave trains or swinging wave packets. The two different limits and the properties of the oscillator equation are discussed at length by Lin and Thurstans (1984).

Pegoraro and Schep (1986) point out that in the spectral representation of an inhomogeneous problem one should allow for a discontinuity at $k = 0$, to be discussed in terms of the appropriate physical boundary conditions.

Note that even though the different approaches are equivalent, the use of swinging wave packets tends to emphasize fast evolution of spiral features, and therefore it is best suited to describe the *early* stages of an evolution process. On the other hand, the *modal* approach brings out the slow evolution of the spiral grand design in the *later* stages of an evolution process. The contrast between early and late stages involves a discussion of the propagation time, the shear-rate time, and the time required for the relevant astrophysical processes (such as star formation) to take place. Indeed, the "modal behavior" of a disturbance is expected to take over very quickly, especially when the role of the boundaries is properly taken into account. To appreciate the generality of these arguments, one may consider an example in the classical theory of heat conduction, as discussed by Jeffreys and Jeffreys (1956, pp. 563–565), or the so-called "Telegraph equation," as discussed by Doetsch (1943, pp. 366–369).

REFERENCES

Aoki, S., Noguchi, M., and Iye, M. 1979, *Pub. Astr. Soc. Japan*, **31**, 737.
Athanassoula, E., and Sellwood, J. A. 1986, *M.N.R.A.S.*, **221**, 213.
Bardeen, J. M. 1975, in *Dynamics of Stellar Systems*, ed. A. Hayli (Dordrecht: Reidel), p. 297.
Bertin, G. 1980, *Phys. Rept.*, **61**, 1.
———. 1983*a*, in *IAU Symposium 100, Internal Kinematics and Dynamics of Galaxies*, ed. E. Athanassoula (Dordrecht: Reidel), pp. 119, 174.
———. 1983*b*, *Astr. Ap.*, **127**, 145.

Bertin, G., and Haass, J. 1982, *Astr. Ap.*, **108**, 265.
Bertin, G., Lau, Y. Y., Lin, C. C., Mark, J. W.-K., and Sugiyama, L. 1977, *Proc. Nat. Acad. Sci.*, **74**, 4726.
Bertin, G., Lin, C. C., and Lowe, S. A. 1984, in *Plasma Astrophysics*, ed. J. Hunt and T. D. Guyenne (ESA SP-207), p. 115.
Bertin, G., Lin, C. C., Lowe, S. A., and Thurstans, R. P. 1989, *Ap. J.*, **338**, 78 (Paper I).
Bertin, G., and Mark, J. W.-K. 1978, *Astr. Ap.*, **64**, 389.
———. 1979, *SIAM J. Appl. Math.*, **36**, 407.
Contopoulos, G. 1983, *Astr. Ap.*, **117**, 89.
Doetsch, G. 1943, *Theory and Applications of Laplace Transforms* (New York: Dover).
Drury, L. O. C. 1980, *M.N.R.A.S.*, **193**, 337.
Erickson, S. A. 1974, Ph.D. thesis, Massachusetts Institute of Technology.
Goldreich, P., and Lynden-Bell, D. 1965, *M.N.R.A.S.*, **130**, 125.
Goldreich, P., and Tremaine, S. 1978, *Ap. J.*, **222**, 850.
Haass, J. 1982, Ph.D. thesis, Massachusetts Institute of Technology.
Haass, J., Bertin, G., and Lin, C. C. 1982, *Proc. Nat. Acad. Sci.*, **73**, 3908.
Hohl, F. 1971, *Ap. J.*, **168**, 343.
Hunter, C. 1983, in *Fluid Dynamics in Astrophysics and Geophysics*, ed. N. R. Lebovitz (Providence: American Mathematical Society), p. 179.
Jeffreys, H., and Jeffreys, B. S. 1956, *Methods of Mathematical Physics* (Cambridge: Cambridge University Press).
Julian, W. H., and Toomre, A. 1966, *Ap. J.*, **146**, 810.
Kalnajs, A. J. 1965, Ph.D. thesis, Harvard University.
———. 1977, *IAU Symposium 77, Structure and Properties of Nearby Galaxies*, ed. E. M. Berkhuijsen and R. Wielebinski (Dordrecht: Reidel), p. 113.
Landau, L. D. 1946, *J. Phys. USSR*, **10**, 25.
Lau, Y. Y., and Bertin, G. 1978, *Ap. J.*, **226**, 508.
Lau, Y. Y., Lin, C. C., and Mark, J. W.-K. 1976, *Proc. Nat. Acad. Sci.*, **73**, 1379.
Lin, C. C., and Bertin, G. 1984, *Adv. Appl. Mech.*, **24**, 155.
———. 1985, in *IAU Symposium 106, The Milky Way Galaxy*, ed. H. van Woerden, R. J. Allou, and W. B. Burton (Dordrecht: Reidel), p. 513.
Lin, C. C., and Lau, Y. Y. 1979, *Studies Appl. Math.*, **60**, 97.
Lin, C. C., and Thurstans, R. P. 1984, in *Plasma Astrophysics* ed. J. Hunt and T. D. Guyenne (ESA SP-207), p. 121.
Lin, C. C., and Yue, Z. Y. 1989, in preparation.
Lowe, S. A. 1988, Ph.D. thesis, Massachusetts Institute of Technology.
Lynden-Bell, D., and Kalnajs, A. J. 1972, *M.N.R.A.S.*, **157**, 1.
Marcus, P., and Press, W. H. 1977, *J. Fluid Mech.*, **70**, 525.
Mark, J. W.-K. 1976, *Ap. J.*, **205**, 363.
Miller, R. H., Prendergast, K. H., and Quirk, W. J. 1970, *Ap. J.*, **161**, 903.
Pannatoni, R. F. 1979, Ph.D. thesis, Massachusetts Institute of Technology.
———. 1983, *Geophys. Ap. Fluid Dyn.*, **24**, 165.
Pannatoni, R. F., and Lau, Y. Y. 1979, *Proc. Nat. Acad. Sci.*, **76**, 4.
Papaloizou, J. C. B., and Pringle, J. E. 1984, *M.N.R.A.S.*, **208**, 721.
Pegoraro, F., and Schep, T. J. 1986, *Plasma Phys. Cont. Fusion*, **28**, 647.
Toomre, A. 1964, *Ap. J.*, **139**, 1217.
———. 1981, in *The Structure and Evolution of Normal Galaxies*, ed. S. M. Fall and D. Lynden-Bell (Cambridge: Cambridge University Press, p. 111.
Tung, K. K. 1983, *J. Fluid Mech.*, **133**, 443.
Zang, T. A. 1976, Ph.D. thesis, Massachusetts Institute of Technology.

G. BERTIN: Scuola Normale Superiore, Pisa 56100 I, Italy

C. C. LIN, S. A. LOWE, and R. P. THURSTANS: Massachusetts Institute of Technology, Room 2–330, Cambridge, MA 02139