

THE POSSIBLE ORIGIN OF THE SPIRAL-ARM INSTABILITY

J. G. HILLS

Department of Astronomy, The University of Michigan, Ann Arbor, Michigan, U.S.A.

(Received 26 February, 1976)

Abstract. Physical arguments suggest the spiral arms may be manifestations of the galaxy not being in dynamical equilibrium – in the sense that the kinetic energy of its stars and gas is less relative to its binding energy than that dictated by the virial theorem. Without constant cooling of the galactic disk (i.e., a progressive increase in the binding energy of the galaxy) such a departure from dynamical equilibrium would be corrected and the spiral arms destroyed in about 10^9 yr due to an increase in the velocity dispersion of the stars in the disk resulting from their interacting with the spiral arms. The rate of cooling required to maintain the spiral arms, about $6 \times 10^4 L_{\odot}$, may be provided by mass loss from stars in the disk population. The cooling arises from the average scale-heights and velocities of these stars being larger than that of the gas in the disk, so that there is a net loss of kinetic energy and an increase in the binding energy of the galaxy due to the ejected gas settling down to a lower terminal velocity and scale-height in the galactic disk.

1. Introduction

In this paper we suggest an instability mechanism that may be responsible for the existence of spiral arms in galaxies. Although the gravitational-wave theory of spiral structure, as formulated by Lin and Shu (1964), has enjoyed considerable success in explaining many properties of spiral arms, there is no satisfactory theory of *why* these arms exist. The situation is similar to the study of stellar pulsations where it is possible to determine many dynamical properties of a pulsating star without knowing the instability mechanism responsible for the pulsations. Even without knowing what produces the spiral arms one can, for example, demonstrate that a standing spiral-wave pattern can exist in a galaxy and relate the angular rotational velocity of this pattern to the radial spacing of the spiral arms. The source of excitation of the spiral arms in galaxies may not be unique. Different mechanisms may excite the waves without affecting their overall appearance just as the permissible standing waves in an open pipe are largely independent of the mechanisms employed to excite them. However, the wide-spread occurrence of spiral arms in disk galaxies suggests that these galaxies have some internal default mechanism which is capable of exciting the waves even in an isolated system. We propose a possible mechanism of this type.

2. Effect of Cooling the Galactic Disk

We suggest that the spiral arms may originate from the galaxy being kept constantly out of dynamical equilibrium by the progressive increase in its binding energy

resulting from the tendency of the gas ejected from dying stars in the galactic disk to lose energy through collisions and to settle into a state of lower kinetic energy and greater binding energy than it had while in the stars. In general, any large-scale (but not necessarily large amplitude) deviation from hydrostatic equilibrium such as that caused by this cooling of the entire galactic disk must be damped out in a time, τ_s , of the order of a few dynamical crossing times of the galaxy or $\tau_s \sim 10^8$ – 10^9 yr. If the cooling is occurring at a rate L_c , in time τ_s the binding of the galaxy is increased by an amount $\Delta E_B = L_c \tau_s$. It is evident that the immediate effect of the cooling is to decrease the kinetic energy of the galaxy by an amount $\Delta E_k = -\Delta E_B$. However, when dynamical equilibrium is restored the virial theorem requires that the potential energy of the galaxy must decrease by an amount $\Delta W = 2\Delta E_B$ and the kinetic energy must *increase* by an amount $\Delta E_k = \Delta E_B$. Thus, in order to restore dynamical equilibrium after the cooling, the galaxy must in time τ_s redistribute its energy by shrinking enough radially to decrease its potential energy by an amount $\Delta W = -2\Delta E_B$ and at the same time increase its kinetic energy by the same amount $\Delta E_k = 2\Delta E_B$. It is evident that the constant cooling of the galactic disk causes its potential energy to exceed its kinetic energy by a steady-state amount $4\Delta E_B$ relative to their values in dynamical equilibrium. Thus, there is not enough pressure (kinetic energy) to support the disk against gravity. As a result the disk must be subject to a gravitational (Jeans) instability with a scale length comparable to the radius of the galaxy. Because the gas and stars in the galactic disk have nearly circular orbits and consequently are severely hampered by the centrifugal force from further radial contraction, much of the resulting gravitational contraction can be expected to be nonradial; i.e., the contraction takes the form of a compressional wave not having rotational symmetry with a self-gravitational potential energy of the order of $W_s \simeq -2\Delta E_B$, the value required to bring the galaxy back in agreement with the virial theorem. The most permanent and consequently the most conspicuous of these waves will be the standing waves permitted by the particular boundary conditions of the system. Since the wave does not show rotational symmetry it is a deviation from dynamical equilibrium and consequently must itself damp out on a dynamical time scale τ_s except for the amount of steady-state deviation allowed by the constant cooling of the galactic disk. To accomplish this damping the self-gravitational potential energy of the wave must be transferred to that of the galaxy in time τ_s . In other words, in the steady-state, the galaxy must still shrink sufficiently to change its potential energy by an amount $\Delta W = -2\Delta E_B$ in time τ_s . To promote this, the disturbance or wave must feed sufficient angular momentum to some of the stars in the disk to allow the rest of the disk to shrink inwards radially. Since the wave does not have rotational symmetry it can feed angular momentum to the stars in the disk by rotating through the galactic disk in the same sense as the galactic rotation. The wave must also increase the kinetic energy of the stars in the disk by the amount $\Delta E_k = 2\Delta E_B$ in time τ_s .

The spiral arms of Lin and Shu are at least qualitatively of the type expected. They are density perturbations having a scale comparable to that of the entire disk, and they are at least one known family of permissible standing waves in a disk galaxy. They also

feed angular momentum to the stars in the disk as required since they do not show rotational symmetry and they rotate at the pattern speed in the same direction as the galactic rotation. They also tend to increase the velocity dispersion of (heat up) the stars that pass through them. The arms have an excess density and provide a deeper potential well than the disk as a whole so that stars are accelerated on flowing through them. This acceleration results in any particular group of stars having a certain characteristic streaming velocity V_s associated with their passage through an arm. Since these flows cannot be strictly coherent the star gas is heated in passing through the arms. This can be thought of, physically, as a tendency for the stars to come to equipartition of kinetic energy with the massive spiral arms. Because of the disparity in mass, the spiral arms must monotonically increase the velocity dispersion of the stars in the galactic disk. This heating of the galactic disk is known to be especially effective at the Lindblad resonance points where the angular velocity of the wave pattern is in resonance with the epicyclic motion of the stars. The energy dissipation involved in this heating would destroy the spiral wave in about $\tau_s = 1 \times 10^9$ yr if nothing acts to rejuvenate it (Toomre, 1969). Thus the spiral arms tend to bring the galactic disk back to dynamical equilibrium in a few dynamical time scales as required if the cooling is occurring over the entire galactic disk.

We must emphasize the point that since the spiral arms are heating up the stars in the galactic disk they must result from a cooling of the galactic disk (i.e., an increase in its binding energy) rather than resulting from an effective heating of the disk imposed by, say, a tidal disturbance by the Magellanic clouds or whatever. This is the irony imposed by the virial theorem. An increase in the binding energy of the galaxy resulting from an initial cooling must after dynamical equilibrium is restored result in an increase in the velocity dispersion of the stars while an initial heating which decreases the binding energy causes the galaxy to expand and results in a decrease in the velocity dispersion. *The well-known fact that the velocity dispersion of the stars in the disk appears to be increasing requires that the binding energy of the galaxy be increasing monotonically with time.*

We have found on physical grounds that the constant cooling of the galactic disk *can* produce a steady-state wave pattern having properties similar to those of the spiral arms. That the cooling actually *does* produce spiral arms as opposed to some other type of wave or disturbance is strongly suggested by computer simulations of disk galaxies (Miller *et al.*, 1970). These investigators find that long-lived spiral patterns form when they arbitrarily cool their systems by constantly reducing the velocity dispersions of a certain subset of the mass units in their models. This method of cooling, which was completely arbitrary in the computer-simulated systems, is highly analogous to the type of cooling provided physically in our model. In our model a certain subsystem, the dying stars, lose mass and this mass in turn ends up in a new subsystem, the interstellar medium, where it has less kinetic energy per unit mass than it had in its parent stars. Thus the computer simulations indicate that the type of cooling provided by our model is capable of producing spiral arms.

3. Cooling Provided by Dying Stars

We now need to calculate the rate of cooling required to maintain the spiral arms and to test whether our hypothetical cooling source is adequate to provide it. Energy conservation dictates that the effective excess depth, ψ , of the potential well due to a spiral arm is given approximately by $\psi \simeq \frac{1}{2}(V_s^2)$, where V_s is the observed flow or streaming velocity of stars of low velocity dispersion in the arm. Thus the rate of cooling required to maintain the spiral arms is given roughly by

$$L_c \sim \frac{1}{2}(f_s M_G) V_s^2 / \tau_s, \quad (1)$$

where $M_G \simeq 2 \times 10^{11} M_\odot$ is the total mass of the galaxy, $f_s \sim 0.1$ is the assumed fraction of M_G in the spiral arms at a given time, $V_s = 6 \text{ km s}^{-1}$ (Humphreys, 1972) is the observed streaming velocity of young stars in the spiral arms, and $\tau_s = 1 \times 10^9 \text{ yr}$ is the dissipation time of the spiral arms. This yields $L_c = 2.3 \times 10^{38} \text{ erg s}^{-1} = 5.0 \times 10^4 L_\odot \simeq 10^{-6}$ the luminosity of the galaxy. The assumed value of f_s may be too high. In this case the calculated L_c is an upper limit on the true L_c .

Cooling of the order of L_c is provided by mass loss from stars in the galactic disk. Over the entire galactic disk the rate of mass loss by the dying members of a group of stars having an average main sequence mass M_i , a lifetime τ_i , a space density in the galactic plane ϕ_i , a scale height above the galactic plane β_i , and a mass loss per star ΔM_i is

$$\frac{dM_i}{dt} = (2\beta_i)(\pi R^2) \left(\Delta M_i \frac{\phi_i}{\tau_i} \right), \quad (2)$$

where $R = 10 \text{ kpc}$ is the characteristic radius of the galactic disk. Table I tabulates the results as a function of the absolute visual magnitude, M_{iV} , of the stars along the main sequence. The values of M_i , ϕ_i , and β_i as a function of M_{iV} were derived from the data in Allen (1963) while τ_i was derived from M_i using the data in Iben (1967). If we assume that the material which has undergone nuclear processing in a star is retained in its stellar remnant while the unprocessed material is ejected from the star on death, we find from the work of Hills and Dale (1973) that $\Delta M_i = (M_i - 0.7 M_\odot)$ for $M_i < 6 M_\odot$ and $\Delta M_i = 0.88 M_i$ for $M_i > 6 M_\odot$. The kinetic energy lost per unit time by the gas ejected from each group of stars is

$$L_i \simeq \frac{1}{2} \frac{dM_i}{dt} [\langle V_i^2 \rangle - \langle V_g^2 \rangle], \quad (3)$$

where $\langle V_i^2 \rangle$ is the mean-squared velocity of the stars with respect to a point in the galactic plane in a circular orbit around the center of the galaxy and $\langle V_g^2 \rangle$ is the corresponding velocity of the gas. Here we use $\langle V_g^2 \rangle^{1/2} = 14 \text{ km s}^{-1}$ (Spitzer, 1968). To find $\langle V_i^2 \rangle^{1/2}$ we first calculate V_{iz} , the average velocity dispersion in the galactic plane in the direction perpendicular to the plane required to produce the observed scale height β_i . This was done by using the force law perpendicular to the galactic plane

TABLE I
Mass loss in the galactic disk

M_V	M_i/M_\odot	τ_i (yr)	ϕ_i (pc ⁻³)	β_i (pc)	$\frac{dM_i}{dt}$ (M_\odot yr ⁻¹)	$\langle V_{iz}^2 \rangle^{1/2}$ (km s ⁻¹)	$\langle V_i^2 \rangle^{1/2}$ (km s ⁻¹)	L_i/L_\odot
-7	56	2.8×10^6	7×10^{-10}	50	3.9×10^{-4}	5	12.8	-1
-6	40	3.5	3×10^{-9}	50	9.5×10^{-4}	5	12.8	-2.5
-5	27	5.2	3×10^{-8}	50	4.3×10^{-3}	5	12.8	-11
-4	19	7.5	2×10^{-7}	50	0.014	5	12.8	-36
-3	14	1.1×10^7	6×10^{-7}	65	0.027	6.5	16.6	+174
-2	10	1.8	2.6×10^{-6}	80	0.064	7.5	19.2	+893
-1	7.1	3.5	1.3×10^{-5}	100	0.146	9	23.0	+3930
0	4.7	7.5	4×10^{-5}	120	0.161	10.5	26.9	+6860
1	3.1	2.1×10^8	1.3×10^{-4}	140	0.131	12	30.7	+7900
2	2.2	5.6	3.6×10^{-4}	160	0.097	13.5	34.6	+7850
3	1.7	1.4×10^9	9.3×10^{-4}	215	0.090	17.5	44.8	+13300
4	1.3	3.4	1.9×10^{-3}	270	0.057	21.5	55.1	+13100
5	1	1.0×10^{10}	2.9×10^{-3}	360	0.020	28	71.7	+7990
sum = $0.81 M_\odot$ yr ⁻¹						sum = $6.2 \times 10^4 L_\odot$		

found by Hill (1960) and Oort (1960), as analyzed by Mihalas (1968). Then we used the fact that the average velocity dispersions along the principal axes of the velocity ellipsoid are approximately in the ratio 8:5:4 (Ogorodnikov, 1965) so that $\langle V_i^2 \rangle^{1/2} = 2.56 \langle V_{iz} \rangle$. The resulting values of V_{iz} , V_i , and L_i are given in Table I. We note that the calculated values of V_i are somewhat larger than the average measured velocity dispersion of main sequence stars within the galactic plane (Mihalas, 1968). This results from a selection effect whereby the stars with large velocities on crossing the galactic plane spend less time in the vicinity of the plane than those having smaller velocities. In the table, the negative values of L_i for the upper Main Sequence stars indicate that for these stars, which have scale heights less than that of the gas, the loss of mass leads to a heating of the galactic disk; after the gas ejected by these stars has interacted with the gas in the interstellar medium, the terminal steady-state velocity of the ejected gas is actually greater than it was while in its parent stars. In our galaxy this heating is small compared to the cooling provided by mass loss from the more abundant, less massive Main Sequence stars. We may note, incidentally, that the kinetic energy per unit mass in the newly formed stars may be somewhat less than in the interstellar medium, and consequently, star formation may provide a further cooling of the galactic disk. However, this cooling is very much less than that provided by the gas being ejected from the older disk population stars.

The table indicates that the net power loss due to the shrinking of the galactic disk is $L_c = 6.2 \times 10^4 L_\odot$ which is adequate to provide the cooling necessary to maintain the spiral arms. In other galaxies, we can expect spiral arms to be present if the system is highly flattened (to meet the requirements of the Lin-Shu theory) and if the average

scale height of the stars suffering mass loss is greater than that of the gaseous disk into which the ejected gas settles. However, our model is not confined to this cooling mechanism. Any other mechanism which steadily increases the binding energy of the galaxy such as the infall of material from the halo or from the intergalactic medium would be just as effective in exciting the spiral arms as an equivalent amount of cooling provided by the infall of gas from dying stars in the galactic disk.

References

- Allen, C. W.: 1963, *Astrophysical Quantities*, Athlone Press, London.
- Hill, E. R. 1960, *Bull. Astron. Inst. Neth.* **15**, 1.
- Hills, J. G. and Dale, T. M.: 1973, *Astrophys. J.* **185**, 937.
- Humphreys, R. M.: 1972, *Astron. Astrophys.* **20**, 29.
- Iben, I.: 1967, *Ann. Rev. Astron. Astrophys.* **5**, 571.
- Lin, C. C. and Shu, F. H.: 1964, *Astrophys. J.* **140**, 646.
- Mihalas, D.: 1968, *Galactic Astronomy*, Freeman, San Francisco.
- Miller, R. H., Predergast, K. H., and Quirk, W. J.: 1970, *Astrophys. J.* **161**, 903.
- Ogorodnikov, K. F.: 1965, *Dynamics of Stellar Systems*, MacMillan, New York, p. 52.
- Oort, J. H.: 1960, *Bull. Astron. Inst. Neth.* **15**, 45.
- Spitzer, L.: 1968, *Diffuse Matter in Space*, Interscience, New York, p. 170.
- Toomre, A.: 1969, *Astrophys. J.* **158**, 899.