CONFLICTS AND DIRECTIONS IN SPIRAL STRUCTURE

C.A. Norman
Institute of Astronomy, Madingley Road, Cambridge.
Sterrewacht, Huygens Laboratorium, Wassenaarseweg 78, Leiden.

ABSTRACT

This is a critical and selective review of current theoretical and observational work related to spiral structure in disk galaxies. Productive future areas of research are suggested.

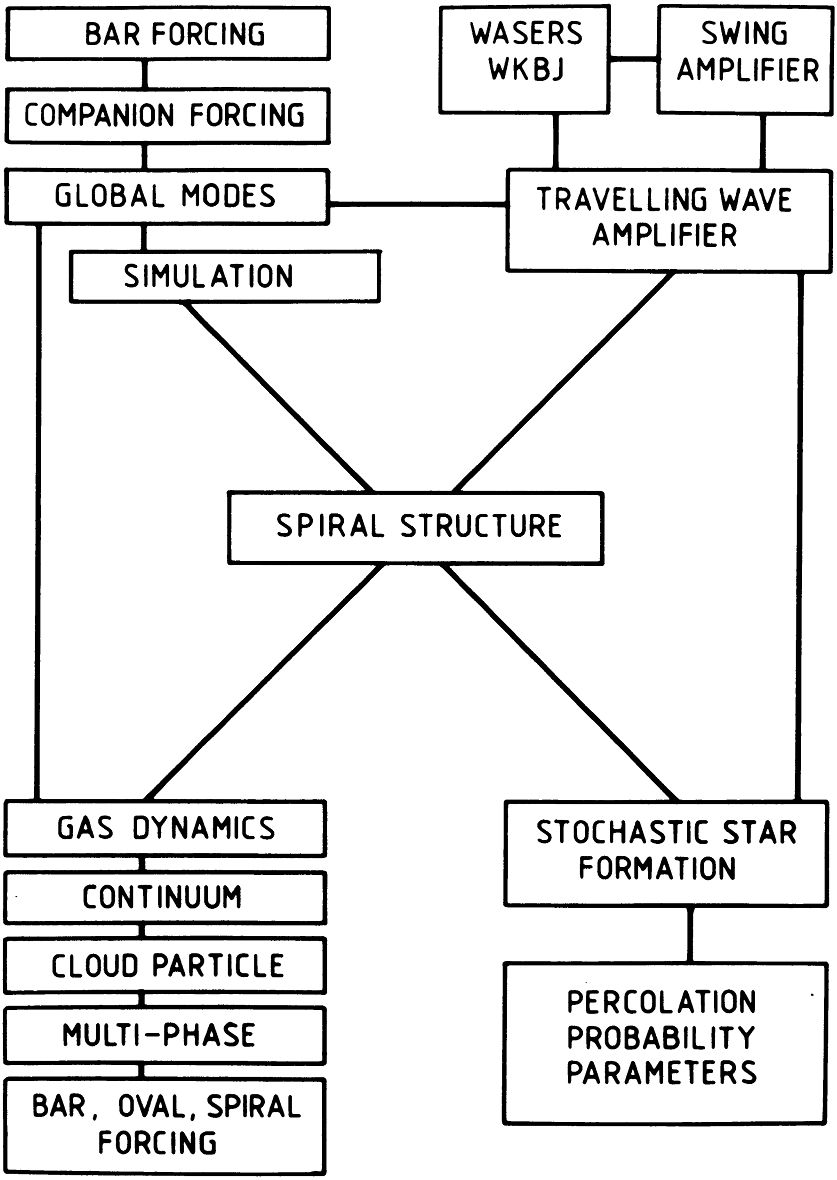
I. INTRODUCTION.

This area of astronomy contains some very interesting interrelated physical approaches that need to be discussed and evaluated in the context of the whole grand attempt to understand spiral structure. I have chosen to illustrate this schematically in Figure 1.

In the top right hand corner we have the physics of travelling wave amplifiers in inhomogeneous systems including the semiclassical WKBJ technique, the theory of instabilities in a shearing, rotating fluid and the phenomena of wave-particle resonant interactions. Moving clockwise we encounter the old percolation problem in a new guise: a shearing disk with the population effect being related to the probability that a burst of massive star formation will induce further massive star formation in neighbouring regions. The gas dynamics studies embody the usual one-fluid continuum approach with various techniques employed to treat the almost inevitable shocked response. Another currently fruitful mode of calculation of fluid response is to use a simulation of clouds or sticky particles that collide inelastically. Underpinning most of these gas dynamic calculations is some forcing potential such as a bar, a spiral or an oval. More sophisticated multiphase media are now under study. The global modes of stellar and gaseous disks present a classic problem in astrophysics that is currently severely ill-posed by our ignorance of that dynamically important component often referred to as dark stuff. This murky business I shall return to later. Global modes are now at least partially understood in terms of the local approximation used in the wave amplifier work but detailed simulations will assist greatly in clarifying this modal physics. The modes may often be a forced response

E. Athanassoula (ed.), Internal Kinematics and Dynamics of Galaxies, 163–174.

Figure 1


BAR FORCING
COMPANION FORCING
GLOBAL MODES
SIMULATION
WASERS WKBJ
SWING AMPLIFIER
TRAVELLING WAVE AMPLIFIER
SPIRAL STRUCTURE
GAS DYNAMICS
CONTINUUM
CLOUD PARTICLE
MULTI-PHASE
BAR, OVAL, SPIRAL FORCING
STOCHASTIC STAR FORMATION
PERCOLATION PROBABILITY PARAMETERS

to an external driving agent, such as a bar or companion, that produces a transient ringing of the global modes of the disk.

II. TRAVELLING WAVE AMPLIFIER

I cannot understand what all the fuss is about! The views of the two principal protagonists are shown in Figure 2. In both cases the cause of the wave generation is that galaxies wish to transfer their angular momentum outwards. The WKBJ approach has been strongly advocated by CC Lin and coworkers (Lin, Haass, Bertin the conference). In the vicinity of corotation there is a three wave interaction between a long trailing wave propagating outwards towards corotation, a short trailing wave propagating inward away from corotation towards the outer Lindblad resonance where it will be damped and consequently allowing the galaxy to effect its radially outward flow of angular momentum. To close the feedback cycle the WKBJ theory has a reflection point outside the inner Lindblad resonance where *short* inwardly propagating trailing waves are reflected as *long* wavelength trailing waves. This does depend rather crucially on how steeply the velocity dispersion, as measured by the Q factor (Toomre, 1977), rises in the inner region of a given galaxy. Furthermore unless Q is of order unity near corotation a WKBJ barrier has to be penetrated by tunelling processes. All the waves are trailing waves in this picture. The original estimates of the growth factor in amplitude per traversal of a cycle is of order $\sim \sqrt{2}$. A more detailed analysis of shearing effects near corotation, incorporating the physics of the Goldrich-Lynden-Bell (1965) instability, gives a growth factor of ~ 4 per cycle.

The second detailed approach is due to Toomre and is significantly behoven to the Goldreich-Lynden-Bell instability. A masterful presentation is given in Toomre (1981). The process feeds on shear and self-gravity. In the region of corotation a short leading wave swings around to become a short trailing wave propagating inwards and emits a shot trailing wave propagating outwards that is damped at the outer Lindblad resonance. Here the feedback cycle is closed in a rather different way. There is assumed to be no inner Lindblad resonance so that the short trailing wave reflects off the centre as a short leading wave propagating outwards. The growth factor can be very large $\sim 10^1\text{-}10^2$ per transit.

In general, the two theories now before us have been brilliantly developed over the last two decades. It all looks basically correct and solved. Probably both processes occur, and swing amplication *could* be incorporated as a type II waser, as Drury (1980) has indicated, or wasering *could* be incorporated as a type II swing amplifier, particularly when the shearing Goldreich-Lynder-Bell response is included, but I suggest that we stick to the nomenclature used in this paper to avoid further subclassification into the mysteries of swinging or percolating wasers!

III. GLOBAL MODES

Here I shall comment on a selection of recent global mode calculations where some attempts have been made to understand the physical growth mechanism in terms of the specific processes discussed in the previous section. The Gaussian disk (Erikson, 1974) has been analysed by Toomre and Kalnajs and the unstable modes have been interpreted as being due to swing amplifier effects. The growth rates are estimated by summing the travel time from the centre to the swing point, the swing time, and the reflection time, and also estimating the amplification factor per feedback cycle. There is agreement to within 10-20% between these estimates and the numerical work. The mode-shapes substantially agree. An important point to note is that if the central region is made into an absorber the whole instability goes away, which is quite consistent with the loss of the feedback cycle. The only other mode is a non-swinging edge mode generated on the sharp edge of the gaussian disk. A beautiful simple interpretation is given by Toomre (1982). The gaseous Kuzmin disk studied by the Iye (this conference) and Aoki et al. (1979) does not yet have a clear identification of the physical mechanism but Toomre (1981) has tentatively identified it with swinging.

Perhaps the most fruitful development in the near future will come from the N-body simulations such as the ones developed by Sellwood (this conference). These have good dynamic range, large N, and good diagnostics for the modes. The physics of travelling wave packets should be obvious if a simple initial disturbance can be propagated that travels, swings, reflects and refracts.

IV. STOCHASTIC STAR FORMATION.

As shown in Figure 3, we *start* with the idea of sequential star formation orginally proposed by Blaauw and summarised in his 1964 review; further developments were added by Elmegreen and Lada (1977). *Then*, add shear and choose a star-burst coherence length or equivalently a probability for continuing the star forming sequence at each point. This is a percolation problem. Simulate this on a large (IBM!) machine (see the review by Seiden and Gerola, 1982) and obtain, for a small coherence length, pictures that are quite reasonable for the large Magellannic clouds, star-bursts in dwarf galaxies such as IZW 18, IIZW 40, spurs in our own galaxy and filamentary structures in 2841-type galaxies.

Now, before considering the large coherence length studies, I should say that since the major proponents of this theory are not here at this meeting, I will not be so hard on the theory as I might otherwise have been! To continue, the large scale percolation phenomenon when applied to the grand design does *not* explain (i) smooth-armed SO's with no young stars (ii) the smooth neutral hydrogen and magnetic field compression correlated with dust lanes (iii) barred and oval galaxies (iv) the two-armed grand design (v) the arm symmetry even in the fine scale structure in M51 (Kalnajs, this conference). However, when there is an underlying

Figure 2

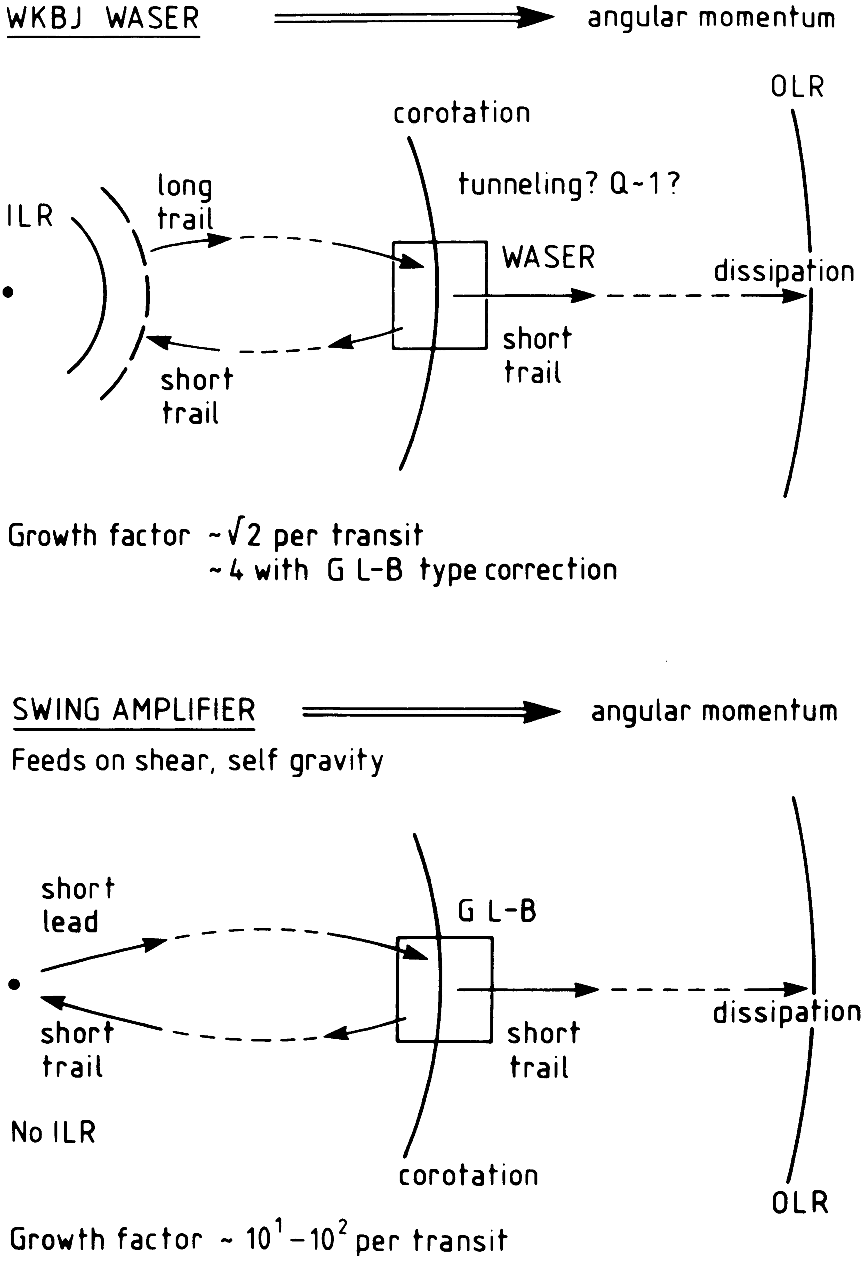

WKBJ WASER
angular momentum
OLR
corotation
tunneling? Q~1?
ILR
long trail
WASER
dissipation
short trail
short trail
Growth factor ~√2 per transit
~4 with G L-B type correction
SWING AMPLIFIER
angular momentum
Feeds on shear, self gravity
short lead
G L-B
short trail
short trail
dissipation
No ILR
corotation
OLR
Growth factor ~ 10¹–10² per transit

Figure 3

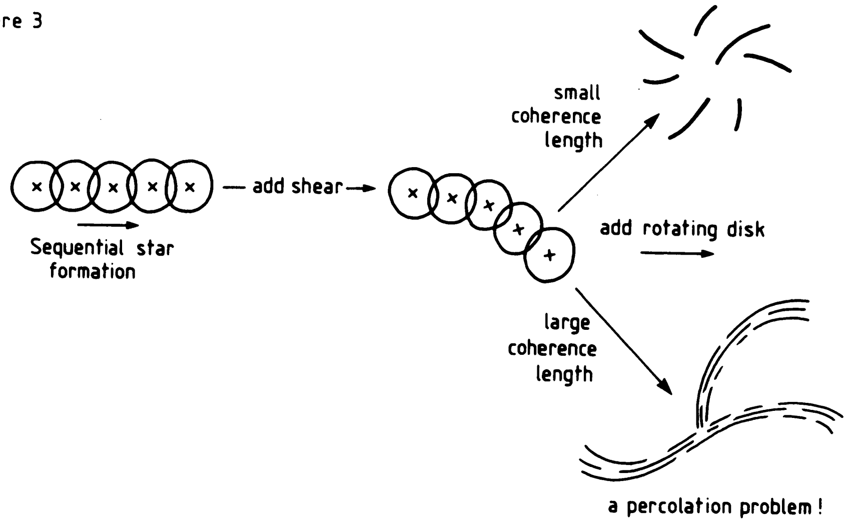
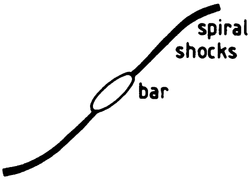
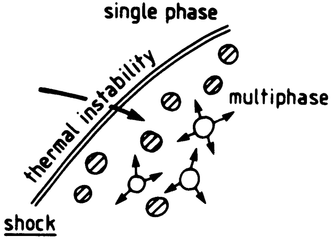


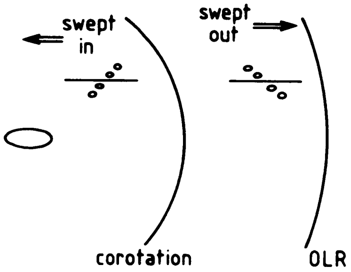
Figure 4(a)



4(b)



4(c)



4(d)

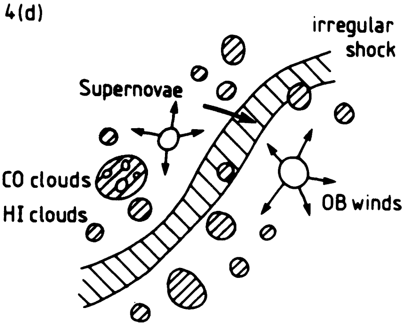
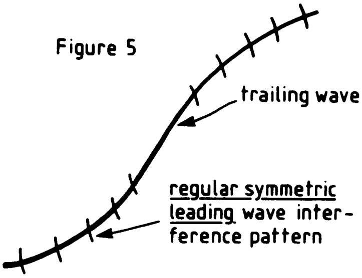


Figure 5



spiral wave such considerations of triggered sequential star formation are extremely germane, as they also are to filamantary 2841-type galaxies.

V. GAS DYNAMICS

This complex field will be partially reviewed by Prendergast (this conference) so I shall restrict myself to noting four aspects. Firstly it is now clear that the one-fluid models responding to an underlying forcing such as a bar, oval or spiral (Fig. 4a) describe very well the overall observed structure of the HI compression, the magnetic field compression and the dust-lane correlation. There is flow of gas into the central regions forced by the dissipation and the angular momentum outflow carried by the spiral response. The efficiency of this process decreases as the effective internal sound speed of the gas is raised.

Marochnik et al. (1983) have sent me details of an extensive multicomponent fluid calculations incorporating thermal balance and selfgravity at the shock where they have isolated the condition for a phase transition to occur at the shock from a one component to a two or multicomponent medium (Fig. 4b).

The calculational tool of simulating the gas response by using an N-body simulation with inelastic cloud collisions has been extensively explored by Schwarz (1979, 1981). As shown in Fig. 4c, the clouds can be regarded as test particles with a small but finite drag. Inside corotation, the particle response leads the bar forcing and the clouds lose angular momentum to the bar and move inwards. Outside corotation the clouds gain angular momentum from the bar and move outwards. The sweeping out time is short ~ 3-10 rotation periods.

More detailed simulations (Fig. 4d) involving a multiphase medium have been done by Roberts and collaborators (Roberts, this conference, Van Albada and Roberts, 1981). The most obvious effects are a longer mean free path with a corresponding larger shock width and weaker shocks with significant fluctuations in amplitude. Post shock heating occured because of the energy input from supernovae and stellar winds from early type stars. Such detailed models are most important for modelling the interstellar medium in M31; say, with its holes, bubbles, varied fine-scale kinematic structure, and also for modelling the detailed distribution and kinematic of the CO distibution, star forming regions and Hl in an near spiral arms (Lieisawitz and Bash, 1982, Kennicutt and Hodge, 1982).

V. OBSERVATIONS.

The relation between morphological properties and dynamic properties has been extensively studied by Kormendy and Norman (1979) with more recent updates by Rubin et al (1980) and Elmegreen and Elmegreen (1982). The basic point is that many grand design spirals have bars and companions and these are quite plausibly associated with driven spiral structure. However, apart from filamentary armed 2841-type spirals where there is no clear forcing of the grand design, there is a class of apparently isolated and non-barred galaxies which show grand design. There are instances of galaxies that could be purely global growing modes (i.e. not forced) due possibly to the absence of an inner Lindblad resonance rendering their liable to pure swing amplifications. There may also be very weak bars such as in NGC 1566 that are clear in the infrared (Hackwell and Schweizer, 1982) but are difficult to detect optically that can play the role of an almost invisible forcing agent.

The Fourier decomposition into logarithmic spirals is an useful tool for analysis of the observed mode structure of a disk. It has serious difficulties since the arms are not regular and large inhomogeneities can generate anomalous structure. A start to such a program is represented in the work shown at this conference by Athanassoula and Considère on NGC 2997 (trailing 3 arms) and M31 (2 trailing and 1 leading); Comte on NGC 1566 (three wave interaction at corotation); Iye on a sample where he finds odd modes; and finally the eyeball decomposition by Haass of M100 (3 trailing arms). Interesting as such studies are, they should be regarded as an initial stage in a long and sophisticated program. I wish to emphasize that swing gives us leading arms although of relatively small amplitude and that regular symmetric interference patterns should be found (Fig. 5). It is quite plausible that such an interference pattern in the underlying potential field could lead to regions of enhanced density in the gas resulting in the periodic beads on a string of OB stars and also possibly regular structure in the positions of giant molecular cloud complexes.

In a similar vein, the film shown of M31 by Brinks (this conference) should inspire the type of detailed investigation I am advocating. The kinematical and spatial information on HI, CO, holes, Z-structures and stellar populations are all a very, rich field for study related to the spiral structure problem. The Fabry Perot work from machines such as Taurus can and will give important detailed constraints on the ionised gas flows (Allen, this conference).

Disk structures as observed by van der Kruit and Searle (1982) with a sharp cut-off at 4 α^{-1} (where α^{-1} is exponential scale of the disk) over a scale $<< \alpha^{-1}$ should, in fact, give an edge mode if sufficient self gravity is present in the disks at such large radii. Observations of velocity dispersions in disk that can be unambiguously related to Q values such as in NGC 916 (Kormendy, this conference) will give radial Q-profiles that will be useful discriminators in differentiating hard or soft (i.e. leading or trailing) reflection in the central parts. The final observation I advocate is to observe simulations for wasering, swiging, edge moding and other!

VI. DIRECTIONS AND CONCLUSIONS.

From an immense array of possibilities for the future development of this field, I have selected the few areas that I consider will be most fruitful between now and the next IAU meeting.

Firstly, global mode calculations and simulations should be understood completely in terms of local effects and consequently more non-linear problems such as disk heating (Carlberg and Sellwood, this conference) and mode saturation can be pursued with confidence.

Secondly, at least some spirals may not have significant dark halos. The fits to the rotation curves shown by Kalnajs (this conference) with constant M/L ~ 2-6 over the whole disk out to the Holmberg radius do not require invisible dynamical dark stuff. The new understanding of disk stability (Toomre, 1982, Sellwood, this conference, Kalnajs, this conference) means that the remarkable Ostriker-Peebles criterion for disk stability may not be so relevant. Self-consistent disks *could* be constructed without dark material for at least some realistic galaxy models. There is no evidence for dark material in ellipticals so why should it not be absent in at least some spirals. The beautiful simulations of T.S. van Albada (this conference) fit the radial density dependence and the velocity dispersion profiles for ellipticals with no dark material. Certainly dark stuff exists, if one inspects the mass-radius relationship on scales of ~ kpc to ~ 10 Mpc (Rubin, this conference) but most of it could be *hot* dark stuff, that will not be associated with all individual galaxies, with Mpc scales intermediate between groups and clusters. There is little hard evidence for dark stuff either in groups with their long crossing times or binaries with their poor statistics. Defending such an hypothesis may be an example of theoretical overshoot but the evidence may be sifted more thoroughly as a direct consequence.

Evolutionary considerations are possibly *the* major new area with incorporation of effects such as large scale radial gas flows, infall and detailed star formation physics. Large scale secular stellar dynamical evolution should be analysed; bars may dissolve as they are torqued to become more and more slender up to the point of a dynamical catastrophe; secular evolution could redistribute the mass sufficiently to cut off the central reflection by repositioning or eliminating the inner Lindblad resonance; hot anisotropic lenses may result from dissolving bars; and a host of other related phenomena.

Finally, associated with the detailed observations of galaxies we have the Fourier decomposition technique; the search for leading arms or at least symmetric regular interference patterns along the arms and the detailed correlation of the HI/HII/CO and stellar population data with realistic multiphased gas dynamics models. There are all excellent bases for the next generation of detailed theories.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge the significant input to this paper from my colleagues at both the Institute of Astronomy, Cambridge and the Sterrewacht, Leiden when we discussed many of these questions in depth. I thank the many scientists who sent me preprints on this subject. Especially helpful were E. Athanassoula, A. Bosma, L. Blitz, E. Brinks, B. Burton, R. Carlberg, D. Lynden-Bell, G. Efstathiou, J. Kormendy, H. Listzt, R. Kennicut, J. Sellwood, A. Toomre and R. Walterbos. I dedicate this paper to the memory of my brother Murray, who liked a good story and enjoyed living at the edge.

REFERENCES

van Albada, G.D. and Roberts, W.W., 1981, Ap. J. 246, 740.
Aoki, S., Noguchi, M. and Iye, M., 1979, Publ. Astr. Soc. Japan, 31, 737.
Blaauw, A., 1964, Ann. Rev. Astron. Astrophys. 2, 213.
Drury, L.O.C., 1980, MNRAS, 193, 337.
Elmegreen, B.G. and Lada, C.H., 1977, Ap. 214, 725.
Elmegreen, D. and Elmegreen, B., 1982, MNRAS, in press.
Erikson, S.A., 1974, Ph.D. Thesis, M.I.T.
Goldreich, P. and Lynden-Bell, D., 1965, MNRAS, 130, 125.
Kennicutt, R. and Hodge, P., 1982, Ap. J. 253, 101.
Kormendy, J. and Norman, C., 1979, Ap. J. 233, 539.
Lieisawitz, D. and Bash, F., 1982, Ap. J. 259, 133.
Marochnik, L.S., Berman, V.G., Mishuror, Y.N. and Suchkov, A.A., 1983, Astr. Space, Sci., 89.
Rubin, V.C., Ford, W.K., Jr. and Thonnard, N., 1980, Astrohys. J. 238, 471.
Schwarz, M.P., 1979, Ph.D. Thesis, Australian National University.
Schwarz, M.P., 1981, Astrophys., J. 247, 77.
Seiden, P.E. and Gerval, H., Fund. Cosmic Physics, 7, 241.
Toomre, A., 1977, Ann. Rev. Astr. Ap. 15, 437.
Toomre, A., 1981, in "Structure and Evlution of Galaxies", ed. S.M. Fall and D. Lynden-Bell, Cambridge, p. 111.
Toomre, A., 1982, in preparation.
Van der Kruit, P.C. and Searle, L., 1982, Astron. Astrophys. 110, 61.

GENERAL DISCUSSION ON SPIRAL STRUCTURE

BLITZ : One thing should be kept in mind when using the results of Blaauw and Elmegreen and Lada as justification for the stochastic star formation model. Both sets of authors have argued that star formation can infect neighboring regions on a scale of tens of parsecs. But for the stochastic star formation model to work, it is necessary for some mechanism to allow the infection to spread from one giant molecular cloud to another -a length scale near the sun of about 500 parsecs-with a propagation time consistent with the theory. This process, which Seiden calls the "microphysics" of the theory, is hidden in the theoretical quantity P_{stim}. It will thus be necessary for proponents of the theory to show that there is a physical mechanism which can and will propagate star formation from one giant molecular cloud to another.

NORMAN : I agree, but Blaauw indicated that a coherence length of ∿ 500 kpc was not unreasonable from the observations of the solar neighbourhood.

IYE : Could you elaborate on your third conclusion or implication namely that large scale evolution may cut off the swing amplification ?

NORMAN : No swing amplification can occur if the mass distribution of the galaxy slowly changes so that the Inner Lindblad Resonance absorbs the inward travelling trailing wave before it can be reflected into an outward propagating leading wave. This effect can occur if, for example, the mass distribution becomes more centrally concentrated, possibly due to strong radial inflow or merging with dwarf companions over a Hubble time.

SIMKIN : Dr. Lin has noted that a Fourier analysis of the "swing amplified" mode shows that its Fourier coefficients can be treated in two different ways, one of which is time independent and the other time dependent. The differential equations (in both cases) are related to the potential distribution in the galaxy (i.e. the density distribution). Thus we should be able to predict what types of density distributions will yield growing modes by analysis of these equations. These predictions can then be compared with "reality" (galaxies). Has this been done ?

LIN : The investigation of the amplification mechanism near the corotation circle is a local discussion, whether it is approached as a swing process or as a WASER process. To fit this mechanism into the description of the global modes is indeed a primary concern of the theoreticians. This has been done, in varying detail, in a number of papers. Specifically, a discussion of the mode calculated for M81 may be found in the references cited at the end of my abstract (Lin and Bertin, 1981 ; Haass, 1982).

IYE : I have Fourier analysed the luminosity distribution of several galaxies including both early and late types and presented the Fourier spectra and the patterns of primary Fourier components of these galaxies in a poster paper in this symposium. We find that the coexistence of trailing and leading spiral components and the coexistence of one-armed, two-armed, and multi-armed spiral components takes place very often in actual galaxies.

RENZ (to LIN) : To decide whether a stationary spiral structure can really develop you have to consider global modes and not just the local mode at corotation. Furthermore, after at best 10^9 years, the growing global modes reach an amplitude where linear theory breaks down and nonlinear effects have then to be taken into account. Without doing nonlinear calculations you cannot pretend to have shown that a stationary spiral structure is possible.

LIN : There appear to be several misunderstandings on the part of the questioner. Specifically, regarding the last statement, we only claim to have demonstrated that quasi-stationary spiral structure is possible and plausible through the calculation of unstable global spirals modes of both the tightly wound type (associated with WASER of Type I) and of the type involving leading waves (associated with WASER of Type II). These have been discussed in our existing publications, in my paper just presented, and in the papers of Dr. Bertin and Dr. Haass presented at this conference. We do not claim that the existence of stationary spiral structure is mandatory. Indeed, it would be wrong to do so. (See Oort's description of the issues on spiral structure in Interstellar Matter in Galaxies, (Benjamin, 1962), p. 234). The occurence of vascillatory large scale modes in an unstable basic state of flow has been repeatedly discussed in classical literature on turbulent flow, weather patterns, etc. Our experience in these subjects strongly suggests that the same conclusions hold in the present case, even though the nonlinear theory has not yet been developed.

The questioner appears to have also misunderstood the purpose of our local analysis. It is intended to clear up possible misunderstandings of the role of "swing amplification". The dramatic transient amplification of wave packets does not imply a corresponding strong amplification of modes or stationary wave trains. The net growth rate is usually much more moderate. Generally speaking, a mechanism of self-regulation of the magnitude of the dispersion of stellar velocities is needed to arrive at favorable conditions for the existence of quasi-stationary spiral structures.

RENZ (to BERTIN) : To select a slowly growing mode as a dominant one over a long time scale, as you argued, you need a mechanism which your equations do not contain. Namely you must assume that fast growing modes will disappear very quickly by enlarging the velocity dispersion. Can you write down the basic equations from which you can obtain both the slowly growing spiral modes and this necessary mechanism ?

BERTIN : The equilibria that we are interested in are subject only to mild instabilities. In choosing such equilibria we use the physical argument that violently unstable disks should be discarded because they would rapidly evolve so as to heal the strong instabilities. Linear stability analysis generally indicates which parameters (such as Q) should be affected in the saturation process, even though detailed equations for nonlinear evolution are not easy to write.