SECTION III.3

SPIRAL STRUCTURE AND STAR FORMATION

Thursday 2 June, 1645 - 1850

Chairman: D. Lynden-Bell

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

Above: Conclave of density-wave theorists; left to right: Shu, Bertin, Lin, Yuan. Background: Alladin and G.D. van Albada. CFD

Below: Confrontation with observers. Foreground: Cohen, Lin, Yuan, with Salukvadze in front. In background: Dame, D.M. Elmegreen, Oort, Ostriker, Kormendy, Van der Laan, Elmegreen, Wouterloot, Kutner, Alladin, Mead. LZ

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

FORMATION AND MAINTENANCE OF SPIRAL STRUCTURE IN GALAXIES

C. C. Lin and G. Bertin*
Massachusetts Institute of Technology
Cambridge, MA. USA
*Permanently at Scuola Normale Superiore, Pisa, Italia

ABSTRACT. The formation and maintenance of spiral structure in galaxies is discussed in terms of spiral modes with a general perception consistent with the hypothesis of quasi-stationary spiral structure. The latter is explained in some detail to give it the original proper perspective and to contrast it with recent studies of non-stationary spiral structures. We again emphasize the possible coexistence of fast-evolving spiral features and slow-evolving spiral grand designs. Spiral modes are described in terms of three categories of propagating waves. The reason is given why isolated non-barred galaxies with spiral grand design are found to be mostly two-armed. Some suggestions are made for future research.

I. INTRODUCTION

Research work on the formation and maintenance of large-scale spiral structure has been pursued along several lines of approach. In this paper, we shall attempt to clarify the relationship among the various approaches by presenting a coherent description of one of them and comparing the conclusions for this approach with those obtained from other alternative approaches on a few crucial issues.

The basic mathematical formalism adopted is the theory of linear spiral modes; these modes are in turn described in terms of steady wave trains. To connect our theoretical analysis with observations, we continue to adopt the hypothesis of quasi-stationary spiral structure (QSSS hypothesis), whose precise nature will be explained in some detail.

To give a proper perspective to our discussions, we again quote from Oort (1962) for his statement of the problem:

> "In systems with strong differential rotation, such as is found in all nonbarred spirals, spiral features are quite natural. Every structural irregularity is likely to be

H. van Woerden et al. (eds.), The Milky Way Galaxy, 513–532.

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

> drawn out into a part of a spiral. But this is not the phenomenon we must consider. We must consider a spiral structure extending over the whole galaxy, from the nucleus to its outermost part, and consisting of two arms starting from diametrically opposite points. Although this structure is often hopelessly irregular and broken up, the general form of the large-scale phenomenon can be recognized in many nebulae."

A few comments will now be made to clarify the above statements.

Galaxies may or may not exhibit a grand design. But it is the grand design that challenges the theorists and the observers alike. The theorists should be able to explain why it is possible for grand designs to exist, in spite of the winding dilemma. The observers should find out how the various features in the grand design are related to each other. The theorists should again explain why. Indeed, such efforts to try to explain the grand-design structure lead us to find out more about the astrophysical processes that prevail in the interstellar medium and in the process of star formation. The key step that made such investigations possible is the adoption of the hypothesis of quasi-stationary spiral structure (QSSS) as a working hypothesis in a semi-empirical context. (See Lin, 1971, pp. 89, 91.)

Some unnecessary controversy has arisen around this hypothesis in recent years. There are at least the following two reasons for this controversy: (1) The hypothesis has sometimes been misunderstood and given a role beyond that of a working hypothesis. This confusion can be easily cleared away by referring to the original literature, as is done in this paper. (2) There have been studies of dynamical mechanisms which suggest the possibility of spiral patterns in rather rapid evolution, not quasi-stationary structures. At the meeting in Besançon last year, only one aspect of the latter issue was addressed. In this paper, we shall attempt to deal with the issues more fully. Before we go on with further details, we should perhaps make a few remarks on the general perception.

From laboratory experiments on hydrodynamic stability and turbulence and from the study of weather patterns, -- both of which can be observed to evolve in time, -- we know that some flow patterns are quasi-stationary while others are transient. There are advantages in focussing our attention on regular quasi-stationary structures. In the study of galaxies, the adoption of the QSSS hypothesis has enabled us to pursue the study of a number of astrophysical phenomena and astrophysical processes (see Section V for a short list of examples). We therefore first focus our attention on the study of quasi-stationary phenomena, and leave the study of transient processes to separate investigations. Clearly, the studies of quasi-stationary and transient phenomena are complementary rather than competitive, unless it can be shown that quasi-stationary structures cannot possibly exist or that they can exist only under exceptional circumstances. Observational evidence gives us assurance that this is not so (cf. paper by D.M. Elmegreen at this conference). At the same time, we shall devote a part of this paper (Section II.3) to the discussion of studies which apparently yield transient impressions and point out why they do not, in fact, conflict with the possible existence of quasi-stationary spiral features.

Actually, this issue was already quite carefully examined by Bertil Lindblad (1963) twenty years ago, when he discussed the implications of some pioneering computational studies by Per Olof Lindblad. Specifically, his paper was entitled "On the possibility of quasi-stationary spiral structure in galaxies", and this possibility was recognized despite indications from the computational studies that the spiral structure may be more likely to be "quasi-periodic" (as described in Lindblad's paper), or regenerative (as described in the paper by Goldreich and Lynden-Bell, 1965).

Besides these general issues, two further points should be noted in the above quotation from Oort. (1) Attention is directed towards non-barred spirals as being the more challenging. (2) Even these galaxies preferentially have two-armed structures. In Section IV, we shall show why this is the case even though we know that there are isolated galaxies with other types of symmetry (cf. Iye et al., 1982 for a specific example). As it turned out, Landau damping of waves at inner Lindblad resonance holds the key; multiple-armed spiral structures are more likely to encounter inner Lindblad resonance as already noted in the early nineteen-sixties. Without inner Lindblad resonance, multiple-armed spiral features may indeed overwhelm two-armed spirals (see Section IV.4).

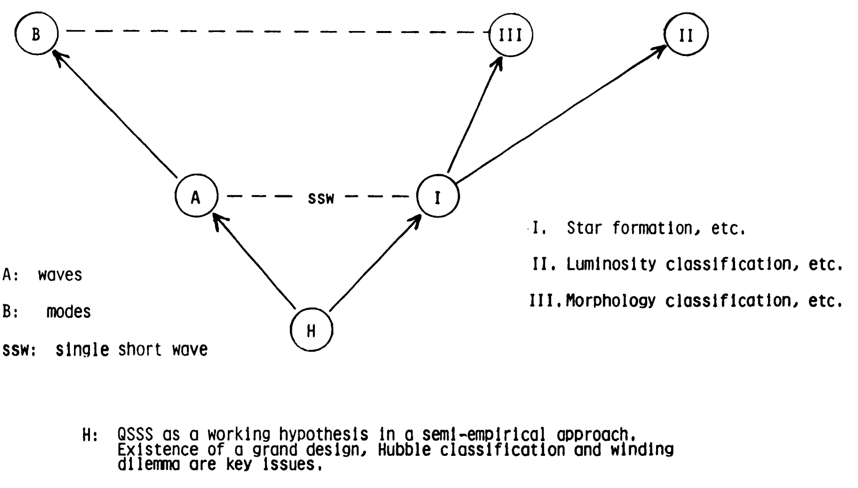


Fig. 1. The hypothesis of quasi-stationary spiral structure.

II. THE HYPOTHESIS OF QUASI-STATIONARY SPIRAL STRUCTURE

1. The Semi-Empirical Approach. As a working hypothesis in a semi-empirical approach, its correctness is primarily to be judged by the inferences made from it. We require consistency both with observations and with internal dynamical mechanisms. As mentioned above, the QSSS hypothesis has led to a number of successful applications to observed phenomena, a partial list of which will be given in Section V. We call special attention to the application to the process of star formation and to the possibility of providing the basis for the luminosity and the morphology classifications of galaxies. On the other hand, we also derive support from dynamical studies, which will be given in some detail in this paper.

The research efforts may, therefore, be represented in terms of a picture of "mountain climbing", cf. Toomre (1977) and Tremaine (1983). Both these authors appear to emphasize a deductive approach from first principles. From our experience with the much simpler system of homogeneous fluids, we know that it would be quite difficult to "deduce" the existence of regular patterns in the full non-linear theory. Rather, we may attempt to describe such patterns by first accepting their plausible existence on the basis of experiments (see examples of regular and irregular patterns given by Figures 128, 131, 137, 158, 177 in van Dyke's Album, 1982). We then attempt to reach a better understanding of the underlying mechanisms that distinguish irregular flow patterns from regular grand designs. These mechanisms are, however, not always easy to describe.

The QSSS hypothesis receives support from the theory of linear spiral modes which has been quite successfully developed in recent years, since the coexistence of a small number of such modes, slowly growing from a certain basic state, provides a natural basis for the eventual existence of a quasi-stationary spiral structure. However, as is well known in the studies of hydrodynamic stability and turbulence, the calculation of linear modes "merely marks the beginning of the dynamical studies", and the same comment applies to the study of galactic spirals (Lin and Lau, 1979, p. 107). The existence of linear modes does not imply that the spiral structures are always necessarily permanent in shape. "The complicated spiral structure of galaxies indicates the coexistence of material arms and density waves -- and, indeed, of the possible coexistence of several wave patterns. These features [in general] influence but do not destroy one another." (Lin, 1971, p. 91.)

The basic time rate in all dynamical studies is of course the shear rate or the epicyclic frequency. However, linear modes in general evolve at a much slower rate. [Indeed, if the time scale for a feed-back cycle in a linear mode were comparable to that of its evolution, the "mode" would hardly be a useful concept from a physical point of view.]

Experience in hydrodynamical studies shows that the observed nonlinear states which are qualitatively similar to slowly evolving linear modes usually also evolve slowly. Thus, the study of linear modes gives a powerful method, perhaps the most convenient, to approach the problems of formation and maintenance of spiral structure on a global scale, since they do represent the intrinsic characteristics of the dynamical system. Thus, even the effect of external excitation may be described in terms of these modes. As an analogy, we may recall that essentially a slowly decaying quasi-stationary mode is excited through the ringing of a church bell.

2. The Single Short Spiral Wave. Here we wish to call attention to one fact which has not always been put in proper perspective. The phenomena listed in Section V have mostly been described by the use of a single short spiral wave. Such a short wave requires a feedback process (must be "replenished", see Toomre, 1969) for its long-term existence as a part of a modal formulation (Shu, 1968). The question is whether this feedback system can be observed. This will be discussed in the section on linear modes (Sections III and IV) where the feedback mechanism will be examined in some detail.

3. Nonstationary Impressions. The plausible existence of quasi-stationary spiral structures can be questioned if all studies suggest that the dynamical processes always imply rapid evolution of spiral structures. Thus, the following three issues deserve our special attention:

a. Rapid dynamical evolution is often found in N-body simulations;
b. strong amplification is associated with a "swinging" process;
c. strong "instability" seems to be present even when the disk is stabilized against axisymmetrical disturbances ($Q \geq 1$).

The first issue will be the main topic of this subsection. The third issue will be taken up in Section III. The second issue, discussed at Besançon (Lin, 1983b), will be briefly reviewed here. As an alternative to the time-dependent "swinging" approach (Toomre, 1981), Lin referred to Drury's work (1980), which does not involve a rapid evolution. It is therefore compatible with growing modes, even though both approaches may still be labelled "swing amplification", in the sense that they both describe the transformation of a leading spiral into an amplified trailing spiral. Bertin (1983b) presented still another analysis of leading-trailing wave amplification in terms of steady wave trains satisfying the local dispersion relationship. Again no transient behavior is implied, i.e., time evolution is not required. Rapid time evolution does occur in certain contexts; for example, as a result of certain special initial conditions (cf. Toomre, 1981). We shall take up this topic again in Sections III and IV in the context of modes.

Let us now turn to the first item. It is well known that there are apparent conflicts between results obtained through the use of numerical simulation and those obtained through analytical methods. There are perhaps three reasons for this. (1) The requirement on the number of particles is highly demanding when simulation of phase-mixing is needed. Often there are simply not enough particles in the phase-space. The simulation of $N_0 \sim 10^6 - 10^8$ particles by a single super-particle severely distorts the distribution function in velocity space. (2) The small number and the large mass of each superparticle magnifies the possible fluctuations of surface density and gravitational field by a factor of the order of $N_0^{1/2}$, which could be as large as $10^3 - 10^4$, thus greatly accelerating the evolutionary processes. These difficulties are not easy to remove. (3) Sometimes artificial physical processes are introduced to control effects such as the rapid rise of velocity dispersion. This could lead us to even worse situations, where it is no longer clear what physical system is being simulated. (The computation may become like a computer game which can be intellectually stimulating and even suggestive, but whose results cannot be accepted at face value. Proper interpretation is needed.)

All these ideas and discussions are not new. Computer experiments in plasma physics have been performed for at least two decades (see an excellent extensive review by Dawson, 1983). Millions of particles have been used, but experts are still demanding more in order to get a proper simulation of processes involving resonances, where phase mixing and Landau damping are important.

4. Some Additional Remarks. In addition to these studies giving nonstationary impressions, there is also the subtle mathematical point of the existence of an infinite number of modes with a continuous spectrum. This issue was examined by Lin and Bertin (1981), and found not to be serious.

In the spirit of the QSSS hypothesis, we therefore adopt the modal approach, because it is mathematically sound and it is convenient for the description of comparatively regular grand designs, including their evolution. The complete description of the evolutionary process must, of course, be based on the use of a nonlinear theory which is yet to be developed. The general nature of such developments can be partly visualized by examining existing theories for some simpler systems. For homogeneous fluids, the mathematical theory can be sketched in the following manner. Any dynamical variable $\psi(\underline{\mathbf{x}},t)$ may be represented in the form

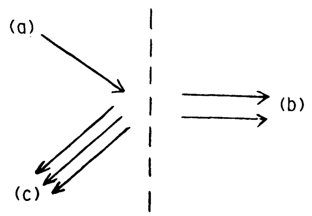
$$\psi(\underline{x},t) = \sum A_n(t)\phi_n(x)$$

in terms of a complete set of eigenfunctions $\phi_n(\underline{x})$. [Integration is implied by Σ when a continuous spectrum exists.] In the linear theory, each $A_n(t)$ is exponential in time t. In the full non-linear theory, the set of variables $\{A_n(t)\}$ satisfies nonlinear differential equations which describe the process of evolution. Extensive investigations along this line have been carried out, especially in meteorological studies. Galaxies are, of course, complicated systems with many components. Non-linear effects in the gaseous component have been shown to be likely to play an important role.

III. DYNAMICAL MECHANISMS: WAVES AND MODES

For the study of dynamical mechanisms, we shall focus our attention on linear processes, with one important exception. This is the issue of angular momentum transfer associated with the spiral patterns. Bertin (1983a) has shown that the angular momentum transfer associated with normal spiral modes is not going to produce a significant impact on the basic state over a period of several billion years even when the amplitude of the disturbance is about one-quarter of the basic distribution. Thus, we need not be unduly concerned with any fast evolution from this mechanism.

Linear modes can be described in terms of waves propagating in opposite directions, together with mechanisms of feedback from the central regions and over-reflection of these waves near the corotation circle. In earlier studies, we recognized two categories of waves: long waves and short waves. With leading and trailing configurations in each case, there are altogether four types of waves. Detailed studies based on the Lau-Bertin (1978) dispersion relationship now lead us to the recognition of a third category of waves: the open waves, again with possibly leading and trailing configurations. Open waves are more open than long waves, but propagate like short waves.



OVER-REFLECTION

Type I:	(a) long trailing wave
	(b) short trailing wave
	(c) short trailing wave
Type II:	(a) open leading wave
	(b) open trailing wave
	(c) open trailing wave

Fig. 2. Two types of over-reflection

In this new perspective, the two principal types of processes of over-reflection near corotation are shown in Fig. 2, which is amended from Fig. 1 of Lin's Besançon paper. The term over-reflection is used to describe a process in which the reflected wave contains more energy than the incoming wave. Over-reflection of Type I is the WASER (Mark, 1976) where only trailing waves are involved; that of Type II corresponds to the "swing amplification" (Toomre, 1981), in the sense that leading waves are converted into amplified trailing waves. Open waves are especially important for ranges of dynamical parameters for which there is only one real solution. For those cases, the distinction between open waves and short waves becomes less apparent.

The Lau-Bertin dispersion relationship is a cubic, with two parameters J and Q (see their paper for definitions of symbols). For the treatment of all three waves, it is therefore convenient to introduce the (J,Q) plane. The same format can be used for the study of the factor of over-reflection after the value of the shear parameter s is specified. In the case of constant linear velocity of circular motion, the shear parameter s equals unity.

Let us now briefly outline some of the principal results which differ from previous conclusions.

1. Waves in the (J,Q) Plane. First, the cubic dispersion relationship leads to the following overall outlook. The two real solutions representing the long/short waves can exist only if

$$JQ^2 \leq 4\sqrt{3}/9.$$

The third real solution represents an open wave, which is often too long to fit into the disk geometry, and should be excluded. However, this is not always the case. We may have to deal with all three categories of waves under certain circumstances (usually for J not too small). On the other hand, for

$$JQ^2 > 4\sqrt{3}/9 \quad ,$$

there is only one real solution, i.e., the open wave. The solutions which represent long/short waves have become complex conjugate pairs. It is in this range of parameters that the "swing" process is usually discussed.

Thus, in general, neither the long/short wave cycle nor the swing cycle represents the complete picture. We shall present examples of calculated spiral patterns in which the feedback processes include all three categories of waves (cf. Fig. 6 in Section IV).

2. Over-Reflection in Moderate Amounts. The second point is that one can identify a range of parameters in the (J,Q) plane such that the amplification factor (over-reflection) is moderate, whether the over-reflection is of Type I or of Type II (see Fig. 2). Moderate over-reflection is typified by an energy ratio of two (and an amplitude ratio $\alpha = \sqrt{2}$). The curve for this condition has been calculated by two methods. The system of curves shown in Fig. 3 is obtained from the approach used by Bertin in his Besançon paper, for different values of the shear parameter $s = -\, d\ln\Omega/d\ln r$. We also

checked Drury's calculations of the amplification factor for the case $s = 1$ by using an equation equivalent to Eq. (35) in Goldreich and Tremaine (1978). (Robert Thurstans did the computer work. Drury used a different, though equivalent, procedure.) The results are shown in Fig. 4 as curves of constant over-reflection (logarithmic scale). The two figures may be compared for $s = 1$, $\ln \alpha = 0.35$ [curve (b) in Fig. 4 versus curve marked $\varepsilon = 1$ in Fig. 3]. The numerical values are not identical because in Bertin's formulation the Weber equation is used, just as in Mark's original WASER formalism. The two approaches are indeed two different approximations, and there is no good reason to decide that one approximation is better than the other. Fortunately, they support each other within the proper limits of these approximations.

3. The (J,Q) Parameters for Moderate Over-Reflection. In the new perspective just discussed, one is led to emphasize the role of the parameter J (or J_1, its value for $m = 1$) in the determination of the stability characteristics of the spiral modes. As shown before (cf. Lin and Lau, 1979, Bertin, 1980), tightly wound spirals with moderate rates of growth are associated with small or moderate values of the J-parameter, and with a value of the Q-parameter close to unity. Such modes are maintained by a specific feedback mechanism.

Let us leave the feedback mechanism aside for the moment and focus our attention on the over-reflection factor at corotation. If one follows the curve of "moderate" over-reflection factor, $\alpha = \sqrt{2}$, in Fig. 4, one finds that $Q \simeq 1$ only if $J \leq 0.5$ (Axisymmetry implies $J = 0$). In contrast, for $J = 1$, very high over-reflection is reached at $(J, Q) = (1,1)$; the factor becomes moderate ($\alpha = \sqrt{2}$) only when Q is raised to about $\sqrt{3}$. Thus, the condition $Q = 1$ does not always have special significance in all galactic disks. It is essentially the condition to be expected in galaxies with relatively low disk mass (low J), typified by normal spirals of types Sa and Sb.

There is another reason for expecting to observe higher values of Q associated with stars. This is the presence of a small amount of gas, which may cause a significant reduction of the effective Q-parameter (see Lin and Shu, 1966). For the specific combination of stellar and gaseous disks considered, marginal stability requires that

$$Q_* \simeq 1 + 2(\sigma_g/\sigma_*) \quad .$$

Thus, for 30% gas-to-star mass ratio (which is not large when disk thickness is taken into account), we find that $Q_* = 1.6$. Combining this with the factor $\sqrt{3}$ discussed above, we see that $Q_* = 2.7$ would not be, in some cases, an unreasonable value. Of course, this calculation gives only an indication of the type of situation to be expected and the result should not yet be directly compared with observations (cf. Kormendy's paper at this meeting).

With respect to observational studies, one should note that the usual definition of Q is quite easily influenced by the presence of a

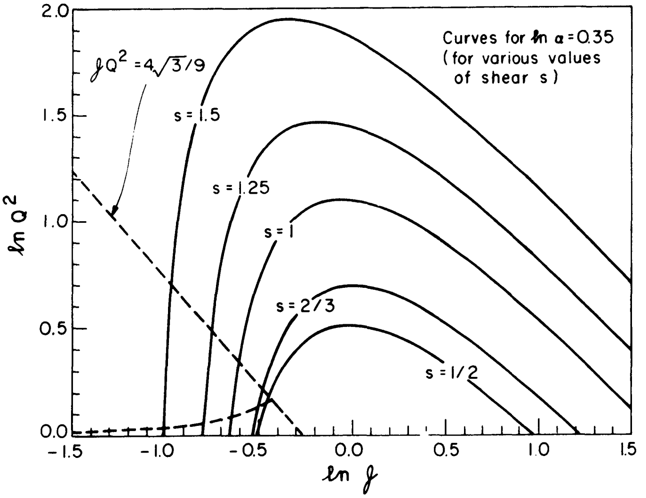


Fig. 3. Curves for a given coefficient of over-reflection ($\alpha = \sqrt{2}$) in the (J,Q) plane for various values of the shear parameter s.

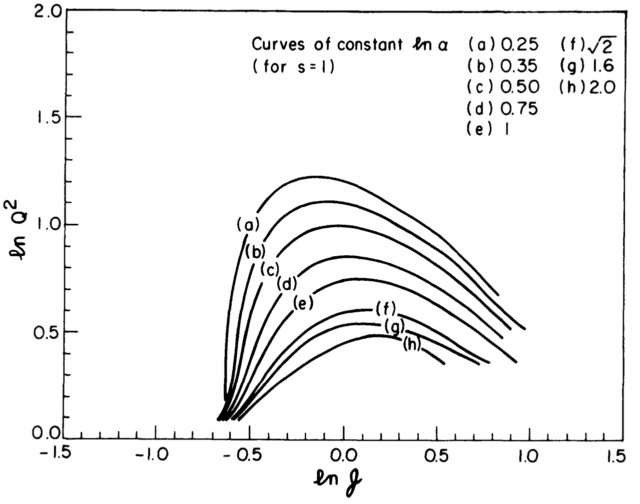


Fig. 4. Curves for various values of the coefficient of over-reflection α in the (J,Q) plane for a given value of the shear parameter (s=1).

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

small amount of stars with high velocity dispersion. This is the "long-tail" problem common to many statistical distributions. After all, the classical Cauchy distribution implies an infinite value for Q, if one applies the same definition as that for quasi-Maxwellian distributions. "Contamination" from other stellar components (e.g., halo stars) has to be removed. Basically, it may be wise to refine the definition of the stability parameter or even to introduce a new parameter that emphasizes the role of the responsive stars with low dispersion velocities.

IV. DYNAMICAL MECHANISMS: SPIRAL MODES AND MORPHOLOGICAL CLASSIFICATION

1. A Process of Self-Regulation. Many modal calculations have been done in the fluid model. The process of Landau damping at inner Lindblad resonance is therefore not incorporated in the calculations. An additional step must be taken to check the result for possible resonance effect. Such a step will make it apparent that the number of unstable spiral modes is expected to be reduced, because the feedback mechanism may be seriously impaired by the Landau damping process. (See subsection 4 below for further details.)

If we combine these considerations with the above discussions of the parameters $\mathcal{J}$ and Q, we are led to the following scenario for the self-regulation of the instability of spiral modes. Suppose that the galactic system, to begin with, has $\mathcal{J} \simeq 1$ (for m = 3, say), and Q = 1. The excitation of spiral waves and modes, possibly violent, will tend to cause the stellar dispersion speed to increase until the disk is "heated up" to such high values of the Q parameter and such a form for its distribution that only a small number of spiral modes may eventually remain moderately unstable. This reduced instability also restricts the increase in the Q-parameter, and the final spiral structure may be expected to be quasi-stationary when nonlinear processes, involving both stars and gas, are included. [In this final state, the over-reflection factor may also be expected to be relatively moderate, but not necessarily very small.] It is also possible that the galaxy will evolve into a state where the dispersion speeds are so large that only a gentle oval distribution of stars results. Such a mass distribution would drive the Population I objects into a bar-like distribution, as shown by Roberts, Huntley and van Albada (1979).

The process of self-regulation just discussed is crucial in the context of the QSSS hypothesis, which can be invalidated by the presence of too many unstable spiral modes. We shall discuss it further in connection with some examples below. Having developed a scenario for the realization of quasi-stationary spiral structures, we are ready to venture to relate our theoretical investigations to one of the major objectives: the search for a dynamical basis for morphological classification. For this purpose, we must first be able to produce modes of various morphological types in the context of the linear theory. Based on experience with hydrodynamical experiments, especially those with rotating fluids, we may then infer that the grand designs in the final state are approximately simulated by the superposition of modes obtained from the linear theory.

2. Modes of Various Morphological Types. In previous publications, (see Lin and Bertin, 1981; Haass, Bertin and Lin, 1982) we have shown how one may use linear modes to simulate the various morphological types of the galaxies by judiciously prescribing three distribution functions: (a) the rotation curve, (b) the mobile component of the disk mass (including the effect of disk thickness) and (c) the distribution of dispersion speed of stars (or the acoustic velocity in the gaseous model). These different morphological types are often associated with different feedback mechanisms. For example, it is clear that the (r), (s) types relate to different degrees of propagation of waves through the center of the galaxies. The time is now ripe for relating the morphological type to dynamical mechanisms and modal calculations. Some examples were published by Haass, Bertin and Lin; some additional examples will now be cited.

Bertin (1983b)contrasted an open mode with a tightly wound spiral mode obtained from the original asymptotic theory. Specifically, their propagation diagrams were compared. The open mode shown by Bertin has an inner turning point for wave propagation. This implies a "reverse swing" of the wave and a rather long feedback cycle. It thus has only a moderate growth rate, smaller than that of the normal mode shown, even though the over-reflection factor is quite high, similar to that of the normal mode. Indeed, this open mode is not represented in Fig. 2 in Norman's Besançon paper (1983). The avoidance of violently unstable modes can be achieved through an increase of the Q-parameter.

For each of the two types of modes discussed by Bertin, we may assign a point in the (J, Q) plane according to the values of these parameters at the corotation circle. In Fig. 5, which shows a portion of Fig. 3, these points are located in the general areas marked I and II, corresponding respectively to the two types of over-reflection. We now note that there is a triangular region in the (J, Q) plane, marked III, corresponding to a gradual transition from one type of feedback to another.

To demonstrate this transition, a sequence of modes has been calculated for which the conditions at the corotation circle vary from Region I to Region II, traversing through Region III. The propagation diagrams show clearly that, in the cases shown in Fig. 6, it is the most open of the three waves that carries the primary signal across and around the corotation circle. The dashed lines show the real part of the complex conjugate pair of comparatively shorter waves. Their roles are, however, not insignificant. Thus, all three categories of waves are present, albeit some in modified forms. The over-reflection process and the feedback process are thus quite complicated; both types are involved. But all of these modes show moderate growth rates. Thus, the presence of over-reflection of Type II does not necessarily imply violent instability. For details, we refer to future publications.

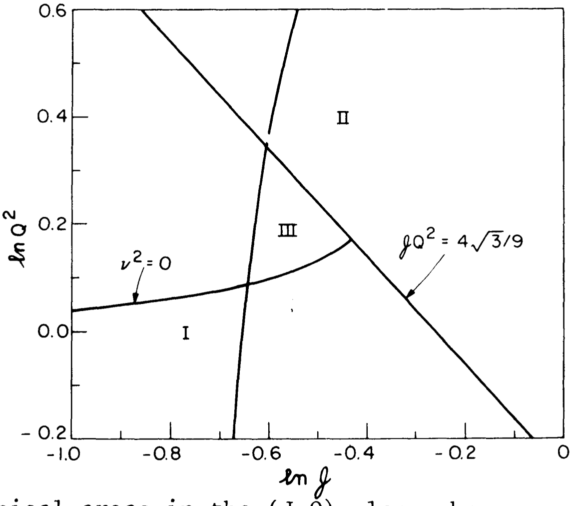


Fig. 5. Typical areas in the (J,Q) plane where over-reflections of Types I and II are important. These are respectively labelled I and II. In domain III, all six waves may be active. The nearly vertical curve is that for $\alpha = \sqrt{2}$ for the open waves. Unstable modes with moderate growth rates have been calculated for all three domains.

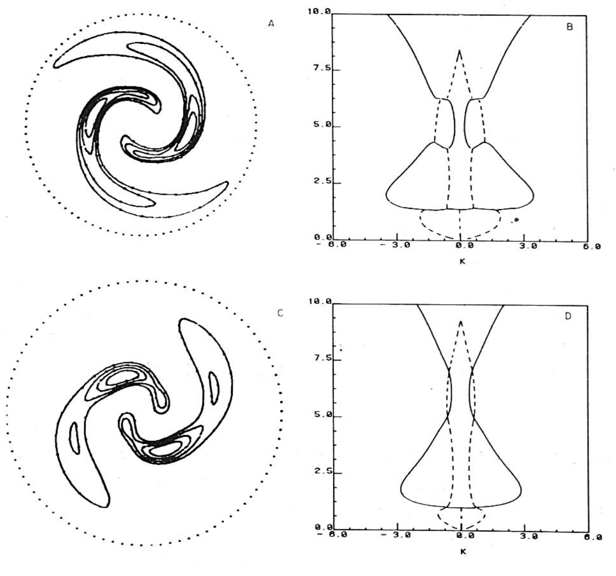


Fig. 6. Spiral patterns and propagation diagrams for two cases, A and C, where both types of over-reflection are active. Note that near co-rotation the open waves, which are real solutions represented by the solid lines, are the most important. For conditions far away from the co-rotation circle, the short waves are the important ones represented by solid lines. In Case A, (ℓn J, ℓn Q^2) = (-.511,+.191) near Region I; in Case C, (ℓn J, ℓn Q^2) = (-.371,+.365) in Region II.

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

3. Are Feedback Processes Observable? The short wave is assigned a predominant role in much of the application theory. What can be expected from the feedback process from an observational point of view? There is no general answer to such a question, but the following comments may be helpful.

In the case of normal spirals, the long trailing wave, which is responsible for the feedback process, produces a modulation on the short trailing wave. (See Bertin et al., 1977, p. 4728, and reference to Mark's earlier work.) For empirical purposes, it simply appears that there is a change of intensity of the spiral field with radial distance from the center. The simplicity of the situation results in part from the fact that in the amplified modes, the over-reflection process gives the short trailing spiral a considerable advantage. For a real galaxy, one can construct a scenario compatible with an over-reflection process of the same type by including the damping of such short waves by the shock wave in the gaseous component.

In the case of open modes such as that discussed by Bertin, the situation is somewhat different. The open waves lead to something like an oval distortion in the central regions, which exhibits itself through the gaseous response to a bar-like forcing. Again, the general impression may be that of only one trailing spiral extending out of a bar.

4. Why Two-Armed Modes? We now turn to the following important issue: why do we not see many three-armed grand designs in galaxies even though the example shown by Haass at the Besançon meeting exhibits higher growth for multiple-armed modes? The key to this puzzle lies with Landau damping at inner Lindblad resonance, which is not included in the fluid code used by Haass (cf. III.3, IV.1 above). In the example shown by Haass, the rotation curve has a nearly constant angular velocity at the center ($\Omega \simeq \Omega_0$, somewhat like that of M33). Thus, the curve for $\Omega - \kappa/2$ has very low values at the center, but the curve for $\Omega - \kappa/3$ has a peak value of $\Omega_0/3$ at the center. Thus, if a three-armed mode has a relatively large corotation circle (and hence a low Ω_p), it would tend to suffer from inner Lindblad resonance. Note that the three-armed modes shown by Haass (1983b) have a rather small corotation circle, since he has deliberately chosen the Q-distribution to have a very small scale, smaller than the scale for the velocity distribution, (V_{max}/Ω_0). In real galaxies, the Q-distribution may have a much larger scale. In such cases, one may regard the corotation circles as being pushed outwards as the scale for the Q-distribution expands. Lindblad resonance follows, and the mode is expected to be damped in the stellar model (even though this is not apparent in the gaseous model calculated). Two-armed modes, being relatively free from Lindblad resonance, thus become predominant. Notice, however, that low frequency two-armed modes would also suffer from inner Lindblad resonance, and consequently the number of unstable two-armed modes can become quite small.

It is obvious that the above line of reasoning applies to four-armed spirals with stronger justification. One-armed spirals do not suffer from inner Lindblad resonance, but they usually have lower growth rates than two-armed spirals. Observationally, their presence together with two-armed spirals is often not as easily detectable. But the well-known asymmetrical motion of the dust lanes in M31 might indeed be due to the presence of a one-armed structure. However, it is not yet entirely clear why galaxies like NGC4254 (Iye et al., 1982) apparently show only a spiral structure with odd harmonics.

V. CONCLUDING REMARKS

We have surveyed the study of dynamical mechanisms from a semi-empirical perspective. The following is a short list of the successful applications of the theory.

A Short List of Some Physical Phenomena Studied

1. Possible dynamical basis for the Hubble classification
2. Plausible dynamical basis for the luminosity classification (Roberts, Roberts and Shu 1975)
3. Process of star formation
4. Dust lanes: distribution
5. Synchrotron emission (e.g., in M51): distribution
6. Distribution and motion of atomic hydrogen (e.g., in the Galaxy and in M81; including the modal theory of Visser and Haass)

On the theoretical side, the following points appear to be worthy of our attention.

1. We should recognize the possible coexistence of (a) spiral grand designs which evolve rather slowly, and (b) less regular spiral features which evolve more rapidly, generally on the dynamical time scale of κ^{-1}, Ω^{-1} or $(s\Omega)^{-1}$. The hypothesis of quasi-stationary spiral structure is thus re-affirmed to provide a basis for the dynamical approach to the morphological classification of galaxies (cf. Fig. 1).

2. The mathematical model adopted and the mathematical formalism used are crucial to the plausible demonstration of the above statements. While the use of wave packets tends to give impressions that favor the less regular spiral features, the equivalent method of using steady wave trains and modes emphasizes the possible existence of long lasting grand designs. Numerical experiments can be misleading unless Landau damping is properly simulated.

3. It is important to recognize different behaviors of the spiral waves and patterns for different physical models and for different ranges of the dynamical parameters. Dynamical processes tend to re-adjust the distribution of the three

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press

basic parameters listed above (Section IV.2), especially the dispersion velocity of stellar motions. This process of self-regulation, aided by the presence of the gaseous component, may be crucial to the quasi-stationary existence of grand designs of various types (cf. Sections IV.1 and IV.2).

Looking into the future, we see that the study of the morphology of galaxies continues to require the collaboration of observers and theorists. In the dynamical approach to the morphological classification of galaxies, we adopt a logical framework based on our understanding of the dynamical mechanisms and on the spiral appearance to be expected under various dynamical situations. We then compare these conclusions with the morphological classification based on observations. Preliminary work (e.g., by Haass, Bertin and Lin, 1982) shows that this approach is highly promising.

Other promising lines of research include the analysis of observed spiral structures into Fourier components and logarithmic spirals, the study of bars and rings, especially in SBa galaxies, and the study of the dispersion of stellar velocities in external galaxies. Some of these studies are being carried out by various researchers. Because of limitation of space, comments on these topics have to be omitted from this paper.

It is perhaps unfortunate that inadequate appreciation of the theoretical points discussed above has led to considerable confusion in the literature. It is hoped that the present paper helps to clear away much of the confusion and to focus our attention on the substantive issues of attempting to explain the observed phenomena such as the different morphology of the galaxies.

VI. THE MILKY WAY

At this conference, various aspects of the structure of the Milky Way have been discussed, including the successful modelling of the outgoing motion of the 3-kpc arm by Yuan as a flow driven by a rotating bar. The resultant picture is quite similar to that appearing in the inner parts of NGC5364. For this paper, we shall restrict ourselves to some short comments and a few references.

We recall (see Lin and Yuan 1978, Lin 1983a) that there have been rather extensive theoretical examinations of the spiral structure in the solar neighborhood from both the observational and the theoretical points of view. Consistency with the two-armed pattern originally suggested by Lin and Shu (1967) has been found. But there is co-existence of other spiral features together with the two-armed pattern, which is found to be compatible with spiral modes calculated by Haass (1983 a). The Carina arm may be an extra feature associated with the condition $|\nu| = 1/2$ (Shu, Milione and Roberts, 1973). It is not surprising that multiple-armed features can exist in the outer parts of the Galaxy in addition to a dominant two-armed structure in the more massive inner part. It is also not surprising that an additional one-armed spiral structure can lead to asymmetries of the kind that appears to be observed in the fourth quadrant. However, the long stretches of spiral arms observed present an overall picture that shows a predominantly two-armed pattern (see Burton, 1976). Finally, it should be emphasized that the structure of individual spiral features such as the high-velocity stream outside of the Sagittarius arm and other streaming motions (see Burton and Shane, 1970) played an important role in the development of the density-wave theory by giving observational support to the basic concept, -- a support that remains valid today.

ACKNOWLEDGEMENT

We wish to thank Jon Haass for many helpful discussions and for permitting us to quote his unpublished results on the sequence of models that are calculated for the demonstration of the various types of over-reflection.

REFERENCES

Bertin, G.: 1980, Phys. Reports 61 (1), pp. 1-69 (217 references)
Bertin, G.: 1983a, Astron. Astrophys. 127, pp. 145-148
Bertin, G.: 1983b, IAU Symposium No. 100, ed. E. Athanassoula (Reidel), pp. 119-120
Bertin, G. et al.: 1977, Proc. Nat. Acad. Sci. 74, pp. 4726-4729
Burton, W.B.: 1976, Ann. Rev. Astron. Astrophys. 14, pp. 275-306
Burton, W.B., and Shane, W.W.: 1971, IAU Symposium 38, eds. W. Becker and G. Contopoulos (Reidel), pp. 397-414
Dawson, J.M.: 1983, Rev. Modern Phys. 55, pp. 403-447
Drury, L. O'C.: 1980, Monthly Notices Roy. Astron. Soc. 193, pp. 337-343
Goldreich, P., and Lynden-Bell, D.: 1965, Monthly Notices Roy. Astron. Soc. 130, 97, pp. 125
Goldreich, P., and Tremaine, S.: 1978, Ap. J. 222, pp. 850-858
Haass, J.: 1983a, in 'Kinematics, Dynamics and Structure of the Milky Way', ed. W.L.H. Shuter (Reidel), pp. 283-287
Haass, J.: 1983b, IAU Symposium No. 100, ed. E. Athanassoula (Reidel), pp. 121-124
Haass, J., Bertin, G., and Lin, C.C.: 1982, Proc. Nat. Acad. Sci. USA 79, pp. 3908-3912
Iye, M., Okamura, S., Hamabe, M., and Watanabe, M.: 1982, Astrophys. J. 256, pp. 103-111
Lau, Y.Y., and Bertin, G.: 1978, Ap. J. 226, pp. 508-520
Lin, C.C.: 1971, in 'Highlights of Astronomy', ed. C. de Jager (International Astron. Union, Reidel, Dordrecht, Netherlands), Vol. 2, pp. 88-121
Lin, C.C.: 1983a, in 'Kinematics, Dynamics and Structure of the Milky Way', ed. W.L.H. Shuter (Reidel), pp. 277-281

Lin, C.C.: 1983b, IAU Symposium No. 100, ed. E. Athanassoula (Reidel), pp. 117-118
Lin, C.C., and Bertin, G.: 1981, in 'Plasma Astrophysics', eds. T.D. Guyenne and G. Levy (European Space Agency, Noordwijk, Netherlands), SP-161, pp. 191-205
Lin, C.C., and Lau, Y.Y.: 1979, Studies in Appl. Math. 60, pp. 97-163
Lin, C.C., and Shu, F.H.: 1966, Proc. Nat. Acad. Sci. 55, pp. 229-234
Lin, C.C., and Shu, F.H.: 1967, IAU Symposium No. 31, ed. H. van Woerden (Academic Press, London), pp. 313-317
Lin, C.C., and Yuan, C.: 1978, 'Astronomical Papers dedicated to Bengt Strömgren', eds. A. Reiz and T. Andersen, published by Copenhagen University Observatory, pp. 369-386
Lindblad, B.: 1963, Stockholm Observ. Ann. 22, No. 5
Mark, J.W.-K.: 1976, Astrophys. J. 205, pp. 363-378
Norman, C.: 1983, IAU Symposium No. 100, ed. E. Athanassoula (Reidel), pp. 163-174
Oort, J.: 1962, in 'Interstellar Matter in Galaxies', ed. L. Woltjer (Benjamin, New York), pp. 234-244
Roberts, W.W., Huntley, J.M., and van Albada, G.D.: 1979, Astrophys. J. 233, pp. 67-84
Roberts, W.W., Roberts, M.S., and Shu, F.H.: 1975, Astrophys. J. 196, pp. 381-405
Shu, F.H.: 1968, Ph. D. Dissertation, Harvard University
Shu, F.H., Milione, V., and Roberts, W.W.: 1973, Astrophys. J. 183, pp. 819-841
Toomre, A.: 1969, Astrophys. J. 158, pp. 899-913
Toomre, A.: 1977, Ann. Rev. Astron. Astrophys. 15, pp. 437-478
Toomre, A.: 1981, in 'Structure and Evolution of Normal Galaxies', eds. S.M. Fall and D. Lynden-Bell (Cambridge University Press), pp. 111-136
Tremaine, S.: 1083, IAU Symposium No. 100, ed. E. Athanassoula (Reidel), pp. 411-416
Visser, H.C.D.: 1978, The dynamics of the spiral galaxy M81. Ph. D. thesis, University of Groningen, Netherlands
Van Dyke, M.: 1982, An Album of Fluid Motion (The Parabolic Press)

DISCUSSION

F.H. Shu: In this conference many people have commented on the destabilizing contribution of interstellar gas. I would like to remind people that interstellar gas can also have an opposite effect, namely to saturate growing normal modes by dissipative effects. A. Kalnajs pointed out in 1972 that galactic shocks introduce a phase shift between the response of the interstellar gas and the background stellar density wave that drives the gas. This tends to damp the wave; and W.W. Roberts and I suggested that this might turn out to be a good thing, if an instability mechanism lies behind spiral structure. The various instability mechanisms subsequently worked out and described in Lin's talk all have the following features in common: an amplifier (at corotation) plus feedback (via long, leading, or open waves). Thus, consider

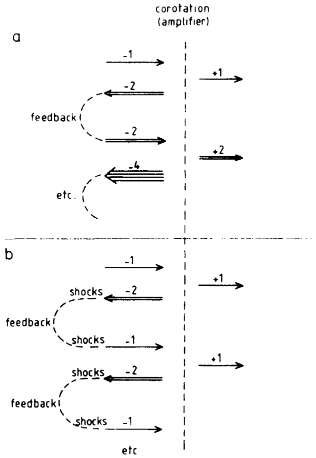


Figure D1 - a) Rapid growth of wave amplitude by amplification at corotation, followed by dissipationless feedback, leading to instability.
b) Quasi-stationary situation: zero growth of wave amplitude; amplification at corotation is followed by dissipative feedback, caused by galactic shocks.

(Figure D1a) the simplest generic example: propagation to corotation of a wave carrying -1 unit of angular momentum, transmission of +1, over-reflection of -2 (to conserve angular momentum), feedback of -2 (if there are no losses), transmission of +2, over-reflection of -4, etc. Thus, the square of the wave amplitude would grow by a factor of 2 per cycle, and may be in danger of growing to catastrophically high values.

Consider, however, how dissipation of the waves by galactic shocks changes this picture (Figure D1b). Start again with a wave approaching corotation and carrying -1 unit of angular momentum. There is transmission of +1, over-reflection of -2, followed now by shock dissipation to cut down the reflected wave as it propagates. By the time the wave is fed back, it might again have only -1 when it comes back to corotation, starting the cycle over again in nearly the same configuration as before. The resulting pattern would be quasi-stationary with evolution only on a slow timescale.

Why should such a quasi-stationary situation obtain? I think it is automatic. Because density waves are dispersive, they do not steepen into shocks until they reach a finite critical amplitude. Until the gaseous density wave reaches this amplitude, the overall wave (stars plus gas) would grow in time. Once the critical amplitude is exceeded significantly, strong shocks develop that halt the growth of the wave, giving a unique solution of finite amplitude.

This consideration may be important in a complementary context. Suppose that in some galaxy there is not enough gas to hold the mode amplitudes to fairly small values. The mode(s) may then grow catastrophically and heat the stellar disk enough to suppress all instabilities. Could something like this have happened to the "flocculent spirals"?

W. Renz: In this connection, I wish to refer to my non-linear calculation of the amplitude to which waves will be limited by damping due to shocks (W. Renz 1982, in "Applications of modern dynamics ...", ed. Szebehely, Reidel, Dordrecht, p. 356). Assuming, with Kalnajs (1972, Astrophys. Letters 11, p. 41), a relaxation time of about 10^9 years and taking amplification at corotation into account, I found the amplitude of a quasi-stationary wave to be about 5%.

P. Pismis: You stated at the beginning of your talk, Dr. Lin, that there would be little problem in having a driving mechanism for spiral arms in a barred galaxy, as the existence of the bar would give rise to such a mechanism. However, a good many barred spirals have a small, well-defined and tight spiral within the nucleus in the middle of the bar. Good examples are NGC 1097 and 4314. How would the existence of such features agree with the scheme of the density-wave theory?

Lin: I must not give the impression that everything is deductive. In the text of my paper I emphasize the semi-empirical approach of the density-wave theory. I would assume that the structures you describe may perhaps also be quasi-stationary. One has to look for the dynamical process that is responsible.

G.D. van Albada: At least in some cases the nuclear spirals observed in barred galaxies may be an artifact produced by the projection of a curved dust lane on the luminous nuclear region. This conjecture is based on the comparison of recent numerical results with e.g. the pictures shown in the Hubble Atlas.

J.V. Feitzinger: A crucial point is the time scale of growth rates.

Lin: Indeed, that is a very important point. We believe that the growth rates are moderate. Wielen's study of the increase of stellar velocity dispersions in the Milky Way points to the same conclusion. The time scale for these changes may indeed be too short in the numerical experiments (for reasons given in Section II.3 of the text).

D. Lynden-Bell: Sellwood has made thorough comparisons of the modes determined analytically by Toomre and by Kalnajs with his own detailed numerical calculations. He went as far as 10^5 bodies, without getting accurate growth rates, when starting with Poisson noise in phase space. However when he started with a special smooth procedure, he managed to get almost exact agreement. This may serve as a warning to those in the audience who make numerical simulations: it is hard work to get exact agreement with analytical results.

Lin: Was Lindblad resonance involved in that case?

D. Lynden-Bell: I must ask Sellwood about that.

J.A. Sellwood (answer supplied after Symposium, on request by Editor): No inner Lindblad resonances were involved in these models. The apparent growth rates were too high, because noise interfered with the measurements at early times. The amplitudes also soon became non-linear.

Lin: I would say that both approaches have their own limitations. The analytical theory certainly is now restricted to the linear case; it is a difficult step to go on to non-linear analysis. As to numerical simulations the plasma physicists, with their decades of experience, are very cautious about involving resonances.

https://doi.org/10.1017/S0074180900243052 Published online by Cambridge University Press