II

SPIRAL STRUCTURE

https://doi.org/10.1017/S007418090003254X Published online by Cambridge University Press

THEORY OF SPIRAL STRUCTURE

Agris J. Kalnajs
Mount Stromlo and Siding Spring Observatories, Research School of Physical Sciences, Australian National University

ABSTRACT

Spiral structure is associated with a slow redistribution of gas, which may already be quite significant over time scales short compared to the age of the galaxy. One has to worry about replacing the gas in order to keep the structure alive.

1. INTRODUCTION

One of the more interesting discoveries of the past decade has been the liveliness of self-gravitating stellar disks. Much of the zip can be directly attributed to differential rotation, a feature which was once thought to be inimical to any coherent structure. Disks show a definite preference for bisymmetrical deviations from axial symmetry, which also happens to be one of the more striking features of large scale spiral structure. Thus there is little doubt in my mind that the understanding of spiral structure is to be found in the study of the dynamics of stellar disks, and what their oscillations can do to the gaseous component.

An alternative approach to spiral structure is to view it as the result of stochastic self-propagating star formation (Mueller and Arnett 1976; Gerola and Seiden 1978) - a galactic version of the computer game Life, played in a shearing world. While the physical mechanisms invoked by this theory are all plausible, it lacks one important ingredient: the large scale organisation and symmetry. I like to illustrate the latter with the help of a photographic trick. Figure 1 shows a well-known galaxy on the left, togather with a symmmetrised version. The latter was obtained from the same negative, except that after the first half of the exposure the paper was rotated 180 degrees around the center and then exposed again. The main thing to note is that the exposures were so chosen as to cancel any feature which did not have a counterpart on the other side of the center, e.g. the companion and the field stars are gone, but the spiral structure remains almost in tact, even the bifurcations in the arms in the 10 and 16 o'clock quadrants. It is the

E. Athanassoula (ed.), Internal Kinematics and Dynamics of Galaxies, 109–116.

explanation of this large scale symmetry and its permanence which should be the primary task of a spiral theory, the irregularities I would consider as secondary.

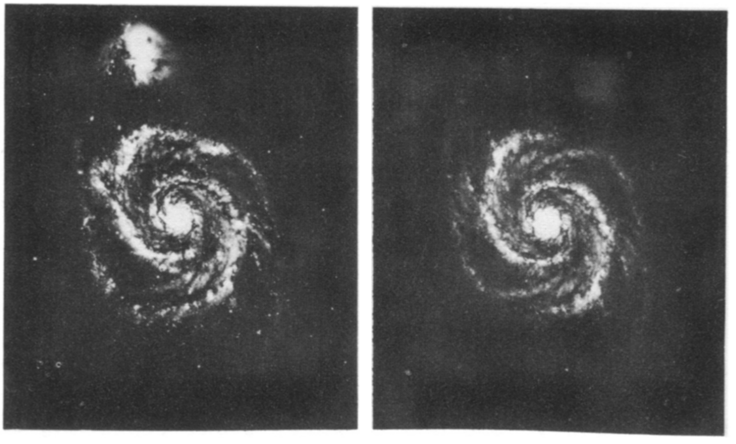


Figure 1. M-51 (left) and its symmetric version (right).

The task of any reviewer of this topic has been lightened considerably by the existence of a very comprehensive review (Toomre 1977). All I can hope to do is to add a few footnotes, pointing out new results and changes in emphasis.

2. STELLAR MODES, GAS, AND TIDES

The theory of spiral structure was born when people began to think of it as a density wave phenomenon. The theory should tell us why spirals

(a) are long lived,
(b) have a large scale, and
(c) require both gas and stars.

A combination which satisfies the above criteria is a stable stellar disk which can support a large scale mode, to which one adds a bit of gas.

In order to be long lived, the mode has to be practically stable, and therefore it can not be noticeably spiral. The spirality arises from the inability of the gas to settle down in the periodic orbits of the non-axisymmetric field, because they often intersect in space. As a result the gas must bump into itself, and this gives rise to a quasi-steady flow with shocks. The bumping is the feature which makes the gas essential and distinguishes it from the stars.

For the above scenario to work there are many parameters such as mode shapes, amplitudes, pattern speeds, gas distribution, etc. which have to be determined, as well as observational constraints to be satisfied (Kormendy and Norman 1979).

The logical place to start is by choosing a disk and calculating the modes. The question is: which disk? All the seemingly reasonable disks turned out to be unstable, and stable ones unreasonable, usually because they were too hot. One way to make cooler disks is to suppose that they are only partly self-gravitating, by imbedding them in a halo. This I find a mixed blessing.

Many studies have by-passed this initial difficulty by simply assuming a mode or a bar (= finite amplitude mode), and proceeded to the second stage: to see what will happen to the gas. The results have been very encouraging, which should provide added incentive to tidy up the disk stability problem.

If we do not insist that spirals be long-lived, we may assume that both stars and gas have a spiral form, as in the case of tidally forced structure. While suggestions and hints that a tidal interaction could generate a fine spiral have been around for at least a decade, the convincing demonstration of how and why it works was provided only recently (Toomre 1981). The key ingredient turns out to be an effective swing-amplifier.

3. INSTABILITIES OF STELLAR DISKS

The stock of global modes of a variety of disks has shown a healthy growth since 1977, most of them demonstrating the well-known fact that disks are unstable, and that haloes can help to make them less so. More important, there has also been considerable growth in understanding of the causes of the fiercest instabilities (Toomre 1981). There appear to be two distinct non-axisymmetric amplifying mechanisms, giving rise to swing-amplified and edge modes. The former depends on the ability of shear to amplify leading outgoing waves while swinging them around and sending them inward as trailing ones. The cycle is closed, when the trailing wave passes through the center, and returns as a leading one. It has been possible to demonstrate the demise of this type of instability by simply placing an absorbing plug in the center of the disk. The same effect could be achieved by redistributing the mass so as to give rise to an inner Lindblad resonance, but the demonstration is not as convincing, since tampering with the mass distribution causes other parameters to change. (This is one of the frustrations inherent in trying to isolate and localise features of a global mode.) Of course we may also turn down the gain of the amplifier by decreasing shear, or by placing some of the mass in a halo. While stability can be achieved by turning down the gain or breaking the cycle, a possible distinction arises when the stable disk experiences a tidal field. Because by breaking the cycle we preserve the lively nature of the disk, it will still respond enthusiastically to any tidal forcing, although in a transient manner.

The edge modes depend on the gradients in angular momentum distribution for their growth. They do not care about central absorbing plugs. Here the obvious cure is to minimise the gradients.

The unstable modes seldom fall neatly into one or the other category, usually both mechanisms operate. However by judicious fiddling one can select almost pure examples to illustrate the two instabilities.

4. EQUILIBRIUM MODELS

The study of spiral structure would be greatly simplified if we knew the equilibrium structure of a disk galaxy. Under the most favourable conditions we may hope to discover the mass distribution from a combination of rotational data and photometry. That still leaves us with the problem of determining the velocity distribution. It is conceivable that some information about the latter could be obtained by requiring the absence of instabilities. The introduction of dark haloes to stabilise disks nullifies that, and certainly the rotation will no longer tell us much about the disk mass. Thus life would be simpler without haloes.

One fairly obvious way of generating stable disks is to let the instabilities run their course. Hohl (1971) showed that it is possible to get rid of the resulting stable bar by simply reshuffling the stars in azimuth. The symmetrised disk remained stable, but was rather hot in the sense that the axisymmetric stability parameter Q was around 4. By a more careful selection of the initial configuration, the final Q could be reduced to around 2. Again it is conceivable that the high Q's might have been due to the violence of the instability. To check this, Hohl tried cooling the disk and in so doing managed to reduce Q to the range 2 to 3 before a slowly growing bar started to heat it up again. It now transpires that the reported radial velocity dispersion was an azimuthal average and hence included all systematic motions. Insofar that the random motions are due to the decay of the systematic flow, the latter could dominate the azimuthal average. Already a reduction of Q by, say 1.5, would make it quite compatible with the (only known) value near the sun. Did we give up too soon?

5. EVOLVING GAS FLOWS

There are two somewhat distinct approaches to generating spiral structure in gas. One is by brute force: you assume a spiral forcing field and get a spiral response (Visser 1978; Roberts et al 1979). The WKBJ theory necessarily falls in this category. The other relies on finesse: you start with a non-spiral field such as a bar or oval distortion, and let the inhomogeneities of the disk - the resonances - give the gas a spiral form. The bar forcing has received most attention simply because the bar phenomenon is so widespread both in nature and theory.

There are several distinct regimes of spiral making. The first is the initial or turn-on transient. Since all calculations seem to start from a circular flow, there is a transition period of the order of a bar revolution. During this phase both stars and gas start out in a similar fashion, and continue until pressure forces part them. The gas remains spiral, while the stars begin to phase mix and finally adopt a distribution which reflects the symmetry of the force field.

Spirals can be generated by an exponentially growing field. As long as the orbital excursions remain small, both stars and gas will show spirality, although the shapes may differ.

Spirality can also be sustained in a steady field by dissipation, due to viscosity or shocks.

The common factor in the above three cases is evolution. Another way of saying this is to note that whenever the gas response to a bar is spiral, the torques exerted on different annuli will not vanish, implying that the angular momentum distribution has to evolve. The torques can be computed from the observed or computed mass distribution, and do not depend on how that distribution was produced.

Clearly the initial transient stage is not the answer. However practical considerations such as numerical stability and accuracy, discourage long integrations, and the slower evolution of the "quasi-steady" state has been largely ignored, or wrongly attributed to numerical effects.

The age of a typical "quasi-steady" gas spiral is about one to two bar revolutions, although in the most recent work (Schempp 1982) stops after half a revolution. The evolution during this shock dissipation phase can be quite rapid in regions close to the bar. Huntley (1980) provides sufficient data to estimate the rate at which the gas in the straight off-set shocks is losing angular momentum. If you assume that the angular momentum at his radius 10L is that corresponding to circular motion in the axisymmetric field, and divide it by the annular torque, the decay time turns out to be about half a rotation period or only 1/6 of a bar period! This is an alarmingly short time, and it may be that the "quasi-steady" pattern survives only because the gas which falls into the center is recycled. Huntley calls the infall "numerical leakage".

A similar rate of infall is evident from tracing the streamlines of the standard model of Sanders and Tubbs (1980). It can be seen that a streamline which leaves the shock at radius a, meets the next shock at half this value.

The above calculations use the beam scheme. The inward spiraling of the streamlines in the vicinity of the bar is also quite evident in the calculations using other hydrodynamical schemes (van Albada and Roberts 1981). The time scale for moving from one end of the bar (6.9 kpc) to the center (0.47 kpc) in their model is about one gigayear, or two bar revolutions. A comparison with the beam scheme shows that both behave the same way.

Because of the numerical viscosity inherent in the above schemes, it was not at all clear how much of the evolution should be attributed to computational artifacts, and how much is real. That shocks imply some evolution was pointed out by Pikel'ner (1970). The first clear demonstration that it was significant over a lifetime of a galaxy came from the work of Schwarz (1979; 1981). In contrast to most studies, he put his effort in trying to understand the flows around the outer Lindblad resonance. (One of the reasons for this choice is the fact that the model parameters become fewer: all bar fields and rotation curves begin to look the same at large distances.) Schwarz modeled the gas with inelastic clouds, and used a Lagrangian scheme to follow their evolution. Once you accept the cloud model, there is little room for numerical misgivings.

The initial evolution is quite similar to that of the continuum schemes. After two bar revolutions a quasi-steady state forms, with a nice spiral collision front. In a typical case, it then slowly evolves into a ring (actually a periodic eccentric orbit) around the outer resonance, over a time scale of ten bar revolutions, or 2.5 gigayears. Once the ring is formed, there are no more collisions, and all evolution stops. One can understand the formation of the shock in terms of intersecting periodic orbits (Kalnajs 1972), and its demise as the result of their drift outwards to the resonance. In order to keep the spiral structure alive, one has to repopulate the depleted orbits, for example by mass loss from stars or infall. (The incipient ring has a distinct shape which is quite common among galaxies, and this identification argues that the pattern speed should be high enough to give rise to an outer resonance.)

The evolution towards a periodic orbit around an eccentricity resonance appears to be quite general, and not peculiar to the outer resonance. There is a hint of this phenomenon at the end of the hydrodynamical calculations of Sorensen and Matsuda (1982).

Since it is the torques that appear to be responsible for the redistribution of the gas, it is natural to ask whether a tightly wrapped spiral field where the ratio of tangential to radial force is small, could slow the drift. Provided it was was tight enough, the answer must be yes. But in the case of M-81, which is very average in terms of the tightness of its arms, Visser's spiral lasts only about four bar revolutions before the redistribution destroys it (Schwarz 1979).

REFERENCES

Gerola, H., and Seiden, P.E. : 1978, Astrophys. J. 223,129.
Hohl, F. : 1971, Astrophys. J. 168, 343.
Huntley, J.M. : 1980, Astrophys. J. 238, 524.
Kalnajs, A.J. : 1973, Proc. Astron. Soc. Austr. 2, 174.
Kormendy, J., and Norman, C.A. : 1979, Astrophys. J. 233, 539.
Mueller, M.W., and Arnett, W.D. : 1976, Astrophys. J. 210, 670.
Pikel'ner, S.B. : 1970, Astron. Zh. 47, 752.
Roberts, W.W., Huntley, J.M., and van Albada, G.D. : 1979, Astrophys. J. 233, 67.
Sanders, R.H., and Tubbs, A.D. : 1980, Astrophys. J. 235, 803.
Schempp, W.V. : 1982, Astrophys. J. 258, 96.
Schwarz, M.P. : 1979, Ph. D. Thesis, Australian National University.
Schwarz, M.P. : 1981, Astrophys. J. 247, 77.
Sorensen, S-A., and Matsuda, T. : 1982, Monthly Notices Roy. Astron. Soc. 198, 865.
Toomre, A. : 1977, Ann. Rev. Astron. Astrophys. 15, 437.
Toomre, A. : 1981, In The Structure and Evolution of Normal Galaxies, ed. S.M. Fall and D. Lynden-Bell, p. 111, (Cambridge : Camb. U. Press)
Van Albada, G.D., and Roberts, W.W. : 1981, Astrophys. J. 246, 740.
Visser, H.C.D. : 1978, Doctoral Thesis, University of Groningen.

DISCUSSION

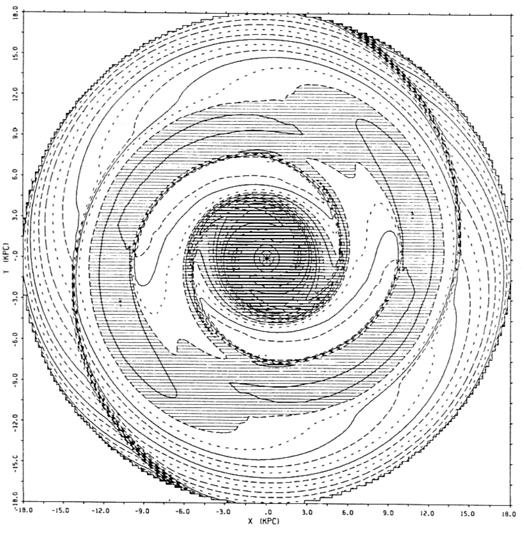
VAN ALBADA : Using the FS2 hydrodynamic code described elsewhere in this volume,a very long computational run was made of a model of M81. The description of the potential was taken from M.P. Schwarz's thesis (1979) ; a velocity of sound of 15 km/s was chosen. The model was computed on a 80 by 160 zone half grid with 200 pc cells ; it was continued to a model time of 4.3 Gyr, or approximately 13 pattern revolutions. As is the case in Schwarz's cloud-particle computations (which were run for approximately 4 pattern revolutions) most of the gas has been removed from the spiral arm region by the dissipation in the shocks. However, the similarity ends here. Instead of narrow rings of matter,the following structures can be identified :

1) An annulus around corotation,attributable to the fact that the gas streaming velocities relative to the pattern are too low to cause shock dissipation in this region. The relatively high gas density in this annulus is due to the large extent of the assumed initial density distribution.

2) A very wide inner annulus having an enhanced gas density caused by the gas transported from the depleted regions somewhat further out in the plane. The width of this annulus is probably real as its appearance remained unaltered in a continuation of the computation on a very high resolution grid. It appears to be caused by the gradual disappearance of shocks near the inner Lindblad resonance.

3) Arms clearly delineated by shocks, having peak densities comparable to those in the corotation annulus. The occurence of shocks in this region, whatever the local gas density, is a direct consequence of modelling the ISM as a perfect gas.

At first sight we might state that Schwarz's and my computations represent two extremes of the possible behaviour of the gas in M81. The ISM is evidently not as viscous and inelastic as modelled by Schwarz, nor is it as free of viscosity as it is assumed to be in my computation. The observed gas distribution, however, does not resemble a cross between Schwarz's and my results. The HI is predominantly concentrated in the region of spiral arms and not in some kind of inner ring, as predicted by both computations. The result presented here is more realistic in that the arms are delineated by shocks. Star formation processes are presumably important in shaping the observed distribution of the gas in M81, preventing the buildup of an inner gas annulus. The observed gas distribution may also be influenced by the acquisition of new gas, as implied by Tinsley and Larson's models of spiral galaxies. Alternatively, we may not exclude the possibility that the observed strong and regular spiral structure in M81 is wholly temporary and caused by the recent interaction with M82 and NGC 3077 evident from the extended HI distribution.



W.W. ROBERTS : The shortness of the time scale suggested by the N-body gas/particle calculations, quoted from Schwarz, for the destruction of gaseous spiral structure and the pile up of gas into a ring, may in part be attributed to the high dissipation adopted for the cloud-cloud collisions in his code. If the dissipational processes in real galaxies are not so severe, then the spiral structure may be expected to persist over longer time scales, and considerably more so if dissipational processes are partially offset by sources of energy input to the cloud system, such as supernovae and stellar winds.

KALNAJS : I agree that decreasing the dissipation in the collision process for particle calculations, or making the gaseous shocks milder by increasing the sound speed, slows the secular radial migration of the gas. But at the same time the spiral arms become broader and less distinct.