II. SPIRAL ARMS AS SHEARED GRAVITATIONAL INSTABILITIES

P. Goldreich and D. Lynden-Bell*

(Received 1964 June 25)

Summary

This paper treats examples of gravitational instability in differentially rotating media. The particular cases dealt with include polytropic, stratified sheets of gas, as well as the infinite homogeneous media. Application of the results is made to the formation of spiral arms in a differentially rotating disk galaxy. Relations are derived which connect the thickness of the galactic disk, its density, and the velocity dispersion perpendicular to the galactic plane with Oort's differential rotation parameters A and B.

Sections 1 and 2 discuss requirements of any theory of spiral arms.

Sections 3 to 8 give a mathematical treatment of gravitational instability in a sheared rotating stratified medium.

Section 9 discusses these results qualitatively and proposes a theory of spiral arm formation based on them.

Section 10 gives the observational tests and consequences of the theory.

Finally Section 11 gives a very brief discussion of barred spirals and points to the many problems left unsolved by the present work.

1. *Introduction*

Since Lord Rosse (1) discovered spiral structure in M 51 the explanation of this beautiful form has been one of the outstanding problems of cosmogony. The straightforward belief that this structure is a natural consequence of a swirling motion was probably held by many of the early observers and it is our hope that the present work goes some distance to establish that belief on a firm theoretical foundation.

Jeans (2) tried to identify the arms with pieces of material that would be shed equatorially as a uniformly-rotating centrally-condensed mass slowly shrank. We know now that a galaxy is not a pressure-supported mass of glowing gas, but a star–gas mixture supported by rotation or stellar motions. Secular shrinking is not therefore a natural form of evolution. The observed rotation is not normally uniform except in barred spirals where a theory reminiscent of Jeans' still looks promising.

Lindblad attempted to give an explanation of spiral arms, first in terms of orbits (3) and then in terms of the complete self-gravitating perturbations of a stellar system (4). His concept of a hierarchy of subsystems with different flattenings was a forerunner of Baade's discovery of different stellar populations, while it is his realization of the dynamical importance of flattening for the stability of the galactic disk that we shall develop here.

* N.A.S.–N.R.C. Fellow 1963–64.

In external galaxies we recognize spiral arms by their brightness on a photographic plate and the knots of H_{II} regions that occur along them. It is sometimes possible to extend the arms inwards towards the nucleus by allowing the eye to follow dust markings. Whether or not such markings are a genuine extension of the same spiral structure we do not know, but we take the primary characteristic of spiral arms to be their photographic brightness. This brightness is caused by very luminous stars, so luminous in fact that they have short lives and can hardly have moved from their birthplaces. We deduce that stars are being formed in spiral arms. Star formation requires considerable condensation of the interstellar medium and if new stars are the end-products of Jeans' gravitational instability then spiral arms *must* be the seat of such instability. This at once raises the question whether the arms themselves can be due to gravitational instability on a slightly grander scale. Whether that is the case or not it is most important to know how gravitational instability occurs in a differentially rotating structure of finite thickness. It is this problem that we shall solve.

2. *Requirements of any theory*

In this section we shall talk only of normal spirals; a discussion of barred spirals is given at the end of the paper.

2.1. *Form.*—Any theory must be wide enough to contain the bewildering variety of galactic forms. The conventional picture of two spiral arms starting symmetrically from the nucleus and winding several times around like continuous threads is wrong in several respects. In only about a third of all normal spirals can it be claimed that just two arms are dominant and although in these there is some tendency to symmetry it is not always very pronounced. The arms are not normally continuous and can be traced without ambiguity once around the nucleus only rarely. In many though not in all cases these arms give the impression of several pieces joined at kinks. But these kinks may be perturbations on the continuous arms of the conventional picture. There are galaxies that fit that picture. The symmetry of their large-scale structure must depend on a more realistic discussion of gravitational instability than we can give here. However we think that this structure could form in a large-scale version of the same type of instability. The remaining two-thirds of normal galaxies are multiply armed structures. In Sc's the arms often branch at unlikely angles and the whole structure is considerably more messy than the conventional picture. A swirling hotch-potch of pieces of spiral arms is a reasonably apt description. A correct theory must have room for neat symmetrical two-armed spirals, but it must not predict that most normal galaxies should be like that. The mechanism of spiral arm formation must be so universal that it can still work under the difficult messy conditions of a typical spiral galaxy.

2.2. *Dynamics.*—S0 galaxies are topographically similar to normal spirals but they have no gas, no dust and no spiral arms. This suggests that stellar dynamics is not alone responsible for arm formation. Gas dynamics differs from stellar dynamics in the following respects:

(1) The nature of the pressure is different. In the interstellar gas turbulent pressure dominates.

(2) Turbulent energy is dissipated and then lost by radiation whereas energy is conserved in stellar dynamics.

(3) The gas is subject to magnetic forces.

(4) Gas may be cooled by the presence of dust grains, etc.

Any theory of spiral arms must depend on at least one of these differences or be merely a reflection of special initial conditions.

2.3. *The winding problem.*—If the arm structure rotates differentially, as the observations indicate that the H_{II} regions do, then the pitch must diminish. In times that are typically a few 10^8 years the arms will become tightly wound. However the proportion of normal spirals with tightly wound arms is small, and it is currently believed that galaxies are typically 10^{10} year-olds. Unless the galaxies have conspired all to be spiral together for a very brief period we must deduce that either:

(1) the spiral structure rotates nearly uniformly although the material rotates differentially, or

(2) the arms are short-lived but reform as open structures, or

(3) that the observations are wrong and spirals rotate nearly uniformly.

To admit (3) is to say that the theorist is bankrupt of ideas. There is little doubt that a large fraction of spirals rotate with considerable shear.

Perhaps the most promising of the theories based on (1) is the density wave theory (**5**) in which a nearly uniformly rotating spiral wave propagates through the star–gas fluid. Stars are most likely to form near the density maxima. To date theoretical discussions of such waves have been limited to thin disk models with infinite density and no pressure. All such models are violently unstable (**6**) since the growth rate of Jeans' gravitational instability is proportional to $(G\rho)^{1/2}$.

A second type of theory based on (1) allows considerable radial streaming and maintains a uniformly rotating structure in a differentially rotating medium by forcing the material to flow out along the arms. Magnetic forces are usually invoked to do this. The weak point of such theories is the large angular momentum transport required to keep the outflowing material at the observed angular momenta. Magnetic fields cannot provide the necessary torque unless they are either impossibly strong, $>10^{-4}$ G, or violently bent on the scale of the thickness rather than the radius of the galactic disk. No fully developed theory based on such ideas has yet appeared.

The present theory is based on (2), the idea that arms are constantly forming and dying. A natural regenerative mechanism based on the turbulent dissipation in the interstellar gas and gravitational instability (**7**) will be proposed after our mathematical discussion of the instability.

3. *A property of differential rotation*

Gravitational condensation in a differentially rotating medium is a complicated process. If we are to understand it we must first solve similar problems which have only a few of the complicating features. In paper I we discussed a series of such simplified problems all of which involved uniform rotation. In this section we discuss the most drastically simplified differentially rotating problem.

We consider an infinite medium which is uniform in the z direction but possibly non-uniform in R ($R^2=x^2+y^2$). We suppose that it rotates at equilibrium with a velocity law $\mathbf{u}_0=+u_0(R)\hat{\boldsymbol{\phi}}$*. We shall consider the behaviour

* The positive sign makes $-u_0(R)$ the circular velocity in the conventional 21 cm picture of the galaxy; $\hat{\boldsymbol{\phi}}$ is the unit vector in the anticlockwise direction in which ϕ increases in R, ϕ, z coordinates.

of small perturbations to such a system when both the gravitational and the pressure perturbations are neglected. In such an approximation (which would be correct for small wave-length disturbances in a cold gas) the perturbations move in nearly circular orbits under the influence of the unperturbed gravity field. We shall constrain the perturbations to be independent of z. Since there are no perturbed forces acting, the particles of the fluid follow Lindblad's elliptic–epicyclic orbits. There is however an essential difference from the stellar dynamical case in that our gas has only one velocity at each point at any one time. Our initial conditions for gas perturbations are thus more organized, and this organization leads to the density becoming large in places purely as a result of propagation along Lindblad orbits.

We write ψ for the gravitational potential, ρ for the density, $\mathbf{u}$ for the perturbation in velocity and we let a suffix 0 or 1 represent an unperturbed quantity or a perturbation, respectively. We assume the unperturbed fluid to be barytropic with the equation of state

$$p_0 = \kappa \rho_0^{\gamma}.$$

The unperturbed equation of motion is

$$(\mathbf{u}_0 . \nabla)\mathbf{u}_0 = \nabla\psi_0 - \frac{1}{\rho_0}\nabla p_0 = \nabla\chi_0, \tag{1}$$

where

$$\chi_0 = \psi_0 - \frac{\kappa\gamma}{\gamma - 1}\rho_0^{\gamma-1}. \tag{2}$$

The complete equation of motion is

$$\frac{\partial \mathbf{u}}{\partial t} + [(\mathbf{u}+\mathbf{u}_0).\nabla](\mathbf{u}+\mathbf{u}_0) = \nabla\chi_0, \tag{3}$$

where we have neglected the perturbed gravity and pressure. Subtracting (1) from (3) and linearizing in $\mathbf{u}$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}_0 . \nabla)\mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u}_0 = 0. \tag{4}$$

Writing out equation (4) in cylindrical polar coordinates (R, ϕ, z) we obtain

$$\frac{\partial u_R}{\partial t} + \Omega \frac{\partial u_R}{\partial \phi} - 2\Omega u_\phi = 0 \tag{5}$$

and

$$\frac{\partial u_\phi}{\partial t} + \Omega \frac{\partial u_\phi}{\partial \phi} + 2Bu_R = 0, \tag{6}$$

where

$$\Omega(R) = + \frac{u_0}{R} \tag{7}$$

and

$$B(R) = + \tfrac{1}{2}\left(\frac{u_0}{R} + \frac{du_0}{dR}\right). \tag{8}$$

Note that $\Omega(R) = -\Omega_g(R)$ where Ω_g is the positive angular velocity of the galaxy in the conventional picture. B is the usual Oort "constant" since $-u_0$ is the circular velocity.

We now transform to co-moving variables by writing

$$\left.\begin{aligned} \phi' &= \phi - \Omega(R)t, \\ t' &= t, \\ R' &= R. \end{aligned}\right\} \tag{9}$$

Then

$$\left.\begin{aligned} \frac{\partial}{\partial t} &= \frac{\partial}{\partial t'} - \Omega \frac{\partial}{\partial \phi'}, \\ \frac{\partial}{\partial \phi} &= \frac{\partial}{\partial \phi'}, \\ \frac{\partial}{\partial R} &= \frac{\partial}{\partial R'} - t \frac{\partial \Omega}{\partial R} \frac{\partial}{\partial \phi'}, \end{aligned}\right\} \tag{10}$$

so

$$\frac{\partial}{\partial t} + \Omega \frac{\partial}{\partial \phi} = \frac{\partial}{\partial t'}. \tag{11}$$

Hence from (5) and (6)

$$\frac{\partial u_R}{\partial t'} = 2\Omega u_\phi, \tag{12}$$

and

$$\frac{\partial u_\phi}{\partial t'} = -2B u_R, \tag{13}$$

therefore

$$\frac{\partial^2 u_R}{\partial t'^2} = -4B\Omega u_R. \tag{14}$$

We define

$$n^2 = 4B\Omega, \tag{15}$$

so the solution of equation (14) is

$$u_R = u \cos(nt' + \alpha), \tag{16}$$

where u and α are arbitrary* functions of R', ϕ' and n is of course a function of R. From equation (12) we deduce

$$u_\phi = -\frac{nu}{2\Omega} \sin(nt' + \alpha). \tag{17}$$

Equations (16) and (17) determine the perturbed velocity which may be used in the perturbed continuity equation to find the density. The perturbed continuity equation reads

$$\frac{\partial \rho_1}{\partial t} + \Omega \frac{\partial \rho_1}{\partial \phi} + \operatorname{div}(\rho_0 \mathbf{u}) = 0, \tag{18}$$

* They must of course be periodic in ϕ'.

which may be written using (10) and (11)

$$\frac{\partial \rho_1}{\partial t'} + \frac{1}{R'}\left(\frac{\partial}{\partial R'} - t'\frac{\partial \Omega}{\partial R}\frac{\partial}{\partial \phi'}\right)(R'\rho_0 u_R) + \frac{\partial}{\partial \phi'}(\rho_0 u_\phi) = 0. \tag{19}$$

Thus

$$\frac{\partial \rho_1}{\partial t'} = (\text{terms periodic in } t')$$
$$- t'\left\{\rho_0 u \frac{\partial n}{\partial R'}\sin(nt'+\alpha) - \rho_0 u \frac{\partial \Omega}{\partial R}\frac{\partial \alpha}{\partial \phi'}\sin(nt'+\alpha)\right.$$
$$\left. + \rho_0 \frac{\partial \Omega}{\partial R}\frac{\partial u}{\partial \phi'}\cos(nt'+\alpha)\right\}. \tag{20}$$

For the secular terms to vanish both u (the initial disturbance amplitude) must be independent of ϕ' and

$$\frac{\partial n}{\partial R} = \frac{\partial \Omega}{\partial R}\frac{\partial \alpha}{\partial \phi'}. \tag{21}$$

(21) is impossible except when both $dn/d\Omega = m$ (an integer) and the disturbance is such that $\alpha = m\phi'$. $dn/d\Omega$ cannot be constant. Hence there are always secular terms in $\partial\rho_1/\partial t'$. Thus ρ_1 will be of the form

$$\rho_1 = \begin{cases} \rho_1(R', \phi') + (\text{terms of period } 2\pi/n \text{ in } t') \\ + t' (\text{terms of period } 2\pi/n \text{ in } t'). \end{cases} \tag{22}$$

If it is remembered that ϕ' is constant for a point moving with the unperturbed fluid motion then it is clear that at such a point the amplitude of the oscillating density is growing with time. The physical reasons for this effect are interesting.

Firstly the dn/dR term arises because the period of Lindblad oscillations depends on radius. Note that this term persists even for axially symmetrical disturbances. The growth corresponds to the initially organized oscillations of the gas becoming progressively more and more out of step at each radius. We show later that the introduction of pressure eliminates this trouble for such disturbances.

Secondly note that the remaining secular terms only appear for non-axially-symmetrical disturbances. The phase of the oscillations in the neighbourhood of a cylinder $R = R_0$ varies with ϕ. The shear therefore brings parts of the disturbance with different phase and slightly different radius to the same azimuth. Thus points on the same azimuth but with slightly different radius have velocities that are progressively more and more out of phase. This leads to the amplitude of the density growing linearly with time. This effect is modified by pressure but the density amplitude still grows (like $t^{1/2}$) (see Section 6).

We do not believe that this interesting behaviour is directly related to spiral arm formation. Although the density becomes very large the wave-length of the disturbances becomes shorter due to the differential rotation. We shall see presently that conditions become less and less favourable to Jeans' instability as the modes become more and more sheared. All non-axially-symmetrical modes are subject to this type of asymptotic behaviour in the linearized theory. Some pioneering investigations of the non-linear theory which we hope to follow up elsewhere strongly suggest that shock waves are formed which dissipate the energy of the disturbance.

These effects are not uninteresting in their own right. They are a mechanism by which perturbations can feed on the energy of the differential rotation of the galaxy and finally convert that energy into shock waves in the interstellar gas. These in turn may be an important source of both thermal and turbulent energy for the gas.

4. *Wave-lengths small compared with the size of the galaxy*

In paper I we found that the critically stable modes of the uniformly rotating sheet had wave-lengths of some 2π times the thickness of the sheet. This scale is smaller than the radius of the galaxy. The large wave-length problem is much more difficult mathematically for a differentially rotating galaxy, so it is expedient to exploit the small scale expected of the critical modes. The present section is devoted to obtaining the non-linear equations which govern the behaviour of disturbances whose wave-lengths are small compared with the distance to the centre of a differentially rotating* galaxy.

Equations.—The equation of motion in rotating axes is

$$\frac{\partial \mathbf{U}}{\partial t} + (\mathbf{U}\,.\,\nabla)\mathbf{U} + 2\boldsymbol{\Omega}_a \times \mathbf{U} - \Omega_a^2 \mathbf{R} = \nabla\psi - \frac{1}{\rho}\nabla p, \tag{23}$$

where $\mathbf{U}$ is the fluid velocity, w.r.t. the rotating axes and $\boldsymbol{\Omega}_a = (0, 0, \Omega_a)$ is the angular velocity of the axes. For our galaxy Ω_a is negative in the right-handed coordinates that we superpose on the 21 cm galactic map. $\mathbf{R} = (x, y, 0)$ is the radius vector from the rotation axis, ψ is the gravitational potential, ρ is the density of the fluid, p is the pressure of the fluid. For polytropic fluid the equation of state reads

$$p = \kappa\rho^{\gamma}, \tag{24}$$

so

$$\left.\begin{aligned} &\frac{1}{\rho}\nabla p = \nabla\left(\frac{\gamma\kappa}{\gamma-1}\rho^{\gamma-1}\right) && \gamma \neq 1, \\ \text{or} \qquad &\\ &\frac{1}{\rho}\nabla p = \nabla(\kappa \log\rho) && \gamma = 1. \end{aligned}\right\} \tag{25}$$

We write

$$\left.\begin{aligned} &\chi = \psi - \frac{\gamma\kappa}{\gamma-1}\rho^{\gamma-1} && \gamma \neq 1, \\ \text{or} \qquad &\\ &\chi = \psi - \kappa\log\rho && \gamma = 1. \end{aligned}\right\} \tag{26}$$

Further we put

$$\mathbf{U} = \mathbf{u}_0 + \mathbf{u}, \tag{27}$$

where

$$\mathbf{u}_0 = \mathbf{u}_0(R)\hat{\boldsymbol{\phi}}, \tag{28}$$

is the velocity in the equilibrium state, w.r.t. the rotating axes and $\mathbf{u}$ is the

* Of course we may then specialize to the uniformly rotating case to obtain the equations solved in paper I.

perturbation. Using the identity $\hat{\mathbf{R}} = -\hat{\mathbf{z}} \times \hat{\boldsymbol{\phi}}$ (hats denote unit vectors) the equation of motion becomes

$$\left.\begin{aligned} &\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}_0 . \nabla)\mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u}_0 + 2\boldsymbol{\Omega}_a \times \mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u} \\ &\quad = \left(\frac{\mathbf{u}_0^{\,2}}{R} + 2\Omega_a \mathbf{u}_0 + \Omega_a^2 R\right)\hat{\mathbf{R}} + \nabla\chi. \end{aligned}\right\} \tag{29}$$

For the equilibrium state itself $\mathbf{u}$ is zero so

$$\left(\frac{u_0^{\,2}}{R} + 2\Omega_a \mathbf{u}_0 + \Omega_a^2 R\right)\hat{\mathbf{R}} + \nabla\chi_0 = 0. \tag{30}$$

Subtracting, we obtain

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}_0 . \nabla)\mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u}_0 + 2\boldsymbol{\Omega}_a \times \mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u} = \nabla\chi_1. \tag{31}$$

So far the development is purely formal and the angular velocity of the axes is arbitrary. We now concentrate our attention on the stability of a small portion of the galaxy near some point at (R_0, ϕ_0, z) initially. We choose the angular velocity of the axes to reduce the unperturbed motion of (R_0, ϕ_0, z) to rest, $u_0(R_0) = 0$. We set up cartesian axes (x, y, z) with origin at (R_0, ϕ_0) and with the x axis pointing outwards from the galactic centre. In the region about (R_0, ϕ_0) we may now expand the components of $\mathbf{u}_0$ in a Taylor series

$$(u_{0x}, u_{0y}, u_{0z}) = R_0 u_0' \left[\left(0, \frac{x}{R_0}, 0\right) + O\left(\frac{x^2 + y^2}{R_0^{\,2}}\right)\right]. \tag{32}$$

Here u_0' is

$$\left.\frac{du_0}{dR}\right|_{R=R_0}.$$

Oort's constant A evaluated at $R = R_0$ may be calculated by remembering that the circular velocity is $-(u_0 + \Omega_a R)$ and that $u_0(R_0)$ is zero. It is

$$A = \tfrac{1}{2} u_0'. \tag{33}$$

Substituting (28) and (29) into equation (27) we obtain

$$\frac{\partial \mathbf{u}}{\partial t} + 2Ax\frac{\partial \mathbf{u}}{\partial y} + u_x 2A\hat{\mathbf{y}} + 2\boldsymbol{\Omega}_a \times \mathbf{u} + (\mathbf{u} . \nabla)\mathbf{u} = \nabla\chi_1, \tag{34}$$

where terms of order x/R_0 have been neglected but those of order x/λ, where λ is the typical scale of the perturbation, have been retained. In a similar manner we may obtain the non-linear continuity equation

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \rho\mathbf{U} = 0; \tag{35}$$

while for the equilibrium

$$\operatorname{div}(\rho_0 \mathbf{u}_0) = 0. \tag{36}$$

Subtracting and using expressions (27), (28) for **U**

$$\frac{\partial \rho_1}{\partial t} + 2Ax \frac{\partial \rho_1}{\partial y} + \operatorname{div}(\rho \mathbf{u}) = 0. \tag{37}$$

Equations (34), (37) together with Poisson's equation and the equation of state (26) are our full set of equations. To within our approximation ρ_0 is independent of R because

$$\rho_0(R, z) = \rho_0(R_0, z) + x\rho_0'(R_0, z) + O(x^2/R_0^2)$$

and presuming that $R_0/\rho_0' \sim \rho_0$ the second term is of order x/R_0 times the first. This is consistent with equation (30) to the order to which we are working. We show this as follows:

$$\left(\frac{u_0^2}{R} + 2\Omega_a u_0 + \Omega_a^2 R\right)_{R_0} = 0 \text{ by our choice of } \Omega_a.$$

Hence $\nabla\chi_0$ vanishes at $R = R_0$ and it is radial elsewhere and proportional to x (to first order). Thus $\chi_0 = \chi_0(0) + O(x^2)$. To the order to which we are working χ_0, ρ_0 and ψ_0 are functions of z alone.

5. *Models*

In Section 4 we made the crudest form of small wave-length approximation to the dynamics of a portion of a differentially rotating galaxy. We wish to note here that those approximate equations are actually exact if they are thought of as describing a certain physical model. The unperturbed state of the model we describe in axes rotating with angular velocity Ω_a. The fluid is homogeneous in x and y but is pressure supported against its own gravity in the z direction. With respect to the rotating axes its unperturbed velocity is parallel to the y axis and of magnitude $2Ax$. In order that this should be an equilibrium velocity in the presence of Coriolis force, there is an imposed tidal field which exactly cancels the Coriolis force. The equations derived in the last section are the exact equations that describe the evolution of arbitrarily large perturbations superimposed on this model. We shall find that we need the full non-linear equations to discover the real behaviour of this system. We shall treat the polytropic sheets with $\gamma = 1$ and $\gamma = 2$ but, following our policy of discussing the simplest problems first, even if they lack some reality, we shall also consider the generalized Chandrasekhar problem. That is, the stability of a uniform medium, of infinite extent vertically, which is uniformly sheared in rotating axes.

The remainder of this section is devoted to developments that are common to these three problems.

We transform to sheared axes, co-moving with the unperturbed flow as follows:

$$\begin{aligned} x' &= x, \\ y' &= y - 2Axt, \\ z' &= z, \\ t' &= t. \end{aligned} \tag{38}$$

Then,

$$\begin{aligned}\frac{\partial}{\partial x} &= \frac{\partial}{\partial x'} - 2At'\frac{\partial}{\partial y'},\\ \frac{\partial}{\partial t} &= \frac{\partial}{\partial t'} - 2Ax'\frac{\partial}{\partial y'},\\ \frac{\partial}{\partial y} &= \frac{\partial}{\partial y'},\\ \frac{\partial}{\partial z} &= \frac{\partial}{\partial z'}.\end{aligned} \tag{39}$$

The time derivative following the unperturbed motion is

$$\frac{\partial}{\partial t'} = \frac{\partial}{\partial t} + 2Ax\frac{\partial}{\partial y}. \tag{40}$$

Under the transformation (38) the equation of motion (34) becomes:

$$\frac{\partial u_x}{\partial t'} - 2\Omega_a u_y + (\mathbf{u}\cdot\nabla)u_x = \left(\frac{\partial}{\partial x'} - 2At'\frac{\partial}{\partial y'}\right)\chi_1, \tag{41}$$

$$\frac{\partial u_y}{\partial t'} + 2Bu_x + (\mathbf{u}\cdot\nabla)u_y = \frac{\partial \chi_1}{\partial y'}, \tag{42}$$

$$\frac{\partial u_z}{\partial t'} + (\mathbf{u}\cdot\nabla)u_z = \frac{\partial \chi_1}{\partial z'}, \tag{43}$$

where $B = A + \Omega_a$, in agreement with Oort's notation, when it is remembered that Ω_a is negative. B is Oort's constant evaluated at R_0.

Similarly the transformed continuity equation (37) reads

$$\frac{\partial \rho_1}{\partial t'} + \left(\frac{\partial}{\partial x'} - 2At'\frac{\partial}{\partial y'}\right)(\rho u_x) + \frac{\partial}{\partial y'}(\rho u_y) + \frac{\partial}{\partial z'}(\rho u_z) = 0 \tag{44}$$

and the perturbed Poisson equation is

$$\left[\left(\frac{\partial}{\partial x'} - 2At'\frac{\partial}{\partial y'}\right)^2 + \frac{\partial^2}{\partial y'^2} + \frac{\partial^2}{\partial z'^2}\right]\psi_1 = -4\pi G\rho_1. \tag{45}$$

Linearization.—Equations (41) to (43) may be linearized by dropping the $(\mathbf{u}\cdot\nabla)\mathbf{u}$ term, while equation (44) is linearized by replacing ρ by ρ_0. The linearized form of equation (26) is

$$\chi_1 = \psi_1 - c^2\frac{\rho_1}{\rho_0}, \tag{46}$$

where c^2 is the *variable* $\kappa\gamma\rho_0^{\gamma-1}$ which happens to be constant when $\gamma = 1$.

These linearized equations have coefficients that depend on t' and on z' (through ρ_0) but which are independent of x' and y'. This is in marked contrast to the untransformed equations whose coefficients depend on x and z and are independent of t and y. It will appear later that our transformation to dashed coordinates (which throws the inhomogeneity from x into t) is essential for a

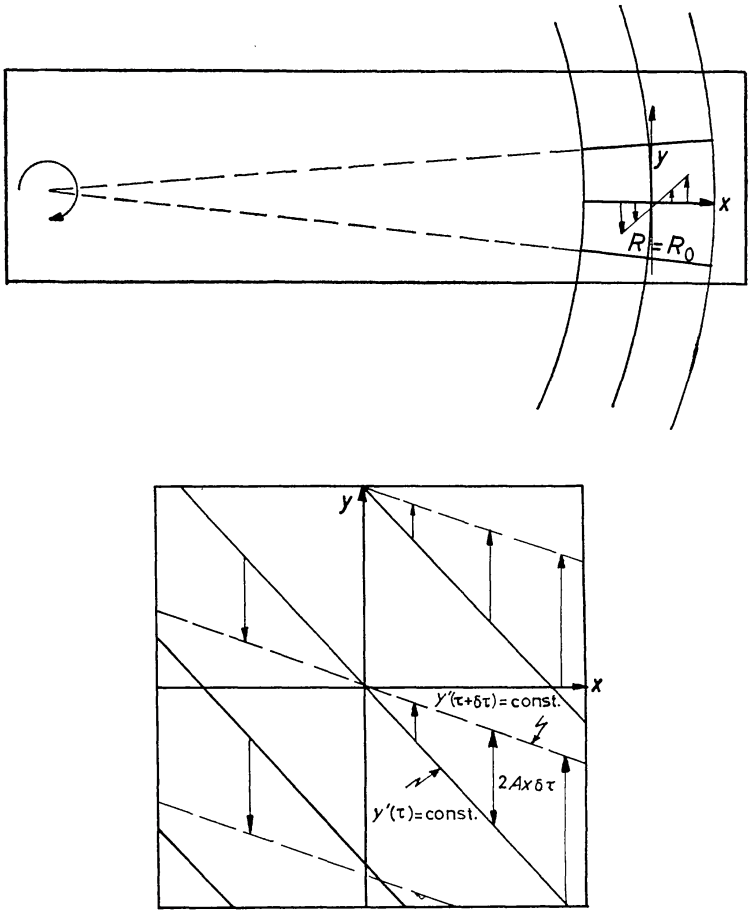

FIG. 1.—*Uniform shear in rotating axes and sheared coordinates.*

correct solution of the problem of real interest. We Fourier analyse in x' and y' by giving the independent variables an $\exp i(k_x x' + k_y y')$ dependence. The linearized equations (41) to (45) then read, dropping the suffix 'a' on Ω:

$$\frac{\partial u_x}{\partial t'} - 2\Omega u_y = i(k_x - 2At'k_y)\chi_1, \tag{47}$$

$$\frac{\partial u_y}{\partial t'} + 2Bu_x = ik_y\chi_1, \tag{48}$$

$$\frac{\partial u_z}{\partial t'} = \frac{\partial \chi_1}{\partial z'}, \tag{49}$$

$$\frac{\partial \rho_1}{\partial t'} + i(k_x - 2At'k_y)\rho_0 u_x + ik_y\rho_0 u_y + \frac{\partial}{\partial z'}(\rho_0 u_z) = 0 \tag{50}$$

and

$$\left[(k_x - 2At'k_y)^2 + k_y^2 - \frac{\partial}{\partial z'^2}\right]\psi_1 = 4\pi G\rho_1. \tag{51}$$

Excepting for the present the special modes with k_y zero we introduce the new "time" variable τ by

$$\tau = 2At' - \frac{k_x}{k_y}. \tag{52}$$

Notice that the time variable now has a different zero for modes with different values of k_x/k_y. There is important physical significance in this zero because one finds by transforming back to x, y that

$$\begin{aligned}\exp\left[i(k_x x' + k_y y')\right] &= \exp\left\{i[k_x x + k_y(y - 2Axt)]\right\} \\ &= \exp\left[ik_y(-\tau x + y)\right].\end{aligned} \tag{53}$$

This shows that our Fourier modes in x' and y' are modes that are sheared with the unperturbed motion of the fluid and that $\tau = 0$ when the (x, y) wave vector of the disturbance is purely in the y direction. What this means physically is that for each mode the time $\tau = 0$ occurs when the contours of perturbed density point radially outwards from the galactic centre. We shall show shortly that density perturbations grow fastest after they pass through this configuration.

Writing equations (47) to (49) in terms of τ we obtain

$$\dot{u}_x - \frac{\Omega}{A} u_y = -ik\tau\chi_1, \tag{54}$$

$$\dot{u}_y + \frac{B}{A} u_x = ik\chi_1, \qquad \text{and} \tag{55}$$

$$\dot{u}_z = \frac{1}{2A} \frac{\partial \chi_1}{\partial z}, \tag{56}$$

where dots denote differentiation, w.r.t. τ and

$$k = \frac{k_y}{2A}. \tag{57}$$

Similarly, equations (50) and (51) now read

$$\dot{\rho}_1 - ik\tau\rho_0 u_x + ik\rho_0 u_y + \frac{1}{2A} \frac{\partial}{\partial z'} (\rho_0 u_z) = 0 \tag{58}$$

and

$$\left[-k_y^2(1 + \tau^2) + \frac{\partial^2}{\partial z'^2}\right] \psi_1 = -4\pi G\rho_1. \tag{59}$$

It is convenient to derive here the z component of the vorticity equation. Multiplying equation (55) by $-ik\tau$ and adding it to $-ik$ times equation (54) we obtain

$$-ik \frac{\partial}{\partial \tau} (u_x + \tau u_y) + ik \frac{B}{A} (-\tau u_x + u_y) = 0, \tag{60}$$

where we have used the relation $\Omega = -\Omega_0 = B - A$. Similarly, we take the two-dimensional divergence of equations (54) and (55) by multiplying (54) by $-ik\tau$ and (55) by $+ik$ and adding:

$$ik \frac{\partial}{\partial \tau} (u_y - \tau u_x) + ik \frac{\Omega}{A} (\tau u_y + u_x) + 2iku_x = -k^2(1 + \tau^2)\chi_1. \tag{61}$$

We cannot proceed further without knowledge of the way in which our variables depend on z. To this end we turn to more specific models.

6. *The sheared infinite medium*

In this section we attack the generalization of Chandrasekhar's problem which includes not only rotation but also uniform shear. The medium is assumed homogeneous and of infinite extent in all directions.

As we have explained in paper I the greatest interest attaches itself to the stability criterion for Chandrasekhar's singular modes, which have wave vectors perpendicular to the axis of rotation. For such modes there are no variations in the z direction and we may put

$$\frac{\partial}{\partial z'} = 0 \qquad \text{and} \qquad u_z = 0. \tag{62}$$

With these restrictions the continuity equation (58) reads

$$ik(-\tau u_x + u_y) = -\frac{\dot{\rho}_1}{\rho_0}. \tag{63}$$

Substituting this value of $-\tau u_x + u_y$ into the vorticity equation (60) we obtain

$$\frac{d}{d\tau}\left[-ik(u_x + \tau u_y) - \frac{B}{A}\frac{\rho_1}{\rho_0}\right] = 0 \tag{64}$$

and therefore

$$-ik(u_x + \tau u_y) - \frac{B}{A}\frac{\rho_1}{\rho_0} = C_1 = \text{const.} \tag{65}$$

The constant C_1 is necessarily small and represents the perturbation of the vorticity per unit mass (multiplied by $\rho_0/2A$). If the perturbations are caused by a time dependent gravity field or by an external pressure, applied to the unperturbed state, then Kelvin's circulation theorem holds exactly and $C_1 = 0$ since it is zero initially in the unperturbed state. If however, the system was never quite in its unperturbed state or if it has had worse perturbations applied to it (which generated vorticity) then C_1 will not be zero (except by conspiracy).

From equations (63) and (65)

$$ik(-\tau u_x + u_y) = -\frac{\dot{\rho}_1}{\rho_0} \equiv -\dot{\theta}_1, \tag{66}$$

$$ik(u_x + \tau u_y) = -C_1 - \frac{B}{A}\frac{\rho_1}{\rho_0} = -C_1 - \frac{B}{A}\theta_1, \tag{67}$$

so

$$ik(1 + \tau^2)u_x = -\tau\dot{\theta}_1 - C_1 - \frac{B}{A}\theta_1. \tag{68}$$

Equations (66), (67), (68) may now be used to eliminate the velocities from equation (61) to obtain

$$-\ddot{\theta}_1 + \frac{\Omega}{A}\left(-C_1 - \frac{B}{A}\theta_1\right) + \frac{2}{1+\tau^2}\left(\tau\dot{\theta}_1 - C_1 - \frac{B}{A}\theta_1\right) = -k^2(1+\tau^2)\chi_1, \tag{69}$$

which simplifies to the form

$$\frac{d}{d\tau}\left(\frac{\dot{\theta}_1}{1+\tau^2}\right) + \left\{\frac{2\frac{B}{A}}{(1+\tau^2)^2} + \frac{\frac{B\Omega}{A^2}}{1+\tau^2}\right\}\left(\theta_1 + \frac{AC_1}{B}\right) = k^2\chi_1. \tag{70}$$

From equations (59) and (46)

$$\chi_1 = \left(\frac{4\pi G\rho_0}{k_y^2(1+\tau^2)} - c^2\right)\theta_1, \tag{71}$$

where

$$c^2 = \kappa\gamma\rho_0{}^{\gamma-1}.$$

For simplicity, taking non-vortical perturbations so that $C_1 = 0$ we have

$$\frac{d}{d\tau}\left(\frac{\dot{\theta}_1}{1+\tau^2}\right) + \left[\frac{2\dfrac{B}{A}}{(1+\tau^2)^2} + \frac{\dfrac{B\Omega}{A^2} - \dfrac{\pi G\rho_0}{A^2}}{1+\tau^2} + \frac{k_y^2c^2}{4A^2}\right]\theta_1 = 0. \tag{72}$$

The equation for the axially-symmetrical "ring" modes with $k_y = 0$ may be obtained from this by writing out τ in full and performing a limiting procedure. With an arbitrary zero for τ one obtains

$$\frac{d}{d\tau}(\dot{\theta}_1) + \left(\frac{B\Omega}{A^2} - \frac{\pi G\rho_0}{A^2} + \frac{k_x^2c^2}{4A^2}\right)\theta_1 = 0. \tag{73}$$

This equation is simple harmonic so the "ring" waves are stable or unstable according as

$$4B\Omega - 4\pi G\rho_0 + k_x^2c^2 > 0 \quad \text{or} \quad \leqslant 0 \tag{74}$$

respectively.

Apart from the factors $1+\tau^2$ which arise from the varying wave number of the non-axially-symmetrical sheared modes, the main difference between equations (72) and (73) is the extra term $(2B/A)/1+\tau^2$ in the former. Since B is negative this term tries to make the solution of equation (72) grow exponentially. This growth will occur even at densities too low to make the "ring" modes unstable (74). If the density is secularly increased the first modes to show considerable growth are sheared. As explained more fully in what follows the period of greatest growth occurs after the time $\tau = 0$ (when the lines of equal density point radially). When growth has occurred those lines will be trailing (to form a spiral pattern if we are prepared to put several of our small-scale analyses side by side).

Equation (72) is the prototype of equations with similar behaviour that we shall be led to discuss for sheets of finite thickness. We shall here discuss the form of its solutions in some detail because we believe them to be closely related to the mechanism of spiral arm formation.

Comparison of the "ring" modes with the others is best made through equations (72) and (73), but a discussion of the form of the solution of equation (72) is best effected by transforming to the new variable

$$\Phi = \theta_1(1+\tau^2)^{-1/2}. \tag{75}$$

Equation (72) then takes the form

$$\ddot{\Phi} + \left[\frac{3}{(1+\tau^2)^2} + \frac{2\dfrac{B}{A} - 2}{(1+\tau^2)} + \frac{B(B-A) - \pi G\rho_0}{A^2} + \frac{k_y^2c^2}{4A^2}(1+\tau^2)\right]\Phi = 0. \tag{76}$$

The factors $(1+\tau^2)$ are only as small as 2 for the period $-1 \leqslant \tau \leqslant +1$. Let us consider the situation when the "ring" modes are just stable so that $[B(B-A)-\pi G\rho_0]/A^2$ is very small. Then for $|\tau|$ large the coefficient of Φ is dominated by the large pressure term, and the solution of the equation oscillates. However, when $|\tau|$ decreases to near zero the $3/(1+\tau^2)^2$ and $-2(-B/A+1)/(1+\tau^2)$ terms become of greater importance. Note that B is negative so that provided $-B/A > \frac{1}{2}$ the second of these terms is always the larger. If we now take long waves so that $k_y^2c^2/4A^2 \ll 1$ then the $-2(-B/A+1)/1+\tau^2$ term is dominant for small τ, so the coefficient of Φ will be negative and the solutions will therefore grow quasi-exponentially. However, the period of growth is limited since $|\tau|$ will again become large which will again change the character of the solutions to oscillation. To sum up, the solution will start by oscillating when the waves are pointing forward. As the differential rotation sweeps them round through the straight out position they go through the period of greatest acceleration (in the sense that $\ddot{\Phi}/\Phi$ is largest). The greatest actual rate of growth of Φ is achieved later when Φ has itself grown and the waves trail. However, considerable trailing is associated with renewed dominance of the pressure term and renewed oscillation at the greatly enhanced amplitude. Since the equation is derived from a linearized analysis it will not remain true if θ_1 ever achieves values of order unity. Thus, if in the period of growth the perturbations in the density become comparable with the unperturbed density, the predicted return to oscillatory character need not occur. For the stratified isothermal sheet discussed later we show that it is energetically possible for the non-linear modes to continue to condense rather than to revert to oscillatory behaviour.

7. *The vertical equilibrium approximation*

In paper I we found that in all modes close to marginal stability the vertical accelerations were small compared with the perturbations of the gravity field:

$$\frac{\partial u_z}{\partial t} \ll g_1 = \frac{\partial \psi_1}{\partial z}. \tag{77}$$

In that problem $\partial/\partial t$ is the time derivative (in the uniformly rotating axes) that follows the unperturbed motion. The analogue in our problem is $2A(\partial/\partial\tau)$. If we write out equation (56), using equation (46), it reads

$$2A\frac{\partial u_z}{\partial \tau} = \frac{\partial \psi_1}{\partial z'} - \frac{\partial}{\partial z'}(\kappa\gamma\rho_0{}^{\gamma-2}\rho_1) = \frac{\partial \chi_1}{\partial z}. \tag{78}$$

Assuming analogously that

$$2A\frac{\partial u_z}{\partial \tau} \ll \frac{\partial \chi_1}{\partial z'}, \tag{79}$$

we have

$$\frac{\partial \chi_1}{\partial z} \simeq 0. \tag{80}$$

Hence

$$\chi_1 \simeq \Lambda_1(\tau). \tag{81}$$

Physically the above approximation neglects any inertia of the fluid to vertical movement. The fluid is thus infinitely responsive to vertical forces.

With χ_1 independent of z it follows from equations (54) and (55) that both u_x and u_y can be independent of z*. We therefore integrate equation (58) from $z=-a$ to $z=+a$ and apply the boundary conditions as in paper I to obtain

$$\dot{\Sigma}_1+\Sigma_0 ik(-\tau u_x+u_y)=0, \tag{82}$$

where

$$\Sigma_0+\Sigma_1=\int_{edge}^{edge}\rho\,dz$$

(the edges being the perturbed edges). Equation (82) is the surface density continuity equation; when written in the form

$$ik(-\tau u_x+u_y)=-\frac{\dot{\Sigma}_1}{\Sigma_0} \tag{83}$$

it is closely analogous to equation (63) for the infinite medium, and the same analysis yields from equation (60) the equation similar to (67)

$$-ik(u_x+\tau u_y)-\frac{B}{A}\frac{\Sigma_1}{\Sigma_0}=C_1^*=\text{const.} \tag{84}$$

Similarly, following the analysis of equations (68)–(71), we arrive at the equation analogous to (72) which reads

$$\frac{d}{d\tau}\left(\frac{\dot{\theta}_1^*}{1+\tau^2}\right)+\left\{\frac{2\dfrac{B}{A}}{(1+\tau^2)^2}+\frac{\dfrac{B\Omega}{A^2}}{1+\tau^2}\right\}\left(\theta_1^*+\frac{AC_1^*}{B}\right)=k^2\chi_1. \tag{85}$$

The only difference is that θ_1^* now stands for Σ_1/Σ_0 rather than ρ_1/ρ_0, that is, θ_1^* is the fractional increase in surface density.

To proceed further we must solve Poisson's equation (59) making use of the boundary conditions and the fact that $\chi_1=\Lambda_1(\tau)$ is independent of z (equation (81)). We have already done this calculation for the cases $\gamma=2$ and $\gamma=1$ in paper I but, comparing equation (59) and paper I equation (38), we see that we now have $k_y^2(1+\tau^2)$ written where k^2 alone stood before. This is hardly surprising because we see from equation (53) that our ordinary space wave number at any time τ is indeed $k_y(1+\tau^2)^{1/2}$. Thus the $|k|$ of the boundary conditions of paper I should also be replaced by this quantity. The solutions for $\gamma=2$ and $\gamma=1$ follow those of Sections 8 and 9 of paper I (except that we now omit the $O(\omega)$ terms owing to the vertical equilibrium approximation). We thus deduce paper I equation (130) which may be written in the form

$$\frac{\kappa}{a}\Sigma_1=K^2F(K)\Lambda_1, \tag{86}$$

where $F(K)$ is defined by the identity (135) of paper I (written again below) and

$$K^2=k_y^2a^2(1+\tau^2) \tag{87}$$

* We assume this to be the case since it is true of the critical modes in the differentially rotating case.

in the present context:

$$F(K)\equiv\frac{(Z^2K^{-1}-f)L^{-2}(T-1)+(Z^{-2}f+1)(K^{-1}+T)}{L^2T+f-K}, \tag{88}$$

where

$$L^2=Z^2-K^2, \tag{89}$$

$$Z^2=\frac{2\pi G}{\kappa}a^2=\left(\frac{\pi^2}{4}\text{ for zero halo pressure}\right), \tag{90}$$

$$f=\frac{2a\rho_0(a)}{\Sigma_0}, \tag{91}$$

and

$$T=\frac{1}{L}\tan L. \tag{92}$$

From equations (57), (87), (86)

$$k^2\Lambda_1=\frac{k_y^2}{4A^2}\Lambda_1=\frac{K^2\Lambda_1}{4A^2a^2(1+\tau^2)}=\frac{\kappa\Sigma_0}{4a^2A^2}\frac{\theta_1^*}{F(K)(1+\tau^2)}, \tag{93}$$

so for $\gamma=2$ equation (85) reads—dropping the vorticity term C_1^*—

$$\frac{d}{d\tau}\left(\frac{\dot\theta_1^*}{1+\tau^2}\right)+\left[\frac{2\dfrac{B}{A}}{(1+\tau^2)^2}+\frac{\dfrac{B\Omega}{A^2}-\dfrac{P}{F(K)}}{(1+\tau^2)}\right]\theta_1^*=0, \tag{94}$$

where

$$P=\frac{\pi G\Sigma_0}{A^2Z^2 2a}=\frac{\kappa\Sigma_0}{4a^3A^2}. \tag{95}$$

In discussing equation (94) it must be remembered that K itself contains a factor $(1+\tau^2)^{1/2}$ and so $1/F(K)$ changes with time. The equation for the "ring" vibrations of the $\gamma=2$ sheet is similar, but simpler, like the situation for the infinite medium. It reads

$$\frac{d}{d\tau}(\dot\theta_1^*)+\left(\frac{B\Omega}{A^2}-\frac{P}{F(k_xa)}\right)\theta_1^*=0. \tag{96}$$

Similarly the equations for $\gamma=1$ isothermal sheet read (for non-radial vibrations)

$$\frac{d}{d\tau}\left(\frac{\dot\theta_1^*}{1+\tau^2}\right)+\left[\frac{2\dfrac{B}{A}}{(1+\tau^2)^2}+\frac{\dfrac{B\Omega}{A^2}-\dfrac{P'}{g(m)}}{(1+\tau^2)}\right]\theta_1^*=0 \tag{97}$$

and for radial vibrations

$$\frac{d}{d\tau}(\dot\theta_1^*)+\left[\frac{B\Omega}{A^2}-\frac{P'}{g(m_x)}\right]\theta_1^*=0, \tag{98}$$

where

$$P' = \frac{\pi G \rho_C}{A^2} = \frac{3}{2}\frac{\pi G \bar{\rho}}{A^2}, \tag{99}$$

$$\frac{1}{g(m)} = \frac{m(1-m^2)}{1+m+\frac{1}{2}m^2\Psi'\left(\frac{m}{2}+1\right)}, \quad \Psi'\left(\frac{m}{2}+1\right) = \sum_{r=1}^{\infty}\frac{1}{\left(\frac{m}{2}+r\right)^2}, \tag{100}$$

$$m = \frac{k_y}{k_0}(1+\tau^2)^{1/2}, \qquad m_x = \frac{k_x}{k_0}, \tag{101}$$

and

$$k_0^2 = \frac{2\pi G \rho_C}{c^2}. \tag{102}$$

F and g suitably normalized define functions $\mathfrak{I}$ (see paper I).

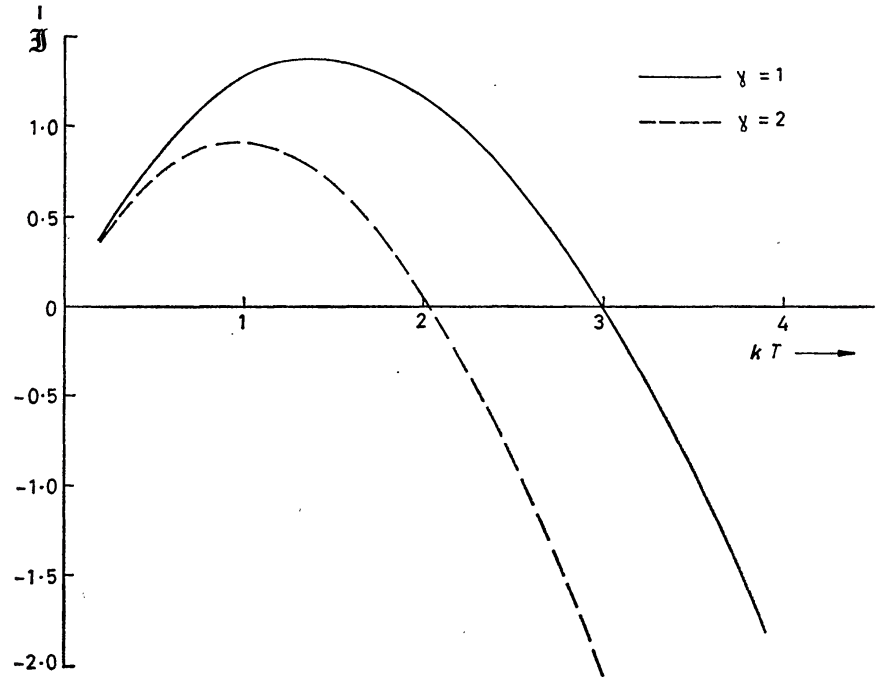

FIG. 2

The functions $1/\mathfrak{I}$ are plotted in Fig. 2. Reference to these graphs enables us to draw qualitative conclusions about the behaviour of the equations for θ_1*. We shall be interested in that particular range of parameters $(A, B, \bar{\rho})$ for which the "ring" modes (as expressed by equations (96) or (98)) are just stable. A discussion of the growth of sheared modes for these parameter values will be the topic of this section. In what follows we shall restrict our discussion of the equations for $\gamma = 2$. This makes the treatment less cumbersome and involves no loss of generality since the $\gamma = 1$ equations behave completely analogously to those for $\gamma = 2$.

We begin by observing (from equation (96)) that some "ring" modes will be unstable unless

$$B\Omega > \max(PA^2/F) = \frac{\kappa\Sigma_0}{4a^3}\max\left(\frac{1}{F(K)}\right) = \frac{\pi G\bar{\rho}}{4{\cdot}44}. \tag{103}$$

We shall denote by K_C the value of $k_x a$ at which $1/F(k_x a)$ attains its maximum value. Hence, we are interested in values of the parameters $(A, B, \bar{\rho})$ for which $B\Omega$ is infinitesimally larger than $\pi G\bar{\rho}/4{\cdot}44$. We now proceed to a discussion of equation (94) for sheared modes using the appropriate parameter values.

K, the argument of $F(K)$ in equation (94), varies with τ as $\sqrt{(1+\tau^2)}$. K starts large when τ is large and negative and decreases to its minimum value of $k_y a$ at $\tau=0$. It subsequently increases again as τ becomes large and positive. If $k_y a < K_C$ then there will be two times, $\pm\tau_C$, when $K=K_C$. At these values of τ and for a range of τ about these values, the coefficient of θ_1^* in equation (94) will be negative (to see this we must remember that both B and Ω are negative quantities). While this coefficient is negative θ_1^* will in general exhibit a rapid growth. As we are interested in the growth of θ_1^* we endeavour to choose $k_y a$ such that this growth is maximized. By a mixture of foresight and hindsight (as provided by computing solutions on EDSAC) we find that the appropriate value of $k_y a$ is about $\frac{1}{2}$ but it varies with both "a" and γ. For this best value of k_y, θ_1^* grows by a factor of $\sim 10^{2{\cdot}7\pm1}$ when we integrate equation (94) from $\tau=-10$ to $\tau=+10$. Graphs of θ_1^* versus τ are given in Figs. 3, 4 and 5.

Growth mainly occurs after $\tau=0$. Since $B\Omega/A^2 - P/F(K)$ is always positive the term that provides the negative part of the coefficient of θ_1^* is $(B/A)(1+\tau^2)^{-2}$. This is greatest in magnitude at $\tau=0$. Maximum growth occurs when $k_y a$ is chosen so that $B\Omega/A^2 - P/F(K)$ reaches its maximum near $\tau=0$ also*.

Computation procedure.—Our procedure was to choose a value of B/A (in the first instance the value $-\frac{2}{3}$ for the solar neighbourhood). For this value we set P so that equation (96) was just stable. We then used these values in equation (94) and computed from $\tau=-10$ to $\tau=+10$ with the starting values $\theta_1^*=1$, $\dot{\theta}_1^*=0$. For the linearized treatment the initial value of θ_1^* is of course unimportant and can be scaled to any required value. We made a number of runs with different values of $k_y a$ seeking that value which gave the greatest growth. We also made runs with oscillations out of phase with the above by choosing θ_1^* to be zero initially and $\dot{\theta}_1^*$ non-zero. The results may be seen from the graphs (Figs. 3, 4 and 5). These are all computed for the isothermal $\gamma=1$ sheet. Similar results were found for the $\gamma=2$ sheet.

8. *Non-linear problem*

We shall again make a vertical equilibrium approximation by neglecting the inertia of the fluid in the vertical direction. With this assumption equation (43) yields

$$\chi_1 = \Lambda_1(t', x', y'). \tag{104}$$

With χ_1 independent of z the velocities u_x and u_y may be taken independent of z from equations (41) and (42).

To proceed further it is easiest to return to unsheared axes and so to use the equation of motion in the form (23). It is also convenient to introduce a two-dimensional notation so that $\mathbf{U}=(U_x, U_y)$, etc. To emphasize this change we shall give all our vector operators a suffix 2. Thus

$$\mathrm{div}_2\,\mathbf{U} = \frac{\partial U_x}{\partial x} + \frac{\partial U_y}{\partial y} \tag{105}$$

* Actually before $\tau=0$ as may be seen by more careful reflection using the graph of $1/F$.

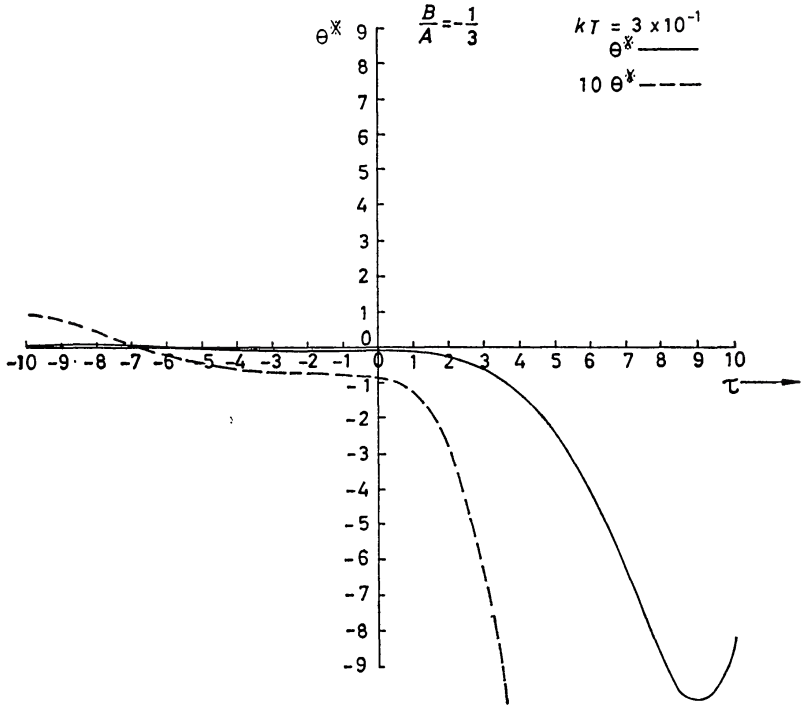

FIG. 3

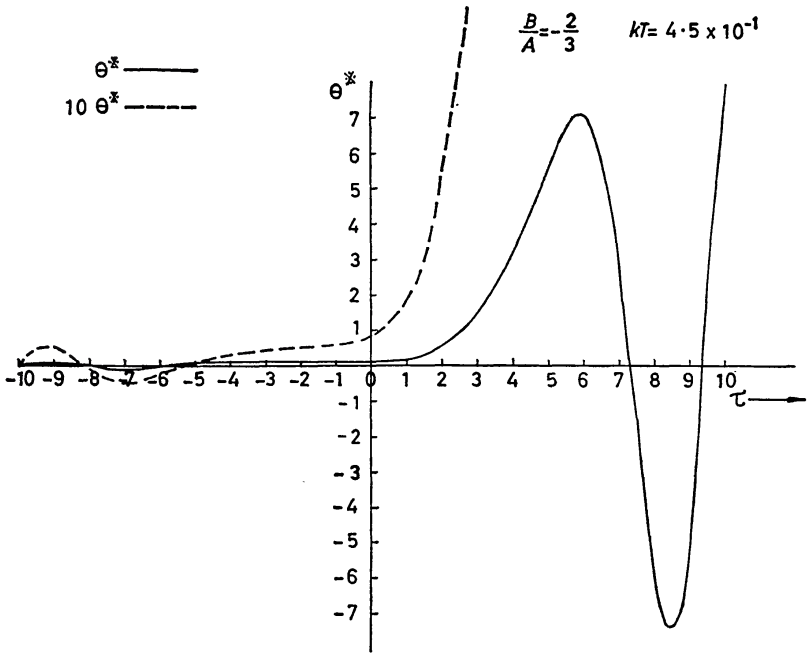

FIG. 4

and

$$\mathrm{curl}_2\,\mathbf{U}=\frac{\partial U_y}{\partial x}-\frac{\partial U_x}{\partial y}. \tag{106}$$

We also use the convective derivative

$$\frac{D_2}{Dt}=\frac{\partial}{\partial t}+U_x\frac{\partial}{\partial x}+U_y\frac{\partial}{\partial y}. \tag{107}$$

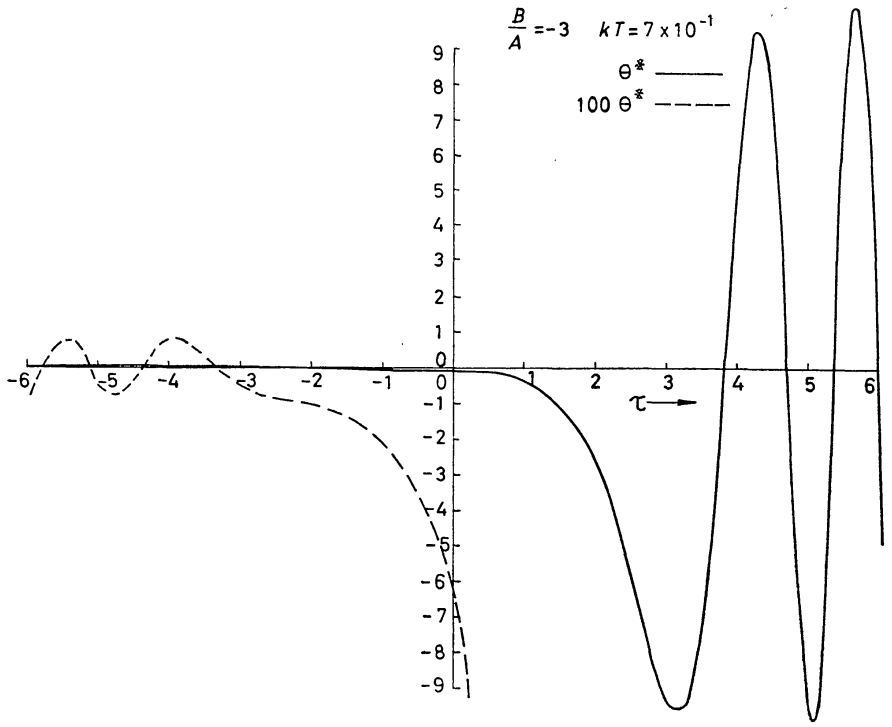

FIG. 5

In this notation equation (23) reads

$$\frac{D_2\mathbf{U}}{Dt} + 2\boldsymbol{\Omega}_a \times \mathbf{U} - \Omega_a{}^2\mathbf{R} = \boldsymbol{\nabla}_2\chi, \tag{108}$$

where $\boldsymbol{\Omega}_a \times \mathbf{U}$ is to mean $(-\Omega_a U_y, \Omega_a U_x)$. The third component of equation (23) reads $0=0$ thanks to vertical equilibrium.

We write

$$\omega = \mathrm{curl}_2\,\mathbf{U} + 2\Omega_a \tag{109}$$

so that ω is the total vorticity in inertial axes. Taking the curl_2 of equation (108) we find

$$\frac{D_2\omega}{Dt} + \omega\,\mathrm{div}_2\,\mathbf{U} = 0, \tag{110}$$

where to derive this equation we have used the identity

$$\mathrm{curl}_2\,[(\mathbf{U}\,.\,\boldsymbol{\nabla}_2)\mathbf{U}] = (\mathbf{U}\,.\,\boldsymbol{\nabla}_2)\,\mathrm{curl}_2\,\mathbf{U} + \mathrm{curl}_2\,\mathbf{U}\,\mathrm{div}_2\,\mathbf{U}. \tag{111}$$

Equation (110) may be written

$$\frac{D_2}{Dt}(\log\omega) + \mathrm{div}_2\,\mathbf{U} = 0. \tag{112}$$

This equation is similar to the surface density continuity equation which we now derive by integrating equation (35) through the sheet:

$$\frac{\partial\Sigma}{\partial t} + \mathrm{div}_2\,(\Sigma\mathbf{U}) = 0, \tag{113}$$

which may be written

$$\frac{D_2}{Dt}(\log\Sigma)+\mathrm{div}_2\,\mathbf{U}=0. \tag{114}$$

We subtract this equation from equation (112) to find

$$\frac{D_2}{Dt}\left(\log\frac{\omega}{\Sigma}\right)=0. \tag{115}$$

This is the convenient form that Kelvin's theorem takes when the horizontal motions are independent of height. For our non-linear discussion we shall consider only perturbation from the equilibrium state caused by perturbing pressure forces or by forces derivable from a time dependent potential. In that case equation (116) holds even when the perturbations are being applied, so we may take as its initial conditions the values in the unperturbed equilibrium. Thus

$$\frac{\omega}{\Sigma}=\frac{2B}{\Sigma_0}, \tag{116}$$

where B is Oort's constant.

We now derive the two-dimensional form of a remarkable equation due to Hunter. Taking the div_2 of equation (108)

$$\frac{D_2}{Dt}(\mathrm{div}_2\,\mathbf{u})+\frac{\partial u_j}{\partial x}\frac{\partial u_i}{\partial x_j}-2\Omega_a\,\mathrm{curl}_2\,\mathbf{u}=\nabla_2{}^2\chi+2\Omega_a{}^2, \tag{117}$$

where we have adopted the summation convention that terms with repeated indices are summed over x and y.

We write

$$E_{ij}=\frac{1}{2}\left(\frac{\partial U_i}{\partial x_j}+\frac{\partial U_j}{\partial x_i}\right), \tag{118}$$

so

$$E_{ij}E_{ij}=\frac{1}{2}\frac{\partial U_i}{\partial x_j}\frac{\partial U_i}{\partial x_j}+\frac{1}{2}\frac{\partial U_i}{\partial x_j}\frac{\partial U_j}{\partial x_i}, \tag{119}$$

but

$$\tfrac{1}{2}(\omega-2\Omega_a)^2=\frac{1}{2}\frac{\partial U_i}{\partial x_j}\frac{\partial U_i}{\partial x_j}-\frac{1}{2}\frac{\partial U_i}{\partial x_j}\frac{\partial U_j}{\partial x_i}, \tag{120}$$

so

$$\frac{\partial U_j}{\partial x_i}\frac{\partial U_i}{\partial x_j}=E_{ij}E_{ij}-\tfrac{1}{2}(\omega-2\Omega_a)^2. \tag{121}$$

Substituting this in equation (117) and evaluating (117) by means of the continuity equation (114) we obtain

$$-\frac{D_2{}^2}{Dt^2}(\log\Sigma)=\tfrac{1}{2}(\omega-2\Omega_a)^2+2\Omega_a(\omega-2\Omega_a)-E_{ij}E_{ij}+2\Omega_a{}^2+\nabla_2{}^2\chi, \tag{122}$$

and therefore

$$\frac{D_2{}^2}{Dt^2}(\log\Sigma)=-\tfrac{1}{2}\omega^2+E_{ij}E_{ij}-\nabla_2{}^2\chi, \tag{123}$$

which is the two-dimensional form of Hunter's equation (8). Using equations (123), (27) and (33) we find

$$\frac{D_2^2}{Dt^2}(\log\Sigma) = -\frac{2B^2\Sigma^2}{\Sigma_0^2} + e_{ij}e_{ij} + 2A^2 - \nabla_2^2\chi, \tag{124}$$

where

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) = E_{ij} - \begin{pmatrix} 0 & A \\ A & 0 \end{pmatrix}. \tag{125}$$

For the unperturbed state

$$0 = -2B^2 + 2A^2 - \nabla_2^2\chi_0, \tag{126}$$

so subtracting

$$\frac{D_2^2}{Dt^2}(\log\Sigma) = 2B^2\left(1 - \frac{\Sigma^2}{\Sigma_0^2}\right) + e_{ij}e_{ij} - \nabla_2^2\chi_1, \tag{127}$$

where

$$\chi_1 = \chi - \chi_0. \tag{128}$$

Now

$$e_{ij}e_{ij} = \mathrm{Tr}\,(\boldsymbol{e}.\boldsymbol{e}) = e_{11}^2 + e_{22} \tag{129}$$

when $\boldsymbol{e}$ is written in principal axes. But

$$e_{11}^2 + e_{22}^2 \geqslant \tfrac{1}{2}(e_{11} + e_{22})^2. \tag{130}$$

Hence,

$$e_{ij}e_{ij} \geqslant \tfrac{1}{2}(e_{ii})^2 = \tfrac{1}{2}(\mathrm{div}_2\,\mathbf{u})^2 = \frac{1}{2}\left[\frac{D_2}{Dt}(\log\Sigma)\right]^2, \tag{131}$$

so

$$\frac{D_2^2}{Dt^2}(\log\Sigma) \geqslant 2B^2\left(1 - \frac{\Sigma^2}{\Sigma_0^2}\right) + \frac{1}{2}\left[\frac{D_2}{Dt}(\log\Sigma)\right]^2 - \nabla_2^2\chi_1. \tag{132}$$

Even with the assumption of vertical equilibrium our form of Hunter's equation is not equivalent to the full equations of motion. In this sense its use in our problem is similar to that of the virial theorem in many dynamical problems. While not enabling us to find explicit solutions of the equations of motion it still allows us to draw general conclusions about the behaviour of such solutions. In particular, we shall use Hunter's equation to ascertain the conditions under which instabilities can continue to grow once non-linear effects become important.

We shall restrict our treatment of the non-linear stability problem to disturbances which are the non-linear generalizations of a sinusoidal plane wave instability. This is not to say that more general disturbances cannot be treated by the same method. We make this restriction simply because this particular class of instability is the most relevant one to spiral arm formation. From our study of the linearized stability problem we know that the unstable plane waves have wave-lengths which are at least several times the characteristic scale height of the sheet. For disturbances with this property we may observe that the vertical behaviour of the density can be approximated by taking the z dependence, at a given (x, y), the same as that for an entire sheet with central density $\rho_C(x, y, 0)$. Thus the scale height of the sheet at any point (x, y) is given by π/k_0 where

$$k_0 = \sqrt{\left[\frac{2\pi G\rho_C(x,y,0)}{c^2}\right]} \quad \text{and} \quad c^2 = \gamma K\rho_C{}^{\gamma-1}(x,y,0) \quad \text{for} \quad p = K\rho^\gamma.$$

Within this approximation it becomes a simple matter to estimate the terms on the right-hand side of Hunter's equation. Before making these estimates there is one further point which must be mentioned. In analogy to the plane wave disturbances discussed in the linearized theory, the disturbances that we shall consider will be constant on lines given by some time dependent linear combination of the x and y coordinates (at fixed z). However, unlike the sinusoidal plane wave instabilities, the scale lengths, perpendicular to the wave fronts, associated with the regions of enhanced and diminished density need not be the same.

We define k^+ and k^- such that π/k^+ and π/k^- are the scale lengths associated with regions of enhanced and diminished density respectively.

The form of Hunter's equation which is most convenient for our purposes is given by equation (132) and rewritten below:

$$\frac{D_2^2}{Dt^2}(\log\Sigma) \geqslant 2B^2\left(1-\frac{\Sigma^2}{\Sigma_0^2}\right)+\frac{1}{2}\left[\frac{D_2}{Dt}\log\Sigma\right]^2-\nabla_2^2\psi_1+\nabla_2^2\begin{Bmatrix} c^2\log\rho & \gamma=1 \\ \dfrac{\gamma\kappa}{\gamma-1}\rho^{\gamma-1} & \gamma\neq 1\end{Bmatrix}. \tag{133}$$

In the form given above, the terms on the right-hand side of Hunter's equation are especially easy to estimate. We shall be interested in the Σ dependence of these terms in regions of enhanced density. However, a brief interpretation of these terms shall be given first.

The $-\nabla_2^2\psi_1$ term contains the condensive force of gravity; the

$$\nabla_2^2\begin{Bmatrix} c^2\log\rho & \gamma=1 \\ \dfrac{\gamma\kappa}{\gamma-1}\rho^{\gamma-1} & \gamma\neq 1\end{Bmatrix}$$

term is the disruptive force due to the pressure; the $-2B^2(\Sigma^2/\Sigma_0^2-1)$ term gives the effect of the centrifugal field in opposing condensation; finally, the $e_{ij}e_{ij}$ term, which is condensive, is likely to be quite small compared with the others. We have inserted $\frac{1}{2}(D_2/Dt\log\Sigma)^2 \leqslant e_{ij}e_{ij}$ in its place and shall not consider it any further.

We now proceed to estimate these terms in the regions of enhanced density. Only these regions are considered since we are interested in the growth of density instabilities. Under the assumptions concerning vertical equilibrium and the scale lengths of perturbations made in this section we see that

$$\nabla_2^2\chi_1=\nabla_2^2\psi-\nabla_2^2\begin{Bmatrix} c^2\log\rho & \gamma=1 \\ \dfrac{\gamma\kappa}{\gamma-1}\rho^{\gamma-1} & \gamma\neq 1\end{Bmatrix}$$

and is independent of z. Its terms can be estimated as follows

$$\nabla_2^2\psi=\nabla^2\psi-\frac{\partial^2\psi}{\partial z^2}.$$

In regions of enhanced density

$$\nabla^2\psi=-4\pi G\rho^+\simeq-(k^{+2}+k_0^{+2})\psi.$$

Hence

$$\nabla_2^2\psi\simeq-\frac{4\pi G\rho k^{+2}}{(k^{+2}+k_0^{+2})}.$$

Similarly

$$\nabla_2^2 \begin{cases} c^2 \log \rho & \gamma = 1 \\ \dfrac{\gamma\kappa}{\gamma - 1} \rho^{\gamma-1} & \gamma \neq 1 \end{cases} \simeq k^{+2}\kappa\rho^{\gamma-1}. \tag{134}$$

In order to compare the two terms above with each other and the term arising from the centrifugal field, we must express ρ^+ and k_0^+ in terms of Σ^+. This is easily done using the definition of

$$\Sigma = \int_{-a}^{a} \rho \, dz \simeq \frac{4\pi G \rho_C}{k_0}.$$

Since

$$k_0 = \sqrt{\left(\frac{2\pi G}{c^2}\right)} \propto \rho_C{}^{(2-\gamma)/2}$$

we have

$$\Sigma \propto \rho_C{}^{\gamma/2} \quad \text{or} \quad \rho_C \propto \Sigma^{2/\gamma}.$$

Hence

$$k_0 \propto \Sigma^{2-\gamma/\gamma}.$$

This tells us that the terms in Hunter's equation due to the pressure, the rotational field and the self-gravity of the disturbance behave as

$$k^{+2}\Sigma^{+2(\gamma-1)/\gamma}, \quad 2B^2\left(1 - \frac{\Sigma^{+2}}{\Sigma_0{}^2}\right) \text{ and } \frac{k^{+2}\Sigma^{+2/\gamma}}{k^{+2} + k_{0i}{}^{+2}\left(\dfrac{\Sigma^+}{\Sigma_0}\right)^{2-\gamma/\gamma}}$$

respectively as Σ^+ increases. ($k_{0i}{}^+$ is the initial value of $k_0{}^+$.)

Until non-linear effects become important, i.e. until $\Sigma^+ - \Sigma_0/\Sigma_0$ becomes of order unity, the terms above can be replaced by their expansions to first order in $\Sigma^+ - \Sigma_0/\Sigma_0$. In this case γ enters Hunter's equation only through the coefficients of these terms. This accounts for the insensitivity of the linear approximation to the value of γ which is used. However, once the non-linear realm is reached the terms above depend critically on the value of γ, since γ now enters into the exponents of the pressure and gravity terms.

For the instability to continue to grow once non-linearity becomes important the gravitational terms must grow at least as rapidly with Σ^+ as the other two terms. This can only happen if k^+ grows like $(\Sigma^+)^S$. Setting the initial value of k^+ equal to k_i^+ we find that (if we restrict ourselves to $\gamma \geqslant 1$) $\gamma = 1$, $1 \geqslant S \geqslant \frac{1}{2}$ is the only solution which allows the gravitational term to dominate the others as Σ^+/Σ_0 increases.

We now see that in isothermal sheets, a mode which is unstable in the linear approximation will continue unstable in the non-linear realm provided its "wave-length" as determined by S, also changes with time.

9. *A physical mechanism for spiral arm formation*

9.1. *Qualitative considerations.*—In the mathematical sections of this paper we have shown that small perturbations of a differentially rotating, stratified sheet of self-gravitating gas may be considered as a superposition of density

waves which are sheared by the differential rotation. The analyses of perturbations in terms of these sheared modes turned out to be very fruitful. It allowed us to extend the concept of instability to include cases in which growth occurs only for a limited time (at least in the linear approximation).

From the equations that we have derived governing the growth of these modes, we have drawn one outstanding conclusion. It is that even when all purely radial disturbances are stable, there are still some sheared waves whose amplitudes grow by factors of more than 100. Moreover, this growth occurs for waves of a well-defined wave-length and begins as the lines of constant density are sheared past the radial direction. If the initial perturbations are so small, that even after this growth has taken place they remain small, then the growth gives way to oscillation as the trailing wave is stretched out by the differential rotation. However, if the initial perturbations are greater than 1 per cent of the unperturbed quantities, then non-linear effects will become important and the oscillations predicted by the linear theory may not arise. In particular, we have shown in our discussion of the non-linear problem (Section 8) that it is energetically advantageous for the growth of perturbations to continue into the non-linear range, provided that the gas is isothermal. However, if the energy of collapse is stored as thermal energy, condensation is eventually halted and the system will "bounce".

We are now in a position to apply the results of our calculations on gravitational instability to the theory of spiral arm formation in normal galaxies.

9.2. *Proposed theory of spiral arm formation.*—The pressure support of the interstellar gas is turbulent in origin. Hence, in the absence of energy sources it will die down. The gas sheet will become thinner and the total density will increase. Eventually $\pi G\bar{\rho}/4B(B-A)$ will become so large that considerable growth of certain sheared perturbations will ensue. Modes of optimum growth potential will have small initial amplitudes and point forwards (with respect to the direction of rotation). As they are swept around by the differential rotation their amplitudes will begin to grow. Greatest acceleration will occur when they point straight out from the galactic centre. The perturbations will grow to a magnitude at which the linear analysis is no longer a good approximation. However the gas will be able to radiate away the gravitational energy released during the collapse enabling the growth of the condensation to continue. At this stage a trailing spiral arm has been formed. We assume that at this point stars are born in the growing condensation. The new stars will stir up the interstellar gas. This extra turbulence will again increase the thickness of the gas layer of the galaxy and reduce its density below the level for instability. When the brightest new stars have died the turbulence of the interstellar gas will begin to diminish, instability will ensue, and the process will be repeated. Thus generation after generation of spiral arms will form, wind up, and disperse. The main secular effects will be the depletion of the gas (which will have to form a slightly thinner sheet each time) and the relaxation of the stellar motions by the gravity fields of the recurrent instabilities.

Although the excess density which causes the instability is produced by the thinning of the gas sheet, nevertheless, a considerable fraction of the total density may reside in the stars (as in our galaxy). Jeans' instability of a uniformly rotating system of stars may be shown to occur, for waves perpendicular to the axis of rotation, if $4\pi G\rho \geqslant 4\Omega^2$, exactly the same criterion as that for instability

in a rotating gas*. Jeans' gravitational instability thus occurs for stars in much the same way as it occurs for gas. Spiral arm formation should not, therefore, be regarded as an instability in the gas but rather as an instability of the whole star–gas mixture which is triggered by an increase in gas density.

Once the instability has developed into the non-linear range the difference between the stars and the gas will become important. The stars will conserve energy so their instability will be resisted by the non-linear terms. By contrast the gas can easily radiate the energy of compression in the time available and can therefore continue to condense.

In the absence of a more refined mathematical theory in which we can treat the stars and gas as separate fluids we must use the approximate criterion that when

$$\frac{\pi G\bar{\rho}_0}{4B(B-A)} > \sim 1{\cdot}0\dagger \tag{135}$$

considerable growth can occur.

The modes that grow most in our investigations have wave-lengths of about $4\pi T$ (when they point radially). Here T, the thickness of the galaxy, is defined as $T(R)=\Sigma(R)/\bar{\rho}(R)$. In the solar neighbourhood $T\sim 800$ pc which makes the wave-length embarrassingly large for something deduced from a small scale approximation. Of course this wave-length will decrease as the shearing proceeds so that when the waves have been swept around to make an angle α with the radius, the wave-length will be about $4\pi T\cos\alpha$. The equivalent length in a more complete theory which allowed for the curvature effects in the galaxy would be related to the distance between spiral arms. In our theory it can do no more than indicate the sort of length scales involved in growing condensations. They are in the right range for spiral arms when suitable angles α are used.

From a local theory we cannot produce any preference for the formation of symmetrical two-arm spirals. However it seems likely that the instability leading to them is a somewhat more organized form of the one discussed here.

10. *Observational consequences and predictions*

Perhaps the most important prediction of the theory is that anywhere in any spiral galaxy (except in the nuclear bulges) the star–gas mixture must be on the borderline of gravitational instability. For stellar velocity distributions whose smaller axis of dispersion in the galactic plane is not considerably greater than the axis normal to the plane this leads to the prediction

$$\frac{\pi G\bar{\rho}}{4B(B-A)} \sim 1. \tag{136}$$

A discussion of the exactness of this number is not inappropriate. For a galaxy that is all isothermal gas it should be 0·7 if ring instabilities are critically stable. However transverse modes would form spiral structure even before that so the number might be slightly further reduced to 0·6 say. For galaxies of stars

* It is a simple matter to make the slight extension of the result proved in (9) to cover waves exactly perpendicular to the axis of rotation.

† This number is rather sensitive to anisotropy in the pressure. For $\gamma=1$ it is 0·7 for isotropy and 1·8 for an anisotropy corresponding to sound velocities in the ratio of 2·1 : 1 between the direction of wave propagation and the vertical. For $\gamma=2$ it is 1·1 for isotropy.

the problem is aggravated by velocity anisotropy and with no velocity dispersion normal to the plane but considerable dispersions in the plane the number would become infinite, the stability being dependent on the surface density as discussed in the note after paper I but with $B(B-A)$ replacing Ω^2 (which takes account of the differential rotation exactly for ring modes). For the ring modes near the Sun with the observed anisotropy of 2·1 : 1 the number would be 1·8 but it is the sheared spiral arm modes that are really relevant to the problem because these are more unstable on two counts. Firstly as discussed above even when pressures are isotropic sheared modes show instability while ring modes are stable. Secondly as discussed in Appendix II of paper I the lower velocity dispersion in the tangential direction should favour sheared modes. From that note and the replacement $4\Omega^2 \to 4B(B-A)$ we see that these transverse sheared modes will be on the verge of instability if $\pi G\bar{\rho}/4B(B-A)$ is slightly less than 1·3(5). This is in striking agreement with Jones' (**10**) discussion of the best value of the total density of gravitating matter derived from observations. Correcting his value for $B=-10$, $A=15$ we obtain from observations

$$\bar{\rho} = \tfrac{2}{3}\rho_C = 6 \cdot 10^{-24} \text{ gm/cm}^3$$

and

$$\frac{\pi G\bar{\rho}}{4B(B-A)} = 1{\cdot}2(5). \tag{136}$$

This is probably better agreement than we deserve.

Unfortunately in external galaxies the velocity anisotropies are not observable but, assuming that they vary from isotropic to somewhat beyond those found at the Sun the expected range of $\pi G\bar{\rho}/4B(B-A)$ would be from 0·6 to 1·5.

In our theory the central line of the condensing material moves with the differential rotation. Hence, we can also make a prediction of the shape of spiral arms. If $(R, \phi(R))$ is a general point on the central line of a spiral arm and if (R_0, ϕ_0) is some reference point of the central line of the same arm then

$$\phi(R) - \phi_0 = (\Omega(R) - \Omega(R_0))t \tag{137}$$

is the equation of the arm at time t after the embryo arm pointed radially. This is a reasonable starting time since the acceleration $\ddot{\theta}_1^*/\theta_1^*$ is greatest about then; it is always the initial stages of growth that are crucial to the development of gravitational instability.

Formula (137) is also open to observational test. Consider an open armed spiral whose planes make an angle of about 45° with the plane of the sky. Furthermore we assume that it has reasonably continuous arms, or at least a number of pieces of arm that overlap in radial distance from the galactic centre. Then, if the tilt of the galaxy can be estimated $\phi(R)-\phi_0$ can also be estimated. If the velocities can be measured then $\Omega(R)-\Omega(R_0)$ can be determined up to a multiplicative "distance scale". Hence a plot of $\phi(R)-\phi_0$ against $\Omega(R)-\Omega(R_0)$ should yield a straight line whose gradient is the product of t and the constant entering the distance scale. Disconnected pieces of spiral arm should have different values of t since we have no reason to suppose that they started to form at the same time.

Other consequences of this theory that have not been exploited in this paper are:

(1) the possibility that the shock waves which seem to be an inevitable consequence of differential rotation, will play an important role in stirring up the interstellar gas;

(2) the strong relaxing effect of the growing spiral arms on stellar motions.

Mechanism (1) may actually provide the feedback which keeps the gas sheet at the right density. A growing spiral arm will provide pressure perturbations for other modes; these will feed on the energy of differential motion before dissipating it in shock waves which produce heat and turbulence in the interstellar gas.

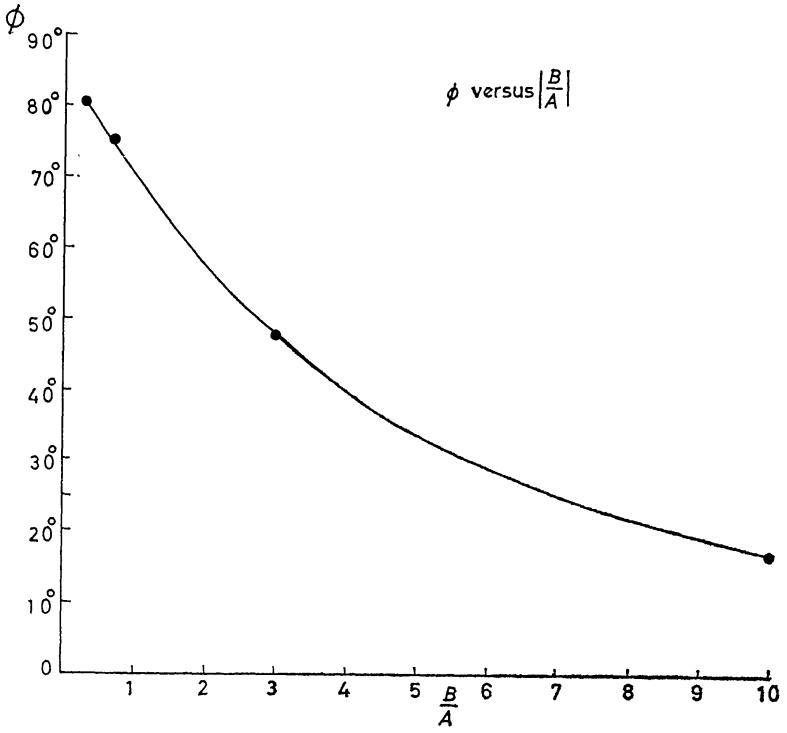

FIG. 6

It is important to realize that equation (135) contains the key to the building of more or less unique models of flat galaxies based on observations of velocity laws. It is well known that the balance of centrifugal force and gravity leads to a distribution of surface density $\Sigma(R)$ for a disk. Equation (135) shows that when the velocity law is known the mean density $\bar{\rho}(R)$ is determined. Thus the thickness $T(R)=\Sigma(R)/\bar{\rho}(R)$ is also determined and the balance between the stellar motions and the gravity on to the disk will determine the vertical velocity dispersion $\sigma_{zz}^2(R)$. It is clear from Figs. 3, 4 and 5 that for different values of $|B/A|$ the wave-lengths associated with the greatest growth vary from about $2\pi T$ (T = thickness of the galaxy) for large $|B/A|$, such as 10 or 3, to about $4\pi T$ for the more greatly sheared cases, $|B/A|=\frac{2}{3}$ or $\frac{1}{3}$. Another result of the computations is that growth by a given factor is attained at smaller values of τ for the larger values of $|B/A|$. Initial perturbations of given size will grow into the non-linear regime before they are very violently sheared when the galaxy is nearly uniformly rotating. They will be more tightly sheared for strong differential rotation. If we take as a representative point the value of the shear angle $\phi=\tan^{-1}\tau$ when growth by a factor of 50 is attained then we find the graph of Fig. 6.

This could be related to the difference between open and tightly wound spirals in which case our prediction would be that the more open spirals rotate more uniformly (i.e. have the larger average values of $|B/A|$). However, this is not an inescapable consequence of the theory because the more open spirals are notably more messy and may therefore generate larger initial perturbations than the tight spirals. Smaller growth factors could therefore bring such perturbations into the non-linear regime. If equation (137) were observationally checked in at least one galaxy one would have some confidence in applying it to face-on spirals in which the velocity law is unmeasurable. By combining results from several pieces of different arms it should be possible to build up a picture of the variation of $\Omega(R)$ with R, determined up to a normalizing constant (and possibly a zero point). This provides a method of determining mass distributions (but not masses) for face-on spirals.

There is also the less certain prediction (discussed in Section 10) that the more open armed spiral galaxies should rotate more uniformly than the tightly wound ones.

Assuming the velocity dispersion tensor

$$\boldsymbol{\sigma}(\mathbf{r}) \equiv \int f \mathbf{cc}\, d^3c \Big/ \int f\, d^3c, \tag{138}$$

where $f(\mathbf{r}, \mathbf{c})$ is the distribution function (weighted with the masses) always has one principal axis vertical, we may deduce Jeans' stellar hydrodynamical equation in the form

$$\frac{\partial}{\partial z}(\rho\sigma_{zz}) = \rho \frac{\partial \psi}{\partial z}. \tag{139}$$

We have also assumed a steady state. Further for a stratified distribution

$$\frac{\partial \psi}{\partial z} = -4\pi G \int_0^z \rho\, dz. \tag{140}$$

Thus

$$[\rho\sigma_{zz}]_0^z = -4\pi G \int_0^z \rho(z) \int_0^z \rho(z')\, dz'\, dz \tag{141}$$

and hence

$$[\rho\sigma_{zz}]_0^z = -\pi G \left[\int_0^z \rho(z)\, dz\right]^2, \tag{142}$$

where the right-hand side has been derived using an integration by parts. Hence

$$\sigma_{zz}(R, 0) = \frac{\pi G}{2} \frac{\left[\int_{-\infty}^{\infty} \rho(R, z)\, dz\right]^2}{\rho_C(R)}, \tag{143}$$

where $\rho_C(R)$ is the density at the centre of the sheet. Now $\rho_C/\bar{\rho} = 1{\cdot}5$ for $\gamma = 1$ and $\rho_C/\bar{\rho} = 4/\pi$ for $\gamma = 2$ which shows that this ratio is insensitive to the value of γ

(in the range of interest). Hence, even if σ_{zz} varies somewhat with height (as it does for a $\gamma=2$ sheet) we may expect

$$\sigma_{zz}(R,0)=\frac{\pi G}{2}\frac{\left[\int_{-\infty}^{\infty}\rho(R,z)\right]^2}{(\pi/4)^{-1}\bar{\rho}} \tag{144}$$

$$=\frac{\pi^2}{8}GT^2\bar{\rho}. \tag{145}$$

Thus not only the surface densities but also the mean densities, thicknesses, and velocity dispersions can be derived from velocity curves for spiral galaxies, provided that we assume the vertical velocity distribution is not much smaller than the least dispersion in the plane.

11. *Further problems*

11.1. *Barred spirals.*—The gravity field of the bar must dominate the dynamics of at least the central regions of barred spirals. Thus only in the outer parts can a theory of the type presented in this paper apply. However the near uniform rotation of barred spirals removes the winding problem which makes it much easier to construct theories that are plausible. In this section we discuss a theory that has been developing at the hands of a number of authors.

If a cloud is falling together under its self-gravity and in the absence of pressure support then any initial inequality of axes will be greatly exaggerated during the motion. This is true even when rotation is present so one might expect objects which have recently fallen together to have very unequal axes, $a\gg b\gg c$ (**11**), (**12**). These necessarily elongated objects will have a natural preference to form their shortest axes along their rotation axes. Ogorodnikov has pointed out that there are such elongated objects among the members of some chains of galaxies catalogued by Vorontsov–Velyaminov, and he suggests that these are protobarred spirals. He and Antonov have suggested that the ends of the bars of barred spirals are neutral points of the total gravitational plus centrifugal field (**13**). A streaming of material from these points could be responsible for the arms of barred spirals. More recently Freeman (**14**) and Prendergast (**15**) have shown that in the rotating axes of the bar, particles leaving the end of a uniform gravitating ellipsoid with a small radial velocity trace out convincing spiral arms. The presence of this material trailing behind the bar must lower its angular momentum. Relieved of some of its angular momentum the bar will shorten. Fujimoto has discussed this process using dynamical models due to Aarseth while Ogorodnikov has pointed to the observation that old barred spirals have stubby arms, in good agreement with this line of thought (**16**)–(**19**).

The origin of the streaming from the ends of the bar and of the continued presence of neutral points there remains obscure, so we present here an idea of a possible mechanism. Suppose the configuration is as postulated at some time, then the bar loses angular momentum to the arms and contracts a little. This contraction increases both the angular velocity and density of the bar. The latter is further increased by the lateral contraction necessary to make the pressure balance the increased lateral gravity (assuming $\gamma<2$). Although the loss of angular momentum eases the problem that the galaxy has in holding itself

together, nevertheless the increased density aggravates the problem. If this aggravation wins then the neutral points will move inwards through the material of the bar leaving more arms behind them.

11.2. *Problems raised by the present work.*—(i) The divergence of the density leading to shocks that we discussed briefly in Section 3 should be further investigated. The physical reason why shocks form in a sheared flow may be compared with the mechanism of shock formation in Riemann's plane non-linear waves. There, different disturbances overtake one another because the wave velocity depends on the density. In our case the behaviour occurs in the linearized theory because the mean flow velocity is added to the wave velocity yielding disturbances that can overtake one another.

It would be important to calculate the energy input into the interstellar gas due to these shocks.

(ii) A non-linear treatment of instability growth using the equations of motion rather than Hunter's equation should be attempted.

(iii) It would be important to determine the critical value of $\pi G\tilde{\rho}/4B\Omega$ for the superposed sheets of gas and stars discussed in Section 10. A better value could then be used to build galactic models.

(iv) Using the formulae given here models of galaxies should be derived.

(v) The stability of such a model without the use of a small scale analysis should be attempted.

(vi) The validity of the vertical equilibrium approximation should be checked.

(vii) Application of this theory of gravitational instability to other differentially rotating systems should be considered, e.g. for the cosmogony of the solar systems.

(viii) A mathematical model of the dynamics of a barred spiral galaxy obeying our mechanism should be worked out.

(ix) The rate of relaxation of stellar orbits due to spiral arm formation should be worked out. This is likely to be the dominant relaxation mechanism in the galactic disk. More generally whenever the mean gravitational field of a whole stellar system is undergoing rapid change the effective relaxing effect is likely to be enormous. The density distributions of elliptical galaxies could be due to rapid relaxation during their chaotic birth stages. This form of relaxation will lead to relaxation times and velocity dispersions that are independent of stellar mass.

Appendix

Validity of the vertical equilibrium approximation.—In paper I we worked out the exact dispersion relation for the incompressible uniformly rotating sheet. It is

$$-\frac{\pi G\rho_0}{\omega^2k^2a^2}\,nka\,th(nka)=\frac{1}{(1+e^{-2|k|a}-2|k|a)2|k|a},$$

where

$$n^{-2}=1-\frac{4\Omega^2}{\omega^2}.$$

However, following the method of the present paper one may also work out the dispersion relation approximately by using the vertical equilibrium approximation and integrating through the sheet. The result is

$$-\frac{\pi G\rho_0}{\omega^2-4\Omega^2}=\frac{1}{(1+e^{-2|k|a}-2|k|a)2|k|a}.$$

Evidently the two expressions coincide (as they must) when ω is small for then n is small and $th(nka)$ may be replaced by nka.

Our present concern is that for the differentially rotating sheet we have used the vertical equilibrium approximation for modes whose wave numbers vary through the critical ones and which may at times be considerably different from them.

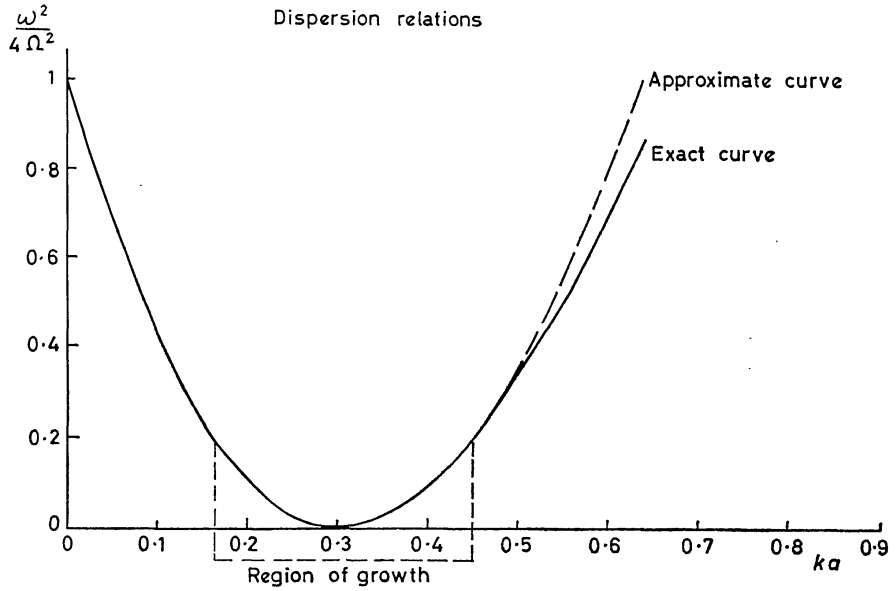

FIG. 7

How bad is the equilibrium vertical approximation when used away from the critical wave numbers? Does it give the correct qualitative behaviour and if so what are the errors quantitatively? What is crucial to the preceding argument is that both the onset and the growth of the instability should be determined accurately. The accuracy of our solutions in their oscillatory regions is irrelevant.

By looking at the computed solutions we determined approximately the range that k swept through during the period of growth. In Fig. 7 we have plotted the exact and approximate dispersion relations and have indicated the region of growth. In this region the approximation is virtually exact; not bad for an approximation which saves six orders of differentiation!

Note added in proof.

We have heard from Dr Toomre and Mr Julian of further work on zero thickness stellar disks including a discussion of sheared modes. These behave very similarly to their gaseous counterparts discussed here. This work was

independent of ours although the same sheared coordinates have been invented by them. Their discussion of truly stellar disks adds to our confidence in applying results obtained for gas to a mainly stellar galaxy (**20**), (**21**).

Department of Applied Mathematics and Theoretical Physics, University of Cambridge: *Clare College, Cambridge.*

1964 *June.*

References

(1) Earl of Rosse, *Phil. Trans. R.S.*, London, 1850, 499.
(2) J. H. Jeans, *Cosmogony and Stellar Dynamics*, Cambridge University Press, 1919, p. 209.
(3) B. Lindblad, e.g. *M.N.*, **97**, 642, 1937, and references given there.
(4) B. Lindblad, e.g. *Stockholm Obs. Ann.*, **20**, No. 4 (1958) and references given there.
(5) C. C. Lin and F. H. Shu, *Ap. J.*, **140**, 646, 1964.
(6) C. Hunter, *M.N.*, **126**, 299, 1964.
(7) W. A. Fowler and F. Hoyle, Star formation, *R.O.B.*, **56**, 1962.
(8) C. Hunter, *Ap. J.*, **139**, 570, 1964.
(9) D. Lynden-Bell, *M.N.*, **124**, 279, 1962.
(10) D. H. P. Jones, *R.O.B.*, **52**, 1962.
(11) D. Lynden-Bell, *Ap. J.*, **139**, 1195, 1964.
(12) D. Lynden-Bell, *M.N.*, **129**, 299, 1965.
(13) V. A. Antonov, *Vestnik Leningrad University, No.* 13, *Series Math. Mech. Astro. Vipusk* **3**, 157, 1961.
(14) K. C. Freeman, *M.N.*, **130**, 63, 1965.
(15) K. Prendergast, *A.J.*, **69**, 147, 1964.
(16) M. Fujimoto, *P.A.S. Japan*, **15**, 107, 1963.
(17) S. Aarseth, *M.N.*, **121**, 525, 1960.
(18) S. Aarseth, *M.N.*, **122**, 535, 1961.
(19) K. Ogorodnikov, (*Lecture in Cambridge*) 1964.
(20) A. Toomre, *Ap. J.*, **139**, 1217, 1964.
(21) A. Toomre, W. H. Julian (in preparation).