ON THE GRAVITATIONAL STABILITY OF A DISK OF STARS

ALAR TOOMRE
Department of Mathematics, Massachusetts Institute of Technology
Received August 17, 1963; revised January 18, 1964

ABSTRACT

This paper considers the question of the large-scale gravitational stability of an arbitrary, highly flattened stellar system, which is assumed initially to rotate in approximate equilibrium between its self-gravitation and the centrifugal forces. It is concluded that no such disk, if fairly smooth or uniform, can be entirely stable against a *tendency* to form massive condensations within its own plane, unless the root-mean-square random velocities of its constituents, in the directions parallel to that plane, are everywhere sufficiently large. Lacking such random motions, it is shown that the system must be vulnerable to numerous unstable disturbances, the dimensions of which may approach its over-all radius, and whose times of growth are to be reckoned in fractions of the typical periods of revolution.

The minimum root-mean-square radial velocity dispersion required in any one vicinity for the complete suppression of all axisymmetric instabilities is calculated (in collaboration with A. Kalnajs) as $3.36\ G\mu/\kappa$, where G is the gravitational constant, and μ and κ are the local values of the projected stellar density and the epicyclic frequency, respectively. From that, and the observed μ and κ, together with their uncertainties, this minimum for the solar neighborhood of our Galaxy is estimated to fall between 20 and 35 km/sec, a range which indeed encompasses the actual radial velocity dispersions of the most predominant types of stars in our vicinity. It is pointed out that both this curious agreement, and also the well-known discrepancy between the z- and r-velocity dispersions at least of the older disk stars, may be explainable in terms of past instabilities of this galactic disk.

I. INTRODUCTION

The well-known instabilities of those Maclaurin spheroids whose rotational flattening exceeds a certain fairly moderate value suggest that other sufficiently flattened, rotating, and self-gravitating systems might in some sense likewise be unstable. At any rate, these instabilities have often been cited as a likely reason why one does not observe elliptical galaxies exceeding a certain degree of oblateness. It is only when we turn to consider what are now thought to be the distributions of all but the youngest stars in the disks of the ordinary (as opposed to the barred) spiral galaxies that this classical result suggests a serious dilemma: How is it conceivable, in spite of these or analogous instabilities, that so much of the fainter stellar matter within such galaxies—and certainly the S0 galaxies—should today appear distributed relatively evenly over disks with something like a ten-to-one flattening?

It is essentially this question which the present investigation will try to answer. In a broader sense, however, this paper forms only a part of a more comprehensive investigation of the large-scale stability of an *entire* galactic disk that was recently envisaged by C. C. Lin (1961) during a discussion with L. Woltjer. Motivated chiefly by a desire to understand to what extent gravitation might be responsible for the spiral phenomenon, Lin asked: What are the circumstances that would be needed for either one or both of the stellar and interstellar parts of a supposedly smooth galactic disk to remain gravitationally stable against all large-scale disturbances? Needless to say, a full answer to Lin's question would demand a simultaneous consideration of both the stellar and the gas dynamics. However, since stars are today believed to comprise the preponderant fractions of the total masses of most of the spiral galaxies, it seemed a legitimate first approximation to ignore the interstellar gas and dust, and to concentrate here on the stability of a thin disk composed only of stars. A discussion of the gravitational stability of a thin layer of gaseous material imbedded within an otherwise stable galaxy will be presented in a later paper (Toomre 1964).

The general character of the present stability problem can perhaps best be appreciated by recalling that no non-rigid, infinitesimally thin, plane sheet of gravitating matter—not even a perfectly uniform, infinite sheet—could long endure near its original state in the presence of the slightest disturbances if it lacked all stabilizing influences. One should therefore inquire: Which other mechanisms could conceivably be present in a rotating disk of stars that would help thwart this inherent tendency toward gravitational clumping? Upon reflection, it becomes evident that at most three such stabilizing effects need be considered. These are (*a*) the Coriolis or centrifugal forces stemming from the rotation, (*b*) the equivalent of a pressure resulting from the random motions of the individual mass elements, and (*c*) the fact that the magnitudes of the gravitational forces themselves would be less if the thickness of the disk were appreciable.

It shall indeed be seen from the simple order-of-magnitude arguments of the following section that no amount of rotation can *by itself* insure the complete stability of a thin, gravitating sheet. The same estimates will, however, also suggest that a combination of a basic rotation and of sufficiently large random velocities in the plane of the disk should be able to cope with all of the instabilities. These general conclusions will subsequently be verified in Sections III–V for several specific examples; those calculations will in addition provide numerical estimates of some of the quantities involved. Finally, in Section VI, we will discuss certain implications of these results.

II. AN ORDER-OF-MAGNITUDE DISCUSSION

a) Instability in the Presence of Rotation

To begin, let us consider an arbitrary thin disk, all parts of which initially revolve about a common center not necessarily with a uniform angular velocity, but without random motions and exactly in equilibrium between the gravitational and centrifugal forces. Now imagine that such a disk somehow suffers a slight contraction over a relatively small region or patch whose characteristic linear dimension is L.

To determine whether the excess gravitational attraction toward such a disturbed region, which by itself would act to increase that local mass concentration, could be overcome solely by Coriolis or centrifugal forces, let us suppose that this local shrinkage has increased the density, μ, per unit surface area in that neighborhood by a fraction ϵ. In that case, a typical material element previously located near the periphery of this region would now find itself roughly an increment of distance ϵL closer to the involved mass of the order μL^2. For that reason, it should feel an additional gravitational force of the order

$$G\mu L^2[(L-\epsilon L)^{-2} - L^{-2}] \approx \epsilon G\mu \tag{1}$$

per unit mass, which would tend to pull it deeper into the affected region. Here G is the gravitational constant.

On the other hand, owing to the tendency to conserve even the detailed angular momentum (about the center of the disturbance), the local or intrinsic angular velocity, say Ω_{local}, of the matter in the contracted region should also have increased by approximately the fraction ϵ. Therefore, as would be judged by an observer moving with the center of the affected region, a typical material element should at the same time experience an increased centrifugal force per unit mass that is roughly equal to the change in the product of its distance, $d = O(L)$, from the center of this region, and of Ω_{local}^2, or

$$\Delta(\Omega_{\text{local}}^2\, d) = \Delta(\Omega_{\text{local}}^2\, d^4/d^3) \approx \Omega_{\text{local}}^2\, d^4\, \Delta(d^{-3}) \approx \epsilon L \Omega_{\text{local}}^2\,. \tag{2}$$

This force would in general try to fling the particular element back out from that region.

It readily follows that the latter force could not be expected to overcome the former imbalance if L were sufficiently shorter than a certain critical length

$$L_{\text{crit}} \approx G\mu/\Omega_{\text{local}}^2\,. \tag{3}$$

Thus, it is the disturbances of sufficiently *short* dimensions which must in general remain unstable in spite of the rotation. This conclusion is akin to one that was arrived at already by Maxwell (1857) during his study of the stability of Saturn's rings; in connection with galaxies, it has recently been explicitly mentioned also by Mestel (1963). It is obviously the opposite of the result found for the Jeans instability proper, but it should be remembered that in the latter case the stabilizing influence consists of pressure forces and not a rotation.

More remarkable than that, however, is the actual magnitude of the length L_{crit}. Clearly, Ω_{local} will in most circumstances be comparable to the angular velocity, Ω, of revolution around the galactic center, although owing to differential rotation the two will not necessarily be equal. Moreover, the latter velocity must depend on the mean surface density, μ_{mean}, of the entire galaxy and on the galactic radius, R, approximately through

$$\Omega^2 R \approx (G\mu_{\text{mean}} R^2)/R^2 = G\mu_{\text{mean}} . \tag{4}$$

Consequently, the critical length L_{crit} from equation (3) may be re-expressed as

$$L_{\text{crit}} \approx (\mu/\mu_{\text{mean}})\,(\Omega/\Omega_{\text{local}})^2\,R\,, \tag{5}$$

thus indicating that whenever the densities μ and μ_{mean} are comparable, and all stabilizing influences other than rotation are absent, the length scale which divides the unstable disturbances from those that are stable is apparently of the same order of magnitude as the galactic radius!

b) Instability in the Presence of Random Motions

One can envisage the role of the virtually collisionless random movements of the stars relative to one another in at least two ways: These random motions can either be thought of as a diffusive mechanism, or else as the equivalent of a pressure. Either way, there seems little doubt that these motions will on the whole tend to suppress, rather than encourage, any given instability.

If we adopt the diffusive interpretation, an obvious criterion for the prevention of an earlier instability by the random motions appears to be that the average mass elements should manage to travel relative to each other at least through a sizable fraction of the linear scale of the perturbed region during the time in which the disturbance amplitude would otherwise have grown, say, by a factor e. Now, for a sheet that is not already in rotation, such a characteristic time of growth in the absence also of random motions can be estimated to be of the order of $(L/G\mu)^{1/2}$. Hence, writing the mean-square random velocity as $\langle u^2\rangle$, we would expect stability only when, approximately,

$$(L/G\mu)^{1/2} > \langle u^2\rangle^{-1/2}\,L\,. \tag{6}$$

Hence, a given velocity dispersion in a non-rotating sheet consisting of encounter-free particles ought to stabilize only those disturbances whose typical dimensions are sufficiently smaller than a second critical length

$$L_J \approx \langle u^2\rangle/G\mu\,. \tag{7}$$

This conclusion, of course, is very analogous (but owing to the different geometry, not entirely equivalent) to the Jeans result.

c) Combined Effect of Random Motions and Rotation

From the above two conclusions we deduce that, under the joint influence of a rotation and a moderate superposed velocity dispersion, the unstable disturbances ought to be confined to a certain intermediate range of scales or "wavelengths." What is more, it appears unavoidable that this range will shrink as the velocity dispersion is increased. Indeed, beyond some finite value of the velocity dispersion, it seems practically certain that these severe instabilities (though perhaps not some slower secular changes) will be avoided altogether. This minimum velocity dispersion for complete stability should roughly correspond to $L_{\text{crit}} = L_J$; hence we estimate it to be something like

$$\langle u^2 \rangle_{\min}^{1/2} \approx G\mu/\Omega_{\text{local}} \approx \Omega R(\mu/\mu_{\text{mean}})\,(\Omega/\Omega_{\text{local}})\,. \tag{8}$$

This evidently is of the same order of magnitude as the typical (linear) velocity of revolution!

III. LOCAL CALCULATIONS INVOLVING ONLY ROTATION

To check on the above rather surprising indications, we shall now embark on several more detailed calculations. The present section will describe some "local" or mathematically small-scale analyses of infinitesimal disturbances to the equilibrium state of an infinitely thin, rotating disk without random motions.

a) Notation and Equations

We introduce non-rotating cylindrical polar coordinates r, θ, z, such that the plane $z = 0$ coincides with that of the disk, the line $r = 0$ with its axis, and the increasing angle θ with the direction of rotation. The undisturbed surface mass density will be denoted as $\mu(r)$, and the corresponding linear velocity of revolution will be $V(r) = r\Omega(r)$, where, as before, $\Omega(r)$ is the undisturbed angular velocity. Furthermore, we shall make use of Oort's "constants"

$$A(r) = \tfrac{1}{2}(\Omega - dV/dr) \qquad \text{and} \qquad B(r) = -\tfrac{1}{2}(\Omega + dV/dr)\,, \tag{9}$$

and will express the radial and circumferential *disturbance* velocity components at a given location as $u'(r, \theta, t)$ and $v'(r, \theta, t)$, and the surface density and the gravitational potential in excess of their unperturbed values as $\mu'(r, \theta, t)$ and $\phi'(r, z, \theta, t)$, respectively.

The following four linearized equations, then, must govern any infinitesimal disturbance which involves no motion in the z-direction:

$$\frac{\partial u'}{\partial t} + \Omega(r)\frac{\partial u'}{\partial \theta} - 2\Omega(r)\,v' = \left.\frac{\partial \phi'}{\partial r}\right|_{z=0}, \tag{10}$$

$$\frac{\partial v'}{\partial t} + \Omega(r)\frac{\partial v'}{\partial \theta} - 2B(r)\,u' = \frac{1}{r}\left.\frac{\partial \phi'}{\partial \theta}\right|_{z=0}, \tag{11}$$

$$\frac{\partial \mu'}{\partial t} + \Omega(r)\frac{\partial \mu'}{\partial \theta} + \frac{1}{r}\frac{\partial}{\partial r}[r\mu(r)u'] + \frac{\mu(r)}{r}\frac{\partial v'}{\partial \theta} = 0\,, \tag{12}$$

and

$$\nabla^2\phi' = -4\pi G\mu'\delta(z)\,. \tag{13}$$

In the last (or Poisson) equation, the symbol $\delta(z)$ denotes the Dirac delta function, and ∇^2 is the *three-dimensional* Laplace operator; the gravitational force per unit mass has here been defined as the positive gradient of ϕ'.

b) Short Axisymmetric Disturbances

For an arbitrary model galaxy these four equations, together with appropriate boundary conditions to be applied at the center and toward large radii, constitute an eigenvalue problem which at present appears too formidable to be solved without recourse to numerical analysis. Nevertheless, some insight into the likely character of their solutions may be had by assuming that a particular oscillatory and axisymmetric eigen-disturbance might be approximated as

$$\left.\begin{aligned} u'(r,t) &= C_1 \\ v'(r,t) &= C_2 \\ \mu'(r,t) &= C_3 \end{aligned}\right\} e^{iar}\, e^{i\omega t} \tag{14}$$

near some radius $r = r_0$, with the C_i being (small) complex constants and with the local "wavelength," $2\pi/a$, being postulated to be much shorter than r_0.

In that case, the corresponding gravitational disturbance potential ϕ' may at once be estimated as follows: Equation (13) shows that if the position on the disk were described by the Cartesian coordinates x and y, instead of r and θ, then a density

$$\mu'(x) = C_3\, e^{iax} \tag{15}$$

would exactly be associated with a potential

$$\phi'(x, z) = C_3\, (2\pi G/a)\, e^{iax}\, e^{-a|z|}\,, \tag{16}$$

where a is to be understood to be positive. Now, having assumed that $ar_0 \gg 1$, we see that the density of equation (14), within a few wavelengths of a point on or near the circle $r = r_0$, must closely resemble that of equation (15). Moreover, since both densities rapidly alternate in sign with changing location, it is also evident that the disturbance gravitational potential must (because of cancellation of phase from larger distances) in either case be determined almost wholly by the disturbance density within an equally small neighborhood of the point in question. Consequently, the radial gravitational force in the present situation must also be approximately

$$(\partial\phi'/\partial r)_{z=0} \cong i\, C_3\, 2\pi G\, e^{iar}\, e^{i\omega t}\,. \tag{17}$$

It now remains only to introduce equations (14) and (17) into equations (10)–(12), and to approximate terms like $r^{-1}\,\partial(r\mu u')/\partial r$ appearing in equation (12) by $\mu(r_0)\partial u'/\partial r$ on the grounds that $r\mu(r)$, for instance, would only change gradually with r compared with the rapidly oscillatory behavior of $u'(r, t)$. From that, it directly follows that the assumed oscillation frequency ω and the local wavenumber a must be related approximately through

$$2\pi G\mu(r_0)a = \kappa^2(r_0) - \omega^2\,, \tag{18}$$

where

$$\kappa^2(r_0) = -4\, B(r_0)\Omega(r_0)\,. \tag{19}$$

In addition, the relative amplitudes are determined as

$$C_1 = (i\omega/2B)C_2 = -(\omega/\mu a)C_3\,. \tag{20}$$

(To be complete, it should be mentioned that the same equations admit a solution also for $\omega = 0$. However, since C_1 vanishes for that solution, meaning that such a steady additional motion is wholly tangential, it is perhaps better to regard that mode not as a disturbance per se but simply as an alternative equilibrium state.)

It is reassuring to note from equations (18) and (19) that a typical particle would gyrate about its mean position or epicenter with very nearly the usual epicyclic frequency, κ, if either the undisturbed surface density, μ, happened to be small or the wavelength very large. More important, however, is that equation (18) generally shows the frequency of oscillation to be decreased through the gravitational interaction between the various elements partaking in the disturbance. Indeed, it implies that disturbances of this axisymmetric and approximately sinusoidal type should be unstable if their local wavenumbers, α, exceeded a certain critical value

$$\alpha_{\text{crit}}(r_0) = \kappa^2/2\pi G\mu \,, \tag{21}$$

or when their approximate wavelengths, λ, as reckoned in the vicinity of $r = r_0$, were less than

$$\lambda_{\text{crit}}(r_0) = 2\pi/\alpha_{\text{crit}} = 4\pi^2 G\mu/\kappa^2 \,. \tag{22}$$

In addition, it is also to be seen from that equation that the characteristic time in which the amplitudes of such unstable disturbances ought to increase by a factor e should be approximately

$$\tau = \kappa^{-1}[(\lambda_{\text{crit}}/\lambda) - 1]^{-1/2} \,. \tag{23}$$

For disturbances of one-half the critical wavelength, for instance, this interval should amount to between about one-twelfth and one-sixth of the period of revolution.

c) Critical Wavelengths for Two Galaxy Models

Figure 1 shows as functions of the radius the magnitudes of the critical wavelengths, λ_{crit}, computed from equation (22) for two of the disklike model galaxies that were devised by the author in another paper (Toomre 1963). These are "Model 2," which had a rotation law

$$V(r) = \Omega_c r(1 + \tfrac{1}{4}r^2/a^2)^{1/2} (1 + r^2/a^2)^{-5/4} \tag{24}$$

and an exactly corresponding surface density

$$\mu(r) = (3\Omega_c^2 a/8\pi G) (1 + r^2/a^2)^{-5/2} \tag{25}$$

(where a denotes a characteristic length and Ω_c the central angular velocity), and the "Gaussian Model," for the particulars of which we refer the reader to equations (31) and (32) of the other paper.

Strictly, the large estimates for λ_{crit} shown in Figure 1 can only be regarded as *extrapolations* from the local theory, inasmuch as they so flagrantly conflict with the small wavelength assumption that was one of the bases for their derivation. Nevertheless, they clearly support the tentative conclusion of Section II*a* by showing that even the longest of those axisymmetric disturbances which one might be willing to regard as "local" are distinctly unstable and grow exponentially with time (except at those large radii at which the surface density has become so small that the mutual interaction has become almost negligible).

d) Short Non-axisymmetric Disturbances

At first, it might be supposed that the non-axisymmetric instabilities of a self-gravitating disk should, if anything, be even more pronounced than the axisymmetric ones, since intuitively it seems likely that the displacement of any given element in longitude should be a simpler task than changing its orbital radius. That such reasoning does not correctly account for the dynamical constraints of the Coriolis forces, however, is indicated by the fact that the above local analysis could easily be extended to a non-axisymmetric disturbance of the form

$$u'(r, \theta, t) = C_1\, e^{i\alpha r} \exp\{i\beta r_0[\theta - \Omega(r_0)t]\}\, e^{i\omega t} \,, \tag{26}$$

if the differential rotation (or the Oort constant A) in a galaxy happened to be negligible. In such a case, and again postulating that one was dealing with relatively short wavelengths, it would be found that

$$2\pi G\mu(r_0)\, (\alpha^2 + \beta^2)^{1/2} = \kappa^2(r_0) - \omega^2 \,, \tag{27}$$

a result which is very analogous to equation (18). Consequently, we conclude that whatever differences there may exist between the shorter axisymmetric and non-axisymmetric disturbances, these must in essence be due only to the circumstance of *differential* rotation.

As regards the general case where the differential rotation is not insignificant, it has been pointed out by Lin (1964) that, formally at least, equations (10)–(13) always admit solutions of the type $u'(r, \theta, t) = U(r) \exp(im\theta) \exp(i\omega t)$, etc. This is somewhat remarkable, for it means that there exist non-axisymmetric solutions which are not sheared at all by the differential rotation. It is interesting too that, from a calculation of some short-wavelength asymptotic forms of those solutions, Lin also obtained a necessary relation between their local wavenumbers and the assumed frequency that is not very different from our equation (18) or (27).

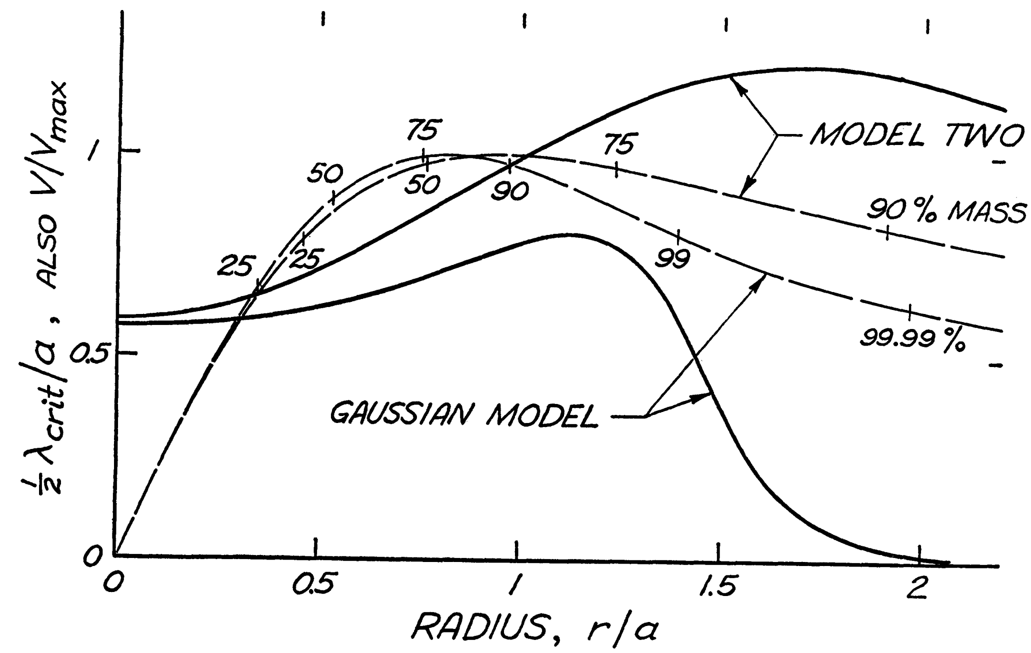

FIG. 1.—Locally estimated critical wavelengths, λ_{crit}, for two galaxy models, as functions of the radius (and divided by 2 for plotting purposes). The broken curves describe the rotation laws for these models, and the vertical marks identify the radii which contain the indicated percentages of their respective total masses.

That the axisymmetric results are not anomalous is further indicated by the following, altogether different family of approximate non-axisymmetric solutions to the governing equations. These disturbances, behaving as

$$\left.\begin{aligned} u'(r, \theta, t) &= S_1(t) \\ v'(r, \theta, t) &= S_2(t) \\ \mu'(r, \theta, t) &= iS_3(t) \end{aligned}\right\} \exp\{i\beta r_0[\theta - \Omega(r)t]\} \tag{28}$$

near some radius $r = r_0$, are indeed of a type which become distorted to an increasing extent by the differential rotation (although, in principle, they should probably be regarded as particular superpositions of Lin's solutions). Their "wrapping up" is made especially evident by the fact that the above phase function may be expanded as

$$\beta r_0[\theta - \Omega(r)t] \cong \beta r_0[\theta - \Omega(r_0)t] + 2\beta A(r_0)t(r - r_0)\,, \tag{29}$$

which also indicates $2\beta At$ to be the instantaneous radial wavenumber.

By again making only such approximations which can strictly be justified in the limit of vanishing wavelength, it can be shown from equations (10)–(13) that the functions $S_i(t)$ must approximately obey the linear equations

$$\frac{dS_1}{dt} = 2\Omega(r_0)S_2 + 2\pi G S_3 \sin\gamma\,,$$

$$\frac{dS_2}{dt} = 2B(r_0)S_1 + 2\pi G S_3 \cos\gamma\,, \tag{30}$$

and

$$\frac{dS_3}{dt} = \mu(r_0)\beta(S_1 \tan\gamma + S_2)\,,$$

where

$$\tan\pi(t) = 2A(r_0)t \tag{31}$$

is the tangent of the instantaneous angle between the crest of these waves and the local radial direction. These coupled differential equations have been integrated numerically

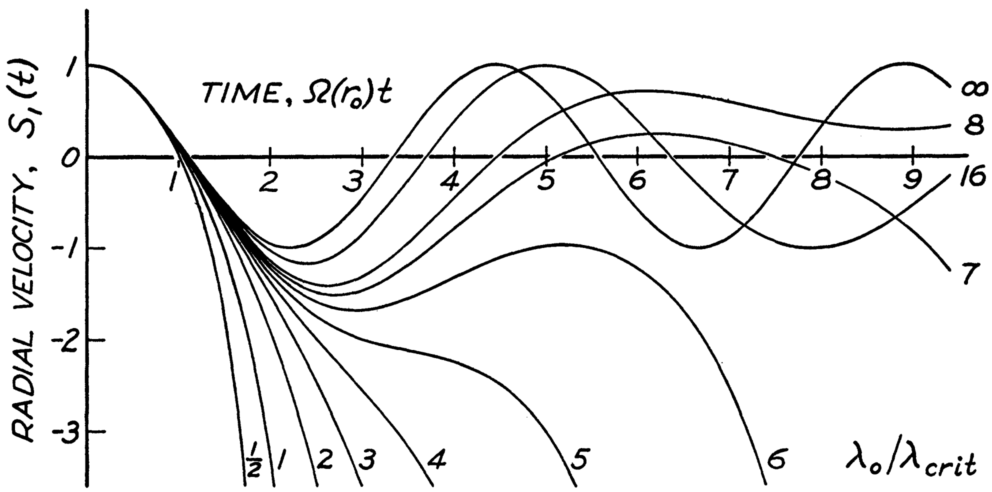

FIG. 2.—The time development of the radial velocities associated with certain non-axisymmetric disturbances of Section III*d*.

for an example where $dV/dr = 0$ [or $A(r_0) = -B(r_0) = \frac{1}{2}\Omega(r_0)$, $\kappa^2(r_0) = 2\Omega^2(r_0)$, and $\tan\gamma(t) = \Omega(r_0)t$], and for the initial conditions: $S_1 \neq 0$, $S_2 = S_3 = 0$ at $t = 0$. Figure 2 shows the subsequent development of the radial velocity, $S_1(t)$, plotted to an arbitrary scale as a function of time. The various curves in that diagram refer to different ratios of the initial wavelength, $\lambda_0 = 2\pi/\beta$, to the critical wavelength defined in equation (22), λ_{crit}.

We observe from Figure 2 that when this ratio is small, an unbounded growth in amplitude ensues almost immediately, just as it would in the axisymmetric case. (That the initial growth is not even more rapid for the shortest disturbances is explained by the fact that both the initial disturbance density and its time derivative happened to be chosen exactly zero for this particular example.) On the other hand, when $\lambda_0/\lambda_{\text{crit}}$ is somewhat larger than unity, the disturbance at first tends to remain small or even oscillatory and "blows up" only later, once the wavelength has been decreased to a value less than λ_{crit} by the shear which accompanies the differential rotation; note, for instance, the curve for $\lambda_0/\lambda_{\text{crit}} = 6$.

e) *The Effect of a Finite Disk Thickness*

The preceding analyses require only one significant modification if they are to apply to a stellar disk whose thickness, though still small compared with the wavelengths in question, is not entirely negligible. This change has to do with the fact that, whereas the Coriolis forces and the excess "dynamical pressures" of the random motions associated with any given amplitude of disturbance might be expected to remain the same as before (for the reason that the z-motions of the stars, to a first approximation, would not be coupled to the motions in the plane), the typical disturbance gravitational force should be smaller if the matter were not concentrated into a single plane.

To estimate the amount of that decrease, let us assume for simplicity that the stars are only to be found between the planes $z = \pm h$, and that the disturbance mass, ρ', per unit volume within this layer is neither a function of the θ- nor the z-coordinate. Specifically, let

$$\rho'(r, t) \cong (C_3/2h)\, e^{iar}\, e^{i\omega t} \qquad \text{when} \qquad |z| < h\,. \tag{32}$$

Now, the radial component of the force at $z = 0$ due to a thin stratum $(z, z + dz)$ with this density ρ' may at once be judged from equation (16) to be proportional to $\exp\{-\alpha|z|\}dz$. Therefore, the disturbance force, say, in the central plane, will in this case be just that of equation (17) multiplied by

$$F(ah) = (1 - e^{-ah})/ah\,. \tag{33}$$

Typically, this factor implies about a 26 per cent reduction in the gravity force in the case where the wavelength of the disturbance, $2\pi/\alpha$, equals 5 times the thickness, $2h$, and about a 14 per cent change when it amounts to 10 times the thickness. Either (wavelength-dependent) reduction might simply be thought of as a decrease in the effective surface mass density.

IV. LARGE-SCALE AXISYMMETRIC DISTURBANCES

This section presents some numerically computed examples of genuinely large-scale axisymmetric eigen-disturbances. These pertain to the fairly realistic equilibrium configuration that has already been referred to as "Model 2" and described in equations (24) and (25), as well as in Figure 1.

a) *Theoretical Considerations*

The following computations were based, for simplicity, on the assumption that the material in Model 2 could be regarded as divided equally among N concentric and coplanar rings of negligible cross-sectional dimensions, each one revolving about the axis of symmetry with a velocity just sufficient to maintain it in equilibrium against the combined gravitational attraction of all of the rings. The undisturbed radii of the individual rings were chosen so as to approximate the mass distribution in Model 2 as closely as possible; this was done by locating the typical or ith ring from the center at an undisturbed radius, r_i, just equal to that which would have encompassed exactly a fraction $(i - \frac{1}{2})/N$ of the total mass, M, of the said model galaxy.

The motive behind this physical approximation was that at least certain of the N strictly axisymmetric and infinitesimal eigen-modes of such a conservative mechanical system, of the type

$$\delta r_i(t) = X_i\, e^{i\omega t}\,, \tag{34}$$

should adequately approximate the less detailed corresponding linear eigen-modes of the continuous model, and that this would avoid dealing with a rather complicated integro-differential equation which would have arisen in a more direct attempt to solve equations (10)–(13). As it turned out, this hope was largely realized—and the continuous problem in essence solved—through the use of an electronic computer, which made it possible to consider radial perturbations to as many as eighty of these rings simultaneously.

It was assumed that each ring preserved its individual angular momentum during a disturbance. Consequently, using $F_{ij}(r_i, r_j)$ to denote the outward-directed gravitational force exerted by the entire jth ring at the radius r_j upon a unit mass of the ith ring at r_i, and letting Ω_i be the unperturbed angular velocity of the ith ring, it was determined that the excess gravitational force on the ith ring at any given instant would be

$$e^{i\omega t}\left[X_i \sum_{j=1}^{N} \frac{\partial F_{ij}}{\partial r_i} + \sum_{j\neq i} X_j \frac{\partial F_{ij}}{\partial r_j}\right],$$

whereas the centrifugal force would have been decreased by

$$3\Omega_i^2 X_i e^{i\omega t} .$$

Balancing the sums of these forces against the radial accelerations, the entire dynamics was summed up in the N simultaneous linear algebraic equations,

$$\sum_{j=1}^{N} A_{ij} X_j = 0 \qquad \text{for all } i , \tag{35}$$

where the coefficients A_{ij} were defined as $\partial F_{ij}/\partial r_j$ when $i \neq j$, and otherwise as

$$\left(\omega^2 + 3\Omega_i^2 + \sum_{j=1}^{N} \frac{\partial F_{ij}}{\partial r_i}\right).$$

In the case when $i \neq j$, a straightforward calculation then revealed that

$$F_{ij} = (2\pi)^{-1} \frac{GM}{N} \frac{\partial}{\partial r_i} \int_0^{2\pi} (r_i^2 + r_j^2 - 2 r_i r_j \cos\theta)^{-1/2} d\theta$$

$$= \begin{cases} \dfrac{2}{\pi}\dfrac{GM}{N}\dfrac{\partial}{\partial r_i}\left[\dfrac{1}{r_i} K\left(\dfrac{r_j}{r_i}\right)\right] & \text{when } r_i > r_j \\ \dfrac{2}{\pi}\dfrac{GM}{N}\dfrac{\partial}{\partial r_i}\left[\dfrac{1}{r_j} K\left(\dfrac{r_i}{r_j}\right)\right] & \text{when } r_i < r_j , \end{cases} \tag{36}$$

where $K(k)$ denotes the first complete elliptic integral, expressible in terms of the hypergeometric function as

$$K(k) = (\pi/2)\, F(\tfrac{1}{2}, \tfrac{1}{2}; 1; k^2) . \tag{37}$$

Note that $\partial F_{ij}/\partial r_j = \partial F_{ji}/\partial r_i$. This meant that $A_{ij} = A_{ji}$, which in turn implied, thanks to a well-known theorem from matrix theory, that the eigenvalues, ω^2, of the set of equations (35) were all necessarily real. In the present context, this guaranteed that there could be no overstability among the radial modes.

The assumption that the cross-section of each ring was infinitesimal proved embarrassing only as regards the attraction of any given ring upon a unit mass of itself. Similarity considerations obviously demanded that this force be of the form

$$F_{ii} = -C_s (GM/N) r_i^{-2} , \tag{38}$$

but, unfortunately, a direct application of equation (36) showed the coefficient C_s to be infinite! This relatively artificial difficulty was overcome, however, by supposing every mass element within a given ring to be acted upon only by the matter located beyond some small angular distance or "cutoff angle," θ_c, from itself; such an approximation was found to imply that

$$C_s = \pi^{-1} \ln \cot (\theta_c/4) . \tag{39}$$

A number of different physically plausible values of θ_c were tried in the course of these calculations; but not surprisingly, in view of the logarithmic character of equation (39) and the fact that except for a few of the innermost rings the self-attraction could not reasonably be expected to account for more than a small fraction of the total force experienced by any one ring, the results were found to be quite insensitive to the exact choice of that cutoff angle. For the examples which follow, the values of C_s (or θ_c) for the different rings were assigned specifically through the requirement that the *total* unperturbed gravitational force on each ring give rise to a rotation velocity identical to that found at the same radius r_i in Model 2.

b) Computed Results

Some idea of the rates of convergence of the results of this discrete analysis with increasing numbers of rings to what might be presumed to be the corresponding features

TABLE 1

SQUARES OF THE EIGEN-FREQUENCIES, ω^2

$N=2$	$N=3$	$N=4$	$N=5$	$N=20$	$N=80$
+0 5252	+0 4534	+0 4219	+0 4062	+0 3719	+0 3659
+0 1124	+ 06778	+ 04801	+ 04043	+ 02780	+ 02423
	+0 01995	+ 01413	+ 01736	+ 00620	+ 00578
		−0 4060	− 3036	− 00243	+ 00355
			−0 9194	−0 05103	+0 00154

of the continuous model may be had from Table 1, which shows the five (algebraically) largest characteristic values ω^2 of equations (35) that were computed for six different numbers of rings, N. The units in Table 1 were chosen to make the angular velocity at the center exactly unity.

Figure 3 portrays the relative amplitudes, X_i, corresponding to the three most rapid, and hence most reliably calculated, oscillatory eigen-modes of the disk composed of 80 rings. The components of those eigen-vectors have there been plotted to an arbitrary vertical scale as functions of the undisturbed radii of the various rings, although some of the points have been omitted to prevent overcrowding. The five additional solid dots in the left half of the diagram indicate the amplitudes similarly obtained for this basic mode in the case where the total number of rings was only *five*.

By definition, every element of the disk would, in the absence of any mutual interactions, only be willing to oscillate about its mean orbital radius, r_0, with a frequency equal to its epicyclic frequency, $\kappa(r_0)$. In such a case, one could indeed conceive of very many distinct oscillatory axisymmetric eigen-disturbances, for each of which the amplitude would everywhere be zero except for a "spike" at the one appropriate radius. What we observe in Figure 3 is evidently that the interaction between the various displaced components of the disk modifies that picture in just two ways: it broadens those spikes, and it restricts the possible frequencies to discrete values. In this connection, note especially the three arrows in that diagram. These point to the radii at which the local epicyclic frequency just matches one of the three eigen-frequencies, ω.

Another feature of Figure 3 which also deserves to be explained in general terms is the mildly oscillatory behavior with the radius of two of the curves inside the radii corresponding to the maxima, as opposed to the purely monotonic decrease of the amplitudes outward from the same peaks. This explanation is based on equation (18), which indicated that at least for any short, sinusoidal disturbances the effect of the gravitational interaction was generally to *decrease* the frequency of oscillation of a typical element. It seems improbable that the same would not also be true qualitatively of those parts of any other disturbance in which the amplitude alternates in sign with varying radius. Coupling that belief with the knowledge that the epicyclic frequency for our Model 2 (as for most other disks) is a monotonically decreasing function of the radius, one can readily appreciate how material inward of the radius where $\kappa(r)$ equals ω could be persuaded to

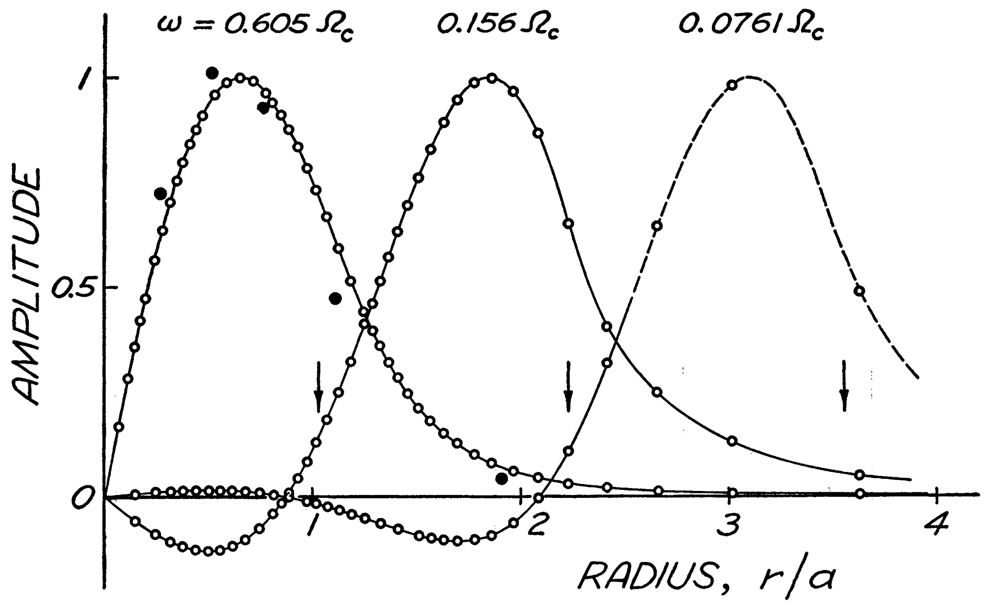

FIG. 3.—The relative amplitudes of three oscillatory eigen-disturbances to Model 2, as functions of the radius. For an explanation of the arrows and the solid dots see text.

oscillate with less than its preferred frequency, and also how, in the absence of any obvious mechanism for *increasing* the frequency of oscillation, the disturbance amplitude would have to trail off to zero quite rapidly beyond that critical radius.

The most important results of this numerical analysis, however, are contained in Figure 4 in the form of four examples chosen from among the numerous *unstable* eigenmodes determined for the set of 80 rings. For convenience in plotting, these amplitudes have been divided by the radii. (It is not meant to be implied here that the particular negative characteristic values ω^2 which were selected have any special significance—as a matter of fact, on the basis of some other results, this author suspects that the spectrum of negative ω^2, unlike that for positive ω^2, would be continuous for the case of a smooth disk without a sharp outer edge.)

Figure 4 clearly supports our earlier contention that disturbances of shorter and shorter wavelengths are increasingly unstable. Moreover, the eigen-mode corresponding to $\omega = -0.0958\, i\Omega_c$, for instance—or for that matter, the curve for $\omega = 0.0751\ \Omega_c$ in Figure 3—furnishes an estimate of the scale of the approximately neutrally stable disturbances: If we define the "wavelength" as the distance from the center of the disk to the next node but one, both these curves are seen to indicate a neutral wavelength close to $2a$. Referring back to Figure 1, we find that this result agrees almost embarrassingly well with the mean value between $r = 0$ and $r = 2a$ that was extrapolated for Model 2 from the theory based on the assumption that the wavelength was small compared with the radius.

No doubt this excellent agreement is largely fortuitous, but it nonetheless suggests that the local theory can probably be trusted to predict within some 20–30 per cent even the values of the neutrally stable "wavelengths," large though these may be. Granted the power of hindsight, it is not difficult to understand why this should be the case: As Safronov (1960) remarked in a similar context, considerably the largest contribution to the disturbance gravity experienced by a given particle stems from material situated within the nearest quarter-wavelength. Hence a more generous criterion for insuring that

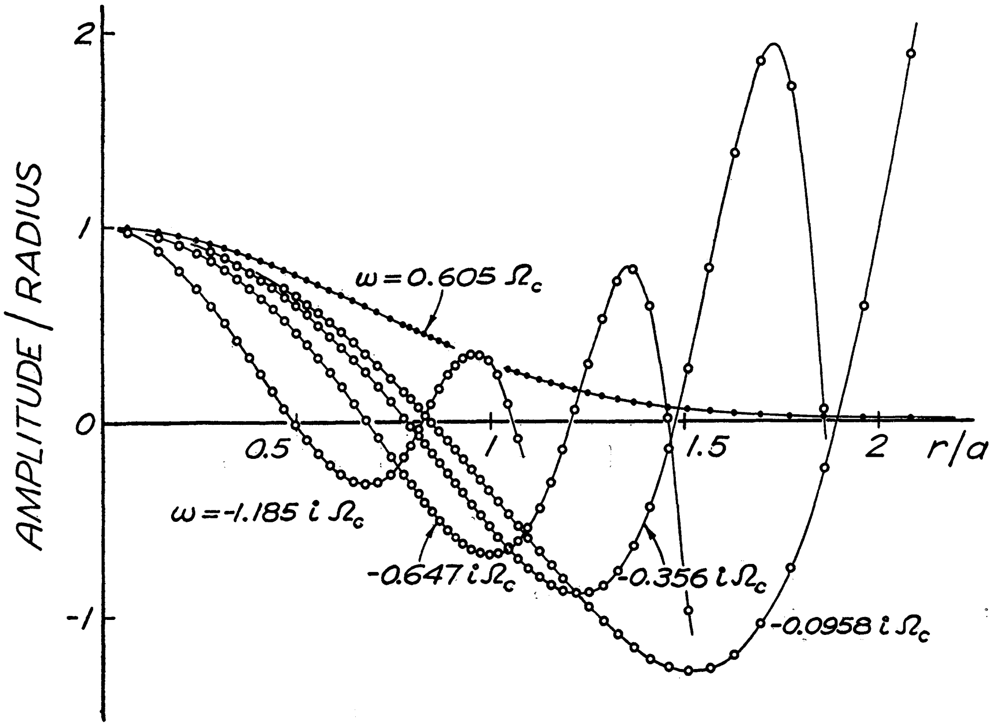

FIG. 4.—The relative amplitudes, divided by the radius, of one stable and several unstable eigen-disturbances to Model 2, as functions of the radius.

the local theory remains of value is probably that *that* fraction of the wavelength should still be small compared with the radius. Application of the latter criterion invalidates the results shown in Figure 1 only from about $r = \frac{1}{2}a$ inward.

Finally, in this connection it should also be noted that Hunter (1963) has recently managed to determine analytic expressions for the entire infinite family of axisymmetric *and* non-axisymmetric infinitesimal eigen-modes to an extremely flattened Maclaurin spheroid—that is to say, to a bounded disk with a uniform undisturbed angular velocity and without random motions. Among other things, Hunter, too, finds that the dimensions of the most extensive unstable disturbances to such a system are quite comparable to the radius, and that the rates of growth of these unstable disturbances (including the non-axisymmetric ones) increase with decreasing "wavelengths."

V. THE INFLUENCE OF RANDOM MOTIONS

We now turn to the question of the gravitational stability of a thin rotating disk composed of a large number (in fact, a continuum) of particles or stars whose random motions may no longer be ignored. For simplicity, the present analysis will essentially be limited to the determination of only the neutrally stable or time-independent disturbances under the same (mathematical, if not physical) assumptions of short wavelengths and axial symmetry as were made in Section III. There can, of course, be no assurance from such a purely local theory that any complete disk would actually admit disturbances which are exactly time-independent. Nevertheless, provided that some local approximations to such postulated neutrally stable over-all disturbances can be found at all, and that these represent the limit of the very slightly *unstable* disturbances, it seems not unreasonable to expect that their properties will constitute a significant dividing line between those local combinations of random motions and rotation which could succeed in overcoming the clumping tendency of the self-gravitation, and those that could not.

a) Collisionless Boltzmann Equation

Since the stars in the disk of an average galaxy can be estimated almost never to suffer close encounters with other stars (e.g., Chandrasekhar 1942, p. 81), this analysis will be based on the presumption that these discrete mass points influence one another's motions only collectively through distant-acting gravitational forces. Therefore, if $f = f(u, v, r, t)$ denotes a velocity distribution function such that the amount of mass dm corresponding to a surface element $r\,dr\,d\theta$ and to the velocity increments du and dv is

$$dm = f(u, v, r, t)\, du\, dv\, r\, dr\, d\theta\,, \tag{40}$$

and if F_r is the total gravitational force per unit mass acting radially outward, the most general axisymmetric disturbances must obey the collisionless Boltzmann equation

$$\frac{\partial f}{\partial t}+u\frac{\partial f}{\partial r}+\frac{v^2}{r}\frac{\partial f}{\partial u}-\frac{uv}{r}\frac{\partial f}{\partial v}+F_r\frac{\partial f}{\partial u}=0\,, \tag{41}$$

where u and v are the radial and the *full* azimuthal velocity components relative to a non-rotating frame of reference.

In what follows, let us think of the distribution function $f(u, v, r, t)$ as consisting of an original or undisturbed part, $f^0(u, v, r)$, and of an infinitesimal disturbance, $f'(u, v, r, t)$. Let us also abbreviate by w the difference between the particular tangential speed v and the speed $V(r)$ that would be required at the radius r for a particle to execute an exactly circular orbit when experiencing only the undisturbed force, $F_r{}^0(r)$. The disturbance force will be written as $F_r{}'(r, t)$. Using this notation, equation (41) may profitably be split into the following four groupings:

$$\begin{aligned}&\left(u\frac{\partial f^0}{\partial r}+\frac{v^2}{r}\frac{\partial f^0}{\partial u}-\frac{uv}{r}\frac{\partial f^0}{\partial v}+F_r{}^0\frac{\partial f^0}{\partial u}\right)\\ &+\left(\frac{\partial f'}{\partial t}+u\frac{\partial f'}{\partial r}+2w\frac{V(r)}{r}\frac{\partial f'}{\partial u}-u\frac{V(r)}{r}\frac{\partial f'}{\partial v}+F_r{}'\frac{\partial f^0}{\partial u}\right)\\ &+\left(\frac{w^2}{r}\frac{\partial f'}{\partial u}-\frac{uw}{r}\frac{\partial f'}{\partial v}+F_r{}'\frac{\partial f'}{\partial u}\right)+\left(\frac{V^2(r)}{r}+F_r{}^0\right)\frac{\partial f'}{\partial u}=0\,.\end{aligned} \tag{42}$$

The contents of the first of these parentheses cancel identically, owing to the equilibrium definition of f^0. The terms in the last parentheses do likewise, by virtue of the definitions of $V(r)$ and $F_r{}^0(r)$. Furthermore, when we now also assume that the velocity dispersion, though not negligible, is small compared with the equilibrium speed, $V(r)$, the third parentheses may be seen to enclose only terms each of which individually is considerably smaller than some member of the second group, at least for all values of u and $w = v - V(r)$ that are apt to be of any interest. The disturbance distribution function, $f'(u, v, r, t)$, must therefore be governed approximately by this twice-linearized Boltzmann equation,

$$\frac{\partial f'}{\partial t}+u\frac{\partial f'}{\partial r}+2w\Omega(r)\frac{\partial f'}{\partial u}-u\Omega(r)\frac{\partial f'}{\partial v}+F_r'\frac{\partial f^0}{du}=0\,, \tag{43}$$

where $\Omega(r)$ again denotes the equilibrium angular velocity of rotation, $V(r)/r$.

b) Time-dependent Disturbances

Consistent with the above assumption that the velocity dispersion is relatively small, we now stipulate the zeroth-order distribution function approximately as

$$f^0(u, v, r) = (\mu/2\pi\sigma_u\sigma_v)\exp(-u^2/2\sigma_u^2 - w^2/2\sigma_v^2)\,, \tag{44}$$

where μ again is the undisturbed surface mass density, here assumed not to vary with the radius. It is, of course, known from the theory of galactic dynamics (e.g., Chandrasekhar 1942, p. 159)—or may be confirmed directly from the zeroth-order Boltzmann equation—that the root-mean-square velocities σ_u and σ_v in this case are not independent but must be related through

$$\sigma_v^2/\sigma_u^2 = \tfrac{1}{2}[1+(r/V)(dV/dr)] = B/(B-A) = -B/\Omega\,, \tag{45}$$

Oort's constants A and B having already been defined in equation (9).

We shall look for (supposedly) short-wavelength, axisymmetric disturbances to this equilibrium state of the form

$$f'(u, v, r, t) = f^0(u, w)\, g(u, w)\, e^{iar} e^{st}\,, \tag{46}$$

where s is assumed to be a small, real, positive constant. Our intent will be to take the limit $s \rightarrow 0$ at a later stage of this analysis, and so to determine the relevant time-independent disturbances and their wavenumbers.

Having assumed that $ar \gg 1$, we may (as in Section III*b*) at once estimate the force corresponding to this disturbance as

$$F_r'(r, t) \cong 2\pi i G e^{iar} e^{st} \int\int f^0(u, w)\, g(u, w)\, du dw\,. \tag{47}$$

On the same grounds, we shall henceforth ignore any variation of $\Omega(r)$ in equation (43), except in the term $u(\partial f'/\partial r)$ in which that angular velocity happens to be differentiated. Consequently, it may be deduced from equations (43)–(47) that the function $g(u, w)$ must to a first approximation satisfy

$$2w\Omega\frac{\partial g}{\partial u}+2uB\frac{\partial g}{\partial w}+(s+iua)\,g=iuK\,, \tag{48}$$

where K abbreviates a relative constant that depends on $g(u, w)$ only in the integral sense

$$K = (2\pi G/\sigma_u^2)\int\int f^0(u, w)\, g(u, w)\, du dw\,. \tag{49}$$

As regards the detailed behavior of $g(u, w)$, equation (48) may be thought of as a first-order partial differential equation for that function, the characteristics of which are described by

$$\frac{du}{2w\Omega}=\frac{dw}{2uB}=\frac{dg}{iuK-(s+iua)\,g}\,. \tag{50}$$

The projections of these characteristics onto the uw-plane are the ellipses

$$-B\,u^2+\Omega w^2 = -B\sigma_u^2(u^2/\sigma_u^2+w^2/\sigma_v^2) = \text{constant}\,. \tag{51}$$

Hence, if we write

$$u/\sigma_u = \rho \cos \psi \quad \text{and} \quad w/\sigma_v = \rho \sin \psi , \tag{52}$$

the rate of change of $g(u, w)$ along any one of the characteristics may be deduced also from equation (50) as

$$\frac{dg}{d\psi} = (a + ib\rho \cos \psi) g - ic\rho \cos \psi , \tag{53}$$

where

$$a = (\sigma_u/2\sigma_v)(s/\Omega) ,$$
$$b = (\sigma_u/2\sigma_v)(\sigma_u \alpha/\Omega) , \tag{54}$$
$$c = (\sigma_u/2\sigma_v)(\sigma_u K/\Omega) .$$

The general solution of equation (53) for any $a > 0$ is

$$g = ic\rho \exp(a\psi + ib\rho \sin \psi) \int_{\psi}^{L(\rho)} \cos p \exp(-ap - ib\rho \sin p)\, dp , \tag{55}$$

where the upper limit, $L(\rho)$, is as yet arbitrary. However, the requirement that $g = g(u, w)$ should in this instance be a single-valued function of u and w, and hence $g(\rho, \psi)$ be *periodic* in ψ with a period 2π, imposes a severe limitation on $L(\rho)$. In fact, it can be shown that the latter condition, which is equivalent to the requirement that

$$e^{-aL(\rho)} \int_0^{2\pi} \cos p \exp(-ap - ib\rho \sin p)\, dp = 0 , \tag{56}$$

cannot be met if a is real and positive, unless the real part of $L(\rho)$ is infinite. Consequently, the function g that is of interest here may be expressed without any loss of generality as

$$g = ic\rho \exp(ib\rho \sin \psi) \int_0^{\infty} \cos(\psi + p') \exp[-ap' - ib\rho \sin(\psi + p')]\, dp' , \tag{57}$$

or, following an integration by parts, as

$$g = (K/\alpha)[1 - C(a, b, \rho, \psi) \exp(ib\rho \sin \psi)] , \tag{58}$$

where

$$C(a, b, \rho, \psi) = a(1 - e^{-2\pi a})^{-1} \int_0^{2\pi} \exp[-ap' - ib\rho \sin(\psi + p')]\, dp' . \tag{59}$$

It must be remembered, however, that the constant K is not arbitrary but must itself depend on $g(u, w)$ as a whole through equation (49); indirectly, that imposes a restriction on α in terms of the other parameters.

c) Time-independent Disturbances

We now proceed to the limit of vanishing instability. By reference to Watson (1944, p. 47), we thus find that

$$\lim_{s \to 0} C = \pi^{-1} \int_0^{\pi} \cos(b\rho \cos p')\, dp' = J_0(b\rho) , \tag{60}$$

or that the neutrally stable form of the function $g(u, w)$ must be

$$g_n(u, w) = (K/\alpha_n)[1 - J_0(b\rho) \exp(ib\rho \sin \psi)] , \tag{61}$$

where the subscripts n denote "neutral." The substitution of this g_n and of the $f^0(u, v, r) = f^0(\rho)$ from equation (44) into equation (49) then results in the following implicit relation for the neutral wavenumber, α_n:

$$\frac{\alpha_n \sigma_u^2}{2\pi G\mu} = 1 - \frac{1}{2\pi}\int_{\rho=0}^{\infty}\int_{\psi=0}^{2\pi} e^{-\rho^2/2} J_0(b\rho)\exp(ib\rho\sin\psi)\,\rho\, d\psi\, d\rho$$

$$= 1 - \int_{\rho=0}^{\infty} e^{-\rho^2/2}[J_0(b\rho)]^2 \rho\, d\rho \tag{62}$$

$$= 1 - \exp(-\sigma_u^2\alpha_n^2/\kappa^2)\, I_0(\sigma_u^2\alpha_n^2/\kappa^2)\,,$$

the two integrals above having been evaluated according to Watson (1944, pp. 47, 395).

The final version of equation (62) is strictly due to A. Kalnajs (1963), who obtained it by another method during an attempt to verify an earlier result of the present author,

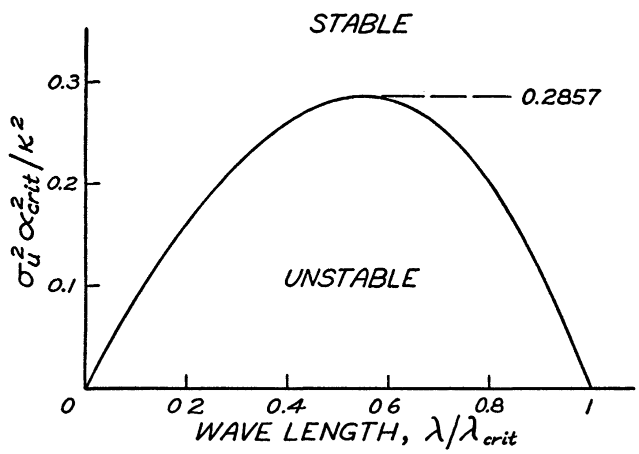

FIG. 5.—A curve describing the wavelengths and radial-velocity dispersions admitting neutrally stable, axisymmetric disturbances, for the case of a disk of stars with random motions and rotation.

from which the modified Bessel function, $I_0(\sigma_u^2\alpha_n^2/\kappa^2)$, had been mistakenly omitted. As it stands, the present analysis is somewhat related also to that of Lynden-Bell (1962) for a three-dimensionally homogeneous example.

As a partial check on equation (62), it should be noted that as the velocity dispersion $\sigma_u \to 0$, while α_n and κ remain finite, the right-hand side of that equation approaches $(\sigma_u^2\alpha_n^2/\kappa^2)$; consequently, the neutral wavenumber then tends to

$$\kappa^2/2\pi G\mu = \alpha_{\text{crit}}\,, \tag{63}$$

in welcome agreement with equation (21). On the other hand, when $(\sigma_u\alpha_n/\kappa)$ tends to infinity, equation (62) yields

$$\alpha_n \approx 2\pi G\mu/\sigma_u^2\,, \tag{64}$$

a result which could have been obtained more directly from the infinitesimal stability analysis of a thin, infinite sheet involving no rotation. As for other values of α_n, equation (62) implies a dependence on σ_u^2 that is fully described by the neutral stability curve in Figure 5; note that the abscissa in that diagram is the neutral wavelength, $\lambda_n = 2\pi/\alpha_n$.

It should be emphasized that the above derivation has strictly only established the inadmissibility of neutrally stable disturbances under circumstances other than those which are described by the curve in Figure 5. However, bearing in mind that that curve was deliberately obtained in the limit of vanishingly *unstable* disturbances, and that it agrees qualitatively with what was to be expected from the order-of-magnitude discussion of Section II*c* as regards the ranges of stable and unstable wavelengths for different velocity dispersions, it seems practically certain that it represents the boundary between the conditions in which an arbitrary sinusoidal disturbance with axial symmetry either would or would not remain *bounded* for all subsequent time; hence the labels "stable" and "unstable" on the respective regions in Figure 5. It thus appears from that diagram that the radial instabilities should be altogether avoided in the presence of root-mean-square radial velocities exceeding

$$\sigma_{u,\,\min} = (0.2857)^{1/2}\,\kappa/\alpha_{\rm crit} = 3.36\,G\mu/\kappa\,, \tag{65}$$

assuming, of course, that the undisturbed velocity distribution is approximately Gaussian.

To give some idea of the magnitudes both of this critical velocity dispersion, and of the corresponding $\sigma_{v,\min}$ implied by equation (45), these two locally estimated quantities

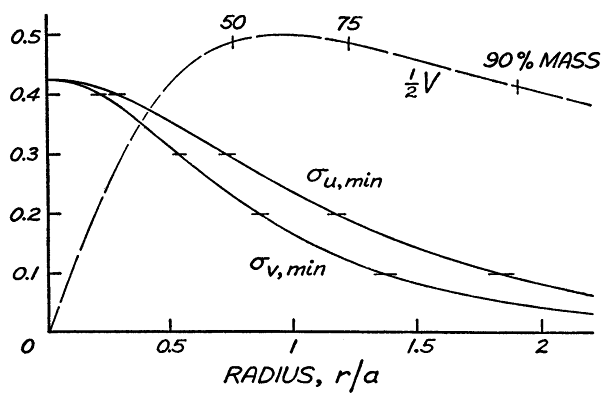

FIG. 6.—The minimum radial and tangential velocity dispersions needed to stabilize Model 2 against all axisymmetric disturbances, expressed as multiples of the maximum linear velocity of rotation and as functions of the radius. The broken curve denotes one-half of the rotation velocity for the same model galaxy, drawn to scale.

have been plotted in Figure 6 as functions of the radius, for a model galaxy with the same surface density and rotation law as Model 2. (Due to the "pressure" of the random motions, Model 2 is, of course, no longer an exact equilibrium configuration, but doubtless we may neglect that inexactitude for the purpose of this illustration.) From Figure 6 we observe that, while these minimum random velocities are indeed roughly of the same magnitude as the typical speeds of revolution, numerically they are not as large as was suggested by the crude estimate of Section II*c*. The comparative smallness of $\sigma_{u,\,\min}$ is further illustrated by the fact that a star having that as its maximum radial velocity would, in the absence of any collective disturbance forces, execute an epicycle with a major axis amounting only to about 0.17 $\lambda_{\rm crit}$.

In retrospect, it therefore seems fair to conclude that the assumption $\sigma_u \ll V(r)$ has been grossly violated only right near the center. As regards the short-wavelength assumption, the violation of which might appear to be more serious, we can only remark (*a*) that Section IV*b* indicated the local theory to remain quite accurate even when the wavelengths in question had become comparable to the radius, and (*b*) that Figure 5 implies our present concern to be chiefly with wavelengths of only about 0.55 $\lambda_{\rm crit}$ which belong to the disturbances most reluctant to be stabilized. All things considered, the systematic increases in Figure 6 of $\sigma_{u,\min}$ and $\sigma_{v,\min}$ toward $r = 0$ appear to be of some significance, even though their central values are mere extrapolations.

d) Non-axisymmetric Disturbances

The present writer has not explored in any detail the effects of random motions upon disturbances without axial symmetry, except to note that a *local* analysis very similar to the foregoing could also have been carried out for non-axisymmetric disturbances had the *differential* rotation happened to be negligible. In that case, the minimum stabilizing σ_u would have been found to be identical with that of equation (65).

This conclusion suggests, but of course does not prove, that the stability even of the comparatively shorter non-axisymmetric disturbances should be assured in general by a radial velocity dispersion approximately equal to $\sigma_{u,\,\text{min}}$. However, a question which the present discussion leaves completely unanswered is to what extent a similar amount of random motion might affect the character of the most extensive non-axisymmetric disturbances, in particular those which ought to determine whether or not a given disk might prefer to develop into a barlike structure.

VI. COMMENTS

To sum up, it has been shown here that any reasonably thin, smooth, and rotating disk of stars should be vulnerable to a variety of remarkably extensive and violent instabilities, if those stars did not already possess sufficiently vigorous random motions relative to one another, superposed on their common rotation. Therefore, any observed smooth distribution of the more common stars in the disk of a galaxy whose age is to be reckoned in tens of revolutions must imply that the stellar velocity dispersions in almost all parts of such a system do now at least equal, if not exceed, the minimum values consistent with stability.

a) A Comparison with Observations in the Solar Vicinity

A lack of observations concerning the random velocities of the *average* stars in the spiral and S0 galaxies at present limits any comparisons between the actual and the theoretically required velocity dispersions to our immediate vicinity of this galactic disk.

Judging from an extrapolation of the volume density of known stars within 20 pc of the Sun, as determined by Gliese (1956), and from Oort's (1960) estimates of the z-component of the gravitational attraction at distances up to 1000 pc above the galactic plane, the mean surface or projected density of *stellar* matter in our neighborhood seems to be bracketed by

$$\mu = 50\text{–}65\ M_{\odot}/\text{pc}^2\ . \tag{66}$$

Our epicyclic frequency, on the other hand, may be estimated as

$$\kappa = 2(-B)^{1/2}\,(A - B)^{1/2} = 27\text{–}32\ (\text{km/sec})/\text{kpc}\ , \tag{67}$$

the lower value being based on the $A = 19.5$ and $B = -6.9$ (km/sec)/kpc advocated by Schmidt (1956), and the upper one on the $A = 15$ determined by Johnson and Svolopoulos (1961) and by Kraft and Schmidt (1963), together with the $B = -10$ deduced by Schmidt (1964).

Equation (65) indicates that any supposedly smooth and very thin disk having the above properties in a given vicinity would require a radial velocity dispersion of at least 22.5–35 km/sec in that same neighborhood, if it was to be stable. Given the same values of μ and κ, it also follows from equation (22) that the longest unstable wavelengths in the absence of random motions would be

$$\lambda_{\text{crit}} = 8.5\text{–}15\ \text{kpc}\ , \tag{68}$$

and the wavele gth of the axisymmetric disturbance most reluctant to be stabilized by a finite amount of random motion,

$$0.55\ \lambda_{\text{crit}} = 5\text{–}8\ \text{kpc}\ . \tag{69}$$

To be realistic, the above estimate for $\sigma_{u,\min}$ requires at least two (largely self-canceling) corrections. The first stems from the fact that we are actually dealing with a disk whose effective thickness of roughly 600 or 700 pc is about one-tenth of the most relevant wavelength given by equation (69); hence the example of Section III*e* suggests that that estimate should be reduced by some 15–20 per cent, or to about 18–30 km/sec. On the other hand, one must also make some allowance for the presence of interstellar material. Even though a gaseous disk of, say, one-tenth the projected stellar density cannot by itself be expected to be unstable with regard to wavelengths of the order of 5–8 kpc, such material should on account of its relatively small random velocities be fairly readily attracted toward any concentrations of stars, thereby augmenting their disturbance gravity. Indeed, if we suppose the gas to be distributed uniformly, and to experience practically no pressure or magnetic forces, it is not difficult to extend the analysis of Section V to obtain the following expression for the neutral wavenumber, α_n, in place of equation (62):

$$\frac{\alpha_n \sigma_u^2}{2\pi G \mu_s} = \left(1 - \frac{\alpha_n}{\alpha_{\mathrm{crit},\,g}}\right)^{-1} \left[1 - \exp\left(-\sigma_u^2 \alpha_n^2/\kappa^2\right) I_0\left(\sigma_u^2 \alpha_n^2/\kappa^2\right)\right] . \tag{70}$$

Here the subscripts *s* and *g* refer to the stars and the gas, respectively. Conservatively assuming that $\mu_g/\mu_s \geq 0.1$ in our vicinity, it may be calculated from equation (70) that a velocity dispersion at least 20 per cent greater than that estimated previously would be required to curb all axisymmetric instabilities involving wavelengths of the order of $0.55\lambda_{\mathrm{crit},s}$. Although such an increase must probably be diminished in actuality by the fact that the gas is already distributed unevenly, it is not clear that the effect is insignificant. Therefore, the best present estimate of the minimum σ_u for stability has to be

$$\sigma_{u,\,\min} = 20\text{–}35 \text{ km/sec} . \tag{71}$$

This theoretical estimate is to be contrasted with an average of the observed velocity dispersions, weighted heavily in favor of (1) the K- and M-type main-sequence stars, and (2) the white dwarfs, which, according to Gliese (1956), respectively account for about two-thirds and some 10–15 per cent of the recognized stellar matter in the solar neighborhood. Unfortunately, no accurate estimate of σ_u for the white dwarfs appears to be available, although their velocity dispersion seems unlikely to be any less than the present-day $\sigma_u \simeq 25$ km/sec observed for the red giants (e.g., Allen 1963, p. 243). There have, however, been a number of determinations of σ_u for the K8–M2-type dwarfs, in which special care has been taken to eliminate a selectional bias of the sort discussed by Woolley (1958): Dyer (1956) found $\sigma_u = 34.5$ km/sec, Mumford (1956) $\sigma_u = 31.4$ km/sec, and Mrs. Wehlau (1957) $\sigma_u = 32.1$ km/sec.

We conclude that the velocity dispersion of the most common stars in our vicinity is not inconsistent with the observationally favored hypothesis that these stars are distributed fairly uniformly over this galactic disk. Of course, owing to the virtual identity of the observed and the theoretically required σ_u's, and the fact that this theory strictly applies only to relatively short axisymmetric disturbances, it is as yet impossible to rule out instabilities altogether. However, should any actually be present, it must be understood that this comparison and Figure 5 also suggest that such instabilities would have to be relatively mild and confined to wavelengths (or similar dimensions) of the order of the 5–8 kpc estimated in equation (69); conversely, all disturbances to this *stellar* disk with wavelengths comparable to the 2-kpc spiral-arm spacing must almost certainly be judged stable. The latter point is important as an argument against any suggestion that the existing spiral structure in this Galaxy might be the result of collective stellar *instabilities* of the sort we have been considering. In fairness, however, it should again be noted that all these remarks of necessity refer only to conditions in the solar neighborhood.

b) Instabilities as a Cause of Increased Random Motions

The question of the eventual fate of any disk of stars originally lacking the random velocities needed to protect itself against all instabilities is not one that can be answered entirely by any linearized stability analysis. All that can surely be said is that such instabilities would *in the short run* (i.e., in one or two revolutions) have caused many excess stars to become attracted to certain areas on the disk at the expense of neighboring regions. However, it must not be presumed that such initial clumpings would necessarily have led to the formation of any *permanent* irregularities.

It is indeed difficult to conceive of any effective means whereby the different stars arriving in any typical region of increased density could have sufficiently dissipated their relative motions to have become gravitationally bound to each other. On the contrary, it seems much more likely that the bulk of the stars involved in any given (generally non-axisymmetric) instability would merely have streamed past each other when their degree of clumping reached a maximum, and that they would eventually have dispersed themselves upon emerging from the opposite sides of the aggregation and upon experiencing the shearing effect of the differential rotation.

Of course, it must be conceded that if the random velocities had not been altered significantly as a result of the clumping, the subsequent smoothing by the differential rotation could not have proceeded too far without again violating the principle that too smooth a disk without adequate random motions must be unstable. On the other hand, even a *net* effect of every instability would surely have had to be some increase in the random or non-circular motions, for it seems improbable that the excess "random" kinetic energy imparted during the growth phase of any instability could ever have been recovered fully.

It follows that an initially unstable disk of stars should probably have undergone not just one but several successive generations of instabilities, after each of which the system would have been left somewhat less unstable than it was previously. In particular, it seems likely that before very many rotation periods had elapsed, the disk would have approached a new equilibrium state that was again fairly regular and quite possibly axisymmetric, but in which the random velocities at the various radii had become—and would henceforth remain—about equal to the minimum values needed for complete stability.

It might seem at first that this suggestion that the system could have again become smooth and yet now possess additional disorganized kinetic energy somehow violates the energy conservation principle; actually, it merely requires some adjustment to have taken place in the gross distribution of material. Curiously, though, the virial theorem indicates that the total gravitational energy in the final state would have had to be the same as that of the original (unstable) equilibrium configuration; hence the said redistribution of stars could not simply have consisted of an over-all contraction, but would have had to entail a contraction perhaps of the inner parts of the disk jointly with a net expansion of the outer portions.

c) An Interpretation of the Observed Velocity Dispersions

The last few remarks obviously invite the query whether the afore-mentioned near agreement between the observed and theoretically required σ_u's of the common stars in our own vicinity might not be more than a simple coincidence. In reply, we can only point to the one strong piece of evidence which suggests that those stars must at least have been created with appreciably less than their present random velocities: This is the well-known fact that the z-velocity dispersions of most classes of nearby stars are considerably smaller than their σ_u's. In the case of the dK8–dM2 stars, for instance, Dyer, Mumford, and Mrs. Wehlau all found σ_w to be only about 19 km/sec, as opposed to their σ_u in excess of 30 km/sec. Although it is now understood from discussions of the approximate third integral (e.g., Contopoulos 1963) that such an anisotropy, once established, could have persisted almost indefinitely, no satisfactory explanation appears forthcoming as to how those stars could actually have *originated* with such distinctly larger radial than *z*-velocities. (For instance, it is difficult to picture a comparable lack of isotropy in the random motions of the original interstellar material.)

Still, one must not jump to the conclusion that the instability mechanism has necessarily been the only process tending to increase the random velocities of the various stars preferentially in the directions parallel to the galactic plane *since* their formation. Schwarzschild and Spitzer (1951, 1953), for instance, considered stellar encounters with massive "cloud complexes" as giving rise to a similar acceleration; conceivably, other mechanisms might also have been operative. Nevertheless, it is probably fair to remark that had all the competing processes together not been able to provide for a sufficient—and a sufficiently rapid—increase in the radial and azimuthal random velocities during that epoch when most of these stars were formed, then the acceleration through their collective gravitational instabilities should almost certainly have supplied the balance.

Much of the credit for this investigation belongs to Professor C. C. Lin, both for asking a very relevant question at the outset and for many subsequent discussions. I am particularly indebted also to Professors B. Strömgren and L. Woltjer for their generous advice and helpful criticism, and to Mr. A. Kalnajs, who called to my attention a substantial error in an earlier version of Section V.

Begun (and completed) at the Massachusetts Institute of Technology, this paper was largely written at the Institute for Advanced Study, during a visit made possible through the kindness of Professors J. R. Oppenheimer and B. Strömgren. In part, this work was also supported by the National Science Foundation. The numerical calculations were performed at the M.I.T. Computation Center.

REFERENCES

Allen, C. W. 1963, *Astrophysical Quantities* (2d ed.; London: Athlone Press), p. 243.
Chandrasekhar, S. 1942, *Principles of Stellar Dynamics* (Chicago: University of Chicago Press), pp. 81, 159, 160.
Contopoulos, G. 1963, *A.J.*, **68**, 1.
Dyer, E. R., Jr. 1956, *A.J.*, **61**, 228.
Gliese, W. 1956, *Zs. f. Ap.*, **39**, 1.
Hunter, C. 1963, *M.N.*, **126**, 299.
Johnson, H. L., and Svolopoulos, S. N. 1961, *Ap. J.*, **134**, 868.
Kalnajs, A. 1963, private communcation.
Kraft, R. P., and Schmidt, M. 1963, *Ap. J.*, **137**, 249.
Lin, C. C. 1961, private communication.
———. 1964, *Ap. J.*, in press.
Lynden-Bell, D. 1962, *M.N.*, **124**, 279.
Maxwell, J. C. 1857, *Adams Prize Essay;* reprinted 1890 in *Scientific Papers* (Cambridge: Cambridge University Press), **1**, 288–376.
Mestel, L. 1963, *M.N.*, **126**, 553.
Mumford, G. S., III. 1956, *A.J.*, **61**, 224.
Oort, J. H. 1960, *B.A.N.*, **15**, 45.
Safronov, V. S. 1960, *Ann. d'ap.*, **23**, 979.
Schmidt, M. 1956, *B.A.N.*, **13**, 15.
———. 1964, in *Stars and Stellar Systems*, ed. A. Blaauw and M. Schmidt (Chicago: University of Chicago Press), in press.
Schwarzschild, M., and Spitzer, L., Jr. 1951, *Ap. J.*, **114**, 385.
———. 1953, *Ap. J.*, **118**, 106.
Toomre, A. 1963, *Ap. J.*, **138**, 385.
———. 1964, *Ap. J.*, to be submitted.
Watson, G. N. 1944, *Theory of Bessel Functions* (2d ed.; Cambridge: Cambridge University Press), pp. 47, 395.
Wehlau, A. W. 1957, *A.J.*, **62**, 169.
Woolley, R. v. d. R. 1958, *M.N.*, **118**, 45.