THE ASTROPHYSICAL JOURNAL, Vol. 160, April 1970

ON THE DENSITY-WAVE THEORY OF GALACTIC SPIRALS. II. THE PROPAGATION OF THE DENSITY OF WAVE ACTION

FRANK H. SHU
State University of New York at Stony Brook
Received 1969 August 5; revised 1969 October 2

ABSTRACT

The properties of galactic density waves are studied in the WKBJ approximation. In the lowest order of approximation, we reproduce the dispersion relation reported by Lin and Shu in an earlier communication. In the next order, we demonstrate explicitly that the density of "wave action" is transported with the group velocity derived by Toomre. Some general implications are drawn for mechanisms proposed for the origin of spiral structure.

I. INTRODUCTION

In the previous paper (Shu 1970, hereinafter called Paper I), we developed an exact formulation for the linear gravitational oscillations which can occur in the plane of a stellar disk with infinitesimal thickness. Studied here are the properties of oscillations whose associated wavelengths are small compared with the radial dimensions of the disk. Such *density waves* allow a relatively complete treatment in the WKBJ approximation (Lin and Shu 1964, 1966).

For mathematical convenience, leading and trailing waves are treated separately. Boundary conditions and stellar resonances are recognized to play an important role in the decision as to whether leading and trailing waves need eventually to be superimposed (Paper I); the detailed consideration of these effects will be left for a future investigation.

II. THE SURFACE DENSITY REQUIRED TO SUPPORT A SPIRAL GRAVITATIONAL FIELD

It is useful to consider the dynamical response separately from the gravitational potential calculation. Thus, we begin by finding from Poisson's equation the surface density required to support a spiral gravitational field.

We adopt an inertial frame referred to cylindrical coordinates (ϖ, θ, z), with $z = 0$ defining the plane of the galaxy and $\varpi = 0$ defining the center. We assume, with Lin and Shu (1964), that in the plane of a galactic disk modeled with infinitesimal thickness there exists a perturbation of the gravitational potential of the form

$$\mathfrak{V}_1(\varpi, \theta, z = 0, t) = V(\varpi)e^{i(\omega t - m\theta)} . \tag{1}$$

We note that this potential is of spiral form with short arm spacing if

$$V(\varpi) = A(\varpi)e^{i\Phi(\varpi)} \quad \text{with} \quad A(\varpi) \quad \text{and} \quad \Phi(\varpi) \text{ real} , \tag{2}$$

and if the variation of the phase $\Phi(\varpi)$ is rapid compared with that of the amplitude $A(\varpi)$.[1] In practice, we require $|\varpi\Phi'(\varpi)| \gg 1$. With the convention that m is a nonnegative integer (its value gives the number of spiral "arms"), the spiral pattern leads if $k = \Phi'(\varpi) > 0$ and trails if $k < 0$.

The gravitational potential $\mathfrak{V}_1(\varpi, \theta, z, t) = V(\varpi, z) \exp [i(\omega t - m\theta)]$ can be supported

[1] The restriction that the rapidly varying part of $V(\varpi)$ be confined to its phase prevents us from considering overstable modes of the type discussed by Hunter (1969).

by the surface density $\sigma_1(\varpi, \theta, t) = S(\varpi) \exp[i(\omega t - m\theta)]$, provided Poisson's equation is satisfied. This requires

$$\left[\frac{\partial^2}{\partial\varpi^2} + \frac{1}{\varpi}\frac{\partial}{\partial\varpi} + \frac{\partial^2}{\partial z^2} - \frac{m^2}{\varpi^2}\right] V(\varpi, z) = 4\pi G S(\varpi)\delta(z) , \tag{3}$$

where $\delta(z)$ is the Dirac δ-function. We wish to solve for $S(\varpi)$, correct to two orders in the small parameter $|k\varpi|^{-1}$, when $V(\varpi, z)$ is given in the plane $z = 0$ by equation (2).

Integration of equation (3) across the galactic plane shows that $S(\varpi)$ can be obtained from the jump condition

$$S(\varpi) = \frac{1}{4\pi G}\left[\frac{\partial V}{\partial z}(\varpi, z)\right]_{z=0-}^{0+} . \tag{4}$$

From the reflection symmetry of the problem, $V(\varpi, z)$ must depend only on the absolute value of z. Further, we multiply V by $\varpi^{1/2}$ to emphasize the cylindrical geometry and write

$$W(\varpi, |z|) = \varpi^{1/2} V(\varpi, z) . \tag{5}$$

After some manipulation, equations (3) and (4) take the forms

$$\frac{\partial^2 W}{\partial\varpi^2} + \frac{\partial^2 W}{\partial|z|^2} - \frac{m^2 - \frac{1}{4}}{\varpi^2} W = 0 \qquad \text{for} \quad |z| > 0 , \tag{6a}$$

$$S(\varpi) = \frac{\varpi^{-1/2}}{2\pi G}\left\{\frac{\partial W}{\partial|z|}(\varpi, |z|)\right\}_{|z|=0} . \tag{6b}$$

Equation (6a) is to be solved under the boundary conditions

$$W(\varpi, 0) = \varpi^{1/2} A(\varpi) e^{i\Phi(\varpi)} , \tag{7a}$$

$$W(\varpi, 0) \to 0 \quad \text{as} \quad |z| \to \infty . \tag{7b}$$

Only the disk part of the galaxy where $|k|\varpi \gg 1$ will be studied; in this region the curvature term $(m^2 - \frac{1}{4})W/\varpi^2$ is of order $(k\varpi)^{-2}$ smaller than the first two terms in equation (6a). Thus, equation (6a), correct to two orders in the WKBJ approximation, is simply Laplace's equation in rectangular coordinates for two dimensions:

$$\frac{\partial^2 W}{\partial\varpi^2} + \frac{\partial^2 W}{\partial|z|^2} = 0 \qquad \text{for} \quad |z| > 0 , \tag{8}$$

which is to be solved under the boundary conditions (7).

In this approximation, the continuation of $W(\varpi, |z|)$ from $|z| = 0$ into the region $|z| > 0$ is trivial when the theory of complex variables is used. We simply replace ϖ in the argument of $W(\varpi, 0)$ by $\varpi \pm i|z|$ to obtain $W(\varpi, |z|)$. The correct choice of sign in $\varpi \pm i|z|$ is dictated by the boundary condition at infinity. Consistent with the adopted approximation, we may replace that condition, given through equation (7b), by the restriction that for small $|z|$ the magnitude of $W(\varpi, |z|)$ falls from its value in the plane $|z| = 0$. If $\varpi + ip|z|$ with $p = \pm 1$ gives the correct choice of sign and if the phase $\Phi(\varpi)$ is a rapidly varying function, this condition can be interpreted with the aid of equation (7a) to imply

$$\text{Im}\{\Phi(\varpi + ip|z|)\} > 0 \qquad \text{for} \quad |z| > 0 .$$

Expanding $(\varpi + ip|z|)$ for very small $|z|$ leads to the requirement

$$p = \pm 1 = \text{sign}\{\Phi'(\varpi)\} = \text{sign}\{k\} . \tag{9}$$

The solution to equation (8), corresponding to the choice of sign (9) and subject to the boundary condition (7a), can now be written

$$W(\varpi, |z|) = (\varpi + ip|z|)^{1/2} A(\varpi + ip|z|) e^{i\Phi(\varpi + ip|z|)} . \tag{10}$$

The surface density required to support the spiral gravitational potential is obtained from equations (6b) and (10) as

$$S(\varpi) = -\frac{|k| V(\varpi)}{2\pi G} \left\{ 1 - \frac{i}{k\varpi} \frac{d \ln}{d \ln \varpi} [\varpi^{1/2} A(\varpi)] \right\} , \tag{11}$$

an expression which contains two orders of the approximation for large $|k|\varpi$. In the lowest order of approximation, equation (11) reproduces the result obtained by Lin and Shu (1964) in which surface-density maxima correspond with potential minima.

III. THE RESPONSE OF THE STARS

Consider the dynamical response of a stellar disk (of infinitesimal thickness) to forcing from an oscillating gravitational potential of the form given by equation (1). In Paper I, the perturbation in the distribution function of stars was shown to have the form

$$\psi_1 = \left(\frac{\partial F_0}{\partial E_0} V + f \right) e^{i(\omega t - m\theta)} , \tag{12}$$

where f is given by equation (22b) of that work:

$$f(\varpi, E_0, J) = -\frac{\omega \partial F_0 / \partial E_0 + m \partial F_0 / \partial J}{2 \sin (\omega \tau_{12} - m\theta_{12})} \int_{-\tau_{12}}^{\tau_{12}} V(\varpi_*(\tau)) e^{i[\omega\tau - m\theta_*(\tau)]} d\tau . \tag{13}$$

In the above $F_0(E_0, J)$ is the distribution function in the basic state. It is assumed to be zero when a star's energy (per unit mass) E_0 is positive or when its angular momentum (per unit mass) J is negative. These assumptions are made to restrict the disk to contain only stars which are bound and which rotate in a given sense.

The other parameters in equation (13) are defined so that $2\tau_{12}(E_0, J)$ is the radial period of oscillation of stars with energy E_0 and angular momentum J, and $2\theta_{12}(E_0, J)$ is the azimuthal angle traversed in time $2\tau_{12}$. The variables $\varpi_*(\tau)$ and $\theta_*(\tau)$ describe, as functions of the time τ, the unperturbed orbit of such a star when its position at $\tau = \pm\tau_{12}$ is $(\varpi, \pm\theta_{12})$:

$$\frac{d\varpi_*}{d\tau} = \Pi_0(\varpi_*, E_0, J) \qquad \text{with} \quad \varpi_* = \varpi \quad \text{at} \quad \tau = \pm\tau_{12} , \tag{14a}$$

$$\frac{d\theta_*}{d\tau} = \frac{J}{\varpi_*^2} \qquad \text{with} \quad \theta_* = \pm\theta_{12} \quad \text{at} \quad \tau = \pm\tau_{12} . \tag{14b}$$

In equation (14a), $\Pi_0(\varpi_*, E_0, J)$ is the radial velocity of the star. The functional dependence of Π_0 on its arguments is given by

$$\Pi_0(\varpi, E_0, J) = \{2[E_0 - \mathfrak{B}_0(\varpi)] - J^2/\varpi^2\}^{1/2} , \tag{15}$$

and $\mathfrak{B}_0$ refers to the steady axisymmetric gravitational potential which characterizes the basic state. The dependence of ϖ_* and θ_* on (ϖ, E_0, J) has been deliberately suppressed.

a) The Modified Schwarzschild Distribution

For the basic distribution of the disk stars we adopt a model which simulates many of the observed features in our own Galaxy. This model, the "modified Schwarzschild distribution," has been described by Shu (1969):

$$F_0(E_0, J) = \Psi_0 = \begin{cases} P_0(r) \exp[-\mathcal{E}/c_0^2(r)], & \mathcal{E} < -E_c(r), \quad r > 0; \\ 0, & \text{otherwise}. \end{cases} \quad (16)$$

The form of the functions $P_0(r)$ and $c_0(r)$ can be chosen to match any reasonable variations, with distance from the galactic center, of the basic stellar surface density σ_* and the rms velocity dispersion in the radial direction $\langle p_\varpi^2 \rangle^{1/2}$. The "epicyclic integrals," r and $\mathcal{E}$, are defined as those functions of J and E_0 which satisfy the relations

$$r^2\Omega(r) = J, \quad \mathcal{E} = E_0 - E_c(r), \quad E_c(r) = \tfrac{1}{2}r^2\Omega^2(r) + \mathfrak{V}_0(r), \quad (17)$$

where $\Omega(\varpi) > 0$ is the circular frequency whose variation with respect to its argument is defined in terms of the basic gravitational potential:

$$\varpi\Omega^2(\varpi) = \frac{d\mathfrak{V}_0}{d\varpi}(\varpi). \quad (18)$$

We may think of equation (16) as giving a parametric representation of the basic distribution function in terms of $(\varpi, c_\varpi, c_\theta)$ when the peculiar velocities (c_ϖ, c_θ) are defined through the relations

$$c_\varpi = \{2[\mathcal{E} - \mathcal{E}_c(r, \varpi)]\}^{1/2}, \quad c_\theta = \frac{r^2}{\varpi}\Omega(r) - \varpi\Omega(\varpi), \quad (19)$$

and when $\mathcal{E}_c(r, \varpi)$ is given by

$$\mathcal{E}_c(r, \varpi) = \mathfrak{V}_0(\varpi) - \mathfrak{V}_0(r) + \frac{r^2\Omega^2(r)}{2}\left(\frac{r^2}{\varpi^2} - 1\right). \quad (20)$$

In the transformation $(E_0, J) \to (\mathcal{E}, r)$, the partial derivatives transform as follows:

$$\frac{\partial}{\partial E_0} = \frac{\partial}{\partial \mathcal{E}}, \quad \frac{\partial}{\partial J} = \frac{2\Omega(r)}{r\kappa^2(r)}\frac{\partial}{\partial r} - \Omega(r)\frac{\partial}{\partial \mathcal{E}}. \quad (21)$$

Here, κ is the epicyclic frequency defined in terms of the circular frequency through

$$\kappa^2(\varpi) = 4\Omega^2(\varpi)\left\{1 + \frac{1}{2}\frac{d \ln \Omega}{d \ln \varpi}\right\}. \quad (22)$$

In what follows, we suppose that the dimensionless parameter

$$\varepsilon(r) = \frac{c_0(r)}{r\kappa(r)} \quad (23)$$

is small compared with unity. This assumption is equivalent to regarding the local dispersion speeds as being small compared with the circular velocity and is a condition which is likely to be satisfied in the disk parts of spiral galaxies.

To two orders of approximation in ε, one of the terms in equation (13) can be seen with the aid of equations (16) and (21) to have the form

$$-\left(\omega\frac{\partial F_0}{\partial E_0} + m\frac{\partial F_0}{\partial J}\right) = \nu(r)\frac{\kappa(r)}{c_0^2(r)}\Psi_0(\mathcal{E}, r) \quad (24)$$

where ν is the frequency of the wave, expressed in units of the epicyclic frequency, viewed relative to the circular motion of the stars:

$$\nu(r) = \frac{\omega - m\Omega(r)}{\kappa(r)}. \quad (25)$$

In the same approximation the rms velocity dispersion is given by

$$\langle c_\varpi^2\rangle^{1/2}(r) = c_0(r) , \tag{26}$$

whereas the basic distribution has the form (see Shu 1969, eq. [33])

$$\Psi_0 = \frac{2\Omega(r)}{\kappa(r)} \frac{\sigma_*(r)}{2\pi m_* \langle c_\varpi^2\rangle(r)} \exp\left[-\epsilon^2/2\varepsilon^2(r)\right] . \tag{27}$$

To derive equation (27) in its present form, it has been assumed that all the stars have the same mass m_*. We have also defined the "eccentricity" of the stellar motion ϵ—not to be confused with $\varepsilon(r)$ appearing in equation (23)—as the dimensionless variable

$$\epsilon = \frac{\sqrt{(2\mathcal{E})}}{r\kappa(r)} . \tag{28}$$

b) Response to Forcing from a Spiral Gravitational Field

To compute the response of the distribution function (27) to an imposed spiral gravitational field (2), we adopt the hypothesis that ε and $|k\varpi|^{-1}$ are of the same order of smallness (about 0.1 and 0.06 for application to the solar neighborhood). To compute the surface-density response to two orders in this asymptotic approximation, we require all formulae to be accurate to two orders in ε or in $|k\varpi|^{-1}$. An exception is the relation giving the radial orbit $\varpi_*(\tau)$ from equation (14a) for which we need accuracy to three orders in ε. (Two orders in ε represent the usual epicyclic approximation.) This exception is required to establish the rapid phase variation of $V(\varpi_*(\tau))$ in equation (13) correctly to two orders.

i) *Stellar Orbits*

For small ε, only the part of velocity space that corresponds to $\epsilon \lesssim \varepsilon(r)$ contains a significant fraction of the stars in the distribution function (27). Hence, in the determination of the stellar orbit defined by equations (14) we need only find expressions which are correct to the corresponding number of orders in ϵ.

A systematic expansion giving the stellar orbits in this manner may be obtained by using the parametric representation introduced in the Appendix of the paper by Shu (1969). In this representation, one radial period $2\tau_{12}$ corresponds to a change of 2π in the radial phase coordinate ϕ defined through the relation

$$\Pi_0(\varpi_*, E_0, J) = r\kappa(r)\epsilon \sin\phi . \tag{29}$$

If the dimensionless time variable

$$s = \kappa(r)\tau \tag{30}$$

is introduced, the parametric representation of the required orbit may be obtained by expansion of equations (14) to yield

$$s - s_0 = \phi - 2B_2(r)\epsilon \sin\phi ; \tag{31a}$$

$$\frac{r}{\varpi_*} = 1 + \epsilon\cos\phi + A_2(r)\epsilon^2\cos^2\phi , \tag{31b}$$

$$\theta_* - \theta_0 = \frac{\Omega(r)}{\kappa(r)}\left[\phi + 2A_2(r)\epsilon\sin\phi\right] . \tag{31c}$$

The coefficients A_2 and B_2 are given by

$$A_2(r) = \tfrac{1}{2}\left[1 + \tfrac{2}{3}\frac{d\ln\kappa}{d\ln r}\right] , \quad B_2(r) = 1 - A_2(r) , \tag{32}$$

and the phase constants s_0 and θ_0 are chosen so that $s = -\pi$, $\theta_* = -[\Omega(r)/\kappa(r)]\pi$ when $\varpi_* = \varpi$ (see eq. [14] and the discussion below).

To two orders in ϵ, equations (31) imply that s and $\kappa(r)\theta_*/\Omega(r)$ vary by 2π when ϕ varies by 2π; therefore, the radial period of oscillation and the azimuthal angle traversed in that time are given by

$$2\tau_{12} = \frac{2\pi}{\kappa(r)}, \quad 2\theta_{12} = \frac{\Omega(r)}{\kappa(r)} 2\pi . \tag{33}$$

It is convenient to express all quantities in terms of (ϖ, ξ, η, s), where ξ and η are defined as

$$\xi = \epsilon \sin s_0 , \quad \eta = \epsilon \cos s_0 . \tag{34}$$

In particular, by eliminating ϕ in equations (31), we may now manipulate the required orbit relations to read

$$\varpi_* = \varpi(1 - R_1 - R_2) , \quad \theta_* = \frac{\Omega(r)}{\kappa(r)} \left[s - 2\left(\frac{\partial R_1}{\partial s} + \xi\right)\right] , \tag{35}$$

where the functions R_1 and R_2 are given by

$$R_1 = \eta(1 + \cos s) + \xi \sin s , \quad R_2 = B_2(\varpi)R_1^2 - [1 + 2B_2(\varpi)]\eta R_1 . \tag{36}$$

ii) *Surface Density Response*

We choose to use the variables ξ and η in place of the momenta variables (p_ϖ, p_θ). The Jacobian of the transformation $(p_\varpi, p_\theta) \rightarrow (\xi, \eta)$, correct to two orders in ϵ, works out to be

$$\left|\frac{\partial(p_\varpi, p_\theta)}{\partial(\xi, \eta)}\right| = \frac{r^4\kappa^3(r)}{2\varpi\Omega(r)} . \tag{37}$$

In the right-hand side of the above and in what follows, we regard r as being given in terms of ϖ and η by

$$r = \varpi(1 - \eta) . \tag{38}$$

In the desired approximation the element of mass distribution of stars in the basic state may be characterized as

$$m_*\Psi_0 dp_\varpi dp_\theta = \frac{r^2}{\varpi} \sigma_*(r) \exp\left[-\frac{\xi^2 + \eta^2}{2\epsilon^2(r)}\right] \frac{d\xi d\eta}{2\pi\epsilon^2(r)} . \tag{39}$$

The range of integration of both ξ and η can be taken to extend from $-\infty$ to $+\infty$. In the process, errors are introduced which are only exponentially small.

The response in the surface density is obtained by multiplying ψ_1 in equation (12) by m_* and integrating over all velocities $(p_\varpi, p_\theta/\varpi)$. The amplitude of the perturbation surface density then reads

$$S_*(\varpi) = \int\int_{-\infty}^{\infty} \frac{d\xi d\eta}{2\pi\epsilon^2(r)} \frac{\sigma_*(r)}{\varpi^2\epsilon^2(r)\kappa^2(r)} \exp\left[-\frac{\xi^2 + \eta^2}{2\epsilon^2(r)}\right]\Bigg\{-V(\varpi) + \frac{\nu(r)\pi}{\sin \nu(r)\pi} \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\varpi_*)e^{i\nu(r)s}\left[1 + im\frac{2\Omega}{\kappa}\left(\frac{\partial R_1}{\partial s} + \xi\right)\right] ds\Bigg\} . \tag{40}$$

Particle resonance, arising from corotation and corresponding to $\nu(r) = 0$, does not occur in this approximation. Only for those stars whose angular momentum is such that

$$\nu(r) = \pm 1, \pm 2, \ldots$$

does resonance occur.[2]

[2] Kalnajs (1965) has given the physical interpretation of these resonances for axisymmetric disturbances in the absence of self-gravitation. Such "free modes" can oscillate only at the fundamental frequency $\omega = \pm\kappa$ and its overtones $\omega = \pm 2\kappa, \pm 3\kappa, \ldots$.

The existence of such resonances is also related to the possibility for Landau damping. It is easily shown that such damping for self-sustained density waves is exponentially small for all radii which are not within a fractional distance ε of a resonance ring $\varpi = \varpi_j$, where

$$\nu(\varpi_j) = j\,, \quad j = \pm 1, \pm 2\,, \ldots . \tag{41}$$

For $\varpi \neq \varpi_j$, the contribution in equation (40) at the (simple) pole $r = \varpi_j$ is proportional to the term

$$e^{-\eta^2/2\epsilon^2} = e^{-(\varpi-\varpi_j)^2/2\epsilon^2\varpi^2}\,, \tag{42}$$

a conclusion which is arrived at *independent* of the assumption of small wavelength scale (Q.E.D.).

For application to spiral structure, only the principal Lindblad resonances $\nu = \pm 1$ need be considered. In the immediate neighborhood of such Lindblad resonance rings, appreciable absorption of the density wave can occur (accompanied, presumably, by conversion of wave energy into random stellar motions). For the present analysis, we confine our attention to the range of ϖ which lie well within the two extreme values of ϖ given by $(\varpi_{-1}, \varpi_{+1})$. This region is referred to by Lin and Shu (1966) as the principal range; inside it the effects of resonances can be neglected.

Now suppose the functional form of $V(\varpi_*)$ is given by equation (2) and ϖ_* is given by equation (35). We expand the potential $V(\varpi_*)$ under the assumption that $|k|\varpi$ is a parameter of largeness comparable to ε^{-1}. In this approximation,

$$V(\varpi_*) = V(\varpi)e^{-ik\varpi R_1}\left[1 - R_1\frac{d \ln A}{d \ln \varpi} + ik\varpi\left(\frac{R_1^2}{2}\frac{d \ln k}{d \ln \varpi} - R_2\right)\right]. \tag{43}$$

We substitute equation (43) into equation (40) and expand about $r = \varpi$, using equation (38), to obtain

$$S_*(\varpi) = -\frac{\sigma_*(\varpi)V(\varpi)}{\varepsilon^2(\varpi)\varpi^2\kappa^2(\varpi)}\left\{1 - \frac{\nu(\varpi)\pi}{\sin \nu(\varpi)\pi}[\langle 1|g_\nu\rangle + \langle G|g_\nu\rangle]\right\}. \tag{44}$$

We have borrowed the Dirac bra-ket notation to mean the integral operation

$$\langle h|g_\nu\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} ds \iint_{-\infty}^{\infty} hg_\nu d\xi d\eta\,, \tag{45}$$

and we have written

$$g_\nu = \frac{e^{i\nu(\varpi)s - ik\varpi R_1}}{2\pi\varepsilon^2(\varpi)}\exp\left[-\frac{\xi^2+\eta^2}{2\varepsilon^2(\varpi)}\right] \tag{46}$$

with

$$\begin{aligned} G = &-\eta\frac{d \ln}{d \ln \varpi}\left(\frac{\sigma_*}{\varepsilon^2\kappa^2}\frac{\nu\pi}{\sin \nu\pi}g_\nu\right) - R_1\frac{d \ln A}{d \ln \varpi} + im\frac{2\Omega}{\kappa}\left(\frac{\partial R_1}{\partial s} + \xi\right) \\ &+ ik\varpi\left[-\eta R_1\frac{d \ln (k\varpi)}{d \ln \varpi} + \frac{R_1^2}{2}\frac{d \ln k}{d \ln \varpi} - R_2\right]. \end{aligned} \tag{47}$$

Since differentiation with respect to ln ϖ commutes with the bra-ket operation, we easily verify the relation

$$\begin{aligned} \langle G|g_\nu\rangle = &-\langle\eta|g_\nu\rangle\frac{d \ln}{d \ln \varpi}\left(\frac{\sigma_*}{\varepsilon^2\kappa^2}\frac{\nu\pi}{\sin \nu\pi}\langle\eta|g_\nu\rangle\right) - \langle R_1|g_\nu\rangle\frac{d \ln A}{d \ln \varpi} \\ &+ im\frac{2\Omega}{\kappa}\left[\left\langle\frac{\partial R_1}{\partial s}\middle|g_\nu\right\rangle + \langle\xi|g_\nu\rangle\right] \\ &+ ik\varpi\left[-\langle\eta R_1|g_\nu\rangle\frac{d \ln (k\varpi)}{d \ln \varpi} + \tfrac{1}{2}\langle R_1^2|g_\nu\rangle\frac{d \ln k}{d \ln \varpi} - \langle R_2|g_\nu\rangle\right]. \end{aligned} \tag{48}$$

The validity of the following formulae can be demonstrated by a straightforward calculation:

$$\langle 1 | g_\nu \rangle = \mathcal{G}_\nu(x) \,, \tag{49a}$$

$$\langle \eta | g_\nu \rangle = \tfrac{1}{2} \langle R_1 | g_\nu \rangle = \frac{i}{k\varpi} x \mathcal{G}'_\nu(x) \,, \tag{49b}$$

$$\langle \eta R_1 | g_\nu \rangle = -\langle R_2 | g_\nu \rangle = \tfrac{1}{2} \langle R_1^2 | g_\nu \rangle = -\frac{1}{(k\varpi)^2} [x\mathcal{G}'_\nu(x) + 2x^2 \mathcal{G}''_\nu(x)] \,, \tag{49c}$$

$$\langle \xi | g_\nu \rangle = -\left\langle \frac{\partial R_1}{\partial s} \middle| g_\nu \right\rangle = \frac{1}{k\varpi} \left[\frac{\sin \nu\pi}{\nu\pi} - \nu \mathcal{G}_\nu(x) \right] , \tag{49d}$$

where $\mathcal{G}_\nu(x)$ is given by the integral

$$\mathcal{G}_\nu(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \nu s \, e^{-x(1+\cos s)} ds \,, \tag{50}$$

and x is defined as

$$x = \varepsilon^2 k^2 \varpi^2 = \frac{k^2 \langle c_\varpi^2 \rangle}{\kappa^2} \,. \tag{51}$$

In equations (49), primes denote differentiation with respect to the argument x, with ν held constant. With the aid of equations (49), the sums of the terms in the square brackets of equation (48) are seen to be identically zero, whereas the first two terms may be combined to read

$$\langle G | g_\nu \rangle = -\frac{i}{k\varpi} x \mathcal{G}'_\nu(x) \frac{d \ln}{d \ln \varpi} \left[\frac{\sigma_*}{\kappa^2} k\varpi A^2 \frac{\nu\pi}{\sin \nu\pi} \mathcal{G}'_\nu(x) \right] . \tag{52}$$

Equation (44) can now be written

$$S_*(\varpi) = -\frac{k^2 V}{\kappa^2(1-\nu^2)} \sigma_* \mathcal{F}_\nu(x) \left\{ 1 - \frac{i}{k\varpi} \mathcal{D}_\nu(x) \frac{d \ln}{d \ln \varpi} \left(\frac{\sigma_*}{\kappa^2} k \frac{\mathcal{F}_\nu}{1-\nu^2} \mathcal{D}_\nu \varpi A^2 \right) \right\} , \tag{53}$$

where $\mathcal{F}_\nu(x)$ is the "reduction factor" found by Lin and Shu (1966):

$$\mathcal{F}_\nu(x) = \frac{1-\nu^2}{x} \left[1 - \frac{\nu\pi}{\sin \nu\pi} \mathcal{G}_\nu(x) \right] , \tag{54}$$

whereas $\mathcal{D}_\nu(x)$ is the function

$$\mathcal{D}_\nu(x) = -(1-\nu^2) \frac{\nu\pi}{\sin \nu\pi} \mathcal{G}'_\nu(x) \Big/ \mathcal{F}_\nu(x) = \frac{\partial \ln}{\partial \ln x} [x \mathcal{F}_\nu(x)] \,. \tag{55}$$

For values of ν of interest for the density-wave theory, $\mathcal{F}_\nu(x)$ can be regarded as a positive definite function for real ν^2. Thus, in the lowest asymptotic approximation for large $|k\varpi|$, the maxima of the surface-density response correspond with the minima of the gravitational potential only in the regions where $\nu^2 < 1$. This result, together with that of potential theory (see § II), means that spiral density waves can be self-sustained only within the principal range.[3]

IV. SELF-SUSTAINED SPIRAL DENSITY WAVES

The properties of self-consistent density waves will now be studied. Apart from the Jeans instability, we shall be primarily interested in neutrally stable waves. This proves

[3] In what follows, we use the term "self-sustained" to mean that the surface-density response is locally consistent with that required to support the spiral gravitational field. The terminology does not imply an ability of a wave packet to persist indefinitely in spite of the propagation of the group.

sufficient to regard ν^2 as being purely real. If we now equate the real and imaginary parts of the right-hand sides of equations (11) and (53), we obtain

$$\frac{k_T}{|k|}(1-\nu^2) = \mathfrak{F}_\nu(x)\,, \tag{56a}$$

$$\frac{1}{2}\frac{d\ln(\varpi A^2)}{d\ln\varpi} = \mathfrak{D}_\nu\frac{d\ln}{d\ln\varpi}\left[\frac{|k|}{k_T}\frac{\mathfrak{F}_\nu}{1-\nu^2}\mathfrak{D}_\nu\varpi A^2\right], \tag{56b}$$

where k_T is defined by

$$k_T = \frac{\kappa^2}{2\pi G\sigma_*}\,. \tag{57}$$

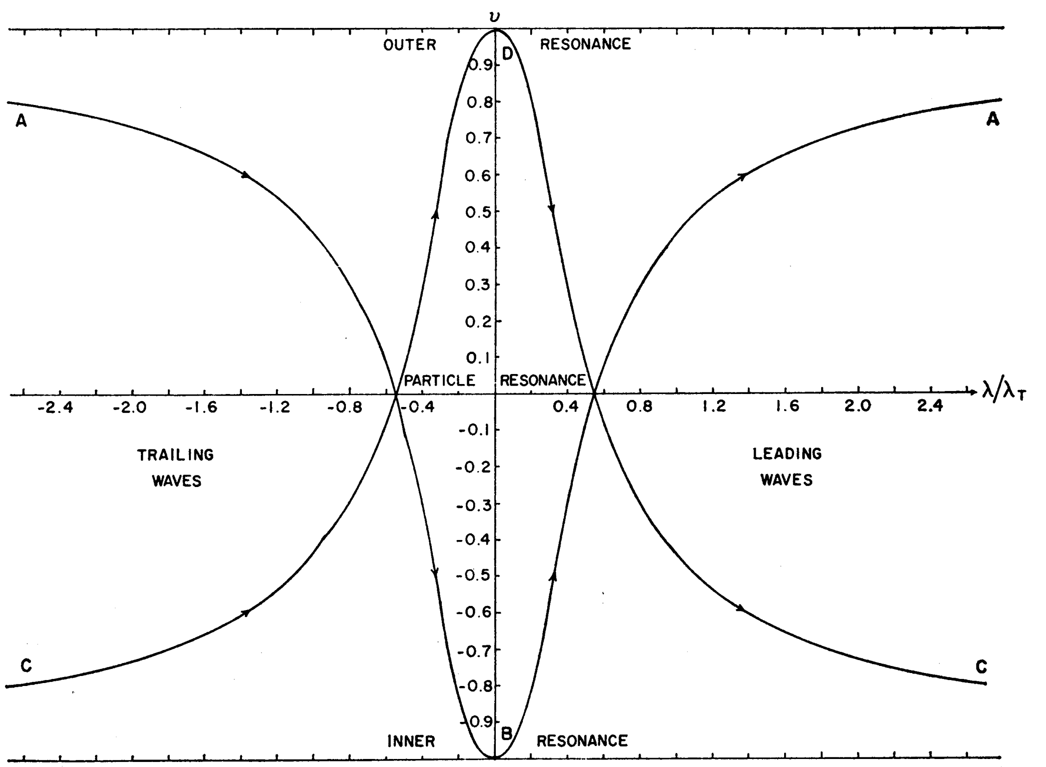

FIG. 1.—Dispersion relation for a marginally stable stellar disk. See text for explanation.

Equation (56a) is the dispersion relation reported by Lin and Shu (1966) (see also Lin 1966) connecting the local wavenumber k to the intrinsic dimensionless frequency ν. It is convenient to refer to the wavelength rather than to the wavenumber. We define the local wavelength λ and Toomre's scale length λ_T by

$$\lambda = \frac{2\pi}{k}\,, \quad \lambda_T = \frac{2\pi}{k_T} = \frac{4\pi^2 G\sigma_*}{\kappa^2}\,. \tag{58}$$

For convenience in the description of propagating waves, we have deviated from the normal practice by assigning to λ the same sign as k.

For a marginally stable disk (see § IV*b*), ν^2 can only be nonnegative, and the dispersion equation (56a) leads to the relationship between ν and λ/λ_T shown in Figure 1. If the usual convention of choosing ω and m nonnegative is followed, regions of algebraically increasing values of ν (generally) correspond to increasing values of ϖ. Negative values of λ/λ_T correspond to trailing spiral waves; positive, to leading ones. Because of the implicit assumption of short scale, the solution for the long waves ($|\lambda|/\lambda_T > 0.55$) can be regarded only as being suggestive.

The arrows in the diagram indicate the sense of propagation of wavenumber information given by Toomre's (1969) analysis. No wave can propagate outside the region where $-1 \leq \nu \leq 1$. However, without a detailed analysis of the WKBJ connection relations, we are not able to specify the behavior of a wave packet after its arrival at one of the Lindblad resonances, $\nu = \pm 1$. The possibility exists, in the linear theory, for reflection or absorption.[4]

Equation (56b) is the corrected version of the relation of Shu (1968) for the variation of the amplitude of the gravitational potential, after the removal of two algebraic errors detected by Toomre (private communication). Toomre suspected these errors

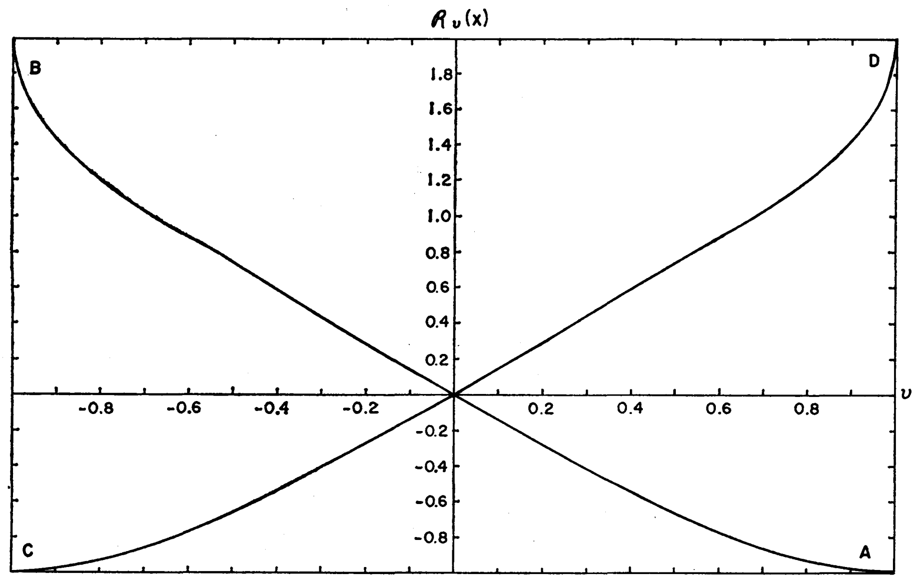

FIG. 2.—The function $\mathcal{R}_\nu(x)$. Labels A, B, C, D, correspond to those given in Fig. 1. Because $\mathcal{R}_\nu(x)$ is an even function of k, these curves show no distinction between leading and trailing waves.

because he felt that the amplitude relation should lead to a conservation relation involving the group velocity. This expectation was subsequently verified in his paper (Toomre 1969).

Equation (56a) may be used to simplify equation (56b) to read

$$\frac{d}{d\varpi}[\varpi A^2 \mathcal{R}_\nu(x)] = 0, \qquad \text{i.e.,} \quad \varpi A^2 \mathcal{R}_\nu(x) = \text{constant}, \tag{59a}$$

$$\mathcal{R}_\nu(x) = 1 - 2\mathcal{D}_\nu(x) = -\left\{1 + 2\frac{\partial \ln}{\partial \ln x}[\mathcal{F}_\nu(x)]\right\}. \tag{59b}$$

These relations follow directly from equations (56a) and (56b) independently of the assumption of marginal stability. Plotted in Figure 2 is the function $\mathcal{R}_\nu(x)$ when the relation between λ/λ_T and ν is assumed to be that given in Figure 1. Because $\mathcal{R}_\nu(x)$ changes sign as ν changes sign, equation (59a) would seem to imply that waves in the regions $\nu < 0$ are not well coupled to those in $\nu > 0$.

[4] Lin (1969) has suggested, however, that a (partial) conversion of inwardly propagating trailing waves of short scale into outwardly propagating trailing waves of long scale may also occur, through *nonlinear* effects, at the inner Lindblad resonance. Here, the amplitude of the surface density becomes infinitely large in the linear asymptotic theory.

The variation in the amplitude of the surface density (relative to the basic surface density) is obtained from equations (11) and (59a). In the lowest order of approximation,

$$\text{mod}\,[S_*(\varpi)]/\sigma_* = \text{constant} \times \varpi^{-1/2}(\lambda_T\kappa)^{-2}\,\mathcal{S}_\nu(x)\,, \tag{60a}$$

$$\mathcal{S}_\nu(x) = \frac{\lambda_T}{\lambda}[|\mathcal{R}_\nu(x)|]^{-1/2} = (1-\nu^2)/[\mathcal{F}_\nu(x)\sqrt{|\mathcal{R}_\nu(x)|}]\,. \tag{60b}$$

Plotted in Figure 3 is the function $\mathcal{S}_\nu(x)$ when the dispersion relationship is assumed to be given again by Figure 1. The large density amplitudes which develop near the particle resonance have already been attributed by Toomre (1969) to the gravitational "near-

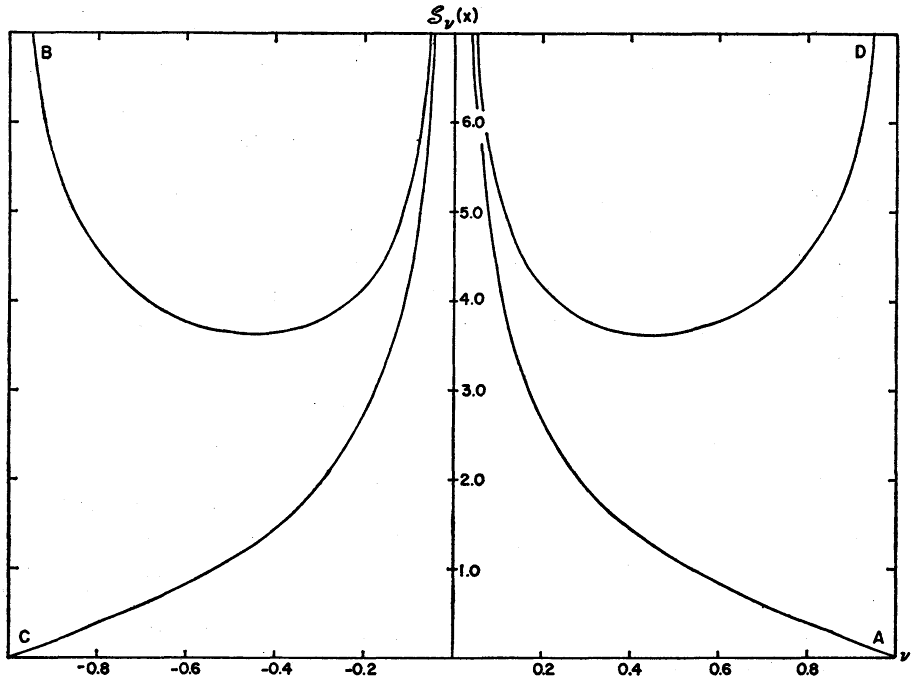

FIG. 3.—The function $\mathcal{S}_\nu(x)$. Labels *A*, *B*, *C*, *D* correspond to those given in Figs. 1 and 2.

instability" of a marginally stable disk. On the other hand, as is discussed below, the large density amplitudes which develop near the Lindblad resonances arise kinematically because of the propagation of a wave group.

a) The Propagation of the Density of Wave Action

Equation (59a) (or eq. [60a]) is remarkably concise; it would be surprising if no simple interpretation prevailed. Such an interpretation has been provided by Toomre (1969). For a monochromatic wave (ω = constant), Toomre showed that equation (59a) can be recovered from physical considerations provided the density of "wave action" $\mathcal{A}$ is propagated radially with the group velocity c_g:

$$\varpi c_g \mathcal{A} = \text{constant}\,. \tag{61}$$

The action density $\mathcal{A}$ is defined as the (time-averaged) energy density $\mathcal{H}$, reckoned by an observer moving with the wave group, divided by the intrinsic frequency of the wave $(\omega - m\Omega)$:

$$\mathcal{A} = \mathcal{H}/(\omega - m\Omega)\,. \tag{62}$$

Kalnajs (private communication) further clarified this interpretation by demonstrating that the energy density $\mathfrak{E}$ and angular momentum density $\mathfrak{J}$, referred to an inertial frame, can be expressed in terms of the action density $\mathfrak{A}$ as

$$\mathfrak{E} = \omega \mathfrak{A} , \quad \mathfrak{J} = m \mathfrak{A} . \tag{63}$$

We note that the relation between $\mathfrak{H}$ (energy density referred to a frame which rotates with the local angular velocity Ω) and $\mathfrak{E}$ and $\mathfrak{J}$ is

$$\mathfrak{H} = \mathfrak{E} - \Omega \mathfrak{J} . \tag{64}$$

This relationship, valid for a wave, is analagous to that, valid for particles, which gives Jacobi's integral in terms of the energy and angular momentum. Here, however, we refer explicitly to a *differentially rotatinq* disk.

The conservation of wave action expressed by equation (61), together with equations (63), now implies the conservation of wave energy and wave angular momentum referred to an inertial frame:

$$\varpi c_g \mathfrak{E} = \text{constant} , \quad \varpi c_g \mathfrak{J} = \text{constant} . \tag{65}$$

The behavior of the density amplitude near the principal Lindblad resonances can now be easily visualized. For short waves, the group velocity c_g tends to zero as the Lindblad resonances are approached; consequently, the energy density of the wave tends to "pile up" there. For long waves, the group velocity tends to a finite value as the Lindblad resonances are approached, but both the wavenumber and the surface density tend toward zero.[5]

b) *Stability of the Disturbance*

The conservation relations (65) imply that, within the approximation considered, spiral density waves are neither overstable nor damped. In particular, trailing and leading waves stand on an equal footing; only the direction of propagation differs. Lin and Shu (1966) based their argument for the preference of trailing waves on the variation of various basic parameters with distance from the galactic center, in particular, on the gradient of the velocity dispersion. Now, it is apparent that the growth (or decay) of the wave amplitude arises because the disturbance is propagated radially in an inhomogeneous medium and not because the disturbance is inherently overstable.

With this interpretation, we know of only one instability mechanism locally operative (in the plane) for a stellar disk whose distribution function is given by the "modified Schwarzschild distribution." This is the Jeans instability discussed for a rotating stellar disk by Toomre (1964). Toomre's result can be reproduced from the present analysis in the following manner.

Unstable disturbances with ν^2 negative in equation (56a) can be found if the local velocity dispersion $\langle c_\varpi^2 \rangle^{1/2}$ is less than the critical value (cf. Toomre 1964, eq. [65])

$$\langle c_\varpi^2 \rangle_{\min}{}^{1/2} = (0.2857)^{1/2} \kappa / k_T .$$

If $\langle c_\varpi^2 \rangle^{1/2}$ is less than $\langle c_\varpi^2 \rangle_{\min}{}^{1/2}$, the short-scale components of such growing disturbances are, in principle, capable of generating more random motion until the stability index

$$Q = \frac{\langle c_\varpi^2 \rangle^{1/2}}{\langle c_\varpi^2 \rangle_{\min}{}^{1/2}}$$

grows equal to (or slightly larger than) unity. At this point all further instabilities are suppressed.

[5] These conclusions are, of course, based on an asymptotic analysis which may need to be modified by a more careful analysis of the conditions near resonance.

Whether the Galaxy is everywhere more than marginally stable is a point of some debate. Julian (1967) is of the opinion that the enhancement by cooperative effects of the irregular forces provided by massive objects (on the order of 10^6–10^7 solar masses each) will inevitably drive Q to values substantially higher than unity. Observations in the plane of the Galaxy show only the "spiral arms" to possess large mass concentrations. In the density-wave theory, a more or less *regular* spiral structure does not lead to appreciable relaxation. Thus, Shu (1968, 1969) argued that the level of stellar velocity dispersion *in the interior of the disk* (excluding the nuclear regions, of course) is determined by gravitational instability alone and that the case of marginal stability $Q = 1$ should apply. *In the outer regions* where depletion of interstellar gas by star formation is yet relatively incomplete, the dissipation of turbulent velocities in the interstellar gas may lead to effective values for Q (for the combined star-gas disk) less than unity. In these regions may still occur the process of gravitational clumping of the interstellar gas described by Goldreich and Lynden-Bell (1965).

V. CONCLUDING REMARKS

Taken together with Toomre's (1969) work, the results of Paper I and the present paper suggest that the origin of spiral structure can, conceivably, be found in the forcing of the disk by some yet undetermined agency.

Lin (1969) has proposed that gaseous condensations produced by the process of "gravitational clumping," aided by the excitation of density "wakes" in the stellar sheet (Julian and Toomre 1966), may serve as a source for trailing spiral waves which propagate into the interior. Even though the forcing of the disk originates in the outer and more rarefied regions, an impression can be made on the interior because the wave energy density tends to "pile up" in the interior (where the group velocity is small). Furthermore, because the waves are initiated as nearly corotating disturbances in the outer parts of the galaxy, pattern speeds characteristic of the material rotation in the outer parts automatically result. This conclusion is consistent with the findings of Lin, Yuan, and Shu (1969) from comparisons of the theory with observations of our own Galaxy.

If the forcing comes from the central regions of the galaxy or if the forcing is external to the galaxy (possibilities which have been discussed by Toomre 1969), the preference for such values of Ω_p would be fortuitous. On the other hand, if the excitation of density waves occurs through corotating disturbances in the outer regions of the galaxy (say, around 15 kpc for our own Galaxy), the selection of such values is quite natural. A partial test of these ideas is possible with external galaxies with well-determined rotation curves and spiral patterns. The test is whether a pattern speed chosen to match the material rotation near the "outer edge" of the observed structure yields an accurate, theoretically deduced spiral pattern for the interior. At Stony Brook, we are in the process of making such determinations.

Once again, I am grateful to Professor C. C. Lin for his encouragement and advice on attacking the problems of spiral structure. Begun at the Harvard College Observatory and completed at Stony Brook, this work has evolved through many fruitful discussions with Dr. Alar Toomre and Dr. A. J. Kalnajs. This work was supported in part by the National Science Foundation.

REFERENCES

Goldreich, P., and Lynden-Bell, D. 1965, *M.N.R.A.S.*, **130**, 125.
Hunter, C. 1969, *Ap. J.*, **157**, 183.
Julian, W. H. 1967, *Ap. J.*, **148**, 175.
Julian, W. H., and Toomre, A. 1966, *Ap. J.*, **146**, 810.
Kalnajs, A. J. 1965, unpublished Ph.D. thesis, Harvard University.

Lin, C. C. 1966, *SIAM J. Appl. Math.*, **14**, 876.
———. 1969, unpublished note, presented in outline at I.A.U. Symp. No. 38 (Basel, Switzerland).
Lin, C.C., and Shu, F. H. 1964, *Ap. J.*, **140**, 646.
———. 1966, *Proc. Nat. Acad. Sci.*, **55**, 229.
Lin, C. C., Yuan, C., and Shu, F. H. 1969, *Ap. J.*, **155**, 721.
Shu, F. H. 1968, unpublished Ph.D. thesis, Harvard University.
———. 1969, *Ap. J.*, **158**, 505.
———. 1970, *Ap. J.*, **160**, 89 (Paper I).
Toomre, A. 1964, *Ap. J.*, **139**, 1217.
———. 1969, *Ap. J.*, **158**, 899.