Mon. Not. R. astr. Soc. (1972) **157**, 1–30.

ON THE GENERATING MECHANISM OF SPIRAL STRUCTURE

D. Lynden-Bell and A. J. Kalnajs

(Communicated by the Astronomer Royal)

(Received 1971 December 29)

SUMMARY

Flat galaxies want to transfer their angular momentum outwards. Only in trailing spiral structures do the gravity torques carry angular momentum outwards.

We show that the presence of waves lowers the angular momentum in the inner parts and increases it in the outer parts. Further there is absorption of angular momentum by stars that resonate with the wave at the outer Lindblad resonance, and at corotation. Emission of angular momentum occurs at the inner Lindblad resonance. The role of spiral structure is to carry angular momentum from the inner parts to the outer parts so that the waves may grow.

We consider bars in galaxies to be standing waves which have grown enough to orient and trap the major axes of orbits with two lobes so that they lie along the bar.

1. INTRODUCTION

If spiral arms were material structures they would wind up in a time scale of a few times 10^8 yr. The large number of reasonably open spirals which must be of greater age suggests that spiral structure is either a recurrent gravitational instability (**1**), or a wave phenomenon. The latter possibility was investigated by Lindblad (**2**) and by Kalnajs (**3**) but the systematic development of the wave idea into a coherent picture is due to Lin, Shu and their collaborators (**5**)–(**10**). There is some direct evidence for propagating waves (**11**), and it seems likely that the wave idea is correct.

However, the theory developed so far puts in a spiral at the very beginning. In a sense it gives no answer to the question 'why are galaxies spiral at all?'. Furthermore in a shatteringly destructive article Toomre (**12**) has shown that if one adopts all the approximations of the Lin–Shu theory and lays down a spiral wave form initially then its propagation will lead to a winding up of the spiral in 10^9 yr or so and to its movement on to a Lindblad resonance. Here it may or may not be absorbed. Thus the theory as originally formulated is seriously incomplete and does not even account for the maintenance of the spiral wave for long times.

To save Lin's wave programme it is necessary to supply a generating mechanism for the waves. In an attempt to do this Lin (**8**) has followed up ideas originally due to Julian & Toomre (**13**) who looked at waves generated by masses in circular orbit. Lin's idea is that such waves are reflected from the inner Lindblad resonance as long waves which return to reinforce the original mass clumps.

Another mechanism has been suggested by Toomre (14) who is impressed by the fact that many of the finest examples of spiral galaxies have quite close companions. In a number of cases he considers that the gravity field of the companion causes the tidal wave which generates the spiral structure. However this idea, which may be traced to Jeans' speculations, probably does not apply to all spiral galaxies. Lindblad (2), and both of the authors (3) and (15) have all looked for a generating mechanism for spiral waves connected with stars that resonate in the spiral wave and we have criticized the work of Lin for neglect of those terms which can lead to angular momentum transference between stars and the wave (16). Contopoulos (7) has also made detailed studies of both stellar orbits close to resonance and the sense of spirality of waves growing there.

Waves themselves have angular momentum density which we show to be positive in the outer parts and negative in the inner parts. For growing waves there must be a transfer of angular momentum and in the limiting case of slow growth, absorption and emission of angular momentum are confined to stars whose orbits resonate with the perturbing field.

Resonances occur where the forces on an orbiting star due to the spiral wave are constant, or where their period is the same as the natural period with which a star oscillates about a circular orbit. The principal forces felt are constant at the corotation circle where the spiral wave rotates with the same angular velocity as the galaxy. As one moves inwards or outwards from that circle the frequency of the forces experienced by a star in circular orbit increases. When we have moved far enough inwards that frequency may achieve the natural epicyclic frequency and there we find the inner Lindblad resonance. The outer Lindblad resonance occurs outside the circular resonance where there is a similar coincidence of frequencies.

Kato (17) recently calculated the crucial angular momentum transfer on resonance by a method that parallels some of our calculations of Section 4, and we agree in his important conclusion that the work done by stars at the resonances may excite the wave which transfers the angular momentum. No appeal to gravitational instability in the outer parts nor to orbiting companions is then needed.

In the current enthusiasm for theories of spiral structure it is often tacitly assumed that the spiral gravitational potential perturbation coincides with the spiral as seen in H II regions and young stars. We would like to record at the outset that such coincidence is by no means an obvious consequence of the gravitational theory. Visible spirals may be a symptom of a more open spiral gravity field.

In this paper we first remark on the sense in which differentially rotating equilibria are minimum energy states for their angular momentum distributions. We then show that this minimum energy can be further lowered if some non-axially-symmetrical disturbance can be found which transfers angular momentum outwards (Section 2). By a study of the gravitational stress tensor (18) of a wave we show that only trailing spiral configurations transfer angular momentum in the right direction (Section 3). In Section 4 we study the angular momentum transfer between a star and a spiral wave. For a steady wave there is no average angular momentum or energy transfer except at resonances. Expressions for the angular momentum density of the wave and the angular momentum exchange between the resonant stars and the wave are then found following Stix's method for calculation of the physical picture of Landau damping in a plasma (19). Section 5 considers the energetics of the angular momentum transference, and produces a mechanistic explanation for what is happening at each of the principal resonances. For those who dislike mathematical details but wish to grasp what happens to enhance the spiral wave, this may be considered as a substitute for Section 4.

Section 6 considers angular momentum transport by the stars that form the wave while Section 7 gives a detailed discussion of the preference for two-armed spirals and a possible picture of the origin of barred spirals. Section 8 considers secular evolution of the shapes of galaxies.

2. MINIMUM ENERGY EQUILIBRIA

Let $h = Rv_\phi$ be the specific angular momentum. We may ask of any galaxy, 'How much mass is there with specific angular momentum in the range $h, h+dh$?'. If we call this mass $\mu(h)\,dh$, then $\mu(h)$ is the distribution of mass with angular momentum. The importance of the function $\mu(h)$ is that it is conserved for general time-dependent axially-symmetrical oscillations of any inviscid system. The reason is that each elemental ring of such a system conserves both its mass and its specific angular momentum during such oscillations. We normally work in terms of the cumulative distribution of mass with specific angular momentum $M(h)$ where

$$M(h) = \int_0^h \mu(h)\,dh.$$

(In systems whose specific angular momentum increases outwards $M(h)$ is the mass within that place where the value h is achieved.)

As in thermodynamics there is a tendency for isolated dynamical systems to evolve in the direction of increasing entropy. This is normally done by increasing the *energy* of their random motions. In the dynamical evolution of a galaxy between quasi-steady states which obey the virial theorem, the total energy, E, is fixed and the total potential and kinetic energies are given by $V = 2E$ and $T = -E$ respectively. We divide the kinetic energy into the energy associated with the systematic rotation about the galactic centre and the remaining 'random' kinetic energy. In particular we may write

$$T_{\text{rot}} = \tfrac{1}{2}\int h^2R^{-2}\mu(h)\,dh$$

where the integration extends over all h. Because T is fixed any increase in random motions requires a decrease in T_{rot}. This can be accomplished if mass with angular momentum moves outwards so that its R increases. This in turn cannot be done for all masses because V must remain constant. However V depends heavily on masses at small radii from the centre whereas T_{rot} is more evenly weighted in radius. A transfer of angular momentum outwards accompanied by an increase in the radius of the outer parts and a decrease in the radius of the inner parts will decrease T_{rot} and the random kinetic energy will increase by the same amount. Non-axially-symmetrical perturbations are capable of transferring angular momentum from mass to mass. The system may then fall toward the lower energy configuration and use the energy so gained to increase its coarse grained entropy (**20**)–(**22**).

It is in this spirit that we ask the important question 'How would a galaxy like to change its angular momentum distribution?'. To study this we ask the subsidiary question 'What changes in $M(h)$ lead to the lowering of the minimum energy attainable?'. To get an intuitive understanding of this problem we discuss the motion of two particles in a fixed galaxy-like potential. Let the masses be m_1 and m_2, the initial specific angular momenta be h_1 and h_2, and the specific energies ϵ_1 and ϵ_2. We ask, what is the minimum value of $E = m_1\epsilon_1 + m_2\epsilon_2$ subject to the constraint that $m_1h_1 + m_2h_2$ is fixed at H? We begin by studying the still more elementary problem: what is the least value of $\epsilon = \frac{1}{2}(v_R^2 + v_z^2 + h^2/R^2) - \Psi(R, z)$ for fixed h? Here $\Psi(R, z)$ is the fixed gravitational potential and for fixed R, Ψ will have its maximum at $z = 0$. Evidently the minimum value of ϵ will occur in the circular orbit of radius R_h which is the value of R at which $\frac{1}{2}h^2R^{-2} - \Psi(R, 0)$ attains its minimum value $\epsilon(h)$.

$$(h^2R^{-3} + \partial\Psi/\partial R)_{R_h} = 0$$

is the condition for a stationary value and so the least energy orbit is circular with centrifugal force balancing gravity. We shall presently need the following results on the minimum specific energy $\epsilon(h)$.

$$\epsilon(h) = \tfrac{1}{2}h^2R_h^{-2} - \Psi(R_h, 0)$$

$$\epsilon'(h) = d\epsilon/dh = hR_h^{-2} - (\partial R_h/\partial h)\,\partial/\partial R_h(\tfrac{1}{2}h^2R_h^{-2} - \Psi(R_h, 0)) = hR_h^{-2} = \Omega(R_h).$$

Let us now return from the one star problem to the two star problem which reduces to the minimization of

$$E = m_1\epsilon(h_1) + m_2\epsilon(h_2)$$

subject to the constraint

$$m_1h_1 + m_2h_2 = H.$$

Evidently

$$dE = m_1dh_1\epsilon'(h_1) + m_2dh_2\epsilon'(h_2)$$

where

$$m_1dh_1 + m_2dh_2 = 0.$$

Hence

$$dE = m_1dh_1(\epsilon'(h_1) - \epsilon'(h_2)) = m_1dh_1(\Omega_1 - \Omega_2).$$

Thus energy can be reduced by exchanging angular momentum in such a way that the orbit of least angular velocity gains angular momentum. Since for galaxies Ω decreases outwards this result means that *the minimum energy is lowered if the angular momentum flows outwards*. Although we have only proved this for two particles the result may be seen to be rather general because introducing viscous friction into any such system clearly leads to transference of angular momentum outwards and to the disappearance of energy in dissipation.

Let us note here that if this sense of angular momentum flow is achieved then the outer parts of a galaxy will move into larger orbits while the inner parts will contract.

3. ANGULAR MOMENTUM TRANSPORT BY GRAVITATIONAL STRESSES

To have access to the lower energy states a galaxy must find a mechanism of transferring angular momentum outwards. This cannot be done by axially symmetrical motions of a stellar system; they produce no gravitational couples between the inner parts and the outer parts. To see what form of gravitational disturbance is necessary we must first introduce the gravitational stress tensor (18). This is the similar object to Maxwell's electromagnetic stress tensor. To find an expression for the stress tensor one expresses the force density as minus the divergence of a tensor. Since the divergence does not define the tensor there are many possible choices, but since all such stress systems give rise to the same forces, any one may be used and one conventionally chooses the simplest. For gravity the force density is

$$\rho\nabla\Psi = -(4\pi G)^{-1}\nabla^2\Psi\nabla\Psi$$

where ρ is the mass density and Ψ is the gravitational potential. Clearly

$$\begin{aligned}\rho\nabla\Psi &= -(4\pi G)^{-1}[\mathrm{div}(\nabla\Psi\nabla\Psi)-(\nabla\Psi.\nabla)(\nabla\Psi)]\\ &= -(4\pi G)^{-1}\mathrm{div}[\nabla\Psi\nabla\Psi-\tfrac{1}{2}\boldsymbol{I}(\nabla\Psi.\nabla\Psi)] = -\mathrm{div}[(\mathbf{gg}/(4\pi G))-(g^2/8\pi G)\boldsymbol{I}]\end{aligned}$$

where $\boldsymbol{I}$ is the unit tensor δ_{ij} and $\mathbf{g} = \nabla\Psi$ the acceleration due to gravity. Hence

$$\rho\nabla\Psi = -\mathrm{div}\ \boldsymbol{T}$$

where $\boldsymbol{T}$, the stress tensor of gravitation field, is given by

$$\boldsymbol{T} = \mathbf{gg}/(4\pi G)-(g^2/8\pi G)\boldsymbol{I}.$$

Evidently $\boldsymbol{T}$ consists of an isotropic tension $g^2/(8\pi G)$, that is the negative of an isotropic pressure, together with a pressure $g^2/(4\pi G)$ along the lines of gravitational acceleration. Alternatively, we may say there are tensions $g^2/(8\pi G)$ perpendicular to the lines of gravitational acceleration together with pressures $g^2/(8\pi G)$ along them.

Consider the gravitational torque produced on the outer part of a galaxy by the inner part. Divide all space by a right circular cylinder centred on the galactic axis. Then the torque couple will be

$$\boldsymbol{C} = \int \boldsymbol{R}\times\boldsymbol{T}.d\boldsymbol{S}$$

where the integration extends over the whole cylinder and $d\boldsymbol{S}$ is along the outward normal to the cylinder. $\boldsymbol{R}$ is the distance from the galactic axis so $\boldsymbol{R} = (x, y, 0)$.

$\boldsymbol{R}\times\mathbf{I}.d\boldsymbol{S} = \boldsymbol{R}\times d\boldsymbol{S} = 0$, because $\boldsymbol{R}$ and $d\boldsymbol{S}$ both lie in the same direction. Hence the isotropic part of $\mathbf{T}$ does not contribute to the angular momentum transfer. Thus

$$C_z = (4\pi G)^{-1}\int Rg_\phi g_R\,dS.$$

Here ϕ is the azimuthal angle about the galactic axis.

Although the couple is a quadratic functional of the field, only the non-axially symmetrical components contribute. We indicate the non-axially-symmetrical part of g by the subscript 1 and write

$$C_z = (4\pi G)^{-1}\int g_{1\phi}g_{1R}\,dS.$$

Hence only the non-isotropic parts of the perturbation stresses contribute to the couple. Notice that to get a couple transferring angular momentum outwards, the average over the cylinder of $g_{1\phi}g_{1R}$ must be positive. Here ϕ is in the sense in which the galaxy is rotating. Thus, on average the gravitational acceleration field must have the sense of a leading spiral, which implies that the perturbation equipotentials have the sense of a trailing spiral (see Fig. 1). Thus *if gravitational torques are to lead a galaxy to a state of differential rotation of lower energy, then it must have a trailing spiral structure.*

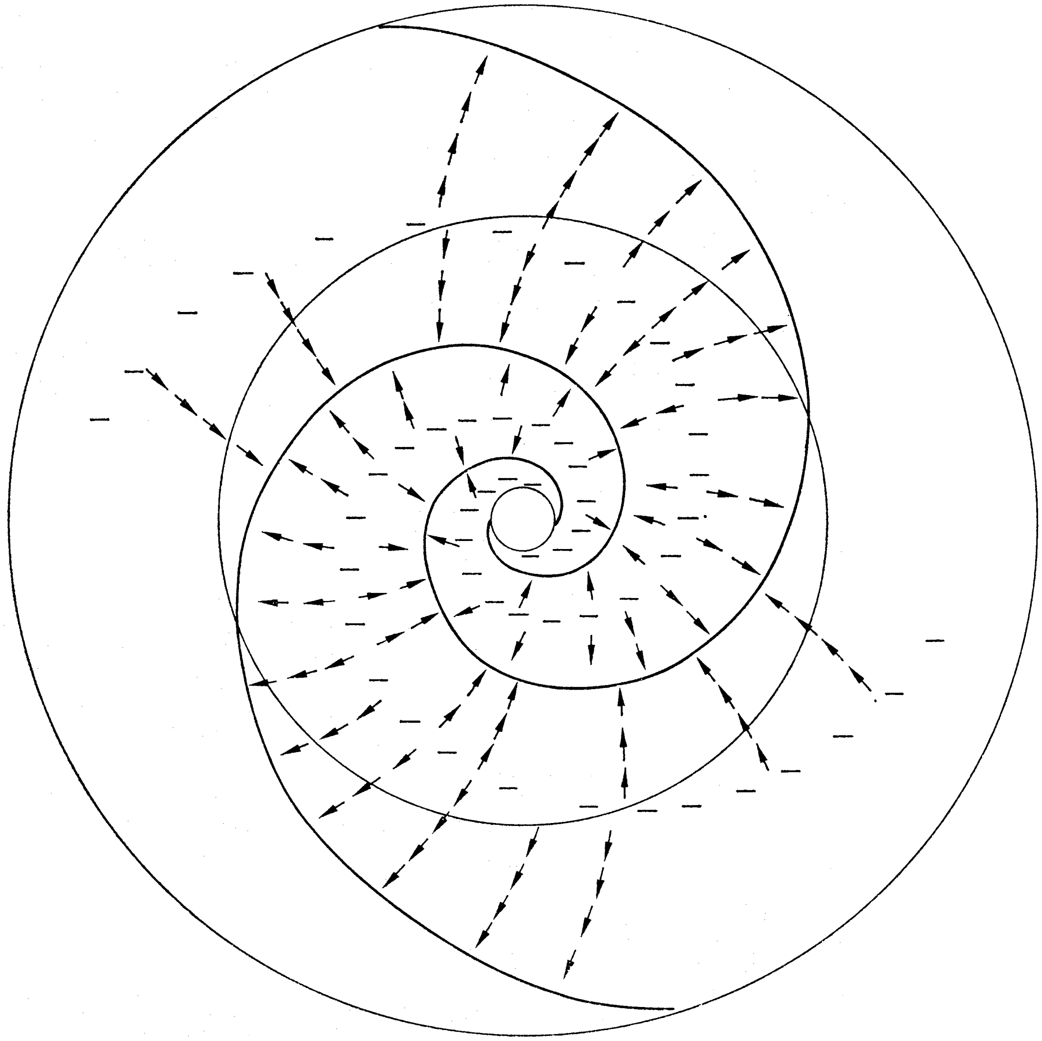

FIG. 1. *Illustrating that the lines of force lie almost orthogonal to the spiral structure. The torque depends on* $g_R g_\phi$ *which is independent of the sense of* g *but does depend on whether its direction trails (like the spiral) or leads. In the diagram* g *clearly 'leads' everywhere because of its approximate orthogonality to the trailing spiral.*

We record here an explicit example of a torque due to a spiral wave in the potential. Suppose the wave has the form

$$\psi = S(R) \cos m[\phi - \Phi(R)] \exp - |k_0 z|$$

postulated in the theory of Lin and Shu.

Then

$$g_{1\phi} = R^{-1} \partial\psi/\partial\phi = -mSR^{-1} \sin m(\phi - \Phi) \exp - |k_0 z|$$

and

$$g_{1R} = \partial\psi/\partial R = \{S' \cos m(\phi - \Phi) + mS\Phi' \sin m(\phi - \Phi)\} \exp - |k_0 z|$$

Hence

$$C_z(R) = -\tfrac{1}{4} m^2 R\Phi' S^2/(G|k_0|) = \tfrac{1}{4} m(k/|k_0|) RS^2/G$$

where $k = -m\Phi'$ and $k \simeq k_0$, the wavenumber, for the tightly wrapped waves of Lin and Shu. Note that if Φ' is negative then the couple on the outer parts, C_z, is positive in the direction of ϕ increasing. Since the form of spiral is given by $\phi = \Phi(R)$, only trailing spirals have Φ' negative and only they transfer angular momentum outwards by their gravitational torques.

4. ANGULAR MOMENTUM EXCHANGE BETWEEN STARS AND A WAVE

The arguments of Sections 2 and 3 have indicated why spiral structure helps a galaxy to lower its rotational energy. In this section we shall assume a weak spiral wave is present in order to see how the stars transfer angular momentum and energy to and from it.

The equations of motion of a single star moving in the galactic plane under the influence of a mean axially symmetrical gravitational potential Ψ and a spiral perturbation ψ are

$$\left.\begin{aligned} \ddot{R} - R\dot{\phi}^2 &= \partial/\partial R(\Psi + \psi) \\ d/dt(R^2\dot{\phi}) &= \partial\psi/\partial\phi \end{aligned}\right\} \tag{1}$$

We write

$$h = R^2\dot{\phi} \tag{2}$$

and

$$E = \tfrac{1}{2}(\dot{R}^2 + h^2/R^2) - \Psi. \tag{3}$$

Unperturbed motion

Notice that E and h change as the motion proceeds but that they are both constants of the unperturbed motion (for which $\psi \equiv 0$). We find it convenient to introduce new variables which incorporate the idiosyncrasies of the unperturbed orbits. For unperturbed motion we integrate equation (3) in the form

$$t - t_0 = \int^R [\dot{R}]^{-1}\, dR$$

where we write $\dot{R}$ in square brackets when it is to be thought of as a function of R, E, h obtained by inverting equation (3); to wit

$$[\dot{R}] = [2(E + \Psi(R)) - h^2R^{-2}]^{1/2}. \tag{4}$$

$[\dot{R}]$ has zeros (of square root type) at the apses of the orbit. On reaching these the other sign of the square root must be taken and the motion retraces its path. Thus the radial motion is periodic with angular frequency Ω_1 given by

$$2\pi/\Omega_1 = \oint [\dot{R}]^{-1}\, dR. \tag{5}$$

It is natural to use the phase of this radial oscillation as a coordinate in place of R itself. We therefore define the phase w_1 by

$$w_1 = \Omega_1 \int^R [\dot{R}]^{-1}\, dR. \tag{6}$$

The conjugate momentum to this angle variable is the radial action which is $1/2\pi$ of the area of the radial oscillation in phase-space.

Thus

$$J_1 = J_1(E, h) = (2\pi)^{-1} \oint [\dot{R}]\, dR. \tag{7}$$

Notice that J_1 is a constant of the unperturbed motion which vanishes when the orbit reduces to a circle about the galactic centre. J_1 is closely related to the square of the amplitude of the radial oscillations about circular motion.

By combining equations (2) and (3) we find ϕ to be given by

$$\phi - \phi_0 = \int^R hR^{-2}[\dot{R}]^{-1}\,dR.$$

The ϕ motion is non-uniform because R oscillates. We therefore define a new angle variable, which changes uniformly, by subtracting off this oscillation.

Thus

$$w_2 = \phi - \int^R (hR^{-2} - \langle hR^{-2}\rangle)[\dot{R}]^{-1}\,dR \tag{8}$$

where the pointed brackets indicate an average over one radial oscillation; that is

$$\langle hR^{-2}\rangle = (2\pi)^{-1}\Omega_1 \oint hR^{-2}[\dot{R}]^{-1}\,dR = \Omega_2. \tag{9}$$

Since w_2 increases with time at the rate $\langle hR^{-2}\rangle$ we have written $\Omega_2 = \langle hR^{-2}\rangle$ for conformity of notation. w_1, w_2, J_1 and h are the angle and action variables that could have been derived by following the Hamilton–Jacobi formalism. The advantage of that formalism is the explicit demonstration that w_1 and J_1, w_2 and h, are independent canonically conjugate coordinates and momenta. To acknowledge this fact we shall write $J_2 \equiv h$. In terms of our new variables w_i, J_i the unperturbed Hamiltonian may be found implicitly by solving equation (7) for E as a function of J_1 and $h(\equiv J_2)$. Calling this function $H_0(J_1, J_2)$ we have the equations of unperturbed motion

$$\dot{J}_i = -\partial H_0/\partial w_i = 0 \tag{10}$$

$$\dot{w}_i = \partial H_0/\partial J_i = \Omega_i(J_1, J_2). \tag{11}$$

The last equality may be formally demonstrated by differentiating equation (7) with respect to E and h and recognizing the results with the help of equations (4), (5) and (9).

$$\left.\begin{aligned}(\partial J_1/\partial E)_h &= \Omega_1^{-1}\\ (\partial J_1/\partial h)_E &= -\Omega_2/\Omega_1,\end{aligned}\right\} \tag{12}$$

where we have used the identity $(\partial E/\partial J_2)_{J_1} = -(\partial J_1/\partial h)_E/(\partial J_1/\partial E)_h$. A simpler verification is obtained by realizing that the solutions to equations (11) must be $w_i = \Omega_i t + w_i(0)$.

The meaning of our variables w_i, J_i is well illustrated in the special case of nearly circular motion. $[\dot{R}]^2$ may then be approximated to be quadratic in $R_1 = R - R_h$, the deviation from a circle. With $\epsilon(h)$ and R_h defined as in Section 2 we write

$$\kappa^2 = (3h^2/R_h^4) - (\partial^2\Psi/\partial R^2)_{R_h} \tag{13}$$

$$a^2 = 2[E - \epsilon(h)]/\kappa^2 \tag{14}$$

and approximate $[\dot{R}]^2$ by

$$[\dot{R}]^2 = \kappa^2[a^2 - R_1^2] + O(R_1^3). \tag{15}$$

This may be integrated into the form $R_1 = a\sin(\kappa t + \alpha)$ and we may evaluate all our other integrals over $[\dot{R}]$ to find explicit expressions. In particular

$$\phi - \phi_0 = \Omega_h t + 2(\Omega_h/\kappa)(a/R_h)\cos(\kappa t + \alpha) \tag{16}$$

where $\Omega_h = hR_h^{-2}$.

The expressions for the frequencies and action variables in this approximation are

$$\left.\begin{aligned} \Omega_1 &= \kappa, \quad \Omega_2 = \Omega_h \\ J_1 &= \tfrac{1}{2}\kappa a^2 = [E-\epsilon(h)]/\kappa, \quad J_2 = h \end{aligned}\right\} \tag{17}$$

whereas the angles w_1 and w_2 are approximately given by

$$\left.\begin{aligned} R_1 &= R-R_h = a \sin w_1 \\ \phi &= w_2+2(\Omega_h/\kappa)(a/R_0) \cos w_1. \end{aligned}\right\} \tag{18}$$

Quite generally w_1 is the phase of the radial oscillation; J_1 is a function of its amplitude; w_2 is the galactocentric angle to a uniformly moving epicentre and Ω_2 is its angular velocity about the galactic centre.

Perturbed motion of one star

Equations (10) and (11) show that in the axially symmetrical field Ψ the J_i remain constant while the w_i increase linearly with time. In the presence of the perturbing potential ψ the orbits will be changed: $J_i \to J_i+\Delta J_i$, $w_i \to w_i+\Delta w_i$. Since we are mainly interested in the *changes* produced we will reserve the symbols J_i, w_i for the unperturbed orbit and express the changes ΔJ_i, Δw_i as functions of the unperturbed position and time. The true position and momentum of any star will be $w_i' = w_i+\Delta w_i$ and $J_i' = J_i+\Delta J_i$.

The perturbed motion is generated by the new Hamiltonian $H_0-\psi$, and now Hamilton's equations read

$$\left.\begin{aligned} \dot{J}_j' &= -\partial/\partial w_j'(H_0-\psi) = \partial/\partial w_j'\psi(J_i', w_i', t) \\ \dot{w}_j' &= +\partial/\partial J_j'(H_0-\psi) = \Omega_j(J_i')-\partial/\partial J_j'\psi(J_i', w_i', t). \end{aligned}\right\} \tag{19}$$

We suppose that the time dependence of ψ is harmonic and that it was turned on slowly in the distant past. Thus we write

$$\psi(J_i, w_i, t) = \mathscr{R}\{\psi(J_i, w_i) \exp i\omega t\}, \tag{20}$$

with ω having a small negative imaginary part. The angle variables are periodic in phase space with period 2π, so we may expand ψ in a Fourier series

$$\psi(J_i, w_i, t) = \mathscr{R}\left\{(4\pi^2)^{-1} \sum_{l,m} \psi_{lm}(J_i) \exp[i(lw_1+mw_2+\omega t)]\right\}. \tag{21}$$

The Fourier coefficients ψ_{lm} are defined in the usual way

$$\psi_{lm}(J_i) = \int_0^{2\pi}\int_0^{2\pi} \psi(J_i, w_i) \exp[-i(lw_i+mw_2)]\, dw_1\, dw_2. \tag{22}$$

For epicyclic orbits the ψ_{lm} are evaluated for a particular potential in Appendix I.

First order orbits

If ψ is small the orbit deflections will be small so we may evaluate them to first order by computing the forces along the unperturbed orbit. Thus

$$\Delta\dot{J}_j \simeq \Delta_1\dot{J}_j = \partial/\partial w_j[\psi(J_i, w_i, t)].$$

The deflection $\Delta_1 J_1$ is found by integrating $\Delta_1\dot{J}_j$ over the past of the orbit

$$\Delta_1 J_j = \int_{-\infty}^{t} \partial/\partial w_j \psi(J_i, w_i-\Omega_i(t-t'), t')\, dt'.$$

With the expansion (21) the integral can be expressed as

$$\Delta_1 J_j = \partial\chi/\partial w_j, \tag{23}$$

where

$$\chi(J_i, w_i, t) = \mathscr{R}\left\{(4\pi^2)^{-1}\sum_{l,m}\psi_{lm}(J_i)\frac{\exp i(lw_1+mw_2+\omega t)}{i(l\Omega_1+m\Omega_2+\omega)}\right\}. \tag{24}$$

Similarly the first order angle deflections are given by integrating

$$\Delta_1\dot{w}_j = \sum_k \partial\Omega_j/\partial J_k\Delta_1 J_k - \partial\psi/\partial w_j = \sum_k \partial\Omega_j/\partial J_k\partial\chi/\partial w_k - \partial\psi/\partial J_j$$

which yield

$$\Delta_1 w_j = -\partial\chi/\partial J_j. \tag{25}$$

Equations (23)–(25) constitute the solution to the first order orbit when one remembers that the J_i are the initial values and that the w_i are the unperturbed angles that vary as $\Omega_i t$. We note that the $\Delta_1 J_j$ are periodic in the initial phases of w_1 and w_2. Thus the average gain of angular momentum of a set of stars that were initially distributed uniformly in the phases is zero to this order. This indicates that the angular momentum exchange between such a group of stars and the wave is of second order in the perturbation potential. There are two different ways of working out these second order terms, exemplified in the different treatments of the similar problem in the physical explanation of Landau damping. Stix (**19**) works out the changes in energy and momentum directly to second order, demanding that the initial positions and velocities be uniformly distributed. He calculates the rate of working on the perturbing potential directly to second order by following the particles. Dawson has argued correctly that going to second order must be unnecessary in a linear problem and he finds a different ' purely linear ' but by no means shorter route to the same answer. He implies that all methods that use non-linear terms must be open to doubt, but this is not the case. The truth is that the Lagrangian method must be taken directly to second order while, due to the rotational invariance of the unperturbed state, the Eulerian approach can obtain a correct second order answer by taking inner products of first order solutions. Since we find Stix's method pleasingly direct we shall follow that method. Eulerian theory gives exactly the same result, see Kalnajs (**3**).

Second order orbits

To calculate the changes to second order we evaluate the forces along the first order orbit. Thus we need $\partial\psi/\partial w_j$ evaluated at $J_i+\Delta_1 J_i$ and $w_i+\Delta_1 w_i$. The extra terms over and above the forces on the unperturbed orbit are clearly those on the right-hand side below

$$\Delta_2\dot{J}_j = \sum_k(\Delta_1 J_k\partial/\partial J_k+\Delta_1 w_k\partial/\partial w_k)\,\partial/\partial w_j\psi(J_i, w_i, t). \tag{26}$$

We have written $\Delta J_j = \Delta_1 J_1+\Delta_2 J_j$, correct to second order. Again the J_j and the w_i are the unperturbed values. Equation (26) may be evaluated using equations (20) and (23)–(25). The result simplifies considerably if we again ask for the change of the actions of a group of stars which were uniformly distributed with the w_i in the unperturbed state. This we obtain by averaging equation (26) over all

$$\langle \dot{h} \rangle = (2\pi)^{-2} \int_0^{2\pi} \int_0^{2\pi} \Delta_2 \dot{J}_2 \, dw_1 \, dw_2 = -\tfrac{1}{2}\mathrm{Im}(\omega) \exp[-2\mathrm{Im}(\omega)t](2\pi)^{-4} \times \sum_{l,m} m(l\partial/\partial J_1 + m\partial/\partial J_2) \frac{|\psi_{lm}|^2}{|l\Omega_1 + m\Omega_2 + \omega|^2}. \quad (27)$$

The details of this elementary reduction are somewhat lengthy. A parallel calculation gives $\langle \dot{J}_1' \rangle$ as the same expression but with the m after the $\sum$ replaced by l. In the limit when the potential perturbation is turned on very slowly $Im(\omega) \to 0$ from below and the angular momentum exchange vanishes except at 'resonances' where $|l\Omega_1 + m\Omega_2 + \omega| \to 0$. Thus a steady potential wave produces no secular change in angular momentum except at these resonances. The changes there are best considered in terms of the whole distribution of stars.

Angular momentum gain of a distribution of stars

For any flat unperturbed system the distribution function is a function of energy and angular momentum. We may therefore write it in the form $F = F(J_1, J_2)$. The total rate of change of the angular momentum of those stars which initially had J_2 in any particular range is

$$\dot{H} = 4\pi^2 \int_{h_1}^{h_2} \int_0^\infty \langle \dot{h} \rangle F(J_1, J_2) \, dJ_1 \, dJ_2$$

where the integration is over all J_1 and the range of J_2 considered. The $4\pi^2$ comes from integration over w_1 and w_2. Substituting expression (27) for $\langle \dot{h} \rangle$ and integrating by parts yields

$$\dot{H} = \frac{1}{8\pi^2} \mathrm{Im}(\omega) \exp[-2\mathrm{Im}(\omega)t] \times \left\{ \int_{h_1}^{h_2} \int_0^\infty \sum_{l,m} m \left(l \frac{\partial F}{\partial J_1} + m \frac{\partial F}{\partial J_2} \right) \frac{|\psi_{lm}|^2 \, dJ_1}{|l\Omega_1 + m\Omega_2 + \omega|^2} dJ_2 - \sum_{l,m} m^2 \int_0^\infty \frac{F|\psi_{lm}|^2 \, dJ_1}{|l\Omega_1 + m\Omega_2 + \omega|^2} \Bigg|_{h_1}^{h_2} \right\}. \quad (28)$$

The first integral corresponds to the change in the angular momentum of those stars that remain in our chosen range h_1 to h_2 while the second (boundary) term corresponds to the angular momentum in that range which is convected through the boundaries. It is interesting to integrate $\dot{H}$ from $t = -\infty$ up to t and to take only the first term so that we get the excess angular momentum for stars in a specified range of J_2 over and above that which they had in the absence of the wave. Denoting this by δH we have

$$\delta H = -\frac{1}{16\pi^2} \exp[-2\mathrm{Im}(\omega)t] \times \int_{h_1}^{h_2} \int_0^\infty \sum_{l,m} m \left(l \frac{\partial F}{\partial J_1} + m \frac{\partial F}{\partial J_2} \right) \frac{|\psi_{lm}|^2 \, dJ_1}{|l\Omega_1 + m\Omega_2 + \omega|^2} dJ_2. \quad (29)$$

Notice that this gives a definite value when $\mathrm{Im}(\omega) \to 0$, even away from resonances. This represents the angular momentum stored in the stars taking part in the wave motion. For end points near the resonances where $|l\Omega_1 + m\Omega_2 + \omega|$ vanishes, the splitting into a boundary term and a density term becomes badly defined as both become infinite. (We shall use expression (29) later for non-resonant ranges of J_2.)

Returning to the expression (28) for $\dot{H}$ we note that in the limit as $\mathrm{Im}(\omega) \to 0$ from below we have

$$-\mathrm{Im}(\omega)\,|l\Omega_1+m\Omega_2+\omega|^{-2} \to \pi\delta(l\Omega_1+m\Omega_2+\omega)$$

where δ is Dirac's delta function. Thus all the contribution to the integral comes from the resonances. The boundary term vanishes. We split $\dot{H}$ into terms arising from the different resonances by writing

$$\dot{H} = \sum_{l,\,m} \dot{H}_{lm}$$

where

$$\dot{H}_{lm} = -\frac{1}{8\pi}\int\!\!\int_0^\infty m\left(l\frac{\partial F}{\partial J_1}+m\frac{\partial F}{\partial J_2}\right)|\psi_{lm}|^2\delta(l\Omega_1+m\Omega_2+\omega)\,dJ_1\,dJ_2. \qquad (30)$$

Kalnajs already derived this expression from the Eulerian approach.

Locations of resonances and the senses of the angular momentum exchanges

To see where the resonances are and what they mean it is best to consider a single m-component of the potential

$$\psi = \sum_l \psi_{lm} \exp i(lw_1+mw_2+\omega t)$$

and to appeal to the epicyclic approximation in which $\Omega_1 = \kappa$, and $\Omega_2 = \Omega$, the angular velocity of circular orbit. Since $\psi \propto \exp i(m\phi+\omega t)$ (note that w_2 involves ϕ) our general expression is a $|m|$-armed perturbation whose pattern speed $\Omega_p = -\omega/m$. We see that there are resonances whenever $\Omega-\Omega_p = -l\kappa/m$ where l is the positive, negative or zero integer defined by the Fourier decomposition of ψ. The $l = 0$ resonance occurs where the circular velocity Ω coincides with the pattern speed Ω_p. Proceeding inwards from this circle one may eventually meet a circle on which Ω exceeds the pattern speed by $\kappa/|m|$. This will be one of the $|l| = 1$ resonances and the other will be found outside the corotation circle where Ω falls below Ω_p by $\kappa/|m|$. Indeed the angular frequency with which a star in circular orbit encounters the potential wave is just $\omega+m\Omega_2$ and when the star encounters it at zero frequency or at its natural frequency of vibration about circular orbit we expect resonances. These epicyclic resonances are called after B. Lindblad who discovered them. At the higher resonances $|l| \geqslant 2$ (which are of lesser dynamical importance) only stars with non-circular motions are involved. Stars in exact circular motion would encounter the wave at two or more times their epicyclic frequency but stars already in epicyclic motion still feel a component at exact epicyclic frequency as they move. Such higher resonances occur on circles even further from the corotation radius than the Lindblad resonances. It is likely that the outer one of any such pair will lie outside the galaxy altogether while the inner one would lie close to the nucleus, if it existed at all. For $|m| = 1$ the inner Lindblad resonance occurs only for retrograde waves. For $|m| = 2$ the resonances are spaced apart by something like galactic dimensions while for $|m| \geqslant 3$ they approach the corotation circle with increasing $|m|$. The preference for two-armed spirals is considered in Section 7. We now consider at which resonance stars gain angular momentum and at which they lose it. We first note that any reasonable distribution function will decrease as the epicyclic amplitude increases so $\partial F/\partial J_1$ will be negative. $\partial F/\partial J_2$ will also be negative due to the general outward fall-off of surface density in galaxies. Typical values of J_2 ($\equiv h$) are the radius of the galaxy times the circular velocity, whereas typical J_1 values are the size of an epicycle times the non-circular velocity, so $J_2 \gg J_1$ for typical stars away from the centre. This implies that $|\partial F/\partial J_1|$ is typically greater than $|\partial F/\partial J_2|$. From equation (30) we therefore deduce that for $l \neq 0$ the sign of the angular momentum exchange between stars and the wave is the sign of lm. For $l = 0$ the resonant stars gain angular momentum. Now lm is negative at the inner Lindblad resonance and therefore quite generally the stars at this resonance give out angular momentum. Similarly stars at the corotation resonance and the outer Lindblad resonance absorb angular momentum. We deduce that the arrangement of angular momentum emitters and absorbers is the right way round to enable a galaxy to lower its rotational energy by transferring angular momentum outwards.

Turning now to the minor terms at the Lindblad resonances we see that $m\partial F/\partial J_2$ opposes $l\partial F/\partial J_1$ when l and m are of opposite sign. Thus the secondary term opposes the primary angular momentum emission at the inner Lindblad resonance but reinforces the absorption of it at the outer Lindblad resonance.

The emission and absorption of angular momentum by resonant stars is *not* dependent on the sense (leading or trailing) of the spiral structure. The gravitational torques can only communicate angular momentum outwards if the spiral trails, so only trailing spirals can communicate the angular momentum between emitter and absorbers. This is not quite the full story of angular momentum transport which is more fully discussed in Sections 6 and 7, but it does give a qualitatively correct picture.

5. ENERGY TRANSFER AND THE MECHANISM AT THE RESONANCES OF A STEADY WAVE

Since the components of ψ with different m separate we may consider just one component at a time. The spiral wave we have considered is steady as viewed from axes that rotate with angular velocity $\Omega_p = -\omega/m$. Thus in those axes the total potential $\Psi + \psi$ is time independent. Each star will therefore conserve its energy with respect to these rotating axes. In celestial mechanics this energy with respect to rotating axes is known as the Jacobi constant. For a star with velocities v_R, v_ϕ with respect to non-rotating axes the Jacobi constant is

$$j = \tfrac{1}{2}[v_R^2 + (v_\phi - \Omega_p R)^2] - [\Psi + \psi + \tfrac{1}{2}\Omega_p^2 R^2]$$

per unit mass; here the $\frac{1}{2}\Omega_p^2 R^2$ is the potential of the centrifugal force which is seen in the rotating axes. Evidently

$$j = \tfrac{1}{2}(v_R^2 + v_\phi^2) - (\Psi + \psi) - \Omega_p R v_\phi = \epsilon_T - \Omega_p h$$

where ϵ_T is the total specific energy of the star in non-rotating axes and h is its specific angular momentum. Since each star has $dj/dt = 0$ we deduce that

$$d\epsilon_T/dt = \Omega_p\, dh/dt$$

or summing over any set of masses

$$dE/dt = \Omega_p\, dH/dt$$

where E is the energy of that set of masses and H is their angular momentum.

Thus to obtain the rate of working of the stars on the wave at any resonance one multiplies the rate of angular momentum loss *from* the stars by Ω_p.

The process of steady angular momentum transfer by a spiral wave is well illustrated in the diagram in which we plot specific energy against specific angular momentum (Fig. 2).

The straight line drawn is $\epsilon = \Omega_p h + \text{const.}$, where Ω_p is the pattern speed. The solid curve is the minimum energy possible in the unperturbed gravity field of the galaxy as a function of angular momentum. These minimum energies are attained for circular orbits. Further one may show that on this curve $d\epsilon/dh = \Omega(R)$ the angular velocity of circular motion with angular momentum h. A star that gains angular momentum δh at the corotation resonance gains energy $\Omega_p \delta h$. However since $d\epsilon/dh = \Omega_p$ at that resonance the star's energy of vibration about circular motion remain unchanged. However a star that loses angular momentum δh at the

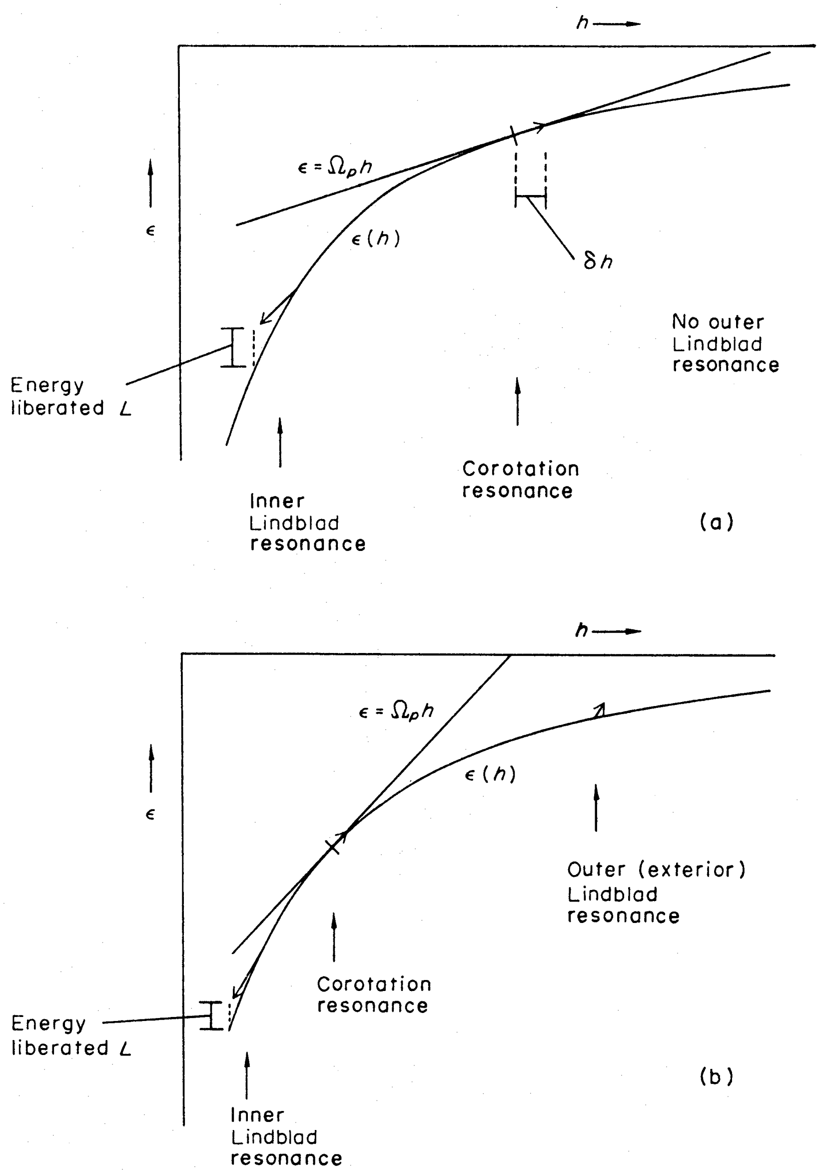

FIG. 2. *Illustrating the transfer of angular momentum* δh *and energy* $\Omega_p\delta$h *between stars in circular motion at the resonances.* $\epsilon = \epsilon$(h) *is the energy of circular unperturbed motion of angular momentum* h.

inner Lindblad resonance loses energy $\Omega_p \delta h$ but sinks to an angular momentum for which still less energy is required for circular motion. As a result the specific energy L in Fig. 2 is liberated into non-circular motion. Considerable enlightment will be found by those readers who construct the diagram like Fig. 2(b) for the case in which both the corotation and the exterior Lindblad resonance absorb angular momentum given out from the inner Lindblad resonance.

Mechanism at the resonances

To carry real conviction the angular momentum transfers at the resonances must not only be calculated but also understood.

Notice that the *major* terms at the Lindblad resonances do not involve angular momentum gradients at the resonance circle, and persist even for circular orbits.

The mechanism giving rise to these major terms may therefore be studied by considering the behaviour of an initially circular orbit exactly at the resonance. Let us study this simple case without the clutter of complicated mathematical apparatus. On a uniform circle there is no gravitational torque so the angular momentum exchange arises from the couple that the perturbation gravity field exerts on the distortion of the orbit. We get the couple correct to second order if we work out the displacement from the unperturbed orbit to first order; and so the displacements due to different causes add linearly. A resonant star in circular orbit feels the forces due to the perturbation potential $\psi = S \sin(kR_1 + m\phi + \omega t)$ at the epicyclic frequency. If we take S to be constant, the radial and transverse forces are in phase so we may write them $F_R \cos(\kappa t + \gamma)$ and $F_\phi \cos(\kappa t + \gamma)$. Writing suffices 1 on perturbations of the stellar coordinates and 0 on unperturbed quantities the equations of motion read

$$\ddot{R}_1 + \kappa^2 R_1 = (2\Omega_0/R_0)h_1 + F_R \cos(\kappa t + \gamma)$$
$$\dot{h}_1 = R_0 F_\phi \cos(\kappa t + \gamma)$$
$$\dot{\phi}_1 = h_1/R_0^2 - 2\Omega R_1/R_0.$$

Here F_R and F_ϕ are the amplitudes of the resonant forces and both phases are γ at the star at $t = 0$. Integrating, with zero initial conditions, one finds

$$h_1 = R_0 F_\phi \kappa^{-1}[\sin(\kappa t + \gamma) + \sin\gamma]$$
$$R_1 = \begin{Bmatrix} -\Omega_0 F_\phi \kappa^{-2} t \cos(\kappa t + \gamma) + \tfrac{1}{2} F_R \kappa^{-1} t \sin(\kappa t + \gamma) \\ + \text{terms of constant amplitude} \end{Bmatrix}.$$

We deduce that the secular effect of the resonant forces is to produce a forced eccentricity on the stellar orbit which, in the approximation of persistent exact resonance, will grow linearly with time. This oscillation consists of two parts due to the radial and transverse resonant forces. The larger radial forces produce secularly growing displacements in R_1 that lag the forces by one-quarter of the Lindblad period $2\pi/\kappa$, while the tangential forces produce secularly growing displacements in R_1 that lag their forces by half a Lindblad period. We now draw what is happening in the axes that rotate with the spiral structure and, to make everything concrete, we take the important case of a two-armed structure. We draw the situation in which we only consider the radial forces due to spiral structure and consider the inner Lindblad resonance.

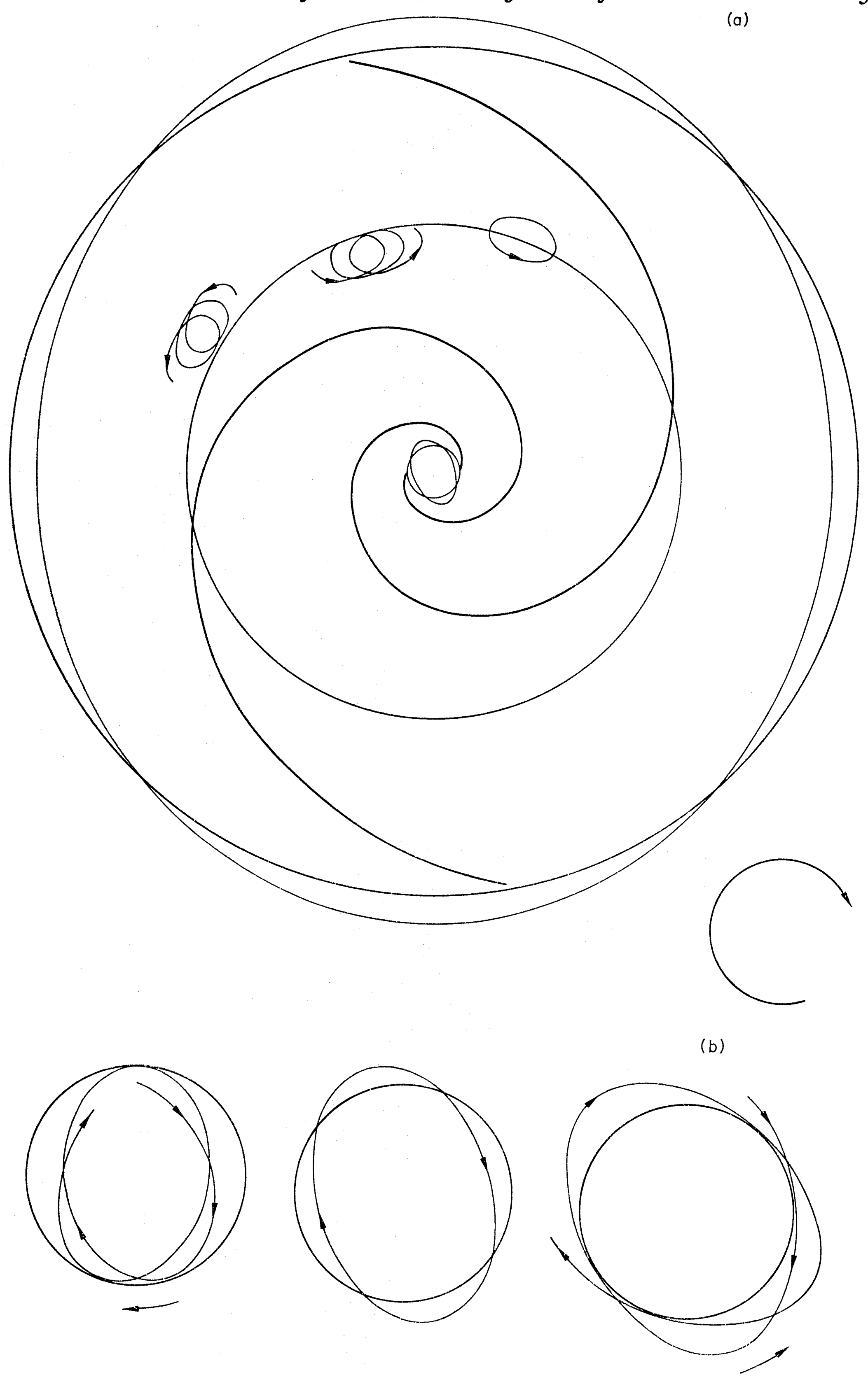

FIG. 3. (a) *Unperturbed resonant orbits at the three principal resonances, the inner Lindblad resonance, the corotation resonance and the exterior Lindblad resonance. Also shown are nearly resonant orbits close to the corotation circle. The figure is drawn in axes that rotate with the spiral structure.* (b) *An enlarged version of the region of the inner Lindblad resonance showing the rotation of the lines of apses for the nearly resonant orbits. The figure is drawn in axes that rotate with the spiral structure.*

In these axes unperturbed orbits close exactly at the resonances. At the inner Lindblad resonance $\Omega - \Omega_p = \kappa/2$ the orbits go in and out just twice for each time around the spiral structure, and the stars describe these orbits in the same sense as that in which the galaxy rotates. At the exterior resonance unperturbed orbits are of a similar shape, but the movement as seen from our rotating axes appears to be retrograde. At the corotation resonance an epicyclic unperturbed orbit reduces in our axes to Lindblad's little ellipse, there being no mean motion of its epicentre (Fig. 3).

Radial forcing at the Lindblad resonances (Fig. 4)

The effect of the radial forcing is to produce a forced eccentricity whose line of apsides coincides with the azimuth at which the spiral structure reaches the resonant circle. Consider the couple produced on such a slightly eccentric orbit due to the spiral structure. The couple is due to the radial displacements of the orbit from a circle. Consider the inner Lindblad resonance. The major axis is displaced from its position on the circle where there is no tangential force to a position just outside that circle where it slightly leads the arm. The resulting tangential force is pulling the arm forwards and the orbit backwards. Similarly, the displacement at the minor axis places it just behind a region where the arm structure has a negative density. This region repels the orbit and again angular momentum is taken from the orbit and fed into the spiral structure. It is simple to calculate the couple on the orbit due to this effect and to show that it varies as $\sin^2 m\phi$ where $\sin m\phi$ is a maximum on the spiral structure. This couple is independent of the sense of the spiral structure but only a trailing spiral structure can steadily transport the torque to the other resonances outside (cf. Section 3).

Consider now the same effects at the outer Lindblad resonance. Here the stars move backwards through the spiral structure so that a trailing spiral pattern generates inward displacements at the azimuths of the arms. Thus the minor axis lags the arms while the major axis leads the negative arms. By the mechanism above angular momentum is therefore transferred from the spiral wave to the stellar orbits.

Tangential forcing at the Lindblad resonances

Added to the radial forcing is the secular effect of the tangential forces. These produce amplitudes out of phase with the forces, so the major axis due to this term alone would lead the spiral structure in the diagram by $\pi/4$. In practice the combined effect of radial and tangential forcing produces a major axis a little ahead of the spiral structure. However the major effect of the tangential forces is not due to the eccentricity that they force on to the orbits, but rather to the slowing down and speeding up produced at different azimuths. Such effects are also present in the reaction to the radial forces but we have neglected them above because for those forces they give no net couple. To save argument, we will consider the effects of the tangential forces in isolation as though the radial forces were not present. In our axes the azimuth is swept out at a rate

$$|\dot{\phi}_A| = |\dot{\phi} - \Omega_p| = |hR^{-2} - \Omega_p|.$$

Since there are no secular effects in h the secular effects in $|\dot{\phi}_A|$ are a slowing down when R is large and a speeding up when R is small. This only holds at the inner resonance where the stars are going forwards in our axes. At the outer

2

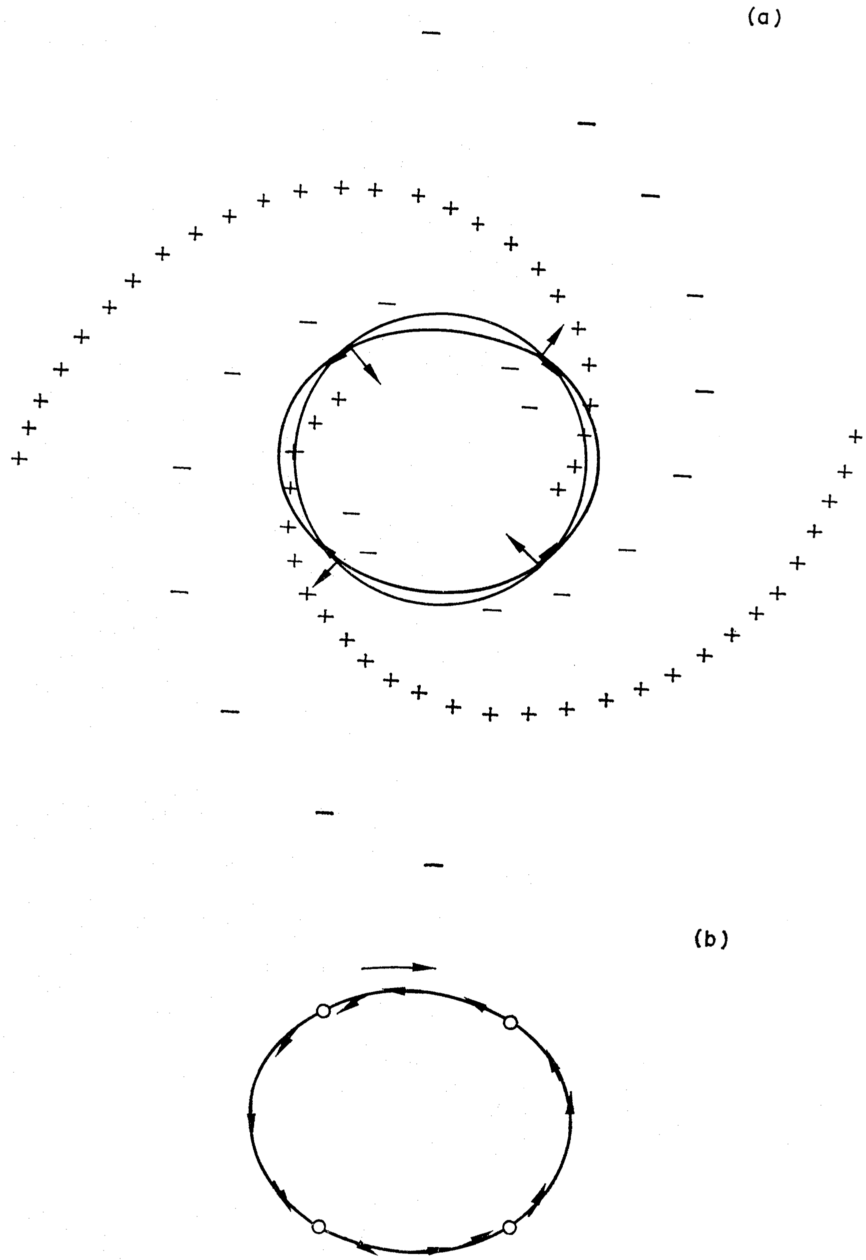

FIG. 4. (a) *The perturbation of the circular orbit forced by the radial forces due to the spiral at the inner Lindblad resonance. The forces felt on the circular orbit are illustrated.* (b) *The excess transverse forces felt on the distorted orbit over and above those felt on the circular orbit. It is clear that the star loses angular momentum and energy as it travels because the tangential forces are against its direct motion. The figures are drawn in the axes that rotate with the spiral structure.*

resonance $|\dot{\phi}_A|$ is smallest near pericentre as seen from our axes, for there $\Omega_p > h/R^2$. At the inner Lindblad resonance the tangentially forced major axis lies $\pi/4$ in front of the spiral structure, the excess density at those azimuths is attracted backwards towards the arm. Similarly the lack of density associated with the minor axis is attracted backwards to the lack of density in the spiral structure that follows it. Thus the couple due to tangential forcing reinforces the couple due to the radially forced eccentricity. The last statement is true at both Lindblad resonances and whatever the sense of spiral structure.

Although exactly on resonance the above physical arguments give a couple which increases linearly with time, this is not the full answer. Stars on orbits close to, but not quite on resonance, contribute for a long time, but eventually reverse their contributions. When one integrates over all near resonant contributions the net effect is time independent. An exactly similar state of affairs prevails in Dirac's treatment of emission and absorption by an atom (**25**).

In expression (30) for the rate of loss of stellar angular momentum on resonance, there are also minor terms which depend on gradients at the resonances. At the Lindblad resonances they vanish in the absence of eccentricity in the unperturbed orbit. We shall reserve their physical discussion until we have discussed the mechanism at the corotation circle, because the mechanisms are related.

Mechanism of angular momentum transfer at the corotation resonance and its relationship with Landau damping

Let us start with the physical explanation so often given for Landau damping so that we may see the relationship with the corotation resonance.

In Landau damping we consider the interaction of an electrostatic wave with particles travelling at velocities very close to the wave velocity. We take axes moving with the velocity of the wave. We consider a uniform distribution of particles which at the initial time have some velocity just a little faster than the wave. Those initially at and near the bottom of a potential trough slow down in trying to climb out, while those on the down grade accelerate into the trough. Thus: (i) there is an excess of particles to be found on the uphill climb and a deficit on the downhill slope.

Now on the uphill climb the particles push the wave in the direction that it is going while on the downhill slope they ride along at its expense. The small excess of particles found on the uphill climb produces a net pushing of the wave forwards. Thus (ii) particles going just faster than the wave feed momentum and energy into the wave. The similar argument applied to particles going just slower than the wave which move backwards in our axes is (iii) particles going just slower than the wave take momentum and energy from the wave. (iv) In normal situations there are more particles moving at the lower velocities so that effect (iii) dominates over effect (ii) and the energy and momentum of the wave are sapped at a rate that depends on the gradient of the distribution function at the velocity of the wave.

Now consider what is happening at the corotation resonance of spiral structure. Take axes rotating at the angular velocity of the wave. In these axes stars with epicentres just within the corotation circle move, on average, with angular velocity just greater than the wave, while stars with epicentres just outside the corotation circle have lesser average angular velocities. Let us unthinkingly take over the analogues of statements (ii) and (iii) above.

Because there is a general tendency for more density in the inner parts of the galaxy we deduce that there will be more particles with angular velocity just greater than the wave than those with less. Thus on our naive hypothesis the particles close to the corotation circle will feed angular momentum and energy into the wave. This simple argument, which I published earlier (**15**), is WRONG and contradicts our mathematical result of Section 4. To find the true physical argument we must consider more deeply. Statements (ii) and (iii) are true in the Landau damping case because of statement (i). Is statement (i) true in the context of the corotation resonance?

The unperturbed orbits close to corotation have very small mean motions in our axes, those on the inside drift slowly forwards those on the outside drift slowly backwards. On feeling the forward tug of a spiral arm a forward moving star will get on to an epicycle with a slightly greater angular momentum and its mean motion in the diagram far from speeding up will slow down. Thus in *azimuth* our stars act like donkeys slowing down when pulled forwards and speeding up when held back. Thus in place of statement (i) above we have (i′) there is an excess of stars to be found on the downhill side and a deficit on the uphill climb. (ii′) Stars with angular velocities just faster than the wave will on average be found on the downhill side and so will take angular momentum and energy from the wave. (iii′) Stars with angular velocities just slower than the wave give angular momentum and energy to the wave. (iv′) In normal situations there are more particles with the faster angular velocities and so the corotation resonance absorbs angular momentum and energy from the wave and gives them to the stars.

Mechanism of the gradient terms at the Lindblad resonances

The physical mechanism described earlier only accounts for the dominant terms at the Lindblad resonances, but our calculation shows that there are also terms involving angular momentum gradients. These terms vanish for circular orbits. We therefore consider non-circular unperturbed orbits and leave out the forced eccentricity whose effects we have already considered. Only exactly at the Lindblad resonances do the orbits close exactly in the rotating frame of the spiral structure. Associated with each orbit near the inner Lindblad resonance is a slight density excess associated with the line of its major axis. Orbits with angular momenta just less than the resonant one will not quite close but may be considered to be closed orbits rotating slowly forwards. The movement of the orbits is slow, while the particles move much more rapidly around the orbit. Orbits with angular momenta just greater than the resonant one will not close but may be considered as closed orbits rotating slowly backwards. If a forward rotating major axis is subject to a torque pulling forward, the particles in the orbit gain angular momentum and the precession of the orbit slows down. Thus the major axes of the orbits act like donkeys. We may now apply all of the arguments that we applied to donkey stars at the circular resonance to the density excesses associated with the major axes of the orbits at the inner Lindblad resonance. As before there is a slight excess of major axes in any region in which the torque is seeking to accelerate the motion of the major axis. Thus just outside resonance there is a slight excess density of the major axes at azimuths just lagging the spiral structure, while just inside resonance the major axes have a slight excess just leading the structure. Allowing for a basic outward fall-off of density we may expect a gradient term which is normally small but absorbs angular momentum. This effect cancels out a small part of the angular momentum given to the wave at the inner resonance by the dominant terms.

The similar explanation for the behaviour of minor axes at the outer Lindblad resonance is an exercise for clear thinking readers who should check their logic against the sign of the gradient term of equation (30).

6. LORRY TRANSPORT OF ANGULAR MOMENTUM

The angular momentum flow through the imaginary cylinder of Section 3 is not solely due to direct gravitational stresses. Some angular momentum is convected by the stars that cross the cylinder. Indeed if $\boldsymbol{S}^*$ is the stress tensor of the stellar motions defined in terms of the distribution function f by

$$\boldsymbol{S}^* = \int f\boldsymbol{v}\boldsymbol{v}\, d^3v$$

then the rate of convection of angular momentum is

$$C^* = \left(\int \boldsymbol{R}\times \boldsymbol{S}^*.d\boldsymbol{S}\right)_z = R^2 \int S_{\phi R}{}^*\, d\phi.$$

Here the integration $d\boldsymbol{S}$ is in principle over the cylinder but for a flat system reduces to an integration over the circle of radius R at $z = 0$. The stress tensor $\boldsymbol{S}^*$ is usually split into two terms, one due to mean motions $\bar{\boldsymbol{v}}$, and the other due to the pressure tensor. Thus

$$\boldsymbol{S}^* = \rho\bar{\boldsymbol{v}}\bar{\boldsymbol{v}} + \boldsymbol{P}^*$$

where

$$\boldsymbol{P}^* = \int f(\boldsymbol{v}-\bar{\boldsymbol{v}})(\boldsymbol{v}-\bar{\boldsymbol{v}})\, d^3v,$$

$$\bar{\boldsymbol{v}} = \rho^{-1}\int f\boldsymbol{v}\, d^3v, \quad \text{and} \quad \rho = \int f\, d^3v.$$

Notice that the couple C^* will contain a term $R^2 \int P^*_{\phi R}\, d\phi$ which is directly dependent on the deviation of the vertex of the velocity ellipsoid. While C^* vanishes in the unperturbed state, Dr Toomre insisted to us that the presence of the perturbation gravity field might inveigle stars into carrying more angular momentum on their outward journeys than on their return. Although we had already shown that in the presence of a steady wave the stars away from resonances neither gain nor lose angular momentum on average, he explained that they might nevertheless transport angular momentum just as a system of lorries can transport coal without accumulating a growing store on the lorries themselves. To prove that a star is allowing itself to be used as a lorry, we need to show that the time average of $R\dot{h}$ is non-zero, so that loading and unloading is correlated with the radius at which a star finds itself.

Evaluation of $\langle R\dot{h}\rangle$ *for a steady wave* ($Im(\omega) = 0$)

We can hereafter replace time averages by averages over unperturbed phases. To second order we have

$$\langle R\dot{h}\rangle = \langle (R_h + R_1 + \Delta_1 R + \Delta_2 R)(\dot{J}_2 + \Delta_1\dot{J}_2 + \Delta_2\dot{J}_2)\rangle$$
$$= \langle R_1\Delta_1\dot{J}_2\rangle + \langle R_1\Delta_2\dot{J}_2\rangle + \langle \Delta_1 R\Delta_1\dot{J}_2\rangle.$$

The remaining terms vanish to this order because $\dot{J}_2 = 0$ and $\langle\dot{h}\rangle = 0$ away from resonances. Now R_1 is independent of w_2 and so

$$\langle R_1\Delta_1\dot{J}_2\rangle = \langle R_1\partial\chi/\partial w_2\rangle = \langle\partial/\partial w_2(R_1\chi)\rangle = 0.$$

The remaining terms may be written

$$\langle R\dot{h}\rangle = \langle R_1\Delta_1\boldsymbol{J}.\,\partial^2\psi/\partial\boldsymbol{J}\partial w_2 + \Delta_1\boldsymbol{w}\,\partial^2\psi/\partial\boldsymbol{w}\,\partial w_2\rangle + \left\langle\left(\frac{\partial R_1}{\partial\boldsymbol{J}}.\Delta_1\boldsymbol{J} + \frac{\partial R_1}{\partial\boldsymbol{w}}.\Delta\boldsymbol{w}\right)\frac{\partial\psi}{\partial w_2}\right\rangle$$

$$= \left\langle R_1\left[\frac{\partial\psi}{\partial w_2}, \chi\right] + [R_1, \chi]\frac{\partial\psi}{\partial w_2}\right\rangle = \left\langle\left[R_1\frac{\partial\psi}{\partial w_2}, \chi\right]\right\rangle$$

where the square brackets are Poisson Brackets. In performing averages of Poisson Brackets the following lemma is useful

$$\int [a, b]\, d^2w = \int \left(\frac{\partial a}{\partial \boldsymbol{w}}\cdot\frac{\partial b}{\partial \boldsymbol{J}} - \frac{\partial a}{\partial \boldsymbol{J}}\cdot\frac{\partial b}{\partial \boldsymbol{w}}\right) d^2w = \int \left\{\frac{\partial a}{\partial \boldsymbol{w}}\cdot\frac{\partial b}{\partial \boldsymbol{J}} + b\,\frac{\partial}{\partial \boldsymbol{J}}\cdot\left(\frac{\partial a}{\partial \boldsymbol{w}}\right)\right\} d^2w$$

where the boundary term generated by the partial integration vanishes due to the periodicity of the w in phase space. Thus

$$\int [a, b]\, d^2w = \int \partial/\partial \boldsymbol{J}.(b\ \partial a/\partial \boldsymbol{w})\, d^2w.$$

Writing χ for a and $R_1\ \partial\psi/\partial w_2$ for b we have

$$\langle R\dot{h}\rangle = -\left\langle\frac{\partial}{\partial \boldsymbol{J}}\cdot\left(\frac{\partial \chi}{\partial \boldsymbol{w}} R_1 \frac{\partial \psi}{\partial w_2}\right)\right\rangle. \tag{31}$$

We evaluate this for specific examples of spiral waves in the Appendix. For very small eccentricities and trailing waves $\langle R\dot{h}\rangle$ is positive between the Lindblad resonances. Such stars gain angular momentum at large R and lose it at small R. Thus lorry transport of angular momentum by such stars opposes the transport by gravity torques. However the situation is more complicated for stars whose radial amplitudes, a, are comparable with the wavelength of the spiral structure $2\pi/k'$. In particular when $k'a \sim 1$ the lorry transport becomes small and may have either sign. When summed over many stars the total helps the gravity torque when the typical amplitude a_0 is such that $k'a_0 > 1$.

7. PREFERENCE FOR TWO ARMS AND THE ORIGIN OF BARRED SPIRALS

We owe to B. Lindblad the discovery that probably accounts for the observed preference for bisymmetry. To explain his idea we must first discuss the form of the unperturbed rosette orbits seen in rotating axes. An observer in a frame that rotates with angular velocity Ω_A still finds that the radial oscillation frequency is Ω_1 but he reckons the mean angular velocity of the star about the galactic centre to be not Ω_2 but rather $\Omega_2 - \Omega_A$. When Ω_1 and $\Omega_2 - \Omega_A$ are commensurable, i.e. when

$$p\Omega_1 = q(\Omega_2 - \Omega_A) \tag{32}$$

(where $|p|$ and q are relatively prime integers) then his plot of the orbit closes exactly. This occurs on the completion of q radial oscillations of the star at which time it also completes p turns about the galactic centre in his rotating frame. To our observer the orbit has q lobes and p turns. However another observer whose axes rotate with a different angular velocity will in general deny that the orbit ever closes because the radial and circulatory angular velocities seen by him will be incommensurable. Equation (32) may be solved for Ω_A for any given values of p and $q \neq 0$, and hence any non-circular orbit can be seen from suitable rotating axes as a closed q-lobe p-turn orbit. The angular velocity of the axes that give it this form is $\Omega_A = \Omega_2 - (p/q)\Omega_1 = \Omega - (p/q)\kappa$ where at the last equality we have used the epicyclic approximation. A stationary observer would see the lobes of the complete orbit rotating at this rate. We notice that in the epicyclic approximation the angular velocity of the axes that make each orbit a p-turn q-lobe orbit is just a function of radius, $\Omega_A(R_h)$, just as $\Omega = \Omega(R_h)$ and $\kappa = \kappa(R_h)$. More generally Ω_A is, like Ω_1 and Ω_2, a function of J_1 and J_2. Notice that our earlier condition for a resonance takes the form $\Omega_A(R_h) = \Omega_p$ the pattern speed. To get this we consider orbits with q, the number of lobes, equal to $|m|$ the number of arms, and p equal to $\pm l$. Our primary resonances have $l = \pm 1$ or 0 and the corresponding lobe angular velocities of the l-turn $|m|$-lobe orbits are $\Omega \mp \kappa/|m|$ and Ω, all of which are functions of R_h. Each in turn equals the pattern speed when R_h is set equal to the radii of the corresponding resonant circles. Lindblad noticed that $\Omega + \kappa/|m|$, $\Omega - \kappa/|m|$ and Ω all changed quite rapidly with radius except in the special case $|m| = 2$ where $\Omega - \kappa/2$ is nearly constant over quite wide regions of many spirals. He pointed out that the angular velocity of the 2-lobe 1-turn orbits is nearly constant over wide regions and so any density distribution made up from uniformly populated orbits of this type will be only weakly sheared by the unperturbed kinematics of the galaxy. The self-gravity of the perturbation does not have to be very great to overcome the weak shearing to make a uniformly rotating pattern.

Trapping of major axes

In Section 5 we discussed the analogy between what happens at the corotation resonance and the theory of Landau damping. We also pointed out a fascinating analogy between what happens to particles at the corotation resonance and what happens to the orientations of whole orbits due to the effects of the gradient terms at the Lindblad resonances. Contopoulos discovered these effects on the orbits and explored them with the help of detailed computer calculations. In the non-linear calculations of Landau damping and two stream instability we find the trapping of particles in the wave troughs. Similarly we may expect the trapping of the orientations of the major axes of near resonant orbits to occur and this is what Contopoulos finds near the Lindblad resonances. It is reasonable that in a non-linear treatment even non-resonant major axes will be caught provided that the major axes cannot cross the hills and valleys provided by the perturbing potential. This may be true over a wide area of the disc when the natural angular velocity of the lines of apsides $\Omega - \kappa/2$ does not vary rapidly with radius.

On the origin of bars

Bars are not an isolated problem different from the density wave theory, rather they form the interesting case of an almost steady standing wave. In the inner parts of galaxies the eccentricities of the stellar orbits have to be larger to ensure stability of axially symmetrical modes. The strengths of the resonances become small when the stellar orbits are eccentric and the modes of systems without resonances obey an anti-spiral theorem so that the basic two armed disturbance is a bar. The resulting tendency to bar making which is found in the linear theory can be further elucidated by considering the trapping of major axes of 2-lobe 1-turn orbits. Following Lindblad we consider a galaxy in which $\Omega - \kappa/2$ does not vary rapidly. Then a non-linear potential perturbation can trap major axes to oscillate about the azimuth of the potential trough like those found near Lindblad resonance by Contopoulos. The density of such trapped orbits will augment the potential and further enhance the trapping. Eccentricities are enhanced for such trapped orbits as at the inner Lindblad resonance so that near-circular orbits are rare. Thus like Lindblad we think that bars are made up of stars in rather eccentric orbits with aligned lines of apsides. However the rotation period of the bar so formed will be increased by the action of its gravitation on the orbits. Yet it must remain appreciably slower than the angular velocity of the stars that compose it.

8. CONDITIONS FOR SPIRAL WAVE GROWTH—SECULAR EFFECTS

It is natural to guess that provided the resonant stars do net work on the wave then the wave will grow. This would lead us to the condition for growth $\Omega_p \Sigma \dot{H}_{lm} < 0$. But this presupposes that the wave has positive energy. Since the energy of the wave is Ω_p times its angular momentum (3), we must examine the sign of δH. Taking the spiral wave of the Appendix, equation (29) may be written

$$\delta H = -\frac{1}{16\pi^2} \int_{h_1}^{h_2} \left\{ \sum_{l=1}^{\infty} \frac{4l^2 m^2 \Omega_1 (\Omega_2 - \Omega_p)(-\partial F/\partial J_1)}{|l^2 \Omega_1{}^2 - m^2 (\Omega_2 - \Omega_p)^2|^2} |\psi_{lm}|^2 - \sum_{l=-\infty}^{\infty} m^2 \left(-\frac{\partial F}{\partial J_2}\right) \frac{|\psi_{lm}|^2}{|l\Omega_1 + m(\Omega_2 - \Omega_p)|^2} \right\} dJ_1\, dJ_2.$$

The second term is small and can be neglected in the epicyclic approximation because $-\partial F/\partial J_1 \gg -\partial F/\partial J_2$. We have written δH in this form to demonstrate that within the corotation radius where $\Omega_2 = \Omega_p$, the dominant term in δH is negative whereas outside corotation both terms are positive. Thus the spiral structure within the corotation radius is a disturbance with negative angular momentum and negative energy. Feeding such a disturbance with energy or angular momentum will damp it—taking angular momentum or energy away from the disturbance will excite it. The corotation resonance which absorbs angular momentum and energy can enhance such a wave while the inner Lindblad resonance which emits angular momentum and energy will damp it. These ideas appear to conflict with the fact that the phase velocity of trailing spiral waves is outwards, but Toomre has already shown us that this difficulty is illusory because the group velocity of short waves is inwards in this region. Thus a possible picture is that waves of negative angular momentum and energy whose wave velocity is outwards are emitted from near corotation from where the group velocity carries them inwards until their negative angular momentum and energy are absorbed at the inner Lindblad resonance. (On going twice through the looking glass we find ourselves in a world that Lewis Carroll would enjoy.) Before we leave our expression for δH it is interesting to note that the mere presence of the wave, quite independent of resonances, lowers the angular momentum of the inner parts and increases that of the outer parts, independently of the sense of the spiral structure. Even without an inner Lindblad resonance a galaxy could evolve transferring negative angular momentum into a central standing wave which would be in the form of a bar.

On semi-observational grounds Lin has proposed that the pattern speeds of spiral waves may be found by putting corotation close to the edge of the galaxy. It is interesting that our inspection of the physical mechanism of angular momentum exchange showed that stars on the inner side of the corotation resonance emitted negative angular momentum strongly but that most of this was cancelled by stars on the outer side of corotation. Wave excitation might occur for waves with corotation at the edge for they get only the strong emission without the cancellation.

However if we base our arguments solely on the major terms then we are led to the picture of an inner region within corotation containing sources of angular momentum and a region outside corotation containing sinks of it. These sources and sinks are due both to the growth of wave angular momentum and to the resonant interactions. The communication of angular momentum between these regions, which is crucial for growth, is impeded because there is a region around corotation where short waves obeying the dispersion relation (Toomre (**12**) equation (10)) cannot propagate. (This statement assumes a sensible galactic model with Toomre's $Q > 1$ ($= 1 \cdot 5$ say). Even were corotation at the edge there would still be a considerable barrier against communication from that resonance by short-waves.) Through this barrier the waves have to tunnel and there will be much less attenuation if they are long or open. Such an open wave with pattern angular velocity somewhat less than half the angular velocity of the central region is the basic mode of a nearly uniformly rotating system and it will become unstable when coupled across the barrier to the outer region. Long waves are strongly favoured when we consider the non-linear effects that limit the maximum amplitudes attainable. A wave ceases to promote a supporting response from the stars once the velocities it generates outstrip its phase velocity. The radial displacement of a star due to a force per unit mass kS is $kS[\kappa^2-(\omega+m\Omega)^2]^{-1}$ and so the condition for no outstripping is $\kappa k S[\kappa^2-(\omega+m\Omega)^2]^{-1} < (\omega+m\Omega)/k$. Writing $\nu = (\omega+m\Omega)/k$, the time taken by a wave of amplitude S in trans-shipping a large fraction of the total angular momentum of a galaxy, $MR^2\Omega = H$, is

$$H/C_2 = MR^2\Omega/(\tfrac{1}{4}mRS^2/G) = 2\Omega^3R^4/S^2 = \frac{2\Omega^3(kR)^4}{\kappa^4\nu^2(1-\nu^2)} \sim (kR)^4/\Omega.$$

Thus there will be significant change in the angular momentum distribution after $(kR)^4/(2\pi)$ rotations. For this to happen in 100 revolutions we must have $kR < 5$; that is waves of inclination $i = \tan^{-1}(m/kR) > 23°$. The angular momentum structure of a galaxy can be significantly changed by such open waves, and therefore galaxies can change shape. The outer parts would expand, the inner parts contract and the orbits would become eccentric especially near the Lindblad resonances. If that resonance does not sop up all the negative angular momentum propagated inwards then standing waves of large amplitude—bars might be built, the evolution being through de Vaucouleurs types $SA \to SAB \to SB$. Normal galactic evolution might simultaneously change the classification from *c* to *b*.

9. POSTSCRIPT

When the wave idea was first proposed by B. Lindblad he laid great emphasis on the (Lindblad) resonances. Kalnajs also has laid stress on angular momentum and resonant interactions, and it is most interesting that the passing on of angular momentum and the forced eccentricity described here have close analogues in Godreich's beautiful work on the commensurable satellite systems of Jupiter and Saturn. Lin's programme for developing Lindblad's idea into a full theory has up to now led to a theory of waves with neither a convincing dynamical purpose nor a certain cause. It is hoped that this paper together with Kato's provide both a reason why trailing spiral waves help a galaxy to evolve dynamically, and a conceivable mechanism for generating waves. We emphasize that we have not proved that this mechanism will work—we have not shown that the excitation can overcome the damping. There is still plenty of work to do here. It was Lin's realization that tightly wrapped waves could be simply described which has founded an analytical theory, and he has led the way in looking to observation to demonstrate that waves are the important process of spiral structure.

Finally it would be wrong to reject Toomre's idea that quite a large number of spirals have been promoted by tides between galaxies. These can certainly excite strong short lived waves and a satellite in orbit could perfectly well act as a recipient of angular momentum.

ACKNOWLEDGMENTS

The first edition of this paper has been considerably changed particularly because of helpful criticism by Shu and Contopoulos and especially by Toomre. We would also like to thank Kato, Lin and Mark for sending their preprints, and both them and the referee, for comments.

Royal Greenwich Observatory

Received in original form 1971 *May* 14

REFERENCES

(1) Goldreich, P. & Lynden-Bell, D., 1965. *Mon. Not. R. astr. Soc.*, **130,** 125.
(2) Lindblad, B., 1958. *Stock. Obs. Ann.*, **20,** No. 6.
(3) Kalnajs, A. J., 1971. *Astrophys. J.*, **166,** 275, also 1965, *Thesis and Ref.* (8), p. 318.
(4) Lin, C. C. & Shu, F. H., 1964. *Astrophys. J.*, **140,** 646.
(5) Lin, C. C., Yuan, C. & Shu, F. H., 1969. *Astrophys. J.*, **155,** 721.
(6) Roberts, W. W., 1969. *Astrophys. J.*, **158,** 123, and *Ref.* (8), p. 415.
(7) Contopoulos, G., 1970. *Ref.* (8), p. 303. *Astrophys. J.*, **160,** 113; **163,** 181.
(8) Lin, C. C., 1970. *I.A.U. Symposium No.* 38, *The Spiral Structure of our Galaxy*, p. 377, eds Becker and Contopoulos. Reidel.
(9) Yuan, C., 1969. *Astrophys. J.*, **158,** 871 and *Ref.* (8), p. 391.
(10) Shu, F. H., 1970. *Astrophys. J.*, **160,** 89 and 99.
(11) Courtes, G. & Dubout-Crillon, R., 1971. *Astr. Astrophys.*, **11,** 468.
(12) Toomre, A., 1969. *Astrophys. J.*, **158,** 899.
(13) Julian, W. H. & Toomre, A., 1966. *Astrophys. J.*, **146,** 810.
(14) Toomre, A., *Refs* (12) and (8) and private communication.
(15) Lynden-Bell, D., 1969. *Ref.* (8), p. 331.
(16) Report by Bok, *Ref.* (8), pp. 464, 465.
(17) Kato, S., 1971. *Publ. astr. Soc. Japan*, **23,** 467
(18) Morgan, T. & Bondi, H., 1970. *Proc. R. Soc. A.*, **320,** 277.
(19) Stix, T. H., 1962. *The Theory of Plasma Waves*, pp. 132–136, McGraw-Hill, New York.
(20) Lynden-Bell, D., 1960. *Thesis*, Cambridge University.
(21) Lynden-Bell, D., 1962. *Mon. Not. R. astr. Soc.*, **124,** 279.
(22) Lynden-Bell, D., 1967. *Mon. Not. R. astr. Soc.*, **136,** 101.
(23) Shu, F. H., 1969. *Astrophys. J.*, **158,** 505.
(24) Erdélyi, A., Magnus, W., Oberhettinger, F. & Triomi, F. G., 1954. *Tables of Integral Transforms*, Vol. I, p. 186, Formula 39. Bateman Manuscript Project. McGraw-Hill.
(25) Dirac, P. A. M., 1947. *Principles of Quantum Mechanics 3rd edn.* Section 53, p. 205, Oxford.
(26) Goldreich, P., 1965. *Mon. Not. R. astr. Soc.*, **130,** 159.

APPENDIX I

EVALUATION FOR SPECIFIC FORMS OF SPIRAL WAVE

We shall consider the spiral potentials of Section 3, but for simplicity we take the complex form

$$\psi = S(R) \exp \{im[\phi - \Phi(R)]\}. \tag{A1}$$

We shall assume that $S(R)$ and $d\Phi/dR$ vary slowly and we shall evaluate our expressions for the nearly circular epicyclic orbits of Lindblad. For each orbit we replace $S(R)$ by $S(R_h)$ and Φ by $\Phi(R_h) - kR_1/m$ where

$$k(R_h) = -m\, \partial\Phi/\partial R \quad \text{evaluated at } R = R_h.$$

The Fourier coefficients ψ_{lm} are obtained by substituting the epicyclic expressions (18) into their definition (22) and using the above form of ψ. Thus

$$\psi_{lm}(R_h, a) = S(R_h) \exp[-im\Phi(R_h)] \times \int_0^{2\pi}\int_0^{2\pi} \exp[ika \sin w_1 + im(2\Omega/\kappa)(a/R_h) \cos w_1 - ilw_1]\, dw_1\, dw_2.$$

We write

$$k \sin w_1 + (2\Omega m/(\kappa R_h)) \cos w_1 = k' \sin(w_1 + \alpha)$$

so that

$$k'^2 = k^2 + [2\Omega m/(\kappa R_h)]^2 \tag{A2}$$

and

$$\alpha = \tan^{-1}[2\Omega m/(\kappa k R_h)] \tag{A3}$$

and we find ψ_{lm} reduces to

$$\psi_{lm} = 2\pi S(R_h) \exp i[l\alpha - m\Phi(R_h)] \int_0^{2\pi} \exp i[k'a \sin(w_1+\alpha) - l(w_1+\alpha)]\, dw_1.$$

The integral is

$$\int_{-\pi}^{+\pi} \exp i[k'a \sin x - lx]\, dx = 2\pi J_l(k'a)$$

where $J_l(k'a)$ is the Bessel function. It can be distinguished from the action variables by its argument. Notice that under our assumptions $|\psi_{lm}|$ is the same as $|\psi_{-lm}|$ because $|J_l|$ is independent of the sign of l. Our result for ψ_{lm} is

$$\psi_{lm} = 4\pi^2 S(R_h) \exp i[l\alpha - m\Phi(R_h)] J_l(k'a). \tag{A4}$$

We also evaluate our expression (31) for the lorry transport of angular momentum. This is made easier if we realize in advance that for our orbits and potential

$$m\, \partial\psi/\partial k = ima \sin w_1 \psi = R_1\, \partial\psi/\partial w_2.$$

With the complex form of ψ we remember that $\langle \mathscr{R}(a)\mathscr{R}(b)\rangle = \frac{1}{2}\mathscr{R}\langle ab^*\rangle$.

Thus

$$\langle R\dot{h}\rangle = -\tfrac{1}{2}\mathscr{R}\left\langle \frac{\partial}{\partial J}\cdot\left(\frac{\partial\chi}{\partial w}\, m\, \frac{\partial\psi^*}{\partial k}\right)\right\rangle$$

and using the definition of χ

$$\langle R\dot{h}\rangle = (2\pi)^{-6}\,\tfrac{1}{2}\mathscr{R}\int_0^{2\pi}\int_0^{2\pi}\sum_l\sum_{l'}(l\,\partial/\partial J_1 + m\,\partial/\partial J_2) \times\left(\frac{\psi_{lm} m\,\partial/\partial k\psi_{l'm}^*}{l\Omega_1+m\Omega_2+\omega}\right)\exp\,[i(l-l')w_1]\,dw_1\,dw_2$$

$$= (2\pi)^{-4}\sum_l \tfrac{1}{4}m(l\,\partial/\partial J_1+m\,\partial/\partial J_2)[(l\Omega_1+m\Omega_2+\omega)^{-1}\,\partial/\partial k\,|\psi_{lm}|^2]. \quad \text{(A5)}$$

In terms of epicyclic variables this may be rewritten

$$\langle R\dot{h}\rangle = \sum_l \tfrac{1}{4}m(l\kappa^{-1}a^{-1}\,\partial/\partial a+m\,\partial/\partial h)\{(l\kappa+m\Omega+\omega)^{-1}S^2\,\partial/\partial k[J_l(k'a)]^2\}. \quad \text{(A6)}$$

To understand this expression it is interesting to look at the limit $k'a \ll 1$ when the wavelengths of the spiral structure are considerably greater than the epicycle size. In this limit only the $l = \pm 1$ terms contribute and the expression reduces to

$$\langle R\dot{h}\rangle = \tfrac{1}{2}kmS^2[\kappa^2-(\omega+m\Omega)^2]^{-1}+O(k'a)^2. \quad \text{(A7)}$$

Using the properties of Bessel functions, one may check that the major terms keep this same sign up to $k'a = 1$ but their magnitude decreases as $k'a$ increases.

We notice that between the Lindblad resonances stars pick up angular momentum from a trailing wave near their apocentres and on average transport it inwards. Since the resulting transport opposes the gravity torque we must work out which is biggest and so we must calculate the lorry transport due to a distribution of stars. To find this flux of angular momentum we must work out the sum of contributions $h\dot{R}$ from all stars currently crossing a reference cylinder such as that of Section 3. We shall approximate assuming the galaxy varies sufficiently slowly with R so that the average of $h\dot{R}$ on the cylinder is no different from the average $h\dot{R}$ for all stars whose initial angular momenta put their epicentres on the cylinder. In the same approximation the unperturbed surface density σ at R is given by thinking of every particle as localized at its epicentre since the random motions merely cause a smoothing of that distribution which is unimportant if the distribution is already smooth. Thus

$$2\pi R_h\sigma(R_h)\,dR_h = \mu(h)\,dh \quad \text{(A8)}$$

where

$$\mu(h) = \iiint F_0\,dw_1\,dw_2\,dJ_1$$

which is the angular momentum distribution of Section 2. With these approximations the net couple due to lorry transport is

$$C^* = 2\pi R\sigma[\langle h\dot{R}\rangle] = \mu(dh/dR_h)[\langle h\dot{R}\rangle]$$

where

$$[\langle h\dot{R}\rangle] = \iiint F_0\langle h\dot{R}\rangle\,dw_1\,dw_2\,dJ_1/\mu(h).$$

Now since we have already excluded resonances our stars do not gain or lose angular momentum secularly, $\langle\dot{h}\rangle = 0$. Similarly $\langle\dot{R}\rangle = 0$ and $\langle h\dot{R}\rangle+\langle R\dot{h}\rangle = 0$ for each star. We therefore rewrite our expression for C^* to use our earlier value for $\langle R\dot{h}\rangle$

$$C^* = -4\pi^2(dh/dR_h)\int F_0\langle R\dot{h}\rangle\,dJ_1 \quad \text{(A9)}$$

For very long waves we may use expressions (A7) and (A8) to obtain

$$C^* = -\tfrac{1}{2}km2\pi\sigma RS^2/[\kappa^2-(\omega+m\Omega)^2].$$

Unlike our expression for the gravity couple this vanishes as $k \to 0$. If long waves are imposed the gravitational torques dominate. However this result must be qualified because we find presently that for the longest *self-sustaining* waves propagating according to the *approximate* dispersion relation the result is not clear cut and there can be circumstances in which lorry transport by long waves beats the gravity torque.

To perform the integral in (A9) for more general waves we need a specific distribution function so we choose the usual form

$$F_0(J_1, J_2) = \tfrac{1}{4}\pi^{-2}\langle J_1\rangle^{-1}\exp(-J_1/\langle J_1\rangle)\,\mu(J_2)$$
$$= \tfrac{1}{4}\pi^{-2}(\kappa a_0^2)^{-1}\exp(-\tfrac{1}{2}a^2/a_0^2)\,\mu(h)$$

here a is the radial amplitude of epicyclic motion and a_0 is its average. Evaluating expression (A9) with this F_0 and writing (A6) we find after integrating by parts on J_1

$$C^* = \mu(dh/dR_h)\sum \tfrac{1}{4}m[l(\kappa a_0^2)^{-1}+m\,\partial/\partial h]$$
$$\times\left\{S^2(R_h)[l\kappa+m\Omega+\omega]^{-1}\int_0^\infty(\kappa a_0^2)^{-1}\exp(-\tfrac{1}{2}a^2/a_0^2)[J_l(k'a)]^2\,d(\tfrac{1}{2}\kappa a^2)\right\}.$$

Since we have already omitted small gradients across an epicycle we must for consistency omit the $m\,\partial/\partial h$ term as $\kappa a_0^2 \ll h$. The remaining integral is well known (**24**). Indeed

$$\int_0^\infty \exp(-\tfrac{1}{2}x^2)[J_l(zx)]^2x\,dx = \mathrm{I}_l(z^2) = \exp(-z^2)I_l(z^2)$$

where $I_l(z)$ is the Bessel function of imaginary argument. Hence using equation (A7) with R written for R_h we have

$$C^* = -\tfrac{1}{2}\pi R\sigma S^2(\kappa a_0)^{-2}\,\partial/\partial k\left[\sum_{l=-\infty}^{\infty} l\kappa\mathrm{I}_l((k'a_0)^2)/(l\kappa+m\Omega+\omega)\right].$$

We must compare this couple with the gravity torque of Section 3 which for this wave gives $C_z = \tfrac{1}{4}m(k/|k_0|)RS^2/G$ where $k_0^2 \simeq k^2+m^2/R^2$. The total couple is

$$C = C^*+C_z = \tfrac{1}{4}mRS^2G^{-1}\,\partial/\partial k\{|k_0|[1-D(k,\omega)]\} \tag{A10}$$

where

$$D(k,\omega) = 2\pi G\sigma(\kappa^2a_0)^{-1}|k_0a_0|^{-1}\sum_1^\infty 2\mathrm{I}_l(k'^2a_0^2)[1-\nu^2/l^2] \tag{A11}$$

and

$$\nu = \kappa^{-1}(m\Omega+\omega).$$

Comparison with Toomre's equations (10) and (11) for tightly wrapped waves shows that $D(k,\omega) = 1$ is the dispersion relation. Our replacement of k by k_0 and k' gives a dispersion relation which is reasonably correct for much more open waves. Comparison of (A11) with equation (29) evaluated in the present approximation shows us that $L = \tfrac{1}{4}mRS^2G^{-1}|k_0|\,\partial D/\partial\omega$ is the angular momentum density so that

$$\int_{h_1}^{h_2} L(\partial R_h/\partial h)\,dh = \delta H.$$

For waves obeying the dispersion relation we have $\partial D/\partial k = -(\partial D/\partial \omega)(\partial \omega/\partial k)$ and so

$$C = L\,\partial\omega/\partial k.$$

The total couple is the angular momentum density times the group velocity. For tightly wrapped waves both L and $\partial\omega/\partial k$ reverse sign at the corotation circle so C is positive. Furthermore since no angular momentum is gained by non-resonant stars we find that steady waves will have C constant between resonances. This equation determines how $S(R)$ varies with R.

Notice from (A10) that the sign of C for *any* waves *obeying the dispersion relation* is the sign of $-m\,\partial D/\partial k$ evaluated when $D = 1$. Although this is definitely positive when $k'a_0 > 1$ it is possible that it may become negative for long waves with $k'a < 1$. This certainly happens if following Lin we write k for both k' and $|k_0|$ in our dispersion relation, but those are bad approximations for the open waves involved. In practice we think it likely that, in so far as open waves can be described by a dispersion relation at all, $\partial D/\partial k$ is everywhere positive at $D = 1$. This leads to the outward propagation of angular momentum even by long waves since C is then positive.